Humanistic Computation and Intelligence

CAADRIA 2026

The 31st International Conference on Computer-Aided Architectural Design Research in Asia

Volume 3

Edited by:

Mohammed Makki
Jielin Chen
Sheng-Fen Nik Chien
Jie Liu
Likai Wang
Shih-Yuan Wang

Humanistic Computation & Intelligence

Proceedings of the 31st International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2026)

Volume 3

Edited by

Mohammed Makki
University of Technology Sydney, Australia

Jielin Chen
National University of Singapore, Singapore

Sheng-Fen Nik Chien
National Cheng Kung University, Taiwan

Jie Liu
Beijing University of Posts and Tele-communications, China

Likai Wang
Xi'an Jiaotong-Liverpool University, China

Shih-Yuan Wang
National Yang Ming Chiao Tung University, Taiwan

Humanistic Computation and Intelligence
31st International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2026)

26 April - 2 May 2026

Hosted By:

National Yang Ming Chiao Tung University (NYCU),
TAIWAN

Co-Hosted By:

CAADRIA, NYUST, FCU

©2026 All rights reserved and published by
The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong

ISBN: 978-988-71359-0-6
ISSN: 2710-4257 (print)
2710-4265 (online)

Cover design by Yu-Yun Chang and cover pattern by Chen-An Lee
Printed in Hsinchu, TAIWAN

Foreword

The annual Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) conference provides an international platform for researchers and practitioners to exchange, discuss, and publish work in computer-aided architectural design. These proceedings, presented across three volumes, include papers accepted for the 31st International CAADRIA Conference, hosted by the Institute of Architecture at the National Yang Ming Chiao Tung University (NYCU) in Hsinchu, Taiwan, and co-hosted by CAADRIA, NYUST and FCU. All papers are also available through the open-access CumInCAD database.

CAADRIA 2026 is framed by the theme *Humanistic Computation and Intelligence*, reflecting a period of rapid technological development alongside growing attention to its broader implications. The theme positions computation not only as a technical tool, but as a collaborator shaped by human judgement, cultural context, ethics, and responsibility. The papers in these volumes explore how computational methods can support and challenge human creativity and agency in architectural design and research.

These proceedings result from a substantial collaborative effort involving the International Review Committee and the Paper Selection Committee. A total of 730 abstracts were submitted by 2,081 contributors from 48 countries across six continents, demonstrating the global reach of the CAADRIA community. Following an initial double-blind review, 340 full papers were invited for further consideration and underwent a second round of rigorous peer review. In total, 215 reviewers from 35 countries contributed their expertise. From this process, 216 papers were accepted for presentation and publication.

We congratulate all authors on their contributions and thank the International Review Committee for their commitment to scholarly quality. We also thank the hosting and organising teams, as well as the broader CAADRIA community, whose collective efforts make the conference possible. It is an honour to serve on the Paper Selection Committee, and we hope these proceedings support ongoing discussions at the intersection of computation, design, and human values.

PAPER SELECTION COMMITTEE

Mohammed Makki	*University of Technology Sydney*
Jielin Chen	*National University of Singapore*
Sheng-Fen Nik Chien	*National Cheng Kung University*
Jie Liu	*Beijing University of Posts and Tele-communications*
Likai Wang	*Xi'an Jiaotong - Liverpool University*
Shih-Yuan Wang	*National Yang Ming Chiao Tung University*

Conference Theme: 'Humanistic Computation & Intelligence'

As the boundaries between machine intelligence and human intuition blur, we stand at a pivotal moment in computer-aided architectural design. This year's conference explores Humanistic Computation and Intelligence—a call to embrace computation not as a detached instrument, but as a creative and ethical collaborator. We invite researchers, designers, and technologists to reimagine intelligence in the service of meaning, ethics, and humanity.

"*Humanistic*" refers to an approach that centers human values, lived experiences, and ethics within technological fields. In CAAD, this entails prioritizing creativity and cultural meaning, enhancing human agency rather than replacing it, addressing social and ethical dimensions, and embracing subjectivity, diversity, and critical thinking alongside data and algorithms.

To guide this year's discourse, we propose four manifestos:

- *The Human at the Core:* Computation should not replace the designer—it must empower human creativity, judgment, and imagination.
- *Design as Dialogue:* Computation is a conversation between human and machine, intuition and analysis, ethics and aesthetics.
- *Reclaiming Agency:* Intelligence is shared and emergent; human agency should be amplified, not overwritten.
- *Culture is Code:* Architecture encodes culture and meaning; computational design must engage critically with this legacy.

As new computational technologies emerge, we recognize the need to bridge machine rationality with the depth of human intuition, culture, and values.

June-Hao Hou, Chair of CAADRIA 2026

About CAADRIA

The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) promotes teaching and research in CAAD in Asia and has members on six continents. It organizes an annual conference, the first of which was held in 1996 in Hong Kong. Since then, 31 conferences have been held in Australia, China, Hong Kong, India, Japan, Korea, Malaysia, New Zealand, Singapore, Taiwan, and Thailand. The conferences provide an opportunity for teachers, students, researchers, and practitioners to meet each other and learn about the latest research in the field. The proceedings of the conferences are available online and and indexed in Web of Science, Scopus and CumInCAD.

CAADRIA was founded with the following objectives:

- To facilitate the dissemination of information about CAAD among Asian schools of architecture, planning, engineering, and building sciences.
- To encourage the exchange of staff, students, experience, courseware, and software among schools.
- To identify research and develop needs in CAAD education and to initiate collaboration to satisfy them.
- To promote research and teaching in CAAD that enhances creativity rather than production.

CAADRIA is one of the four founding organizations of the International Journal of Architectural Computing (IJAC), and guest-edits one issue each year. IJAC is published by Sage in both paper and electronic versions.

Rudi Stouffs, President of CAADRIA

CAADRIA Officers

Rudi Stouffs (President)
National University of Singapore, Singapore
Nic Bao (Secretary)
RMIT University, Australia
Provides Ng (Treasurer)
The Chinese University of Hong Kong, Hong Kong SAR
Jeroen van Ameijde (Membership Officer)
The Chinese University of Hong Kong, Hong Kong SAR
Chen Hao (Webmaster)
Agency for Science, Technology and Research, Singapore
Anastasia Globa (Election Officer)
The University of Sydney, Australia
Toshiki Hirano (Outreach Officer)
The University of Tokyo, Japan
Dan Luo (Publication Officer)
University of Queensland, Australia
Marc Aurel Schnabel (Administrative Officer)
Xi'an Jiaotong-Liverpool University, China
Likai Wang (Organization Officer)
Xi'an Jiaotong-Liverpool University, China
Chitraj Bissoonauth (Communication Officer)
Xi'an Jiaotong-Liverpool University, China

CAADRIA Fellows

Tomohiro Fukuda
Osaka University, Japan
John Gero
University of North Carolina, USA
Christiane M. Herr
Southern University of Science and Technology, China
Tom Kvan
Southern University of Science and Technology, China
Walaiporn Napakan
Chulalongkorn University, Thailand
Marc Aurel Schnabel
Xi'an Jiaotong-Liverpool University, China

CAADRIA 2026 Conference Committees

Organizers & Hosts

June-Hao Hou (Chair)
National Yang Ming Chiao Tung University, Taiwan
Teng-Wen Chang
National Yunlin University of Science and Technology, Taiwan
Sheng-Fen Nik Chien
National Cheng Kung University, Taiwan
Shih-Yuan Wang (Workshop Coordinator)
National Yang Ming Chiao Tung University, Taiwan
Chor-Kheng Lim
Yuan Ze University, Taiwan
Sheng-Kai Tseng
National Yang Ming Chiao Tung University, Taiwan
Tien Ling
National Yang Ming Chiao Tung University, Taiwan
Pei-Hsien Hsu
National Yang Ming Chiao Tung University, Taiwan
Chen-An Lee (Conference Manager)
National Yang Ming Chiao Tung University, Taiwan
Jinbao Zhong (Program Manager)
National Yang Ming Chiao Tung University, Taiwan

Paper Selection Committee

Mohammed Makki (Chair), Jielin Chen, Sheng-Fen Nik Chien, Jie Liu, Likai Wang, and Shih-Yuan Wang

Post-Graduate Students Consortium

Sky Tiantian Lo (Chair)

Project Exhibition Committee

June-Hao Hou (Chair), Lee Kent, Jen-Hwang Ho, Chen-An Lee, and Ling-Li Tseng

Young CAADRIA Award Committee

Christopher Bamborough (Chair), Nic Dingwen Bao, Christiane M. Herr, June-Hao Hou, and Walaiporn Nakapan

Sasada Prize Committee

Christiane M. Herr (Chair), Anastasia Globa, Immanuel Koh, and Marc Aurel Schnabel

Proceedings Technical Support

Gabriel Wurzer

CUMINCAD Team

Bob Martens and Tomo Cerovsek

Special Thanks

Bo-Xuan Lu (Web Manager), Cheng-An Wang (Service Manager), Yu-Yun Chang (Media Designer), Rou-Han Chen (Resource Manager), Ting-Yu Yang (Resource Manager), Yu-Tse Hsieh (Printing and Registration Manager), Xiang-Jun Fu, Ping Ho
National Yang Ming Chiao Tung University, Taiwan

Jia-Rong Li, Shih-Ting Tsai, Ching-Chih Chang
National Yunlin University of Science and Technology, Taiwan

James Gomez, Marlin Hannam, Laiba Jamil, Josef Sfara, Nancy Sun
University of Technology Sydney, Australia

International Review Committee

Amin Adelzadeh,
CITA - Royal Danish Academy
Asli Agirbas,
Ozyegin University
Asterios Agkathidis,
University of Liverpool
Aysegul Akcay Kavakoglu,
Istanbul Technical University
Mohammad AlAbbasi,
Umm Al-Qura University
Miktha Farid Alkadri,
Universitas Indonesia
Abdulrahman Alymani,
Alfaisal Univrsty
Pierandrea Angius,
Architectural Association School of Architecture
Abdullah All Mamun Anik,
University of Huddersfield
Handan Aş Çemrek,
Istanbul Health and Technology University
Guy Austern,
Technion
Pedro Azambuja Varela,
Faculdade de Arquitectura da Universidade do Porto
Daniel Baerlecken,
N/A
Zishen Bai,
University College London
Christopher Bamborough,
University of New South Wales
Dingwen Nic Bao,
Royal Melbourne Institute of Technology
Jack Barton,
Voxelmates
Muge Belek Fialho Teixeira,
Queensland University of Technology
Sarah Blahut,
University of Kassel
Alexandra Bogdanova,
N/A
Peter Booth,
University of Tasmania
Luis Borunda,
Viriginia Tech
Johannes Braumann,
Association for Robots in Architecture / Creative Robotics
Luca Breseghello,
Technical University of Denmark
Katya Bryskina,
Katya Bryskina Studio
Peter Búš,
Institute of Future Human Habitats, Tsinghua Shenzhen International Graduate School
Tomas Cabezon Pedroso,
Carnegie Mellon University
Inês Caetano,
INESC-ID / IST - Instituto Universitario de Lisboa
Glenda Caldwell,
Queensland University of Technology
Abhinav Chaudhary,
Architectural Association School of Architecture
Jielin Chen,
National University of Singapore
Ting Han Chen,
Yuan-Ze University
Sihua Chen,
Osaka University
Qinchuan Chen,
Yonsei University
Zi-Ru Chen,
Southern Taiwan University of Science and Technology
Tiffany Cheng,
Cornell University

Lok Hang Cheung,
Xi'an Jiaotong - Liverpool University
Ling Kit Cheung,
University of Technology Sydney
Sheng-Fen Chien,
National Cheng Kung University
Thian-Siong Choo,
Singapore Polytechnic
I-Ting Chuang,
University of Auckland
Kristof Crolla,
The University of Hong Kong
Camilo Cruz Gambardella,
Monash University
Robson da Silva,
State University of Campinas
Shuyao Dai,
The University of Sydney
Burak Delikanlı,
Istanbul Technical University
Danilo Di Mascio,
The University of Huddersfield
Ben Doherty,
BVN
Leonor Domingos,
Iscte - University Institute of Lisbon
Lei Dongxue,
Ningbo University
Jonathan Dortheimer,
Ariel University
Theodoros Dounas,
Heriot Watt University / Adventurous systems
Eftihis Efthimiou,
AUTh, UIbk, Decode fab lab, Deca architecture
Valeria Ena,
Universidade de Lisboa
Benjamin Ennemoser,
Texas A&M University
Sarvin Eshaghi,
University of Illinois Urbana-Champaign
Shayani Fernando,
Swiss Federal Institute for Forest, Snow and Landscape
Fred Fialho Teixeira,
University of Queensland
Thomas Fischer,
Southern University of Science and Technology
Wilson Florio,
State University of Campinas
Anna Font,
Architectural Association School of Architecture
Tomohiro Fukuda,
The University of Osaka
Song Gao,
Shanghai Jiguang Polytechnic College
Fernando García Amen,
FADU/UDELAR - Universidad de la República
Rodrigo Garcia-Alvarado,
University of Bio-Bio
Nicole Gardner,
University of New South Wales
Nadja Gaudillière-Jami,
Royal Danish Academy
Gregorius A Gegana A,
Universitas Pelita Harapan
Ali Ghazvinian,
Texas Tech University
Anastasia Globa,
The University of Sydney
Alexander Grasser-Parger,
Graz University of Technology
Derya Gulec Ozer,
Istanbul Technical University
M. Hank Haeusler,
University of New South Wales
Adonis Haidar,
University of Liverpool
Gilles Halin,
Lorraine University / URM MAP Laboratory

Zhen Han,
Tianjin University
Xinyuan Hao,
University of Washington
Christiane M. Herr,
Southern University of Science and Technology
Pablo C Herrera,
Universidad Peruana de Ciencias Aplicadas
Toshiki Hirano,
The University of Texas at Austin
Tzu-Chieh Hong,
University of Kansas
Anca-Simona Horvath,
Aalborg University
Kensuke Hotta,
POllc. / Keio University
June-Hao Hou,
National Yang Ming Chiao Tung University
Xiaoran Huang,
North China University of Technology / Swinburne University of Technology
Chien-hua Huang,
N/A
Sheng-Yang Huang,
University College London
Chenyu Huang,
Tongji University
Markus Hudert,
Aarhus University
Erin Hunt,
Texas Tech University
Magdy Ibrahim Mahmoud,
Zayed UNiversity
Taysheng Jeng,
National Cheng Kung University
Muxin Jia,
New Jersey Institute of Technology
Leyuan Jiang,
Xi'an Jiaotong - Liverpool University
Wanzhu Jiang,
South China University of Technology
Iestyn Jowers,
The Open University
Adarsh Joy Joseph,
CEPT University
Sam Conrad Joyce,
Singapore University of Technology and Design
Anetta Kepczynska-Walczak,
Lodz University of Technology
Chin Koi Khoo,
Deakin University
Daiki Kido,
Kajima Corporation
Takuya Kikuchi,
Kajima Corporation
Axel Kilian,
N/A
Dongyun Kim,
Harvard Graduate School of Design
Frederick Chando Kim,
National University of Singapore
Lora Kim,
Wentworth Institute of Technology
Il Hwan Kim,
Illinois Institute of Technology
Yuki Kobayashi,
Osaka Metropolitan University
Immanuel Koh,
Singapore University of Technology and Design
Odysseas Kontovourkis,
University of Cyprus
Ih-Cheng Lai,
Tamkang University
Thanh Luan Le,
MAP group Korea
António Leitão,
Instituto Superior Tecnico

Surapong Lertsithichai,
N/A
Shuyang Li,
National University of Singapore
Chenxiao Li,
The University of Hong Kong
Bin Li,
South China University of Technology
Yang Li,
National University of Singapore
Fernando Lima,
Belmont University
Yin-Shan Lin,
Northeastern University
Chieh-Jen Lin,
Tainan University of Technology
Jingyang Liu,
Xi'an Jiaotong - Liverpool University
Jie Liu,
Beijing University of Posts and Telecommunications
Yiming Liu,
Southern University of Science and Technology
Yongkang Liu,
Tsinghua University
Tian Tian Sky Lo,
The Hong Kong Polytechnic University
Paul Loh,
Bond University
Davide Lombardi,
Xi'an Jiaotong - Liverpool University
Alvaro Lopez Rodriguez,
The Bartlett School of Architecture
Siliang Lu,
Bosch
Rafael Luna Zelaya,
University of Technology Sydney
Jiaxiang Luo,
Tsinghua University
Mohammed Makki,
University of Technology Sydney
Monalisa Malani,
N/A
Bob Martens,
TU Wien
Iasef Md Rian,
Xi'an Jiaotong - Liverpool University
James Melsom,
University of Technology Sydney
Leo Meng,
University of New South Wales
Shuhan Miao,
Massachusetts Institute of Technology
Joy Mondal,
WEsearch Lab
Zayad Motlib,
University of New South Wales
Mohammadmehdi Moulaii,
Bu-Ali Sina University
Simmi Mourya,
PayPal
John Naylor,
Architectural Association School of Architecture
Mine Özkar,
Istanbul Tecnical University
Kamile Ozturk Kosencig,
Istanbul Okan University
Burak Pak,
Deakin University
James Park,
Montana State University
Marios C. Phocas,
University of Cyprus
Charles Portelli,
Perkins & Will
Hamidreza Rafizadeh,
Queensland University of Technology

Balaji Rajasekaran,
dMac Group
Rabee Reffat,
Nottingham University Ningbo China
Dagmar Reinhardt,
The University of Sydney
Jinmo Rhee,
University of Calgary
Eduardo Rico Carranza,
N/A
Gabriella Rossi,
CITA - Royal Danish Academy
Sagar Varma Samanthapudi,
N/A
Marc Aurel Schnabel,
Xi'an Jiaotong - Liverpool University
Gerhard Schubert,
Technical University of Munich
Kyle Schumann,
University of Virginia School of Architecture
Maryam Shafiei,
American University in Dubai
YuTing Sheng,
Feng Chia University
Yinyi Shi,
Tongji University
Hiroyuki Shinohara,
Chinese University of Hong Kong
Milad Showkatbakhsh,
Architectural Association School of Architecture
Daniela Silva,
ISCTE - Instituto Universitario de Lisboa
Eleni Skevaki,
Lab for Creative Computation
Aditya Soman,
TPbennett Architects
Yang Song,
Xi'an Jiaotong - Liverpool University
Hadas Sopher,
Ariel University
Stefania Stellacci,
Information Sciences, Technologies and Architecture Research Centre - Iscte
Rudi Stouffs,
National University of Singapore
Woongki Sung,
Massachusetts Institute of Technology
Ying Yi Tan,
Singapore University of Technology and Design
Chuheng Tan,
University of Cambridge
Teng Teng,
University Of Pennsylvania
Helena Theodoropoulou,
Hellenic Open University
Faruk Can Unal,
Yeditepe University
Sepehr Vaez Afshar,
University of Illinois Urbana-Champaign
Jeroen van Ameijde,
The Chinese University of Hong Kong
Marco Vanucci,
N/A
Likai Wang,
Xi'an Jiaotong - Liverpool University
Sining Wang,
Soochow University
Jiaqi Wang,
The University of Hong Kong
Yuchen Wang,
Tongji University
Xiao Wang,
Southeast University
Shih-Yuan Wang,
National Yang Ming Chiao Tung University
Daxu Wei,
University College London

Albert Wiltsche,
Graz University of Technology
Yen-Liang Wu,
National Taichung University of Science and Technology
Kaicong Wu,
University of Arkansas
Yuchen Xie,
The University of Hong Kong
Weishun Xu,
Zhejiang University
Yuanyuan Xu,
Tongji University
Qianbin Xu,
University of Queensland
Hainan Yan,
N/A
Yang Yang,
Xi'an Jiaotong - Liverpool University
Xuyou Yang,
Singapore University of Technology and Design
Sevil Yazici,
Istanbul Technical University
Didem Gunes Yilmaz,
İstanbul Bilgi University
Boyuan Yu,
The Chinese University of Hong Kong
Hasin Ahmad Zahin,
Midstream and Terminal Services
Baizhou Zhang,
Southeast University
Yicheng Zhang,
Tianjin University
Yuqing Zhang,
South China University of Technology
Yue Ying Zhang,
Southeast University
Wei Zhang,
Huazhong Agricultural University
Yafei Zhang,
Kohn Pedersen Fox (KPF) / Kean University
Jiangyang Zhao,
Xi'an Jiaotong - Liverpool University
Mingming Zhao,
The Chinese Univerisity of Hong Kong
Hanbing Zhao,
The Chinese University of Hong Kong
Hao Zheng,
City University of Hong Kong
Guanqi Zhu,
N/A
Junling Zhuang,
Columbia University

Keynote Speakers

'The Rise of Agentic and Transcalar Architectures'

Jeffrey Huang
Professor and Director, Media x Design Lab, EPFL Switzerland

Jeffrey Huang is Director of the Media x Design Laboratory and Full Professor of Architecture and Computer Science at EPFL, with joint appointments in the Faculties of Computer and Communication Sciences (IC) and Architecture, Civil and Environmental Engineering (ENAC). He received his diploma from ETH Zurich and earned both his Master's and Doctoral degrees from Harvard University, where he was awarded the Gerald McCue Medal for academic excellence. He began his academic career at MIT's Sloan School of Management and later taught at the Harvard Graduate School of Design before joining EPFL in 2006. He has held visiting positions at Tsinghua University and Stanford's d.school, and was a Berkman Fellow at Harvard. From 2014 to 2017, he served as Founding Dean at the Singapore University of Technology and Design (SUTD). His research explores the convergence of architecture, artificial intelligence, and urban systems, advancing new forms of design intelligence across physical and digital environments.

'From CAADRIA through Peter Eisenman to Frank Gehry: Ongoing Projects of NCTU (NYCU)'

Yu-Tung Liu
Chief Advisor of Architecture and Art, China Medical University and Hospital

Yu-Tung Liu received his Doctor of Design from Harvard University in association with MIT. He is the founding professor of the Graduate Institute of Architecture at National Chiao Tung University and former Vice President of Asia University, and currently serves as Chief Advisor of Architecture and Art at China Medical University and Hospital. He has held visiting positions and delivered lectures at leading institutions worldwide, including Harvard, Cambridge, Princeton, Yale, Berkeley, Roma La Sapienza, Oslo, Technion, and Beijing. Liu served as Conference Chair of CAADRIA in 1997 and its President in 2002-2004. He has published eight books with Birkhäuser and numerous articles in ISI-indexed journals, and received the inaugural Ivan Petrovic Prize from eCAADe. His work has been exhibited internationally, including at the Venice Biennale. He has also led major initiatives such as the Next-Gene project, bringing together global and Taiwanese architects and gaining international recognition.

'Biological Intelligence (BI): The Origin of Architectural Creativity'

Ying-Chao Kuo
Principal Architect, Bio-architecture Formosana (BaF)

Ying-Chao Kuo is Principal Architect of Bio-architecture Formosana (BaF), a Taipei-based firm recognized as one of Taiwan's leading architectural practices. BaF's portfolio spans cultural, educational, transportation, hospitality, workplace, and housing projects across both public and private sectors. His notable work, the Bei-Tou Library (2006), is Taiwan's first certified green building and has become one of the most widely published examples of sustainable architecture in the country; it was also selected by Flavorwire as one of the world's 25 most beautiful libraries. Kuo's work focuses on sustainability and circular design, integrating a biological-intelligence (BI) design approach that informs the firm's philosophy. He has received numerous awards for his contributions to sustainable architecture, including the Outstanding Architect Award of Taiwan in 2011. In addition to practice, he has contributed to architectural education as an associate professor at several universities.

'Humanly Possible Worlds'

Andrew Witt
Kavita and Krishna Bharat Professor of Design and AI, Washington University in St. Louis

Andrew Witt is a designer working across digital and physical domains, trained in both architecture and mathematics, with a focus on rigorous, technically synthetic approaches to form. He is the inaugural Kavita and Krishna Bharat Professor of Design and AI at Washington University in St. Louis, where he researches the relationship between computation, artificial intelligence, and design culture. He is co-founder of Certain Measures, a design studio integrating data-driven and scalable approaches to spatial problems; its work is held in the permanent collection of the Centre Pompidou and exhibited internationally, including at the Venice Biennale and the Barbican Centre. Witt is the author of Formulations: Architecture, Mathematics, Culture (MIT Press) and the forthcoming The Technological Home. He previously taught at Harvard GSD, where he co-directed the Master in Design Engineering program, advancing transdisciplinary approaches to systems design.

TABLE OF CONTENTS

Heritage Reconstruction and Documentation **9**

Reviving Icelandic Vernacular: Computational Optimization of Raw Basalt Stone Arches 11
Nof Nathansohn, Keith J. Lee, Skylar Tibbits, Caitlin Mueller

Detecting Traditional Chinese Window Patterns in Suzhou Gardens Using YOLOv8: A Computer Vision Approach to Architectural Heritage Recognition 21
Jiangyang Zhao, Jingyang Liu, Davide Lombardi

Fine-Tuned Flux Kontext Model for Aerial Image Enhancement and 3D Reconstruction Improvement 31
Yang Yu, Yu Qian Liu, Ding Wen Bao

Human-In-The-Loop As-Built Structure Analysis Through Motion Capture and Virtual Reality 41
Yuxuan Guo, Zizheng Yu, Xinyue Xia, Miaoying Cao, Sining Wang

Computational Manifestation of the Sri Yantra: Bridging Hindu Sacred Symbolism and Contemporary Architectural Design 51
Chethana M Jangid, Pradipta Roy Choudhury

Visual Impact Assessment of High-Rise Buildings Surrounding Cultural Heritage Sites in the Historic Urban District of Hanzhong 61
Yuan Liu

Low-Cost Mobile LiDAR and RTK GNSS Integration for Georeferenced HBIM of Cultural Heritage: A Scan-to-BIM Workflow Based on Multi-Sensor Data Fusion for Accuracy Enhancement 71
Handan Aş Çemrek, Ümit Işıkdağ

Perception-Driven 3D Printing-A Study on Adaptive Toolpath Generation for Complex Surfaces 81
Yujiao Li, Hao Wu, Jiapan Sun, Philip F Yuan

XR and Spatial Cognition **91**

Simulating Interactive Architecture: A Behaviour-Based Interactive Design Platform 93
Chao Yan, Te Li, Ruyi Yang, Ziyi Chen, Yanzi Jin, Jiaxin Yao, Wenqin Cai, Aize Han, Yuxuan Deng

Streets as Shared Canvases: An In-Situ Mixed Reality Platform for Citizen-Led Urban Restorative Design 103
Yuhao He, Tian Tian Lo, Fenggui Rao

From Automation to Interaction: A Representation Framework for Integrating Occupant Control in Intelligent Environments 113
Ilayda Memis, Yuval Kahlon, Frank Petzold, Davide Schaumann, Christiane M Herr

Computing on the Physical Model: Interactive Visual Feedback of Quantitative Criteria in Design through Physical Models 123
Artemis Pantopoulou, Dimitris Papanikolaou

Reimagining Injustice Heritage through Virtual Reality: Designing Immersive Experiences of Memory and Space 133
Zi-Ru Chen, Kai-Hsiang Liang, Yu-Jie Liu, Wen-Xin Chen, Yu-Tzu Tseng

Perceptual Environments for Regulating Interaction Balance in Cross-Disciplinary Teams 143
Jia-Rong Li, Teng-Wen Chang

From Interaction to Synergy: Human–AI–Robot Collaborative Design in Immersive Mixed-Reality Environments 153
Zhiyong Dong, Peter Búš, Qi Wang, Maowei Jiang

A Vision-Perception-Based Adaptive Building Façade: Research and Prototype Validation 163
Di Zhang, Yunsong Han, Yuxiao Wang, Feixue Chen

Embodied Navigation and HCI **173**

Interactive Multi-Modal Sensory Open-Plan Workspaces: A Methodological Pilot Study in Virtual Open-Plan Offices 175
Yuqing Du, Anastasia Globa, Thomas Parkinson

Merging Graph Theory and VR: Human-Centred Assessment of Fire Evacuation Efficiency 185
Yixuan Wang, Marc Aurel Schnabel, Rong Fu, Cheng Zhang

Gamified Urban Design in Virtual Reality: Embodied Play as a Mode of Creative Participation 195
Mengdie Wang

Pedagogical Strategy Combining VR and Traditional Drawings to Depen Architectural Case Study Understanding: Merging Virtual Immersion and Symbolic Encoding: Validating the Synergistic Effect of Mixed Media via Dual Coding Theory 205
Kai-Hsiang Liang

Effects of the Biophilic Design Density of Transitional Spaces on Human Stress Recovery in a Virtual Reality Environment 215
Zhixian Li, Sisi Zlatanova

Rhizomatic Learning in Architectural Education: Integrating Humanistic Computation and Cultural Contexts through Case-Based Inquiry 225
Anetta Kepczynska-Walczak, Bartosz Marek Walczak

Evaluating a Simulation Game for Energy-Efficient Design: Computational Thinking and Usability Outcomes in a University Course 235
Jayedi Aman, Delaney O'Brien, Katey Lam, Laura B Cole, Jong Bum Kim, Laura Zangori, Anindita Laz Banti, Pegah Sepehri

Transfer Learning for Cross-Context Climate Action Insights, From Digital Citizen Participation to Climate-Aware Decision-Making 245
Cem Ataman

Computational Sensory Environments **255**

Comparing Photo-Textured LOD2 and LOD3-Equivalent Models in Urban VR: A Preliminary Study on Gaze Behavior and Spatial Impressions 257
Kosuke Kobayashi, Masatoyo Ogasawara

Linking Spatial Preference to Egocentric Information Availability: An Information Scape Approach to User-Centered Design 267
Ye Yint Aung, Shen-Guan Shih, Lucky Shin-Jyun Tsaih, Kevin Harsono

Sketching Beyond The Page: Hybrid Design Workflows Using Augmented Reality 277
Alexandros Konstantinou, Georgios-Marios Bokos, Dimitris Papanikolaou

An Embodied Interaction Framework for Analytical Engagement in Virtual Architectural Exhibitions 287
Yueying Zhang, Tian Tian Lo

Phygital Gamification for Co-creation Environmental Education: An XR-Based System for Urban Waste Cognition and Collaboration in Smart Cities 297
Yuhao Li, Tian Tian Lo, Zilu Wang, Usman Khalid Rafique, Xi Chen

Neuro-Adaptive Lighting Architecture for Emotional State Modulation: Real-Time Facial Expression-Driven Closed-Loop System with EEG Validation 307
Tiansu Wang, Tianyi Bo, Zheng Hou, Chen Hong, Jing Huang, Donghui Dai

Examining Critical Factors of Nighttime Lighting Design for Cultural Heritage Sites using Digital Twins 317
Ting Lu Chiang, Hao-Yun Chi, Tsung-Hsien Wang

Occupant-Driven Building Performance **327**

Designing for Biodiversity in the Anthropocene: Multispecies Building Envelopes as Ecological Infrastructure 329
Narkis Joffe, Oren Agay-Shay, Lilu Golan, Anastasia Slavina, Yasha J Grobman

Parallel Optimization for Competing Design Scheme Exploration: A Competitive Island-Based Evolutionary Algorithm for Early-Stage Design Exploration 339
Likai Wang

Integrated Acoustic Geometry: Towards Multi-Objective Design and Robotic Fabrication of Acoustic Wall Systems 349
Jingyang Liu, Jiangyang Zhao, Yang Song

From Behavior to Office Design Insights: A Framework for Modeling Attractiveness Using Map-matched Trajectories and 3D Tag Information 357
Hiroyuki Hasada, Junya Maruyama, Ayumi Mukai, Hiroki Imai, Ryohei Takahashi, Tomoki Akiyama, Jingya Li, Hiroaki Goto, Shiro Takeshita, Yudai Honma

From Optimisation to Nudging: An Occupant-Centric Simulation Framework for Identifying Energy-Aware Shading Behaviour 367
Bohan Wang, Robert W. Lindeman, Peter Thompson, Farhang Tahmasebi, Yihan Huang, Heide K. Lukosch

Regenerative Cities: A Data-Driven Generative Framework for Adaptive Interventions within Heritage Urban Contexts 377
Rutuja Rode, Sonali Rane, Abhijeet Manjunath, Milad Showkatbakhsh

Surrogate-Based Optimization for Climate-Responsive Design of Thermally Activated Building Envelopes 387
Tianxin Chang, Zhi Zhuang

Multi-Agent Large Language Model Framework for Window Optimization: Bridging human design intent and data-driven environmental performance 397
Reza Taghavifard, Negin Cheraghi

Climate-Embedded Building Systems **407**

A Co-Evolutionary Approach to Multi-Objective Optimization in Architecture 409
Jonathan Wong, Puja Bhagat, Milad Showkatbakhsh, Mohammed Makki

A Space Habitat: Generative Urban Systems 419
Ling Kit Cheung, Mohammed Makki, Milad Showkatbakhsh

Optimizing Multifunctional Space Use in Mission-Critical Facilities: A Multi-Objective Reinforcement Learning Approach 429
Eduard Haiman, Davide Schaumann

The Multi-objective Optimization of Breeze Blocks: Re-thinking modern architecture design feature in a performance-oriented method 439
Jiajian Chen, Hiroyuki Shinohara

Enhancing Urban Ecosystem Services: A Computational Approach to Multispecies Habitat Distribution on Building Envelopes 449
Anastasia Slavina, Surayyn Selvan, Oren Shay, Narkis Joffe, Lilu Golan, Yasha Grobman

Climate-Aware Adaptive Regeneration of Ageing Housing Stock: An Agent-Based Framework for Residential Layout Design and Optimisation 459
Zhang Yin, Frederick Chando Kim

Optimising Pareto Front Selection Through the Use of Artificial Intelligence: A Generative Revitalisation of Rabdan's Pedestrian Landscape 469
James Gomez, Josef Sfara, Mohammed Makki

Urban Cohabitation within Mangrove Ecosystems through Evolutionary Computation: The Case Study of Abu Dhabi 479
Laiba Jamil, Nancy Sun, Mohammed Makki, Apostolos Kyriazis

Adaptive Environmental Control **489**

Reclaiming Agency in Building Performance Prediction: Enhancing Occupant-Centered Simulation Accuracy through Data Assimilation and Explainable Machine Learning 491
Yuxin Zhou, Yudai Tanino, Tomohiro Fukuda, Minoru Etoh

Digital Twins for Urban Public Spaces: a Virtual Reality-based Method for Analysing Micro-scale Environment-Behaviour Relationships 501
Wenting Zhang, Sifan Cheng, Jeroen van Ameijde

Data – Driven Prediction Framework for Natural Ventilation Performance: A Case Study of Vietnamese Modern Tube Houses 511
Trong Tin Tran, Yen Khang Nguyen Tran

From Sensing to Control: loT for Sleep-Aware Architecture 521
Jing Zhong, Ziyang Han, Peilin Li, Jun Yin, Yuxin Zhang, Shuai Lu

Varanasi Heritage Infill: Specification of a Loam-Based Modular Kit-of-Parts 531
Abhijeet Manjunath, Rutuja Rode, Sonali Rane, Milad Showkatbakhsh

Computational Urban Walkability **541**

Integrating Perceptual and Computational Frameworks for Walkability Assessment in Campus Environments Using AI-Based Models 543
Sena Kaynarkaya, Aslı Çekmiş, İsmail Çetin, Yusuf Hüsayin Şahin, Gözde Ünal

Prompts to Perception: A Promptable Segmentation Approach to Urban Transparency and Safety 553
Zheng Fang, Xuechen Li

Reasoning with Generative Vision-Language Models to Enhance Low-quality Street-view Imagery for More Accurate and Interpretable Urban Building Analysis 563
Haoyang He, Shuyang Li, Wanyu Pei, Rudi Stouffs, Yu Ye

Human-Informed Urban Vitality: Integrating Form, perception, and design via Heterogeneous Graph Learning 573
Muxin Jia, Kaiheng Zhang, Taro Narahara

Research on Urban Waste Classification Optimization Integrating Spatial Analysis and Behavioral Intervention: A case study of Kowloon City, Hong Kong 583
Yuanxi Zhang, Tian Tian Lo, Yi Jian

From Evaluation to Design: Vector-Similarity Search for Analogue-Based Land-Use Design Exploration 593
Kiarash Kiany, Jinmo Rhee

Predicting Pedestrian Flow Spatio-temporal Heterogeneity via Conditional GANs: Focusing on Urban Street Market Dynamics 603
Yujing Wu, Yuhan Yao, Waishan Qiu, Yuankai Wang, Jie Shen

Who's Strolling Under the Sun: Study on Adaptive Behaviors in Urban Public Spaces Under High-Temperature Exposure 613
Yujing Wu, Shiqi Dai, Zanwen Ma, Menghan Wang, Xintong Lyu, Waishan Qiu, Chongyang Ren

Agent-Based Urban Simulation **623**

From Pixels to Predicates: Structuring urban perception with scene graphs 625
Yunlong Liu, Shuyang Li, Pengyuan Liu, Yu Zhang, Rudi Stouffs

Gene-up: A Somatic Neural Model for Rebalancing Urban Entropy 635
Vania Arlene, Tasos Varoudis, Provides Ng

A Case for Socially-Driven Pedestrian Simulations in Urban Environments 645
Puja Bhagat, Jonathan Wong, Milad Showkatbakhsh

Moonflower: A Multi-Agent Grasshopper Component for Human Behaviour and Furniture Relocation under Dynamic Solar-Shade Conditions 655
Ayşe Ebrar Kavak, Beyza Gül Topçu, Sema Alaçam, Orkan Zeynel Güzelci

Root-Inspired Generative Network: An Agent-Based Model for the Design of Context-Aware Architectural Structures 665
Kwong Lau Hung, Toni Kotnik

Beyond Traditional Metrics: A High-Fidelity Latent-Representation Framework for Urban Morphological Typology Refinement and Quantification 675
Yu Meng, Xinru Wang, Chengyu Sun

Exploring Pedestrian-Tactile Paving Interactions in High-Density Urban Spaces: A Computer Vision and Spatial Analytics Approach 685
Jiqian Huang, Yean Yang, Xiaoxi Guo, Shiliang Wang, Waishan Qiu, Xintong Lyu, Chongyang Ren

Revealing Post-Pandemic Social Interaction Shifts in Urban Squares: A Deep Learning-Based Dynamic Monitoring Framework 695
Shuyu Huang, Nanshuo Mu, Shuyao Ren, Qi Qin, Xintong Lyu, Waishan Qiu, Chongyang Ren

Urban Morphology and Regeneration **705**

Just Emotional Geographies: Intersecting Decolonial and Post-Gender Design in Garden Cities with AI Audits 707
Yuxuan Liu, Trishla Chadha, Noor Mubesher, Aya Meskawi, Provides Ng

A Geometry-Based Computational Assessment of Parcel-Level Redevelopment Feasibility in Seoul 717
Sanghoon Byun, Dongjoon Park, Bumjoon Kang

Between Numbers and Narratives: Integrating Qualitative and Quantitative Methods for Typological Analysis of Dutch Cross Buildings 727
Jess Chang, Casper van Engelenburg, Marija Mateljan, Seyran Khademi, Georg Vrachliotis

Materialising Latent Hybridity: Interpolating the Natural and the Artificial for Post-Digital Urban Futures 737
Sheng-Yang Huang, Enriqueta Llabres-Valls, Xiao Wang, Yongchun Wang, Hengyu Gong, Junyi Wu, Yanlin Bao

Urban Blocks Regionalization By Integrating Spatial Proximity And Morphological Similarity 747
Ruize Wu, Jiachen Bu, Ziyu Tong

From Cells to Strategies Game: Learning Block Layouts through Policy Networks and Game-Inspired Rewards 757
Han Xie, Mark Balzar, Hans Schartner, Zeynep Aksöz Balzar, Linxue Li

Colour Representations of Urban Space 769
Mohammed Makki, Rafael Luna Zelaya, Linda Matthews, Ling Kit Cheung, Nimish Biloria, Jordan Mathers, James Melsom

Application-Oriented Automatic Identification of Urban Fabric: A Graph-Theoretic Approach to Morphological Unit Delineation 779
Xinru Wang, Yu Meng, Chengyu Sun

Heritage Reconstruction and Documentation

Reviving Icelandic Vernacular

Computational Optimization of Raw Basalt Stone Arches

Nof Nathansohn[1], Keith J. Lee[2], Skylar Tibbits[3] and Caitlin Mueller[4]
[1,2,3,4]*Massachusetts Institute of Technology*
[1]*nof@mit.edu, 0000-0003-0496-0554*

Abstract. This research examines computational optimization techniques for revitalizing traditional Icelandic stone construction in sustainable contemporary architecture. Iceland's harsh climatic conditions and unique geological context have historically shaped sophisticated vernacular architecture, particularly the iconic turf house. However, modern construction increasingly favours steel and glass structures, risking the loss of indigenous architectural knowledge. This study develops an integrated computational framework for designing compression-only stone arches using raw basalt rocks. The methodology encompasses the digital scanning of 43 irregular basalt stones, structural optimization using the Force Density Method, and physical validation through a prototype that demonstrates compression stability. Results reveal two distinct optimization approaches - force-driven and material geometry-driven - with the latter aligning more closely with traditional material-aware design principles. The successful physical prototype validates that computational tools can preserve and extend indigenous architectural knowledge, positioning digital methodologies as enablers of sustainable, locally adapted architecture.

Keywords. Computational Optimization, Vernacular Architecture, Compression Structures, Digital Fabrication, Material Circularity, Force Density Method, Sustainable Construction, Basalt Stone

1. Introduction

Iceland emerges from the North Atlantic Ocean as a geological marvel, situated at the convergence of the North American and Eurasian tectonic plates. This dynamic geological canvas has always demanded architectural practices that directly negotiate with extreme environmental conditions (Sigmundsson, 2006). With over 130 volcanic mountains and approximately 30 active volcanic systems, Iceland's landscape is continually reshaped by geological processes, requiring building practices that can withstand sudden geological transformations.

The absence of extensive forest resources fundamentally shaped Icelandic architectural practices. Unlike contexts where timber serves as a primary construction

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 11-20.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

material, Iceland's vegetation is characterized by low-lying shrubs, mosses, and limited woodland areas. This ecological limitation compelled early settlers to develop sophisticated building techniques using alternative materials, resulting in the iconic turf house - a comprehensive survival strategy developed over centuries of environmental interaction (Van Hoof & Van Dijken, 2008). Traditional Icelandic turf houses utilized locally available materials - primarily stones, earth, and turf - to create dwellings that withstand extreme winds, temperature fluctuations, and limited resources. These



Figure 1. Reconstructed farmstead in Iceland. Source: Wikimedia Creative Commons

structures achieved exceptional thermal insulation through thick layers of turf and strategically placed stones (figure 1). The walls, often exceeding one meter in thickness, provided protection against harsh winds while maintaining interior temperatures through sophisticated passive heating (Zoëga et al., 2023). The stone components worked entirely in compression, a principle allowing irregular rocks to perform structurally without requiring precise cutting or shaping.

Contemporary architectural practices in Iceland now represent a complex negotiation between global technological trends and local ecological imperatives. Modern steel, glass, and concrete structures frequently contrast with traditional building approaches, raising critical questions about architectural identity, sustainability, and cultural continuity. This research examines how computational design methodologies can revive traditional stone construction practices, positioning digital tools not as a departure from tradition, but as enablers of material-aware, climate-responsive architecture (Figure 2).

1.1. A NEW PARADIGM FOR MATERIAL CIRCULARITY

The architectural industry stands at a critical juncture where environmental imperatives intersect with technological innovation. The buildings and construction sector accounted for 36% of final energy use and 39% of energy and process-related CO_2 emissions in 2018 (UNEP & GlobalABC, 2019).

Computational methodologies are particularly significant in addressing material circularity and sustainable construction. Recent research demonstrates innovative approaches to incorporating irregular and reclaimed materials through computational design. Alkhayat et al. (2024) present a methodology for reusing concrete rubble from demolition sites, developing algorithmic tools to support the use of irregular rubble-based inventories as structural components. Their "rubble trusses" exemplify how

digital workflows can transform waste material into sophisticated structural elements through scanning technologies, parametric design, and structural optimization.

Wibranek and Tessmann (2019) demonstrate this through their work on "Digital Rubble," where they combine photogrammetry, 3D printing, and mixed reality to create compression-only structures from unprocessed rock. Van Mele et al. (2016) illustrate comprehensive structural design strategies in their work on the Armadillo Vault, employing Thrust Network Analysis to design a freeform, unreinforced, cut-stone vault that achieves a thickness-to-span ratio comparable to that of an eggshell, covering 75 square meters.



Figure 2. Scaled model of a basalt mono-material arch in compression

The convergence of digital scanning technologies and computational design tools has created new possibilities for working with irregular materials, providing pathways to significantly reduce the environmental impact of construction by transforming materials previously considered too complex into viable building components.

1.2. RESEARCH MOTIVATION

The pressing need to address sustainable construction practices in Iceland emerges from environmental, cultural, and technological factors. As contemporary construction increasingly adopts structures that neglect local sustainable principles, there is an urgent need to reconnect modern architectural practices with ecological wisdom embedded in vernacular building techniques. Iceland's unique geological context provides an abundance of basalt, a locally available material that historically proved resilient in traditional construction. By developing computational methods to optimize raw basalt rocks, this research aims to revitalize a readily available, low-carbon material choice perfectly adapted to local conditions.

However, the sophisticated understanding of structural performance demonstrated in traditional turf houses - particularly stone foundations and walls creating stable compression structures - risks being lost as contemporary construction methods prevail. There exists a significant gap in current research regarding the optimization of

irregular stone structures for extreme climatic conditions. While recent studies have demonstrated the potential of computational tools for irregular materials (Alkhayat et al., 2024) and the possibilities of stone vault construction (Van Mele et al., 2016), few have specifically addressed optimizing raw stone structures for the demands of a subarctic environment. The development of a computational methodology for designing compression-only stone arches using raw basalt represents a critical step toward more sustainable construction practices in Iceland. This approach addresses the environmental impact of construction while offering a pathway to preserve and evolve traditional building techniques through modern technological means.

2. Methodology

The research methodology integrated digital scanning, computational form-finding, and physical validation through a comprehensive workflow developed in three interconnected phases. Each phase builds upon the previous to create a unified framework for optimizing raw basalt stones into stable compression structures. The overarching strategy employs a hanging chain model: rocks are threaded with cable to form an inverted arch that, when flipped, creates a pure compression structure - a principle derived directly from traditional masonry construction methods.

2.1. DIGITAL INVENTORY CREATION

A digital database of 43 raw basalt rocks was established using Polycam software and LiDAR scanning technology (Figure 3). The geometric properties of each rock were analysed in Grasshopper using bounding box analysis to determine length and width parameters. Following Alkhayat et al. (2024), two orientation strategies were tested: a long orientation that maximizes material efficiency but increases fabrication complexity, and a flat orientation that requires more rocks but simplifies the drilling process (Figure 4). For structural analysis, the cross-sectional area of each rock was calculated by contouring the mesh and using the smaller area length. This approach ensures conservative structural estimates as demonstrated in Alkhayat et al.'s research with irregular concrete elements. This digital inventory serves as the foundation for subsequent optimization phases, ensuring computational form-finding works with actual material constraints rather than idealized geometries.

2.2. COMPUTATIONAL FORM-FINDING AND OPTIMIZATION

The structural form-finding process employed the Force Density Method (FDM), a computational approach developed by Linkwitz (2014) for analysing and optimizing structural networks. FDM introduces force densities (q) as the ratio of force (F) to length (l) in each element: $q = F/l$. This leads to a linearized system of equilibrium equations determining the equilibrium equations for each node i:

$$(x_1 - x_0)/l_1 \cdot F_1 + \ldots + (x_4 - x_0)/l_4 \cdot F_4 + p_x = 0$$

The Python implementation in Grasshopper used in this research performed load-point calculations iteratively across network nodes.

Two distinct optimization strategies were implemented to explore different approaches to integrating material properties with structural form.



Figure 3. A basalt rock that was scanned for this research (left) and the digital inventory (right).

2.2.1. Random Forces as Shape Generator

The first strategy used FDM with random force distributions, with the average rock length in the inventory determining the number of segments. Optimization was performed using Galapagos for Grasshopper, with force densities as variables and structural performance as fitness criteria. The matching between optimized form and inventory utilized the Digital Circularity plugin (Lee, 2024), which efficiently paired rock geometries with structural requirements (Figure 5 and 6).

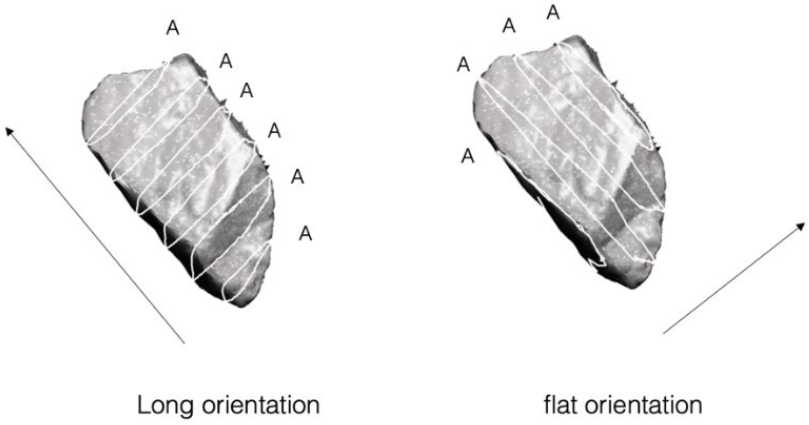


Figure 4. Rock orientation for fabrication and structural analysis.

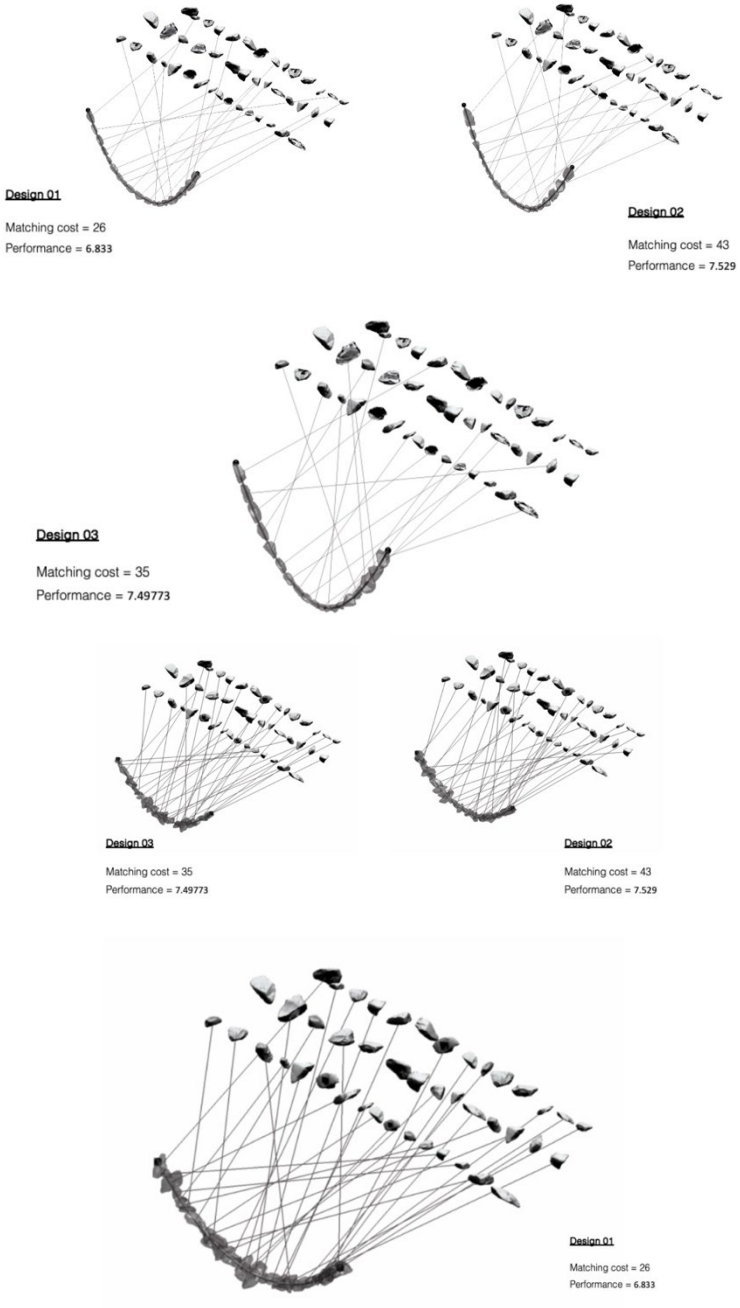


Figure 5, Figure 6. Three designs of long (top) and flat (bottom) orientations are matched using FDM and optimized by Galapagos.

2.2.2. Form Finding Using Inventory Forces

The second strategy derives force distributions directly from rock properties, allowing the rocks' mass and geometry to determine arch configuration. This approach more directly engages with traditional material-aware design principles, where inherent characteristics of available materials guide structural form. Multiple iterations with random rock ordering were tested, and Galapagos was employed to optimize matching costs (Figure 7).

This material-driven approach proved particularly successful in creating forms that naturally accommodate basalt inventory irregularities while maintaining structural integrity, reflecting traditional builders' intuitive understanding of how stones work together in compression.

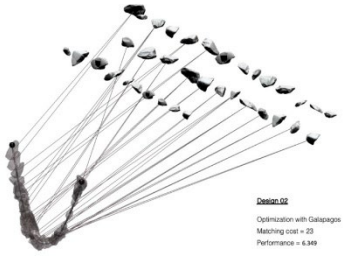


Figure 7. Rocks form finding by matching order.

2.3. PHYSICAL VALIDATION AND ASSEMBLY

To validate the computational framework, a physical prototype was constructed following the optimized design. The scanned basalt rocks were drilled according to computational specifications, and a steel cable was threaded through them to create the hanging chain assembly (Figure 8). Between the stones, basalt-felted fibres were placed to fill gaps caused by surface irregularities, ensuring complete contact and proper compression distribution between the stones. This intermediate tensile state allowed precise positioning of each rock according to optimized geometry.



Figure 8. The hanging chain assembly with threaded basalt rocks is positioned according to computational optimization.

Once the desired geometry was achieved and verified against the computational model, the assembly was carefully inverted, and the cable tension was released. The resulting compression arch demonstrated stable structural behaviour, with each rock

locked in place by compressive forces distributed according to FDM optimization. The arch successfully supported its own weight plus additional loading, validating that the computational framework accurately predicted structural performance when working with irregular natural materials.

This physical validation confirms that the three-phase framework - digital inventory, computational optimization, and physical assembly - functions as an integrated system for reviving traditional stone construction principles through contemporary tools (Figure 9).

3. Results

The optimization results revealed distinct trade-offs between orientation strategies and optimization approaches. For the first method, testing both orientations through Galapagos revealed that long orientation achieved better matching costs but worse structural performance. In contrast, flat orientation demonstrated superior structural performance but required a larger rock inventory for effective matching.

Using the Design Space Exploration plugin for Grasshopper (Digital Structures, MIT, 2019), 100 different designs were sampled, revealing that matching costs increased exponentially with higher forces, particularly in lighter regions (Figure 10). This relationship suggested diminishing returns: as force increased, improvements in performance became incrementally smaller while matching costs rose steeply. The data



Figure 9. Detail of the arch edge and basalt fibres assembly.

Indicated optimal results occurred in the moderate-to-high force range, where performance-maintained stability without excessive matching costs.

The second method, employing rock-driven form-finding, produced notably improved results when optimized with Galapagos compared to non-optimized configurations. The optimization achieved both better matching costs and improved structural performance. However, data points at higher costs showed diminishing returns in performance gains, suggesting a practical limit to the benefits of optimization (Figure 11). Most significantly, the physical prototype validated computational predictions. Load testing confirmed the arch could support its own weight plus additional distributed loading without observable displacement or structural distress. This successful physical validation demonstrates that the computational framework accurately captures the structural behaviours of irregular stone assemblies and can reliably guide compression-only structure design using raw, local materials.

The integration of digital scanning, computational optimization, and physical

assembly proves that traditional stone construction principles can be successfully reinterpreted through contemporary tools while maintaining essential structural logic.

4. Discussion and Future Work

This research establishes a foundation for several lines of investigation. The computational framework should be expanded to incorporate Iceland's specific environmental challenges, including wind-load analysis for extreme conditions, thermal performance optimization, seismic considerations given Iceland's volcanic activity, and multi-objective optimization that simultaneously considers structural and environmental performance.

Physical prototyping at larger architectural scales represents an essential next step. Full-scale testing would enable investigation of various assembly methods and long-term structural performance. Development of efficient construction workflows for larger structures would demonstrate the practical viability of this approach for contemporary building projects.

Building upon this research, the integration of soil-based 3D printing with embedded seeds presents an intriguing future direction. This technology could provide an innovative assembly method for gaps between rocks, creating living connectors that parallel traditional Icelandic turf houses' use of vegetation for structural reinforcement. Such an approach would represent a contemporary interpretation of vernacular building practices, where living materials strengthen structures over time through root growth (Nathansohn et al., 2024). However, this biomimetic extension warrants focused investigation as a separate study to fully explore its potential for bridging traditional and contemporary construction methods.

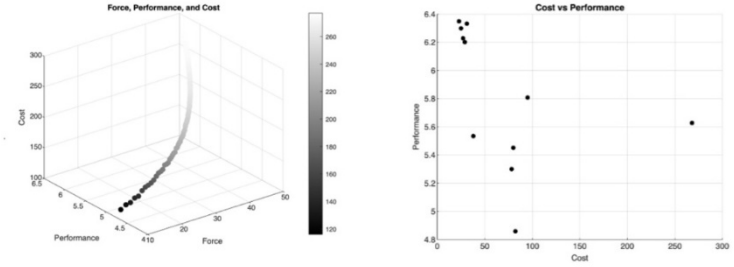


Figure 10, Figure 11. Flat orientation sampling (left), inventory forces method analysis (right)

5. Conclusion

This research demonstrates how computational optimization can revive traditional Icelandic stone construction practices for contemporary building needs. The integrated framework developed here - encompassing digital scanning, computational form-finding, and physical validation - demonstrates that irregular local materials can be successfully optimized to create stable compression structures while preserving the essential ecological wisdom of vernacular techniques. The comparative analysis of force-driven versus geometry-driven optimization revealed essential insights about

working with irregular natural materials. The material-driven approach demonstrated particular promise in its ability to let rock properties guide form generation, directly echoing material-aware design principles of traditional Icelandic construction.

Most significantly, this work establishes a methodological framework for preserving indigenous architectural knowledge through computational means. Rather than seeing digital tools as a departure from tradition, they are positioned as enablers of material-aware, climate-responsive architecture. As the construction industry faces increasing pressure to reduce its environmental impact, approaches that combine local materials, traditional wisdom, and advanced computational tools will become increasingly valuable. This research suggests that the future of sustainable architecture in Iceland lies not in choosing between traditional and contemporary approaches, but in their thoughtful synthesis.

Acknowledgments

This research commenced as part of a project at the Massachusetts Institute of Technology in the fall of 2024, under the supervision of Caitlin Mueller, Skylar Tibbits, and Keith J. Lee. Many thanks to Awen Gronda for scanning the rocks and to Ari Otoo for helping with the fabrication. The digital tools used include Polycam for scanning, Karamba 3D (Preisinger 2013) and Digital Circularity Toolkit (Lee 2024) for inventory analysis.

References

Alkhayat, L., Lee, K. J., & Mueller, C. (2024). The Crown Jewels: Algorithmic design of rubble trusses for the Arabian Gulf. In ACADIA 2024 (pp. 2-13).

Lee, K. J. (2024). Digitalcircularitytoolkit. Zenodo. https://doi.org/10.5281/zenodo.10701234

Linkwitz, K. (2014). Force Density Method. In S. Adriaenssens, P. Block, D. Veenendaal, & C. Williams (Eds.), Shell structures for architecture: Form finding and optimization (pp. 59-70). Routledge.

Nathansohn, N., Gillis, E., Linker, G., Zermati, N., Segal, A., Freedman, D., Partook, R. H., & Naim, O. (2024). To grow a building: Soil and seeds 3D-printing. 3D Printing and Additive Manufacturing. https://doi.org/10.1089/3dp.2023.0357

Sigmundsson, F. (2006). Iceland geodynamics: Crustal deformation and divergent plate tectonics. Springer.

UNEP & GlobalABC. (2019). 2019 Global status report for buildings and construction. United Nations Environment Programme.

Van Hoof, J., & Van Dijken, F. (2008). The historical turf farms of Iceland: Architecture, building technology and the indoor environment. Building and Environment, 43(6), 1023-1030. https://doi.org/10.1016/j.buildenv.2007.03.004

Van Mele, T., Mehrotra, A., Echenagucia, T. M., Frick, U., Ochsendorf, J., Dejong, M., & Block, P. (2016). Form finding and structural analysis of a freeform stone vault. In Proceedings of the IASS Annual Symposium 2016.

Wibranek, B., & Tessmann, O. (2019). Digital rubble: Compression-only structures with irregular rock and 3D printed connectors. In Proceedings of the 37th eCAADe Conference (pp. 547-554).

Zoëga, G., Sigurðardóttir, S., & Zoëga, B. (2023). Turf building in Iceland – Past, present, and future. Open Archaeology, 9(1), 20220345. https://doi.org/10.1515/opar-2022-0345

Detecting Traditional Chinese Window Patterns in Suzhou Gardens Using YOLOv8

A Computer Vision Approach to Architectural Heritage Recognition

Jiangyang Zhao[1], Jingyang Liu[2] and Davide Lombardi[3]
[1,2,3]*Department of Architecture, Design School, Xi'an Jiaotong-Liverpool University*
[1]*Jiangyang.Zhao@xjtlu.edu.cn, 0000-0002-3950-7656*
[2]*Jingyang.Liu@xjtlu.edu.cn, 0000-0002-2671-4015*
[3]*Davide.Lombardi@xjtlu.edu.cn, 0000-0001-7128-4635*

Abstract. Traditional lattice windows in Suzhou gardens, core elements of Chinese architectural heritage, face identification challenges due to intricate patterns and inefficient manual methods. This study develops a YOLOv8n-based approach for automated detection of thirteen lattice types (documented in Yuanye and Yinzhao Fayuan), using a 488-image field dataset. The model achieves mAP50 = 0.7299, effective under controlled conditions. Integrated into H-Scan software with DeepSeek LLM, it provides cultural context (symbolism, craftsmanship) and bilingual interaction, aiding students, tourists, and researchers. Validating lightweight AI for heritage recognition, this work bridges technical detection and digital education, supporting architectural knowledge preservation.

Keywords. Computer Vision, Suzhou Garden, Traditional Chinese Window, Architectural Heritage, Object Detection

1. Introduction

The lattice-windows in traditional Chinese architecture are not only practical components but also integral elements of architectural aesthetics. Window lattices are more than decorative features; they serve multiple functions, including regulating sunlight, enhancing visual appeal, and conveying cultural symbolism (Xu, 2012; Zhang, 2023). Their designs often incorporate artistic forms such as calligraphy and painting, reflecting the profound heritage of traditional Chinese culture. In classical gardens and buildings, lattice patterns frequently embody the artistic spirit of landscape painting. Through techniques such as borrowed scenery (借景) and framed views (框景), they seamlessly integrate natural and humanistic spaces, expressing the philosophical ideal of harmony between man and nature (Xu, 2012).

The classical gardens of Suzhou are exemplary models of Chinese landscape design and serve as tangible repositories of the region's cultural heritage. A defining architectural feature of these Ming and Qing dynasty gardens is their diverse repertoire

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 21-30.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

of lattice windows. These traditional forms continue to influence contemporary Chinese architecture, making their accurate identification and appropriate application a matter of significant value for students and professionals in the field.

The recognition of these lattice windows, however, presents a considerable challenge. Their intricate designs and morphological diversity complicate identification, particularly for non-specialists. Traditional methods, which rely on manual visual inspection and cross-referencing with two historical rule books, are inefficient and often fail to capture subtle formal variations (Dye, 1974). The first rule book is the Yuanye (园冶), authored by Ji Cheng and published in 1634. The second book is Yinzhao Fayuan (营造法原) (Yao Chengzu, 1986). The two books indicate that the Suzhou traditional gardens contain over 100 lattices across dozens of pattern types. Consequently, visitors and students relying on guidebooks or tours frequently struggle to appreciate the specific cultural connotations of individual patterns, leading to an oversight of this critical architectural heritage.

In recent years, deep learning-driven computer vision has emerged as a transformative tool for the automated recognition of architectural heritage. The visual AI models such as You Only Look Once （YOLO) (Redmon et al., 2016) and Graph neural Networks（GNNs) (Jiang et al., 2022) have been increasingly applied in the detection of historic architecture for their efficiency in recognising and classifying complex patterns, including the detection of traditional Chinese roofs (Hou et al., 2023) and the classification of building types in Athens (Janković, 2020) . Other AI models have also advanced image recognition for applications in heritage sites, such as religious buildings (Fesl et al., 2022), monuments (Saadat et al., 2020), and craft patterns (Girsang, 2021; Horn et al., 2022; Liu, 2020), offering valuable insights for developing deployable tools to help students, tourists, and researchers developing a deeper understanding of architectural details. Among these, YOLO is notable for its real-time, accurate object detection, making it suitable for mobile and desktop applications (Redmon et al., 2016). Despite these advances, research on the automated detection of traditional Chinese window lattices remains limited, as existing studies focus on detecting the location of windows (Du & Wang, 2024) rather than the lattice patterns.



Figure 1. A classical tall window in a Suzhou Garden (left); the locations of lattice patterns on tall windows(right).

However, existing research exhibits a critical gap: most studies on traditional Chinese architectural components focus on macro-structures (e.g., window locations, roof types) rather than the fine-grained classification of lattice patterns. For instance, some studies achieve the detection of locations of "windows" as generic components but fail to distinguish between specific patterns like ice cracks and tortoiseshell patterns (Du & Wang, 2024). Furthermore, few models account for practical application needs in garden settings—such as real-time recognition via handheld devices or batch data analysis for researchers—creating a disconnect between technology and end-user requirements. To address this gap, this study presents Heritage-Scan (H-Scan), a software based on YOLOv8n and DeepSeek (DeepSeek, 2025) for the automatic detection and interpretation of traditional window lattices. Particular emphasis is placed on the tall-window lattices (Figure 1) of Chinese classical gardens from the Ming and Qing dynasties. The paper answers the following questions:

- How effective is a YOLOv8n-based object detection model in accurately identifying diverse traditional tall-window lattice patterns in Chinese gardens?
- How can automated pattern recognition enhance access to and understanding of traditional Chinese architectural elements in digital heritage education?

2. Methodology

The methodology of this project comprises four sequential steps (Figure 2): data collection and preprocessing, YOLO model training and optimisation, and the development of interactive software. The first step involves collecting relevant photographs of Ming-Qing style tall windows from local Suzhou gardens. These images will then be used to train a YOLO model, which will subsequently undergo optimisation to enhance detection precision. Finally, an interactive software will be developed, integrating a Large Language Model (LLM) to provide contextual explanations for the elements detected by the trained YOLO model.

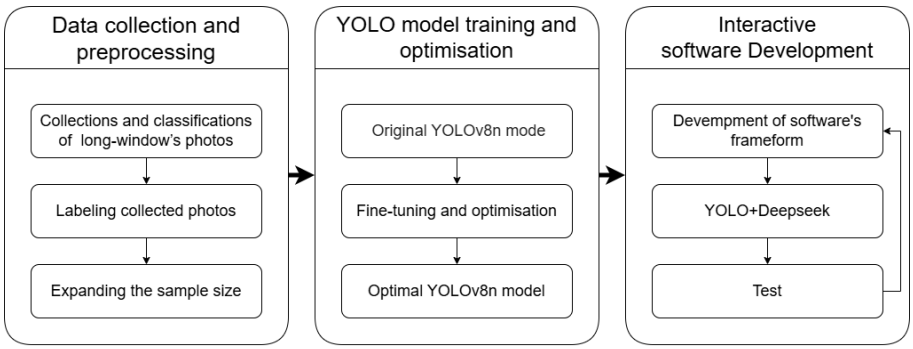


Figure 2. A flowchart illustrating the development and optimisation of the Heritage-Scan (H-Scan) Software.

2.1. DATA COLLECTION AND PREPROCESSING

The data relating to the tall windows was collected by using mobile phones to photograph windows in Zhuozhengyuan (拙政园), Liuyuan (留园), Ouyuan (耦园), and Wangshiyuan (网师园) during both the morning and afternoon. This timing was deliberately chosen to account for the influence of natural light, as variations in lighting conditions can impact detection accuracy. Specifically, if data collection were restricted to the morning, the trained model might fail to detect tall windows effectively in the afternoon due to differences in light intensity and quality. In total, 1,143 photographs were collected, of which 309 could be classified using the reference texts Yuanye (园冶) and Yinzhao Fayuan (营造法原). These 309 images were subsequently categorised into thirteen types of tall windows. This dataset was used to train a YOLO8n model. The specific names of these tall-window types are listed in Table 1.

Table 1. The detection precision for the lattice patterns of thirteen varieties of tall windows

Category ID	Window Category Name	Original data	Expand data
0	Octagonal Scene (八角景)	45	57
1	Ice Crack Pattern (冰裂纹)	6	6
2	Mixed Pattern with Glass Inlay (插角乱纹嵌玻璃)	11	25
3	Palace Style (宫式)	76	92
4	Palace Style Half Window (宫式半窗)	23	37
5	Crabapple Diamond Pattern (海棠菱角式)	48	57
6	Floral Knot with Glass Inlay (花结嵌玻璃)	24	39
7	Well Pattern with Diamond Inlay (井字嵌菱式)	3	9
8	Sunflower Style with Glass Inlay (葵式嵌玻璃)	2	5
9	Hexagonal Panoramic Pattern (六角全景纹)	18	45
10	Ruyi Diamond Flower Pattern (如意菱花式)	10	26
11	Book Strip with Swastika Pattern (书条川万字)	15	39
12	Book Strip Style (书条式)	13	23
Total		309	488

The 309 collected images were subsequently labelled using T-Rex Label—a browser-based, AI-powered auto-labelling tool developed by the IDEA computer vision team—based on the tall windows designated names. In this study, the thirteen types of windows were labelled using their Chinese names, with English translations to be finalised in future work. As shown in Figure 3, coloured bounding boxes indicate the outlines of the tall windows, with each colour corresponding to a specific window

type. The labelling results were cross-validated by two relevant researchers to ensure accuracy.



Figure 3. Two lattice patterns of tall windows within two separate images were labelled through the use of coloured bounding lines.

Given that the number of labelling images was insufficient to ensure the accuracy of the trained YOLO model, data augmentation was applied to expand the sample size to 488. The augmentation methods included geometric transformations (flipping, rotation, scaling), pixel adjustments (brightness and saturation modifications), and Gaussian noise injection. This process was conducted in a Python environment. Finally, labelling images have 488, including thirteen types of tall windows.

2.2. YOLOV8N MODEL TRAINING AND OPTIMISATION

In this research, the YOLO (You Only Look Once) visual AI model was selected for detecting the traditional tall-window type due to its excellent practical performance. Among the various YOLO models, YOLO8n was specifically chosen. Its lightweight nature, high efficiency, and relatively low requirements for computer hardware make it particularly suitable for mobile applications. The training, validation, and testing of the models were carried out on a computer platform (CP). In future work, the trained models will be deployed to mobile devices.

Regarding the hardware configuration in this study, it consists of an NVIDIA GeForce RTX 4050 Laptop GPU, an Intel Core i7-14650HX CPU, and 16GB of memory. The operating system employed is Windows 11. The programming language used is Python, version 3.11.9. The deep-learning framework utilised in this research includes PyTorch 2.0.1 and Torchvision 0.15.2. These specifications ensure the reproducibility of the experiments and provide a solid foundation for the model's development and application.

In this study, we used the default set of YOLOv8n to train the data, and mAP50 is 0.522. To evaluate the quality of a YOLO model, key metrics such as mAP are typically considered. The mean Average Precision (mAP50) is a normalised score ranging from 0 to 1, used to assess the overall detection performance of the model. A

higher mAP50 value indicates greater detection accuracy and a lower rate of missed targets. To advance the detection precious, we fine-tuned the YOLOv8n model (a state-of-the-art object detection neural network) on our custom dataset for tall-window detection. To ensure effective and efficient learning, the training process was configured with several key parameters: the model was set to be trained for 100 epochs, meaning it processed the entire dataset 100 times to thoroughly learn the relevant features. We used a batch size of 16, which determines how many images are processed together in one step, striking a balance between learning stability and computational memory usage. The images were resized to 768x768 pixels to provide the model with sufficient detail for accurate detection. The learning process was carefully controlled with a low initial learning rate of 0.00005, which defines the step size during learning to prevent overshooting the optimal solution, and the AdamW optimiser was used to intelligently adjust this rate during training. To improve the model's ability to generalise to new, unseen data, data augmentation was applied, artificially creating variations of the training images (e.g., through rotation or scaling). Furthermore, mixed-precision training was enabled to speed up the process without sacrificing accuracy, and an early stopping patience of 30 epochs was set to automatically halt training if no improvement was observed, thereby preventing overfitting and conserving computational resources. This combination of parameters was chosen to achieve robust performance efficiently.

2.3. H-SCAN SOFTWARE DEVELOPMENT

The trained YOLO model is only a detection model that can be applied in a Python environment. Therefore, it is not friendly for users with no coding background, such as architecture students, related researchers and tourists. It is necessary to develop a software to fill the application gap. A system, Heritage-Scan (H-Scan), is developed to recognise the lattice-patterns of tall windows based on the trained YOLO model.

In this system, a number of crucial technologies are incorporated. For the graphical user interface, PyQt5 is utilised. It not only supplies a comprehensive set of interface components and an event-handling system, but also supports Chinese display by setting the Microsoft YaHei font. Moreover, it enables internationalisation via the language module, allowing for the switching between Chinese and English. In the domain of image processing and computer vision, OpenCV (cv2) is employed for image reading, processing, and video capture. Meanwhile, NumPy is used for array operations and data processing. Concerning the deep-learning backend, PyTorch is used to load and run deep-learning models, with YOLOv8 acting as the primary object-detection algorithm.

The H-Scan leverages the DeepSeek large language model to elucidate the cultural background of the identified lattice tall windows, covering aspects such as the symbolism of their patterns, functional purposes, and manufacturing techniques. As shown in Figure 4, the process begins by inputting images into the system, where the windows are detected using pre-trained YOLO models. Subsequently, the names of the identified windows are used as input for DeepSeek to retrieve relevant cultural information, which is finally displayed in a slider box. Predefined prompts focusing on symbolic meaning, function, and manufacturing ensure the accuracy and relevance of the generated explanations. The AI interactive system serves as a valuable tool for students, tourists, and researchers, allowing them to effortlessly acquire essential

knowledge about various window styles without the need for expert explanation.

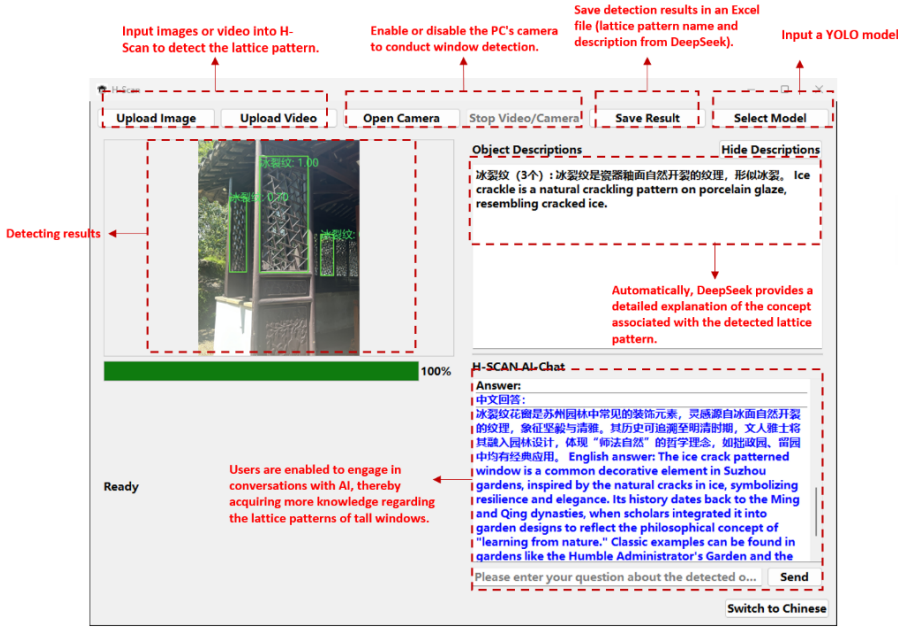


Figure 4.The operation interface of H-Scan.

3. Results and Analysis

The fine-tuned model demonstrated strong performance in detecting lattice patterns in tall windows under favourable conditions. Based on the evaluation results, our Tall-window YOLO model achieved an mAP50 of 0.7299, indicating satisfactory performance according to common benchmarks (where mAP50 $\geq$ 0.5 is considered acceptable). As well as the mAP50 of our fine-tuned YOLO8n is also higher than the trained model by using the default YOLO8n, with an mAP50 of 0.522.

However, its detection accuracy declines in complex real-world scenarios, such as poor lighting, partial occlusion, unusual viewing angles, and limited data availability. As illustrated in the accompanying Figure 5, some tall windows were missed when partially obscured by people or adjacent structures. Additionally, when windows were captured from unconventional angles, resulting in unclear lattice patterns that are challenging even for human observers (Figure 5), detection failures occurred. For instance, one tall window was missed due to poor lighting conditions, as shown in the figure. As shown in the table, the detection precision of Floral Knot with Glass Inlay and Ice Crack Pattern is lower compared to other window types. The primary reasons are the small dataset sizes for these two window types and the insufficient training data available for the YOLO models. Although there were instances of missed detections for certain window types, in practical applications—when images are captured under normal lighting conditions and at reasonable viewing angles—the window patterns of the tall-window category can still be accurately identified.



Figure 5. Windows are not detected because they are obscured by adjacent structures, captured from unconventional angles, and due to poor lighting.

Table 2. The detection precision of thirteen lattice patterns of tall windows.

Category ID	Window Category Name	Average Precision (0-1)
0	Octagonal Scene (八角景)	0.727
1	Ice Crack Pattern (冰裂纹)	0.597
2	Mixed Pattern with Glass Inlay (插角乱纹嵌玻璃)	0.777
3	Palace Style (宫式)	0.840
4	Palace Style Half Window (宫式半窗)	0.590
5	Crabapple Diamond Pattern (海棠菱角式)	0.727
6	Floral Knot with Glass Inlay (花结嵌玻璃)	0.436
7	Well Pattern with Diamond Inlay (井字嵌菱式)	0.512
8	Sunflower Style with Glass Inlay (葵式嵌玻璃)	0.756
9	Hexagonal Panoramic Pattern (六角全景纹)	0.906
10	Ruyi Diamond Flower Pattern (如意菱花式)	0.741
11	Book Strip with Swastika Pattern (书条川万字)	0.945
12	Book Strip Style (书条式)	0.880

4. Conclusions

This study demonstrates the feasibility of applying computer vision and large language model (LLM) technologies to address the longstanding challenge of identifying and interpreting traditional tall-window lattice patterns in Suzhou gardens. By developing a YOLOv8n-based object detection model and integrating it into the H-Scan software, we systematically evaluated the model's performance in recognising thirteen distinct lattice types documented in classical texts, Yuanye (园冶) and Yinzhao Fayuan (营造法原). Quantitative results showed the model achieved an mAP50 of 0.7299 better than the default YOLOv8n model (mAP50 is 0.522), metrics that meet acceptable benchmarks for architectural component detection. While the model exhibited strong performance for patterns with clear structural features (e.g., the "Book Strip with Swastika Pattern (书条川万字)" at 0.945 precision), it faced limitations with complex or underrepresented patterns (e.g., the "Floral Knot with Glass Inlay" at 0.436 precision) due to small sample sizes and low visual distinguishability. These findings confirm that YOLOv8n, despite its lightweight design, is effective for automated lattice recognition in controlled conditions (e.g., normal lighting, frontal angles) and provides a practical foundation for digital heritage preservation.

Beyond technical validation, this research bridges the gap between AI-driven detection and cultural interpretation by embedding the Deepseek LLM into H-Scan. The software enables end-users—including architecture students, tourists, and researchers without coding backgrounds—to not only identify lattice patterns in real time or via batch processing but also access contextual cultural information (e.g., symbolic meanings, manufacturing techniques). This integration transforms passive "observation" of architectural heritage into active "learning," addressing the inefficiency of traditional manual identification and the oversight of cultural connotations. While challenges remain, such as improving model robustness in occluded or low-light scenarios and expanding the dataset to include more regional lattice variations, this work establishes a replicable framework for merging computer vision with digital heritage education. Future iterations of the model and software will focus on these limitations, with the goal of creating a more comprehensive tool for safeguarding and disseminating traditional Chinese architectural knowledge. Another limitation is that the H-Scan is only applied by developers, so it will be introduced to students, tourists, and related researchers to apply in architectural learning and gather feedback to optimise the systems to match the users' requirements. Currently, the H-Scan is only available on PC, and a mobile version will be developed in the future.

Attribution Statement

Doubao (ByteDance, 2025) and Grammarly (Grammarly Inc., 2025) were used to correct errors in spelling and grammar.

References

ByteDance. (2025). Doubao: Yunque [Large language model]. https://www.doubao.com/chat/

Du, L., & Wang, Y. (2024). Bi-YOLO: A novel object detection network and dataset for components of China heritage buildings. Journal of Building Engineering, 97, Article 110817. https://doi.org/10.1016/j.jobe.2024.110817

Dye, D. S. (1974). Chinese lattice designs (Vol. 5). Courier Corporation.

DeepSeek. (2025). DeepSeek-VL2 [Large language model]. https://www.deepseek.com/

Fesl, J., Jelinek, J., Hornickova, K., Nevarilova, Z., Konopa, M., & Feslova, M. (2022). AI-based system for cultural heritage objects identification from real photos. In Proceedings of the International Conference on Advanced Computer Information Technologies (ACIT) (pp. 476–479). IEEE. https://doi.org/10.1109/ACIT54803.2022.9912752

Girsang, N. D. (2021). Literature study of convolutional neural network algorithm for batik classification. Briliance Journal, 1(1). https://doi.org/10.47709/briliance.v1i1.1069

Grammarly Inc. (2025). Grammarly: AI writing and grammar checker app (Version 14.1198.0) [Writing assistance tool]. https://app.grammarly.com/

Horn, C., Ivarsson, O., Lindhé, C., Potter, R., Green, A., & Ling, J. (2022). Artificial intelligence, 3D documentation, and rock art—Approaching and reflecting on the automation of identification and classification of rock art images. Journal of Archaeological Method and Theory, 29(1), 188–213. https://doi.org/10.1007/s10816-021-09518-6

Hou, M., Hao, W., Dong, Y., & Ji, Y. (2023). A detection method for the ridge beast based on improved YOLOv3 algorithm. Heritage Science, 11(1), Article 167. https://doi.org/10.1186/s40494-023-01023-5

Janković, R. (2020). Machine learning models for cultural heritage image classification: Comparison based on attribute selection. Information (Switzerland), 11(1). https://doi.org/10.3390/info11010012

Jiang, B., Chen, S., Wang, B., & Luo, B. (2022). MGLNN: Semi-supervised learning via multiple graph cooperative learning neural networks. Neural Networks, 153, 204–214. https://doi.org/10.1016/j.neunet.2022.05.024

Liu, E. (2020). Research on image recognition of intangible cultural heritage based on CNN and wireless network. EURASIP Journal on Wireless Communications and Networking, 2020(1). https://doi.org/10.1186/s13638-020-01859-2

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 779–788). IEEE.

Saadat, M. A., Hossain, M. S., Karim, R., & Mustafa, R. (2020). Classification of cultural heritage mosques of Bangladesh using CNN and Keras model. In International Conference on Intelligent Computing and Optimisation (pp. 647–658). Springer.

Xu, Y. (2012). Theory landscape architecture skit: The flower window design culture. Applied Mechanics and Materials, 226, 2418–2421.

Yao, C. (1986). Yingzao Fayuan [The principles of architectural standards]. Beijing, China: China Architecture & Building Press.

Zhang, Z. (2023). Designing the ice-crack pattern of traditional Chinese windows. Nexus Network Journal, 25(4). https://doi.org/10.1007/s00004-023-00721-4

Fine-Tuned Flux Kontext Model for Aerial Image Enhancement and 3D Reconstruction Improvement

Yang Yu[1], Yu Qian Liu[2] and Ding Wen Bao[3]
[1,2,3] *RMIT University, School of Architecture and Urban Design*
[1] *yuyang.architest@gmail.com, 0009-0007-6985-2411*
[2] *liulily9610@gmail.com, 0009-0000-3298-3492*
[3] *nic.bao@rmit.edu.au, 0000-0003-1395-8747*

Abstract. The quality of photogrammetric 3D reconstruction is strongly dependent on the fidelity of source imagery, yet in UAV-based architectural and heritage documentation, continuous video capture—while safer and more efficient than stop-and-shoot photography—often produces frames degraded by motion blur and reduced sharpness, leading to compromised geometric reconstruction. This paper proposes a diffusion-based image enhancement workflow that restores drone video frames to near still-photo quality for photogrammetric applications. A four-stage pipeline is developed, comprising automated architectural image curation, generation of paired training data with controlled degradation, Low-Rank Adaptation (LoRA) fine-tuning of the Flux Kontext diffusion model, and quantitative evaluation through 3D reconstruction. Trained on 2,000 paired architectural images with an emphasis on geometric fidelity, the fine-tuned model significantly improves image sharpness and structural consistency. Experimental results show that enhanced video frames achieve 95.8% of the SSIM of still photographs (0.68 vs. 0.71) and a 116% increase in Tenengrad sharpness, while reconstructed models exhibit a substantial increase in point cloud density, confirming improved feature matching and geometric completeness. By enabling continuous UAV flight without sacrificing reconstruction quality, the proposed workflow reduces survey time and collision risk while maintaining measurement reliability, demonstrating the potential of geometry-aware diffusion model fine-tuning as a practical and scalable solution for efficient architectural and heritage 3D documentation.

Keywords. Drone Photography; 3D Reconstruction, Diffusion Model Fine-tuning, Architectural Image Enhancement, Heritage Preservation

1. Introduction

3D reconstruction technology has developed substantially in recent years, especially, Drone-integrated 3D reconstruction has become an essential tool for architectural surveys and heritage documentation (He et al., 2024; Partama, 2025). While algorithms are shifting from, h-based approaches toward point cloud modelling and Gaussian

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 31-40.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

splatting[1] (Li et al., 2024a, 2024b), reconstruction quality remains heavily dependent on source image clarity (Łabędź et al., 2021; Leem et al., 2025). However, acquiring high-quality aerial data is challenging.

Drone images, in the majority of cases, seem to present problems like motion blur originating from vibration, uneven lighting (Burdziakowski & Bobkowska, 2021), and atmospheric disturbances (Liao et al., 2025). This suggests that these issues harm both the matching of features during processing and the ultimate accuracy of the final model, thus tpointing toward a need for more focused attention.

Considering the nuanced nature of these findings, taking sharp photos by keeping the drone still takes substantially longer and increases the potential risk of crashing into buildings or trees (Chodura et al., 2025). This challenge appears to be particularly difficult for large heritage projects, given the multifaceted nature of their complexity. Therefore, this analysis indicates that the majority of current work typically employs stop-and-shoot photography rather than extracting frames from continuous video, since motion blur from moving drones lowers reconstruction quality.

From this particular interpretive perspective, traditional image restoration tools are often not predominantly effective. Common enhancement methods usually have difficulty recovering what appear to be lost high-frequency details (Motayyeb et al., 2022a; Scarmana, 2023). What also appears significant in this context is that GAN-based methods also appear to present problems, as they can sometimes create artificial textures that may conceivably reduce geometric accuracy based on previous results (Laroche et al., 2023a). What this tends to indicate is a major concern for architectural measurement and heritage preservation.

Recent progress in diffusion-based generative models, such as Flux Kontext's reference-guided generation (Labs et al., 2025), indicates promise for improving image quality while ostensibly keeping structures accurate. These findings point toward an investigation into whether fine-tuning these models can potentially restore video frames to the perceived quality of still photos, thereby making it possible to utilize continuous video capture without losing what seems to be reconstruction accuracy.

2. Related Work

Existing scholarship shows that photogrammetric reconstruction accuracy primarily depends on image quality. Motion blur caused by drone vibration reduces point cloud density and geometric accuracy (Sieberth et al., 2014) Compression artifacts in video also make it more challenging to detect features reliably. Traditional enhancement methods, such as histogram equalization (Łabędź et al., 2021) and contrast adjustment (Motayyeb et al., 2022b), remain limited since they work only with existing pixel values.

Recent neural network methods show promise for recovering degraded images. Neural Radiance Fields (NeRFs) can handle some imperfect inputs (Lin et al., 2025), but they struggle with the uneven blur found in drone videos. GAN-based super-

1 Gaussian splatting: A novel 3D rendering technique that represents scenes using 3D Gaussians instead of meshes, enabling high-quality, real-time novel view synthesis.

resolution methods can create realistic details but sometimes add artificial textures that reduce geometric accuracy based on previous results (Laroche et al., 2023b). Conventional enhancement can even introduce new artifacts that reduce reconstruction quality (Arif et al., 2025), indicating the need for geometry-aware restoration.

Diffusion models offer a new direction in generative AI, providing better control over output quality than GANs. The Flux Kontext architecture (Labs et al., 2025) uses reference-guided generation to maintain structural consistency, which is vital for photogrammetry. By fine-tuning these models on architectural image pairs, we aim to restore video frames to near-photo quality while maintaining the geometric accuracy required for 3D reconstruction. However, existing studies rarely evaluate whether image enhancement methods improve downstream photogrammetric geometry rather than perceptual image quality alone.

3. Problem Validation and Motivation through Empirical Test

To test why optimizing image quality is necessary, we conducted a controlled experiment using a DJI Mini 4 Pro drone to capture the same building in two ways:

- Dataset A (Capture Images): 45 separate photos (8K downscaled to 4K), each taken with the drone stopped at a fixed viewpoint.
- Dataset B (Video Frames): 45 frames extracted from continuous video, captured while flying at 10 m/s.

This experiment was designed to quantitatively demonstrate the quality gap between video frames and still images.

Even though the drone features advanced stabilization, the 3D models reconstructed from the video frames (Dataset B) were much poorer—showing missing details, unclear geometry, and evident deformation (Figure 1). Taking 45 still photos took about 12 minutes (approximately 15 seconds per photo), while continuous video capture took only 2 minutes (around 2.6 seconds per frame).

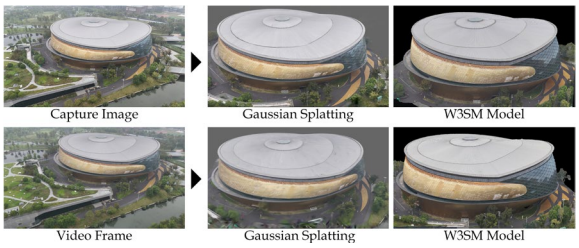


Figure 1. Reconstruction Model by Dataset A&B

Getting closer to the target could improve resolution but would also increase motion blur, collision risk, and flight duration. Importantly, the issues in video frames arise primarily from fixable image quality degradation rather than lost information. We therefore hypothesize that deep learning–based image enhancement can make video frames nearly as good as still photos, enabling faster, continuous drone capture without sacrificing reconstruction quality.

This study contributes the following:

- A fine-tuning method for Flux Kontext focused on architectural images.
- A workflow that restores details while keeping geometric accuracy for measurement.
- An efficient process that saves flight time and reduces collision risk compared with traditional stop-and-shoot methods.

4. Methodology

Based on our previous validation study, we developed a four-step enhancement workflow that combines Flux Kontext fine-tuning with drone photogrammetry. This workflow converts regular drone images into high-quality 3D models through the following stages:

- Image Classification: Automated dataset curation using multimodal classification.
- Image Procession: Generating paired training data sorted by quality.
- Fine-tuning: Text-guided LoRA fine-tuning of the Flux Kontext model.
- Model Reconstruction: Photogrammetric validation using iTwin Capture software.

This approach gives a full framework for making domain-specific training data, including how to preprocess data and set key parameters, and ends with validation in real architectural reconstruction tasks (Shown as Figure 2).

4.1. STAGE 1: IMAGE CLASSIFICATION

To effectively restore blurred architectural images of different types, we needed a large, well-organized training dataset. We collected over 100,000 images from public architectural sources, covering the main types of images used in architectural design.

We used a stratified filtering process with the multimodal large language model LLaMA 3.2-Vision (Grattafiori et al., 2024). First, images were sorted into two main groups: architectural renderings and real-world photographs. Then, real-world photos were further divided into interior and exterior scenes using content analysis. This step-by-step screening resulted in about 20,000 high-quality outdoor architectural photos for our dataset (shown in Figure 2. Stage 1).

4.2. STAGE 2: IMAGE PROCESSION

Flux Kontext works by taking pairs of images and a text prompt that describes what transformation to perform. In our study:

- Target images: High-quality reference photos.
- Reference images: Blurred or degraded versions of the targets.
- Text prompts: Describe the process of making blurry images clear.

LoRA training needs enough diversity to generalize well, but not so much that the model overfits. Based on experience, we found that using 2,000 image pairs is

optimal—it's diverse enough but doesn't cause overfitting.

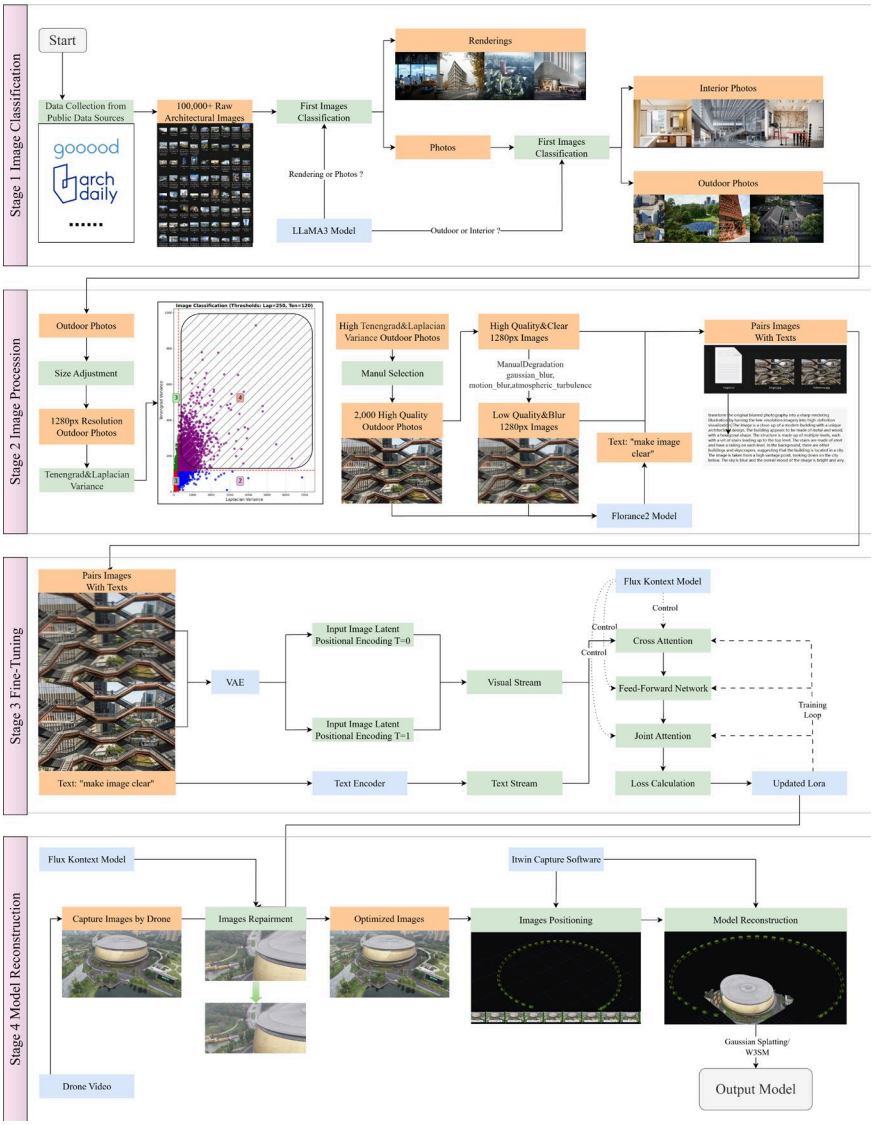


Figure 2. Enhancement Workflow

To create these pairs, we first evaluated 20,000 outdoor photos from Stage 1 using Tenengrad and Laplacian Variance to measure sharpness. In Stage 2 (Figure 2), we sorted images into four quality zones. About 2,000 images in Zone 3, with the best sharpness scores, were chosen as Target images. We then made Reference images by adding controlled blur to these high-quality photos, simulating what happens in long-range drone shots.For clear transformation descriptions, we used the Florence-2 model (Xiao et al., 2023) to generate image-to-text prompts (Figure 2).

4.3. STAGE 3: FLUX KONTEXT MODEL FINE-TUNING

We adopted Low-Rank Adaptation (LoRA) for model fine-tuning rather than full parameter training, as comprehensive fine-tuning of the Flux Kontext foundation model would require at least 200,000 high-quality paired images with detailed annotation—impractical for this study.

The hyperparameters were determined based on a trade-off between computational constraints and the specific requirements of photogrammetric reconstruction. Given that 3D reconstruction algorithms are highly sensitive to image resolution and high-frequency details, we prioritized a higher input resolution of 1,280 pixels. Hardware limitations constrained training feasibility. Due to the high resolution required for preserving architectural details, training was conducted at 1280px resolution under constrained batch settings, requiring approximately 42GB VRAM (Hu et al., 2021). This configuration ensures the recovery of fine architectural details essential for geometric accuracy, which would be compromised at lower resolutions typically used with larger batch sizes.

The training workflow (Figure 2) Stage 3 follows a streamlined architecture. Training samples comprise input images paired with editing instructions (e.g., "make image clear"). The VAE encoder generates latent representations with distinct positional encodings (T=0 for noised target, T=1 for reference), while the Text Encoder processes instructions into embeddings. Cross-Attention layers inject LoRA parameters to learn text-image correspondences, with features fused through Joint Attention mechanisms. The base model remains frozen while only low-rank matrices (A, B) are updated via backpropagation, with loss combining diffusion loss, text region loss, and background preservation loss. This enables efficient fine-tuning (~10M trainable parameters versus 12B frozen) while learning precise instruction-following capabilities.

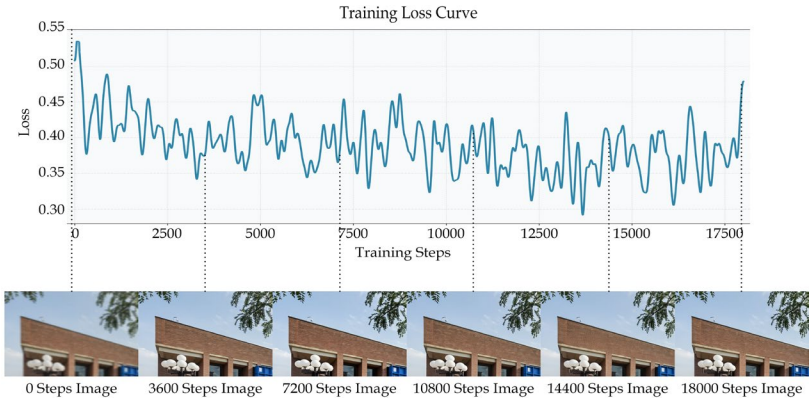


Figure 3. LoRA training process and test results

Table 1. The result of Lora test

Steps	PSNR	SSIM	Tenengrad	Laplacian	Loss
0	13.63	0.40	936.78	1080.45	/
3600	16.48	0.47	26117.79	9643.48	0.38
7200	14.85	0.39	26869.62	8081.17	0.30
10800	16.55	0.47	28397.89	11557.38	0.37
14400	17.89	0.51	30033.85	11463.60	0.36
18000	16.38	0.49	28468.51	10624.15	0.56

The trained LoRA models were evaluated on blurred images. Model selection prioritized restoration quality metrics while monitoring loss values to detect overfitting (shown in Figure 3). Analysis revealed optimal performance at 14,400 training steps, achieving PSNR [2] of 17.89 and SSIM [3] of 0.51 with significant sharpness improvements, while loss values remained stable without overfitting indicators. This checkpoint was selected for deployment based on its balanced restoration quality and generalization capability.

4.4. STAGE 4: IMAGE ENHANCEMENT AND 3D RECONSTRUCTION

Enhanced images generated through the fine-tuned Flux Kontext workflow were processed through iTwin Capture software for photogrammetric reconstruction (Figure 4). The workflow first computed spatial registration through structure-from-motion, followed by dense point cloud generation and surface reconstruction.

5. Results

After obtaining the optimal LoRA model, this experiment restored the 45 images mentioned in Chapter 3 and reused them for 3D model reconstruction. Figure 4 illustrates the reconstructed models from three datasets—captured photos, video frames, and enhanced video frames into W3SM Model [4]and Gaussian Splatiing while Table 2 presents the corresponding quantitative metrics.

The dataset includes Captured Images, Video Frames, and Enhanced Images, each with their respective reconstruction results. For quantitative analysis (shown in Table 2), we compared PSNR and SSIM values between screenshots of the reconstructed models and the original images. We also calculated Tenengrad and Laplacian variance for same-sized images to assess both fidelity and reconstruction quality.

Video frames exhibited a 62–C67% drop in sharpness metrics, establishing a baseline for evaluating the proposed enhancement method. After applying the fine-tuned Flux Kontext model, the Enhanced Images showed significant improvements:

2 PSNR: A metric measuring image quality by comparing the maximum signal power to the noise error; higher values indicate better quality.

3 SSIM: A metric measuring the similarity between two images based on luminance, contrast, and structure, matching human perception.

4 W3SM Model: A global 3D surface model type provided by Google Earth, derived from satellite and aerial imagery, used here as a macroscopic contextual reference.

- 95.8% of the SSIM value of photographs (0.68 vs 0.71).
- 69–81% quality-gap closure across PSNR, SSIM, and Tenengrad metrics.
- Tenengrad doubled (+116%), indicating clear detail recovery.

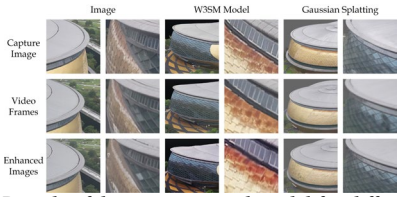


Figure 4. Details of the reconstructed model for different images

Table 2. Average value of image parameters from multiple angles of the reconstruction model

Name		PSNR	SSIM	Tenengrad	Laplacian
Capture Image	Capture Image	\	\	1240.42	112.37
	Gaussian Splatting（221,950 points）	18.2	0.71	774.27	8.19
	W3SM	11.05	0.42	755.12	9.6
Video Frame	Video Frame	\	\	1124.44	40.27
	Gaussian Splatting（38,667points）	12.4	0.55	289.24	2.7
	W3SM	9.58	0.35	555.22	7.2
Enhance Images	Adjusted Images	\	\	1164.44	65.28
	Gaussian Splatting（149,182points）	16.7	0.68	625.18	5.6
	W3SM	9.58	0.35	603.38	8.3

The quantitative results demonstrate that the Flux Kontext model significantly enhances the intrinsic quality of the video frames. Specifically, the enhanced frames achieved the target reconstruction quality thresholds (PSNR = 16.7 ≥ 16, SSIM = 0.68 ≥ 0.65), which enabled continuous flight capture and reduced total flight time by approximately 60% compared to traditional stationary photo capture. Regarding the 3D reconstruction evaluation, point cloud density[5] was selected as a primary metric because it directly reflects the success rate of feature matching and the geometric completeness of the reconstruction. Our experiments show that the point cloud density increased substantially from 38,667 to 149,182 points. This fourfold increase indicates a significant improvement in feature extraction capabilities, leading to greater geometric accuracy in the final reconstructed model.

These results demonstrate that computational image restoration can effectively balance efficiency and quality in UAV-based 3D reconstruction. The workflow combining continuous video capture with AI enhancement not only saves time in the field but also lowers the risk of drone collisions. This approach is particularly valuable for complex architectural sites and cultural heritage documentation.

5 Point Cloud Density: The number of points per unit area in a 3D model; higher density implies more detailed surface reconstruction.

6. Discussion & Conclusion

Although the proposed enhancement method does not yet fully match the absolute reconstruction quality of high-resolution static photography, it achieves a critical balance: improving reconstruction fidelity by ~60% while reducing survey time by 60%. This indicates strong potential for real-world deployment in heritage documentation and construction monitoring, where efficiency is paramount. Furthermore, the approach is well suited to upgrading legacy 3D models generated from early, low-density data sampling (e.g., those seen in platforms such as Google Earth). Given the prohibitive time and cost of re-surveying such large-scale datasets, our method offers a practical pathway to computationally enhance existing low-quality models without physical re-sampling.

The fine-tuning strategy can be extended to multi-structure environments—such as dense urban districts or dynamic outdoor scenes—by integrating multimodal diffusion frameworks to improve robustness across diverse architectural geometries and challenging lighting conditions. Importantly, this direction addresses "unreachable" data, enabling improved reconstructions where precise physical sampling is infeasible due to inaccessibility or safety constraints. By computationally compensating for data deficiencies, the method supports reconstruction in projects that would otherwise be stalled by acquisition limitations.

Overall, the proposed workflow provides a comprehensive fine-tuning pipeline spanning problem definition, data preparation, training optimization, and quantitative validation. By systematically addressing the trade-off between acquisition speed and model quality, it offers a promising route toward efficient, geometry-aware AI enhancement for drone-based architectural documentation, and a robust foundation for future AI-assisted photogrammetry.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Australian Research Council (IH220100016).

References

Arif, R., Harintaka, & Maharani, M. (2025). Investigating the Impact of Enhanced Images on 3D Reconstruction of non-Lambertian Object Using Neural Radiance Fields. IOP Conference Series: Earth and Environmental Science, 1486(1), 012024. https://doi.org/10.1088/1755-1315/1486/1/012024

Burdziakowski, P., & Bobkowska, K. (2021). UAV Photogrammetry under Poor Lighting Conditions—Accuracy Considerations. Sensors, 21(10), 3531. https://doi.org/10.3390/s21103531

Chodura, N., Greeff, M., & Woods, J. (2025). Evaluation of Flight Parameters in UAV-based 3D Reconstruction for Rooftop Infrastructure Assessment (No. arXiv:2504.02084). arXiv. https://doi.org/10.48550/arXiv.2504.02084

Grattafiori, A., Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Vaughan, A., Yang, A., Fan, A., Goyal, A., Hartshorn, A., Yang, A., Mitra, A., Sravankumar, A., Korenev, A., Hinsvark, A., … Ma, Z. (2024). The Llama 3 Herd of Models (No. arXiv:2407.21783). arXiv. https://doi.org/10.48550/arXiv.2407.21783

He, T., Chen, K., Jazizadeh, F., & Reichard, G. (2024). Unmanned aerial vehicle-based as-built surveys of buildings. Automation in Construction, 161, 105323. https://doi.org/10.1016/j.autcon.2024.105323

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models (No. arXiv:2106.09685). arXiv. https://doi.org/10.48550/arXiv.2106.09685

Łabędź, P., Skabek, K., Ozimek, P., & Nytko, M. (2021). Histogram Adjustment of Images for Improving Photogrammetric Reconstruction. Sensors, 21(14), 4654. https://doi.org/10.3390/s21144654

Labs, B. F., Batifol, S., Blattmann, A., Boesel, F., Consul, S., Diagne, C., Dockhorn, T., English, J., English, Z., Esser, P., Kulal, S., Lacey, K., Levi, Y., Li, C., Lorenz, D., Müller, J., Podell, D., Rombach, R., Saini, H., … Smith, L. (2025). FLUX.1 Kontext: Flow Matching for In-Context Image Generation and Editing in Latent Space (No. arXiv:2506.15742). arXiv. https://doi.org/10.48550/arXiv.2506.15742

Laroche, C., Almansa, A., & Tassano, M. (2023a). Deep Model-Based Super-Resolution with Non-uniform Blur. 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 1797–1808. https://doi.org/10.1109/WACV56688.2023.00184

Leem, J., Kim, J., Kang, I.-S., Choi, J., Song, J.-J., Mehrishal, S., & Shao, Y. (2025). Practical error prediction in UAV imagery-based 3D reconstruction: Assessing the impact of image quality factors. International Journal of Remote Sensing, 46(3), 1000–1030. https://doi.org/10.1080/01431161.2025.2450561

Li, Q., Yang, G., Gao, C., Huang, Y., Zhang, J., Huang, D., Zhao, B., Chen, X., & Chen, B. M. (2024a). Single drone-based 3D reconstruction approach to improve public engagement in conservation of heritage buildings: A case of Hakka Tulou. Journal of Building Engineering, 87, 108954. https://doi.org/10.1016/j.jobe.2024.108954

Liao, B., Zhou, B., Qiu, G., Li, X., & Wang, M. (2025). MBF-SR: A Single-Image Super-Resolution Method for Remote Sensing Images in Dynamic Imaging Scenarios. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 23633–23645. https://doi.org/10.1109/JSTARS.2025.3605248

Lin, Z., Zheng, Y., & Yao, L. (2025). Novel View Synthesis of Defocused Blur Scenes Based on Neural Radiance Fields. IEEE Access, 13, 95122–95136. https://doi.org/10.1109/ACCESS.2025.3574881

Motayyeb, S., Fakhri, S. A., Varshosaz, M., & Pirasteh, S. (2022a). ENHANCING CONTRAST OF IMAGES TO IMPROVE GEOMETRIC ACCURACY OF A UAV PHOTOGRAMMETRY PROJECT. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B1-2022, 389–398. https://doi.org/10.5194/isprs-archives-XLIII-B1-2022-389-2022

Partama, I. G. Y. (2025). 3D MODELING USING UAV-PHOTOGRAMMETRY TECHNIQUE FOR DIGITAL DOCUMENTATION OF CULTURAL HERITAGE BUILDINGS. International Journal of GEOMATE, 28(126). https://doi.org/10.21660/2025.126.4768

Scarmana, G. (2023). A METHOD FOR REVERSING THE EFFECT OF DIGITAL IMAGE BLURRING: A SPACE DOMAIN INVESTIGATION. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-1/W1-2023, 431–435. https://doi.org/10.5194/isprs-archives-XLVIII-1-W1-2023-431-2023

Sieberth, T., Wackrow, R., & Chandler, J. H. (2014). Motion blur disturbs – the influence of motion-blurred images in photogrammetry. The Photogrammetric Record, 29(148), 434–453. https://doi.org/10.1111/phor.12082

Meta. (2024). Llama 3 [Vision Model]. Image Reorganization. https://ai.meta.com/llama/

Microsoft. (2024). Florence 2 [Vision Model]. Image Reorganization https://www.microsoft.com/en-us/research/project/florence/

Human-In-The-Loop As-Built Structure Analysis Through Motion Capture and Virtual Reality

Yuxuan Guo[1], Zizheng Yu[2], Xinyue Xia[3], Miaoying Cao[4] and Sining Wang[5]
[1,2,3,4]*School of Architecture, Soochow University*
[5]*School of Architecture & Urban Planning, Shenzhen University*
[5]*siningwang@szu.edu.cn*

Abstract. This study presents a human-in-the-loop (HITL) method for analyzing as-built structures, addressing the limitations of fully automated surveying in capturing irregular or occluded elements. Traditional techniques like laser scanning and photogrammetry often face blind spots and complex data processing. The system uses motion-capture gloves and a virtual reality (VR) interface, allowing operators to collect spatial data through gestures such as tapping, pinching, and grabbing. Developed in Unity3D, the virtual environment mimics the physical site for real-time measurement, annotation, and reconstruction. The resulting 3D model is compared to the design to identify deviations and guide adjustments. Validation with bamboo construction, involving natural materials and manual assembly, showed improvements in accuracy, workflow simplicity, and on-site integration.

Keywords. Human-In-The-Loop, As-Built Analysis, Virtual Reality, Bamboo Structure, Motion Capture

1. Introduction

As-built structural analysis is essential for realizing complex architectural projects. By identifying potential discrepancies between the actual construction and the initial design intent, such analysis provides valuable insights into areas that may require design revisions or construction modifications. Recent advancements in this field have primarily focused on improving data acquisition techniques, including laser scanning, depth-sensing cameras, and unmanned aerial vehicle (UAV) photogrammetry. The three-dimensional geometric data obtained from these technologies are then compared with digital design models to assess and quantify any deviations.

Despite their widespread use, current as-built data collection methods present notable limitations. Point cloud scanning, for instance, is highly sensitive to environmental conditions and physical obstructions. Reflective environments, overlapping components, or occlusions caused by fixed structures like walls and beams, or by transient elements such as equipment, workers, and scaffolding, frequently block the scanner's line of sight, resulting in significant data loss. This may lead to incomplete datasets, potentially resulting in a low-quality model reconstruction.

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 41-50.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Thus, such accessibility constraints usually require substantial manual intervention to restore missing information (Kim et al., 2021). Also, conventional 3D scan approaches typically involve data post-processing steps, including denoising, registration, model reconstruction, and feature extraction, which are computationally intensive and time-consuming. Prior studies indicate that such a workflow remains highly parameter-sensitive and only partially automated, demanding repetitive calibration to ensure accuracy (Li et al., 2025). These challenges may therefore influence the efficiency, scalability, and real-time applicability of using as-built data for further design revisions.

Recent architectural research increasingly emphasizes integrating machine precision with human perceptual intelligence in construction. Studies by Wang et al. (2023) and Kyaw et al. (2024) propose hybrid workflows combining human, cyber, and physical systems to improve design realization. In these frameworks, human operators intervene in robotic fabrication and assembly, using their spatial awareness and contextual understanding to complement automated systems. Such HITL setups aim to boost adaptability, reduce errors, and enhance responsiveness in complex environments, showcasing the potential of collaborative intelligence in modern construction automation. This work introduces a tactile, gesture-based as-built structural analysis method using wearable motion-capture gloves and a VR headset. Continuous finger tracking and custom gestures like tapping, pinching, and grabbing allow operators to access real-time spatial data of building components. The data are processed to generate 3D geometric models for comparing with the original design, enabling detection and assessment of construction deviations.

2. Related works

The core principle of the HITL is the integration of human cognitive flexibility and perceptual intelligence into computational systems. Humans exhibit superior judgment, decision-making, and problem-solving capabilities, particularly in complex, uncertain, and creative contexts (Bensch et al., 2025). These can help to complement automated processes and ensure reliable performance.

2.1. HITL TECHNOLOGIES

Enabling technologies of Industry 4.0, such as the Internet of Things, cloud computing, smart sensors, VR, and AR, can establish robust channels for human–machine interaction, enabling systems to interpret and respond to human intentions in real time (Cimini et al., 2020). In the construction domain, Human–Robot Collaboration (HRC) has become a more adaptable approach than full automation. Construction tasks take place in complex, unstructured environments, where preprogrammed robotic systems often struggle to respond to unexpected conditions. The HRC framework combines human intuition and decision-making with robotic precision and efficiency, enabling flexible, context-aware responses to changing site scenarios (Zhang et al., 2023). With advancements in wearable sensing and computer vision techniques, the entire bodies of human operators can now be captured and simulated in virtual environments for interaction with tools, machines, and built contents (Kurien et al., 2018).

With head-mounted devices, AR technologies can enhance human-machine synergy during construction. By incorporating imperfection and wide tolerances into

both digital design tools and on-site construction, HITL's digital design-to-build emphasizes craft and flexibility over precision (Gwyllim et al., 2022). Similarly, VR expands human–machine collaboration into design and simulation. Using Grasshopper VR, geometric models created in Rhinoceros can be streamed directly into an immersive environment, enabling designers to perform real-time parametric modeling and full-scale spatial evaluations. For instance, when combined with Karamba3D and Kangaroo 2, such HITL workflows allow designers to adjust member dimensions or joint configurations interactively and immediately see deformation and optimization results (Crolla et al., 2025).

2.2. GESTURE-BASED HITL IN ARCHITECTURE

Recent studies have highlighted the growing importance of gesture recognition and motion capture technologies for enabling intuitive, embodied communication between humans and machines. Sensor-based and vision-based systems are the two major motion capture techniques. The former typically use cameras and computer vision algorithms to interpret hand and body movements without requiring wearable devices, enabling more natural, unrestricted interaction. Common examples include RGB-D cameras and Leap Motion sensors (Qi et al., 2023).

Compared with static image recognition methods, AR-based gesture systems can capture continuous hand movements, such as transitioning from a clenched fist to an open hand. This preserves temporal information critical for facilitating dynamic interactions (Kerdvibulvech et al., 2019). For example, Kanter et al. (2023) employed the MediaPipe Hands machine-learning framework alongside standard RGB cameras to translate users' gestures into digital geometric models, which were subsequently manufactured through 3D printing. This "physical gesture–digital model–physical artifact" process establishes a tangible connection between human motion and architectural form. Building upon this, Kyaw et al. (2024) integrated gesture recognition into digital fabrication workflows via Gesture-based Mixed Reality (GBMR). During the construction of a 36-foot-high timber structure, GBMR demonstrated an effective HITL approach that leveraged human-guided object positioning, recognition, and calibration.

HITL offers a conceptual foundation for integrating human cognitive intelligence into computational systems, while the interface between human cognition and machine execution remains a challenging task. Current studies on gesture-based interaction have yet to meet the needs of direct, human-guided geometric data collection in real-world construction environments. Therefore, this work presents a new HITL workflow that employs real-time wearable gesture recognition for intuitive, on-site capture of as-built geometric information. By making human spatial perception the main sensing mechanism within the data collection process, this method improves measurement accuracy in occluded or irregular areas and increases the efficiency of as-built modeling workflows.

3. Methods

3.1. SYSTEM SETUP

This work introduces a HITL setup for as-built analysis that integrates motion capture and VR technologies to enable intuitive data recording and digital twin reconstruction of on-site structural geometries. As illustrated in Figure 1, the system employs a VIRDYN mHand Pro inertial motion-capture glove and a PICO 4 VR headset as its primary hardware components, while the interactive logic and data processing modules are developed in Unity3D (version 2022.3.60f1c1). A customized C# control script governs real-time gesture recognition, spatial point registration, rod generation, virtual interaction, component colorization, and model exportation. Hand position and gesture data are transmitted to the headset via a plugin-based signal receiver. By registering spatial features, such as the endpoints of linear elements or control points of curved components, the operator can define structural geometries directly within the virtual environment. This configuration establishes a seamless integration between human motion input and computational modeling, enhancing the precision and interactivity of as-built analysis.

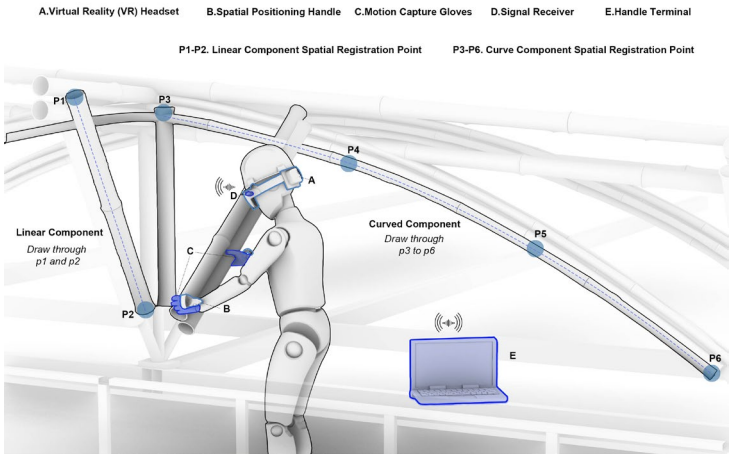


Figure 1. Motion capture and VR system setup

The proposed system is divided into three hierarchical layers: the perception layer, the intermediate processing layer, and the application layer. The perception layer captures hand-pose and gesture data, while the intermediate processing layer handles data communication, coordinate transformation, and signal filtering within the Unity environment. The application layer manages interaction logic and geometric modeling. The mHand glove transmits fused joint-angle and pose data via SDK or UDP broadcast at 120 Hz, while the PICO handles provide six degrees of freedom (6DoF) spatial tracking. Unity receives and processes these data in real time to calculate hand poses, interpret gesture states, and map operations within the virtual environment for spatial point registration and geometry creation. The generated geometries are then exported as .obj format files before being imported into Rhinoceros and Grasshopper for

deviation analysis.

3.2. SYSTEM OPERATION

To perform glove spatial calibration for hand motion capture in Unity3D, the operator used the "Start()" function to initialize all virtual objects, load the gesture input management module, and carry out spatial calibration. The calibration includes two procedures: 1) To perform the spatial alignment calibration, the operator first identifies a known reference point in the physical environment, confirms its location in the virtual space, and establishes a transformation matrix between the real-world and virtual coordinates. 2) To calibrate the gesture, the operator then extends the palm into a standardized pose to record baseline joint angles. These data are used to compensate for joint bias and correct inertial sensor drift. Following calibration, glove pose data are unified with the PICO spatial coordinate system, ensuring spatial accuracy and visual consistency within the virtual display.

Table 1. Gesture signals and their modeling operations

	Gesture	Function	Description
1		Tap	The system creat a new point (blue sphere) at the fingertip position.
2		Grab	The system make the point movable and follows the hand
3		Release	The system fix the registered point is in position
4		Lock	The system marks the point as locked
5		Linear Generation	The system uses a LineRenderer to draw a linear rod
6		Curve Generation	The system generates a curved rod using Catmull–Rom spline interpolation
7		Export	The system exporting vertex coordinates, connections, and attributes into a standard

The operator defined seven gesture signals for on-site virtual modeling. Using the "Update()" function in Unity3D, the system continuously captures hand pose and gesture inputs per frame and determines the current operation mode using a finite state machine. Table 1 lists the key gesture signals and their corresponding modeling operations. To register a point in virtual space, for example, the “tap” gesture creates a new spatial point (a blue sphere) at the fingertip location, which is used to mark key points on structural components. The “grab” gesture calculates the distance between the palm and the nearest registered point; if the operator's hand nears a spatial point (<0.06m), the point becomes active and follows hand movement. The “release” gesture then fixes the grabbed point in space, and the "lock" gesture anchors it to prevent accidental touch or movement. The sphere's color changes from blue to red (Figure 2). In this work, gesture recognition primarily relies on the built-in functions of the mHand

glove. To improve robustness, a debounce mechanism and confidence threshold are

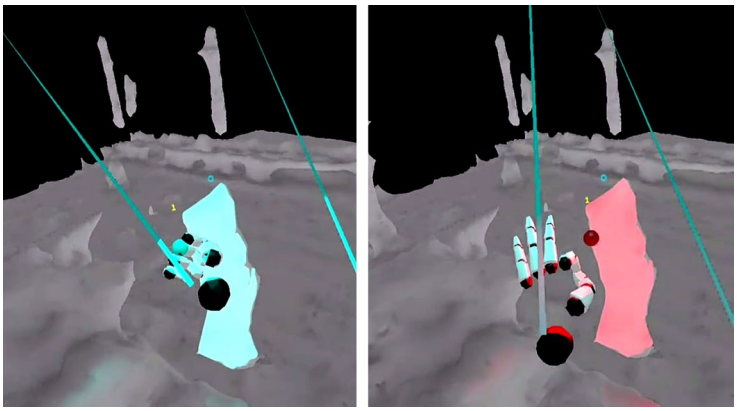


Figure 2. Tap to register a point in space (left) and lock it in place (right).

applied to prevent false triggers.

The script detects interactive targets using raycasting and distance-based selection, allowing accurate creation and manipulation of spatial points in the virtual environment. Each registered point appears as a colored spherical object with a unique ID and timestamp for data tracking. During interaction, the point updates its position to follow the operator's hand when "grabbed" and remains fixed in space when "released." A trajectory line visualizes the path of motion. To improve measurement stability, the system uses smooth interpolation, thereby reducing inertial drift and measurement noise.

3.3. GESTURE-BASED MODELING

Following point registration, the operator generates linear or curved structural elements using predefined gestures or commands. When two points are selected and the "linear generation gesture" is detected, a 3D cylindrical geometry is created to represent the linear segment, automatically adjusting its orientation, position, and scale based on the

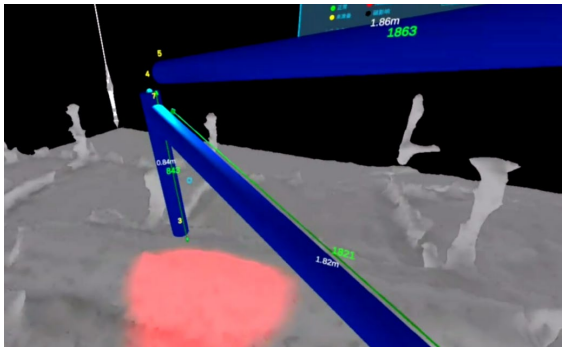


Figure 3. Generation of linear rod geometries and dimension annotations.

selected points' spatial coordinates. For curved elements, when more than two points are selected, the system applies spline interpolation to reconstruct a smooth curve rod, and increasing the number of control points improves geometric accuracy. All generated components are visualized as semi-transparent geometries with real-time length annotations (Figure 3), enabling the operator to perform intuitive spatial assessment and facilitating interactive feedback within the virtual environment.

The final step involves the operator using the predefined "export" gesture signal to automatically output all registered spatial entities, vertex coordinates, annotations, and attribute data into a standardized .obj file. Each exported model is labelled with a site identifier and timestamp to ensure traceability and version control. The file is then imported into Rhinoceros, where a Grasshopper-based analysis script assesses deviations from the design model. This analysis aims to detect hidden construction errors or discrepancies in structural components. Additionally, to reduce inertial measurement unit (IMU) drift, the system supports periodic recalibration and external pose fusion via PICO spatial tracking. In safety-critical or spatially constrained construction environments, the virtual scene can be mirrored to mobile devices, enabling synchronized visualization and control without a VR headset. This feature enhances operational flexibility and field adaptability. Figure 4 highlights the system workflow for HITL as-built analysis.

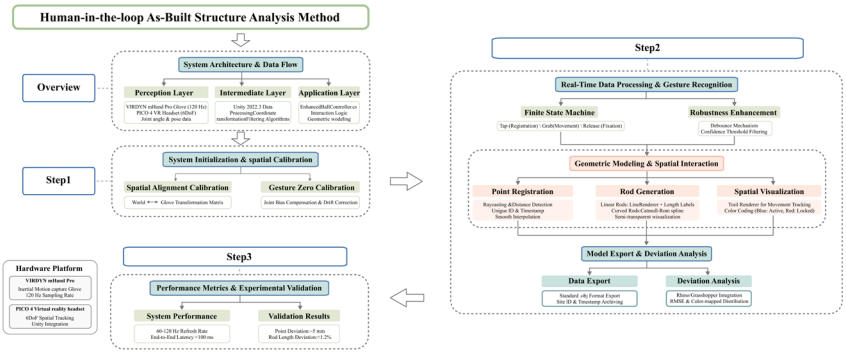


Figure 4. System workflow for HITL as-built analysis.

4. Results

4.1. PROJECT M.E.S.H.

M.E.S.H. is a student-led design-to-build project developed for the 2025 Sustainable Construction Competition in Zhang Jiakou, Hebei Province, China (Figure 5). The project features a non-standard sloped roof with bamboo structures designed to incorporate photovoltaic panels. The roof is composed of raw bamboo trusses of varying heights along the north–south axis, connected by bent bamboo arches along



Figure 5. Exterior view of M.E.S.H..

the east–west direction. All bamboo materials were pre-processed before being transported to the construction site. Due to bamboo's inherent material variability, such as differences in diameter, length, node spacing, and mechanical properties, each structural component had to be selected, cut, and assembled on-site.

Under these construction conditions, efficient as-built analysis was essential for enabling rapid, data-informed decision-making on-site by the bamboo contractors. Conventional measurements for complex bamboo assemblies require repeated manual verification and descriptions of components' spatial information, which are labor-intensive and prone to human error. However, high-precision digital methods, such as drone-based photogrammetry or LiDAR scanning, require multiple viewpoints and extensive post-processing. The rendering and data interpretation processes were impractical for providing immediate feedback during construction activities.

4.2. AS-BUILT BAMBOO STRUCTURE ANALYSIS

The authors employed the aforementioned HITL method to evaluate the as-built geometric conditions of the bamboo structure. A student operator, wearing motion-



Figure 6. On-site data recording using motion capture gloves and VR device.

capture gloves and a VR headset, recorded the component's spatial data in real time as he touched the bamboo. Another operator held the mobile device streaming the results while comparing the as-built structural geometries with the designed ones. By physically interacting with the structure, the operator could register key geometric data without interrupting the assembly process. To do this, the operator's hand moved to a target position and used predefined gesture commands to trigger spatial point registration. Linear truss elements were defined by start–end point pairs, while curved arch geometries were captured by sequentially tracing multiple control points and then reconstructed using splines (Figure 6). Compared with drone scanning, which is prone to occlusion and limited access, the glove-based HITL approach allowed direct, tactile data collection, effectively eliminating blind spots, especially at densely intersecting joints where laser scanning often cannot capture holistic component geometries.

Within the customized mobile interface, the reconstructed as-built model was overlaid with the original design model for deviation analysis, with blue geometries representing the reconstructed model and green geometries representing the original design. For deviations exceeding 100mm, the augmented visualization highlighted geometric discrepancies and automatically illustrated deviations, such as nodal displacements and member length differences, using red dots (Figure 7). This provided quantitative feedback for on-site adjustments.

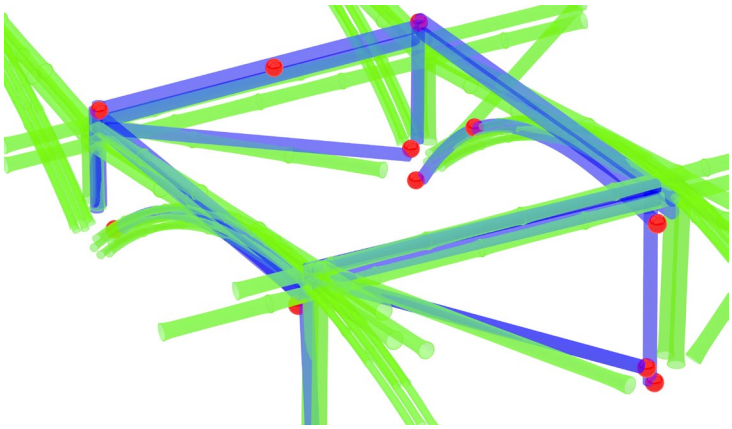


Figure 7. As-built and designed model comparison.

5. Conclusion

This work presents a HITL workflow for as-built structural analysis that integrates motion capture and VR technologies. The results demonstrate that this human–machine collaborative framework enhances the flexibility and intuitiveness of geometric documentation. The findings validate the feasibility of incorporating gesture-based interaction into architectural design and construction. Despite its advantages, several limitations remain. The current system is optimized primarily for linear components and has not yet been extended to free-form or curved geometries. The measurement range is constrained by the VR device's spatial tracking capacity,

and minor positional drift may affect point registration accuracy. Additionally, the stability of gesture recognition can influence the consistency of command execution. Future work will focus on expanding sensing modalities, such as eye tracking and voice control, and transitioning from VR to AR to enable real-time feedback during data acquisition. Addressing these challenges will advance the HITL workflow toward a scalable, real-time, and human-augmented solution for construction quality control and as-built analysis.

Acknowledgments

This work was funded by the Young Scientists Fund of the National Natural Science Foundation of China (grant No. 52408032).

References

Bensch, L., Bensch, O., Nilsson, T., Paradiso, J. A., Davidoff, S., Blanchard, L., Pataranutaporn, P., Weber, M., Kutz, J. N., & Flemisch, F. (2025). Human-in/on-the-Loop AI: Enabling Human Controllability and Decision-Making in Spaceflight. In Handbook of Human-Centered Artificial Intelligence (pp. 1–46). Springer, Singapore. https://doi.org/10.1007/978-981-97-8440-0_40-1.

Cimini, C., Pirola, F., Pinto, R., & Cavalieri, S. (2020). A human-in-the-loop manufacturing control architecture for the next generation of production systems. Journal of Manufacturing Systems, 54, 258–271. https://doi.org/10.1016/j.jmsy.2020.01.002

Crolla, K., & Sheikh, A. T. (2025). Grasshopper VR: Development of an immersive visualisation interface for parametric design workspaces. Virtual Reality, 29(3), 133. https://doi.org/10.1007/s10055-025-01174-5.

Jahn, G., Newnham, C., & Berg, N. (2022). Augmented Reality for construction from steam bent timber. In J. van Ameijde, N. Gardner, K. H. Hyun, D. Luo, & U. Sheth (Eds.), POST-CARBON - Proceedings of the 26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (pp. 191–200). Computer-Aided Architectural Design Research in Asia (CAADRIA). https://doi.org/10.52842/conf.caadria.2022.2.191.

Kanter, J., & Quinteros, K. (2023). Gestural design—Hand tracking for digital drawing. Architecture and Planning Journal, 28(3).

Kerdvibulvech, C. (2019). A review of Augmented Reality-based human-computer interaction applications of gesture-based interaction. In C. Stephanidis (Ed.), HCI International 2019 – Late Breaking Papers (pp. 233–242). Springer International Publishing. https://doi.org/10.1007/978-3-030-30033-3_18.

Kim, H., & Kim, C. (2021). 3D as-built modeling from incomplete point clouds using connectivity relations. Automation in Construction, 130, 103855. https://doi.org/10.1016/j.autcon.2021.103855.

Li, X., Liu, C., Yang, X., & Yan, H. (2025). 3D surface defect detection and measurement using point cloud unwrapping via dimension reduction. Automation in Construction, 177, 106349. https://doi.org/10.1016/j.autcon.2025.106349.

Qi, J., Ma, L., Cui, Z., & Yu, Y. (2024). Computer vision-based hand gesture recognition for human-robot interaction: A review. Complex & Intelligent Systems, 10(1), 1581–1606. https://doi.org/10.1007/s40747-023-01173-6.

Zhang, M., Xu, R., Wu, H., Pan, J., & Luo, X. (2023). Human–robot collaboration for on-site construction. Automation in Construction, 150, 104812. https://doi.org/10.1016/j.autcon.2023.104812.

Computational Manifestation of the Sri Yantra: Bridging Hindu Sacred Symbolism and Contemporary Architectural Design

Chethana M Jangid[1] and Pradipta Roy Choudhury[2]
[1,2]*Reva University*
[2]*Indian Institute of Technology*
[1]*chethana.jangid@gmail.com, 0009-0002-8996-1512*
[2]*pradiptarch@gmail.com, 0000-0001-9443-8860*

Abstract. The Sri Yantra (Sri Cakra) is one of the most intensely interpreted figures in Hindu sacred geometry, operating simultaneously as a metaphysical diagram, ritual instrument, and latent spatial order. While many authors have attempted to "solve" its geometry—ranging from analytic constructions by Huet and Maxwell to Marma Stana–based and image-derived approaches—these efforts largely remain confined to two-dimensional representation. This paper proposes a computational framework in Rhino–Grasshopper that reconceives the Sri Yantra not as a static diagram but as a parametric and architectural system. The framework reconstructs three geometric families of the Sri Yantra: (1) an analytic Huet–Maxwell reconstruction, (2) a Marma Stana–driven integer-ratio model based on 48 vertical divisions of the enclosing circle, and (3) a golden-ratio / golden-triangle variant aligned with architectural proportion theories. These 2D geometries are systematically translated into three-dimensional architectural assemblies through controlled parameters of selection, repetition, and independent height assignment. The primary findings demonstrate that the under-determined nature of the Sri Yantra enables the generation of stable yet diverse spatial families; that Marma Sthana–based integer scaffolding produces the most architecturally robust and symbolically legible outcomes; and that selective activation of yantric elements allows ritual hierarchies to be mapped onto architectural gradients of light, enclosure, and movement. The resulting morphological atlas positions the Sri Yantra as a metaphysical algorithm capable of supporting contemporary, human-centred computational architectural design while remaining grounded in sacred symbolism and cultural meaning.

Keywords. Sri Yantra, Sacred Geometry, Parametric Modelling, Algorithmic Architecture, Hindu Symbolism, Digital Fabrication, Cultural Heritage

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 51-60.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

1. Introduction

The Sri Yantra (Sri Cakra) sits at a unique crossroads: it is simultaneously a Cosmo gram, a ritual device traced on floors and altars, and an abstract geometric figure that has long intrigued mathematicians, philosophers, and computer scientists. Over recent decades, several geometric "solutions" have been proposed—from analytic constructions by Huet and Kulaichev, through Maxwell's reconstructions, to Marma Stana–based models and image-derived techniques. However, most of these studies remain limited to the resolution of a precise two-dimensional diagram. For architects and computational designers, a different question becomes critical: if the Sri Yantra is understood not only as an image but as a generative rule system, what architectural possibilities emerge from its internal logic? This question shifts attention from correctness of geometry to productive ambiguity, from diagram as object to diagram as spatial generator. In response, this paper develops a computational framework in Rhino–Grasshopper that reconstructs three distinct geometric families of the Sri Yantra and systematically translates them into three-dimensional architectural typologies. The work is intentionally double-sided: it tests the precision and flexibility of parametric scripting, while also examining how sacred symbolism can be spatialised without reducing it to mere formal abstraction. The contribution lies both in the technical workflow and in demonstrating how culturally embedded diagrams can operate as metaphysical algorithms within contemporary architectural practice.

Translating the Sri Yantra into architectural systems is not intended as an act of formal appropriation, but as a continuation of its historical role as a spatial, ritual, and bodily organiser. Across Hindu practice, the Sri Yantra has always operated beyond representation—structuring movement, orientation, and experiential hierarchy. By rendering its under-determined geometry computationally explicit, this framework allows contemporary architectural design to engage with cultural diagrams as living, adaptable systems rather than frozen icons. The advantage of this translation lies in enabling symbolic continuity while supporting variation, contextual tuning, and human-centred spatial interpretation through parametric means.

2. Literature Review

2.1. SRI YANTRA AS SACRED DIAGRAM AND SPATIAL ARCHETYPE

The Sri Yantra (Figure 1) is traditionally understood as a diagram of Sri Lalita Tripurasundari and the cosmic order she inhabits. Its nine interpenetrating triangles—four upward-facing (Śiva) and five downward-facing (Sakti)—are enclosed by lotus petals, concentric rings, and a square bhupura with gateways, forming a sequence of ritual avaraṇas. Although often represented as a flat diagram, the Sri Yantra has always been spatial in practice. It is inscribed on floors, carved into stone, embedded within temple layouts, and activated through circumambulatory movement. Even as linework, it implicitly organises body, orientation, and progression, suggesting architectural behaviour rather than pictorial representation.

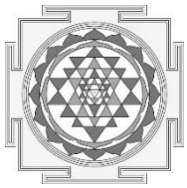


Figure 1: Historical representations of Sri Yantra

2.2. GEOMETRIC RECONSTRUCTIONS: ANALYTIC AND INTEGER MODELS

Huet's analytic approach treats the Sri Yantra as a system of intersecting lines constrained by non-linear equations, while Maxwell demonstrates that the diagram is under-determined, allowing multiple valid configurations. This insight is crucial: it suggests that the Sri Yantra is not a single resolved figure but a family of related geometries. Santhi et al. (2012) generate Sri Yantra images algorithmically, focusing on reproducible 2D outputs. More recently, a Marma Stana–based model anchors triangle vertices to bodily energy points along a vertical axis divided into 48 integer levels. This approach introduces symbolic embodiment into the geometry, allowing integer ratios to stabilise the system while retaining cultural legibility.

2.3. PROPORTION THEORY, GOLDEN RATIOS, AND TEMPLE GEOMETRY

Table 1: Summary of key geometric approaches—analytic, Marma Sthana, image-based, golden-ratio—with their main assumptions, strengths, and limitations. (Table summarises representative approaches rather than an exhaustive taxonomy.)

Approach	**Construction logic**	**Strengths**	**Limitations**
Analytic (Huet / Maxwell)	Solves line–line intersections via non-linear constraints to fix triangle vertices and overlaps.	Very precise; good baseline for checking intersections and sensitivities.	Under-determined; numerically fragile; weak link to body/ritual scales.
Marma Stana (48 divisions)	Uses a 48-step vertical axis (d_0 = 360) with triangle apices/bases anchored to Marma Stana levels.	Integer ratios; easy to script; symbolically legible; backbone of this work.	Depends on one interpretive scheme; not a unique "canonical" solution.
Image-based generation	Algorithmic drawing from target images using rules, thresholds, and iterative refinement.	Fast 2D reproduction; useful for visual validation and graphic outputs.	Raster/illustrative; no guaranteed exact geometry; weak 3D translation.
Golden-ratio / φ-based	Constrains radii, slopes, and grids using φ and golden triangles, linked to proportion theory.	Bridges yantra with architectural proportion and temple overlays.	Historical link speculative; φ-tuning can drift from canonical layouts.

Proportion theory in architecture has long mediated between perception, structure, and

cultural meaning. Singh (2018) highlights the tension between fixed proportional canons and adaptable systems. Within Sri Yantra discourse, some authors align its structure with golden ratios or golden triangles, linking the diagram to temple planning traditions and cosmological grids. While such links are often speculative, they open a productive space for architectural reinterpretation—treating the Sri Yantra not as a closed symbol but as an energetic topology that can be sampled, scaled, and spatially reconfigured.

2.4. GAPS FOR ARCHITECTURAL AND PARAMETRIC PRACTICE

From a computational design perspective, three critical gaps emerge:

- Diagram–space gap: Most reconstructions stop at 2D representation, without translation into inhabitable space. In Western contemporary practice, diagrammatic and symbolic drawings such as Daniel Libeskind's Micromegas series have functioned not as literal forms but as generative devices through which spatial experience, light, and narrative emerge beyond the originating graphic (Libeskind, 1983)
- Parametric family gap: The under-determined nature of the Sri Yantra is rarely exploited as a generative design family.
- Humanistic computation gap: Few studies integrate symbolic meaning, bodily experience, and spatial perception with algorithmic rigor, addressing how symbolic diagrams can operate as generative systems producing experiential and cultural meaning rather than purely optimised form (Oxman, 2017). This project situates itself in these gaps: using Grasshopper and scripting to construct parametric families of Sri Yantra geometries, giving them architectural thickness, and doing so in a way that keeps the figure's symbolic and ritual roles in view rather than treating it as a purely abstract dataset.

3. Methodology

3.1. OVERALL WORKFLOW

- Module 1 – Core 2D generator: Implements three geometric families of the Sri Yantra—analytic (Huet–Maxwell), Marma Stana 48-division, and a golden-ratio variant. In the Marma family, geometry is scripted in Python on a reference diameter
- Module 2 – Selector and height pattern: Receives triangles with metadata (names, centroids, Bindu), filters them (e.g., by planetary name), applies rotation, mirroring and optional radial copies, and assigns heights from a user-defined list or pattern.
- Module 3 – Petals, rings, and bhupura to 3D: Uses the 2D petals, rings and Trailokya-mohaṇa bhupura loops and extrudes each layer to independent heights in +Z, producing a full 3D Sri Yantra assembly.

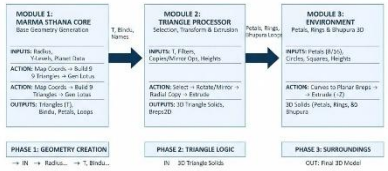


Figure 2: Block diagram of Module 1, Module 2, and Module 3.

3.2. MODULE 1 – MARMA STANA CORE WITH LOTUS AND TRAILOKYA-MOHAṆA

3.2.1. Integer-based vertical framework

The Marma Stana family forms the backbone of the project. The enclosing circle is aligned with a vertical axis of length $d_0 = 360$ units, following the Marma-based Sri Cakra model. Key y-levels for the nine planetary triangles (Ravi, Chandra, Kuja, Budha, Guru, Shukra, Shani, Rahu, Ketu) and the Bindu are encoded as constants taken from the paper.

A simple mapping recentres this axis so that $y = d_0/2$ maps to the origin, with a global scale factor:

$$\text{scale} = 2R / d_0$$

so that any chosen R yields a similar Sri Yantra without altering the internal ratios.

3.2.2. Nine triangles as polyline primitives

Each planetary triangle is specified by a name, an "up"/"down" orientation, top and bottom y-levels, and left/right x-extents. "Down" triangles are constructed with their base at the upper level and apex below; "up" triangles invert this relation. The script builds closed polylines pL, pR, pA, pLand computes centroids using Area Mass Properties.

Module 1 outputs (refer Figure 3):

- C – circle tightly enclosing all triangle vertices,
- T – list of nine Polyline triangles,
- Names – triangle names,
- Centres – centroids,
- Bindu – central point in world coordinates.

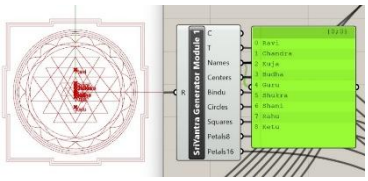


Figure 3: Module 1 output for the Marma Stana family.

3.2.3. Lotus petals, rings, and bhupura

On top of this core, the script constructs:

- Lotus circles: A radius max_r is computed as the maximum distance from centre any triangle vertex. The 8-petal tips sit at

$$R_{lotus8} = max_r\,(1 + \alpha)$$

and the 16-petal tips at

$$R_{lotus16} = R_{lotus8} + \beta \cdot max_r$$

with α and β tuned so petals remain clearly outside the triangles.

- Three outer rings: Radii R_ring1, R_ring2, R_ring3 are spaced outward from R_lotus16 with fixed increments.
- Trailokya-mohaṇa bhupura loops: A base function divides each square side into equal segments and creates a Greek-key-like gateway at the middle third. Offsetting the outer loop inward produces middle and inner tracks with consistent gap and gate depth.
- 8 and 16 petals: Petals are built as smooth, S-curved shapes rather than straight polylines. Two mirrored S-curves meet at common top and bottom points, so neighbouring petals join tangentially, approximated using higher-degree interpolated curves.

The resulting triangles, petals, circles, and bhupura loops form the complete 2D Marma Stana Sri Yantra that feeds into Module 3.

3.3. MODULE 2 – TRIANGLE SELECTOR, TRANSFORMS, AND HEIGHT PATTERNS

Module 2 organises the triangles into parametric patterns that can be spatialised.

3.3.1. Selection by name and optional filters

Triangles T, names Names, and centroids Centres come directly from Module 1. Two basic modes are offered:

- Use All = True: all nine triangles are included.
- Use All = False: a lowercase whitelist is built from Name Filter (e.g., "Ravi", "Ketu", "Shani"), and only matching triangles are selected.

This enables patterns such as "Śiva only," "one planet repeated radially," or custom planetary subsets.

3.3.2. Rotation, mirroring, and radial copies

For each selected triangle, the script can:

- apply a global rotation Angle Offset about the bindu,
- mirror across the Y-axis (MirrorX) and/or X-axis (MirrorY),
- generate Copies number of radial instances by evenly rotating around the centre.

Transforms are implemented explicitly with trigonometric expressions, keeping control over angles and signs. A deliberate "pass-through" case ensures that when Copies = 1, Angle Offset = 0, and no mirrors are active, the triangles are used exactly as defined in Module 1, without resampling. Refer Figure 4

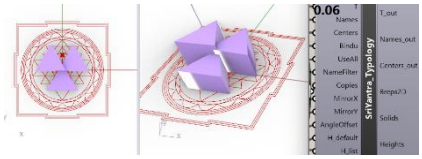


Figure 4: Module 2 examples—single-planet emphasis and multi-copy configurations.

3.3.3. Height patterns and extrusion

Module 2 also assigns height and performs extrusion:

- A base height H_default is used if no list is provided.
- If H_list is present, it is treated as a repeating pattern: each instance uses H_list[idx % len(H_list)]. Designers can script rhythms such as 0.2,0.4,0.1,0.3around the circle.
- Each polyline is extruded along +Z with height |h|, capping planar holes to form closed Breps.

Outputs are:

- T_out, Names_out, Centers_out – transformed 2D triangles,
- Breps2D – planar triangle surfaces,
- Solids – extruded 3D volumes,
- Heights – assigned heights per instance.

3.4. MODULE 3 – PETALS, RINGS, AND BHUPURA FROM 2D TO 3D

Module 3 takes the 2D petals, circles, and bhupura loops from Module 1 and applies a common 3D logic (Refer Figure 5):

- Petals: Each curve (or GH_Curve wrapper) is converted into a planar Brep (Pet8Surf, Pet16Surf) and extruded to heights H_p8 and H_p16 (Pet8Sol, Pet16Sol).
- Rings: The three ring circles are turned into planar discs (RingSurf) and extruded by

H_r1, H_r2, H_r3 (Ring Sol).

- Bhupura loops: Whether supplied as polylines or Breps, the outer boundary is extracted, converted to a planar Brep (SqSurf), and extruded to H_sq1, H_sq2, H_sq3 (SqSol).

Special handling unwraps Grasshopper geometry types and consistently forces extrusions in the +Z direction to avoid mirrored or inverted profiles.

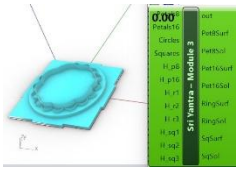


Figure 5: 3D view showing triangles, petals, rings, and bhupura with independent heights.

3.5. OTHER FAMILIES: ANALYTIC AND GOLDEN-RATIO RECONSTRUCTIONS

Alongside the Marma Stana family, two additional reconstructions are implemented:

- Analytic family: Inspired by Huet–Maxwell, triangle vertices are solved numerically to satisfy intersection constraints, generating a more "exact" but algebraically heavy configuration.
- Golden-ratio family: Key radii and triangle slopes are constrained by φ-based ratios and aligned with temple-plan grids, echoing broader proportion theory.

All three families share the same data structure (named triangle polylines with associated centres and Bindu), allowing them to be fed into the same Modules 2 and 3. In this paper, the Marma family is discussed in depth, while the analytic and golden-ratio families primarily appear in comparative overlays. (Final Model refer Figure 6)

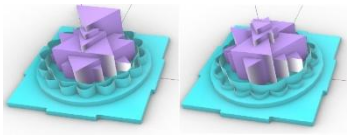


Figure 6: Final Composite model and Parametric Variations.

4. Results And Discussion

4.1. MORPHOLOGICAL ATLAS OF 2D SRI YANTRA VARIATIONS

Across the three geometric families, the project compiles an atlas of 2D Sri Yantra diagrams. The analytic family prioritises exact intersections, giving taut but fragile linework. The Marma Stana family is more forgiving: its integer scaffold stabilises triangle positions and keeps the figure visually "bodied." The golden-ratio family produces slightly stretched or compressed triangles, sometimes recalling temple plans with emphasised axes.

The atlas shows Sri Yantra as a neighbourhood of related diagrams rather than a

single figure. Small shifts in slopes or y-levels redistribute visual weight along particular directions, enabling deliberate emphasis on diagonals, centre, or periphery.

4.2. 3D TYPOLOGIES: PAVILIONS, SANCTA, AND FAÇADES

Through Modules 2 and 3, these diagrams become three main 3D typologies:

- Pavilions: Triangles act as ribs or columns, petals form light roof shells, and rings become benches or stepped plinths; the bhupura reads as a layered perimeter threshold.
- Sancta: Inward-pointing triangles encircle the Bindu; ring heights tune how "deep" or open the core feels.
- Façades: Partial yantra patches are projected onto elevations; triangle heights control shading depth, petals become perforated panels.

Height patterns via H_list make these typologies highly adjustable—Sakti vs. Śiva emphasis, soft outer rings with strong inner ones, or gradual gradients encoding programme and privacy.

4.3. ENVIRONMENTAL AND EXPERIENTIAL OBSERVATIONS

Preliminary readings suggest:

- Light: Raised triangles and petals around the Bindu create a soft central glow in roofs; on façades, alternating heights generate fine-grained shadow rhythms.
- Airflow: Differentiated ring heights form bands of openness and resistance, acting as layered filters for wind and background ventilation.
- Acoustics: Taller central elements with low peripheries gently focus sound towards the Bindu (observed qualitatively).

Though not yet backed by full simulations, these tendencies show that once the yantra has thickness, its layered logic operates as environmental control as well as symbolism.

The evaluation in this study is qualitative and exploratory, relying on comparative spatial reading across parametric variations rather than quantitative simulation. Architectural performance is assessed through visual inspection, sectional logic, and experiential inference related to light, enclosure, movement, and ritual hierarchy. This approach aligns with early-stage computational design research where symbolic and spatial legibility precede optimisation.

4.4. SYMBOLIC LEGIBILITY AND HUMAN-CENTRED COMPUTATION

A persistent question is how recognisable the Sri Yantra should remain once it becomes architecture. In some configurations—especially in plan or roof view—the yantra is still legible; in others, it dissolves into traces such as a stepped bhupura-like threshold, triangular apertures, or a petal rhythm along a façade. The project therefore treats the Sri Yantra as a metaphysical algorithm rather than a fixed icon, using Grasshopper scripts to "listen" to the diagram by selectively activating a single planetary triangle,

exaggerating the bhupura while keeping inner triangles low, or translating ritual inner and outer āvaraṇas into gradients of height, light, and acoustic intimacy. Computation here is kept close to lived experience and ritual narrative, not merely to geometric optimisation.

5. Conclusion

This paper has outlined a Rhino–Grasshopper framework that manifests the Sri Yantra in architectural terms through three linked modules and multiple geometric families. By working with analytic, Marma Stana integer, and golden-ratio reconstructions, the Sri Yantra is treated as an under-determined system rather than a single resolved diagram. The Marma Stana family demonstrates how symbolic points and integer alignments can be encoded into robust parametric scripts, while subsequent modules give triangles, petals, rings, and bhupura loops architectural height and environmental behaviour. The resulting atlas of 2D and 3D variants supports reading the Sri Yantra as a metaphysical algorithm—a layered ruleset that can be instantiated in multiple ways, balancing symbolic clarity and spatial invention. This points toward a broader mode of computational design in which traditional diagrams function as living, parameterised systems, and future work will extend the framework through quantitative environmental analysis, closer links to temple plans and sections, and participatory "tuning" of yantra parameters in communal or ritual settings.

Acknowledgements

This work is carried out under the thesis project titled 'STREE YUGA AARANI' as a part of the requirement of final year thesis at the School of Architecture, REVA University Bandalore, under the guidance of Pradipta Roy Choudhury. Our sincere gratitude goes out to Mr. Manik Chand Jangid whose constant support, encouragement, and belief in our work formed the foundation of this research journey.

Attribution Statement

ChatGPT (OpenAI, 2025) was used for structural composition of the manuscript and refinement of language flow. Grammarly (Grammarly Inc., 2025) was used to correct grammar and syntax.

References

Bhatt, T. (2016). The comparative study of quantum theory and Shree Yantra. International Journal of Information Technology & Computer Sciences Perspectives, 5(2), 98–105.

Huet, G. (2002). Sri Yantra geometry. Theoretical Computer Science, 281(1–2), 647–687.

Libeskind, D. (1983). Micromegas: The architecture of end space. Academy Editions.

Mahesh, K. (2023). Understanding the geometry of Sri Chakra. International Journal of.

Oxman, R. (2017). Thinking difference: Theories and models of parametric design thinking. Design Studies, 52, 4–39.

Santhi, B., Kumar, N. R., Bharathy, C., & Balakrishnan, R. (2012). Generation of divine image – Sri Yantra. Research Journal of Applied Sciences, Engineering and Technology, 4(14), 2241–2246.

Singh, M. R. (2018). Proportions in architecture. International Journal of Research – Granthaalayah,6(3),126–133. https://doi.org/10.29121/granthaalayah.v6.i3.2018.1505

Visual Impact Assessment of High-Rise Buildings Surrounding Cultural Heritage Sites in the Historic Urban Areas of Hanzhong

Yuan Liu[1]
[1]*China State Construction Engineering*
[1]*670807195@qq.com, 0009-0007-6375-5493*

Abstract. High-rise buildings are often assumed to visually threaten historic urban landscapes, yet height alone does not fully explain such impacts—especially in Chinese cities where high-rise development is widespread. This study aims to establish a localized, perception-driven visual impact assessment framework to determine when and how high-rise buildings affect heritage views. Using Hanzhong's historic urban area as a case, the research integrates parametric viewshed modeling, eye-tracking experiments, and Semantic Differential (SD) evaluations within a three-stage workflow. Results indicate that building height primarily determines visibility conditions, while façade color, material contrast, and positional prominence mediate whether visibility translates into perceived disturbance. Empirical analysis identifies 21 buildings with significant impact, 43 with mild impact, and 28 with negligible impact. The findings highlight the need to move beyond two-dimensional height control toward three-dimensional, perceptual-based heritage landscape management.

Keywords. heritage impact assessment, visual impact, historic urban district, eye-tracking, high-rise buildings

1. Introduction

In recent years, the visual conflict between high-rise buildings and historic urban areas has become a major concern in international heritage conservation (World Heritage Centre, 2002, 2003). In China, however, heritage protection started relatively late, and actual planning practices are often influenced by land acquisition and local economic pressures in urban renewal. Consequently, height-control regulations are frequently exceeded, especially in heritage-concentrated core areas, where the form, color, and massing of high-rise buildings disrupt the visual integrity of historic landscapes (Yang et al., 2022; Zhang, 2017). Unlike many countries where early-stage Heritage Impact Assessments effectively regulate high-rise development, China faces the reality that numerous high-rise buildings have already been constructed and are difficult to remove. This makes visual mitigation and micro-scale design adjustments an increasingly relevant approach.

These conditions highlight a fundamental question: Do high-rise buildings inevitably exert negative visual impacts on historic urban areas? Recent studies indicate that their

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 61-70.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

influence is not determined by height alone; color, massing, façade composition, and positional relationships in the visual frame all shape public perception of heritage settings (Liu, 2022; Serra et al., 2021). Thus, high-rise buildings should be scientifically evaluated rather than uniformly rejected. Key issues arise: which visual factors generate disturbance, which are manageable, and how can existing high-rise buildings be better integrated into heritage views?

To address these questions, this study develops a context-specific visual impact assessment framework for Chinese historic urban areas. Taking Hanzhong as a case study, the research draws upon the Guidance on Heritage Impact Assessments (2011, 2022) and Seeing the History in the View (2012), integrating Grasshopper-based viewshed analysis, eye-tracking experiments, and Semantic Differential (SD) surveys. The proposed perception-centered workflow identifies the visual characteristics and disturbance patterns of high-rise buildings within heritage viewsheds, providing actionable guidance for high-rise management and visual optimization in historic urban areas.

2. Methodology

2.1. METHODOLOGICAL FRAMEWORK AND RESEARCH DESIGN

According to relevant international guidelines, the standard HIA process consists of three stages: investigation and understanding, assessment and diagnosis, and mitigation and strategy development. Based on this framework, the present study develops a localized operational process accordingly (see Fig. 1).

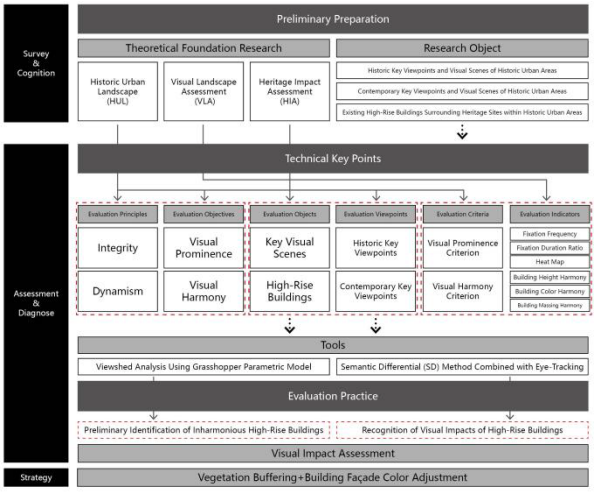


Figure 1.Procedure of Heritage Visual Impact Assessment

In the first stage, statutory protection plans, heritage inventories, building height data, and GIS information are collected to construct a 3D model of Hanzhong's historic urban area. An ideal height-and-form model is then created based on protection requirements to identify buildings exceeding allowable limits. Historical images, field

surveys, and planning documents are used to determine evaluation point locations, followed by standardized photography. In the second stage, over-height buildings are overlaid with evaluation-point viewsheds. Parametric viewshed analysis in Grasshopper removes those not visually observable. Standardized view images are produced, and eye-tracking experiments capture public visual attention toward both high-rise buildings and heritage objects. Combined with SD questionnaire results, this stage identifies the locations and characteristics of buildings with varying levels of visual impact. In the third stage, a visual impact classification system is established to diagnose disturbances at key viewpoints, and targeted mitigation strategies are proposed accordingly.

2.2. THEORETICAL FOUNDATIONS

VIA originates from landscape perception and visual cognition theories, emphasizing that urban landscape evaluation is shaped by both spatial form and human perception (Qi, 2013). In heritage conservation, previous studies have widely recognized VIA as an effective tool for managing visual conflicts caused by high-rise development and safeguarding visual integrity in historic settings (Ashrafi et al., 2021; Serra et al., 2021).

Traditional perception-based methods, including questionnaires and expert scoring, rely heavily on subjective judgments and offer limited insight into how specific spatial elements shape visual perception (Mundher, 2022). Meanwhile, digital techniques such as GIS-based viewshed and sightline analyses quantify physical visibility and visual prominence but remain insufficient in capturing observers' actual perceptual experience (Kloos et al., 2019; Shen et al., 2020).

Eye-tracking methods provide objective measurements of visual attention by capturing fixation distribution and salience patterns (Li, 2022; Li Yuan et al., 2022). Recent studies combining eye-tracking with visual computation further demonstrate that façade attributes—such as color contrast, brightness, and surface continuity—play a more decisive role in attracting visual attention than building height alone (Liu et al., 2025). However, eye-tracking alone cannot explain why specific elements attract attention or distinguish the perceptual attributes underlying visual disturbance. To address this limitation, Semantic Differential (SD) questionnaires are employed as a complementary interpretive layer. By integrating objective attention metrics with subjective perception, this combined approach enables clearer interpretation of visual impact mechanisms.

Therefore, under the frameworks of HIA and VIA, this study integrates parametric modeling, eye-tracking analysis, and SD questionnaires to form a comprehensive methodology that links spatial quantification with perceptual and psychological validation. This approach offers more precise and operational tools for understanding the visual impact mechanisms of high-rise buildings in historic urban areas.

2.3. DATA ACQUISITION AND MODEL CONSTRUCTION

The study draws on four data sources: statutory planning documents, urban geographic and building datasets, historical and contemporary images, and on-site surveys. Using these materials, a 3D urban model was constructed in Rhino and Grasshopper, and all GIS datasets were unified under the WGS84 coordinate system to ensure spatial

consistency across sources, comprising:

- An Ideal Model, reconstructed from protection plans and height-control regulations, representing planned visual conditions;
- an Existing Model, derived from current building heights and recent imagery, reflecting actual urban form and revealing buildings that exceed height limits(he vertical accuracy of building height information is estimated to be within an error range of approximately ±2 m).

Comparing the two models allowed preliminary identification of potential visual conflicts. Field surveys and in-situ photographs were further used to validate model accuracy and supplement spatial information.

2.4. EVALUATION POINT SELECTION AND VIEWSHED ANALYSIS

2.4.1. Evaluation Point Selection

Following the approach recommended in Seeing the History in the View by Historic England, this study identifies evaluation points based on the landscape control requirements of the Hanzhong Historic City Conservation Plan, historical photographs, and contemporary street-view images, while also considering public activity spaces and major visual corridors. A total of 41 points were selected, classified into: A-type (heritage-dominant viewpoints), B-type (riverfront and mountain-facing viewpoints), and C-type (urban public-space viewpoints).

These points together form a systematic visual observation network, providing the basis for subsequent quantitative analysis and perception experiments. The detailed selection workflow is shown in Fig. 2.

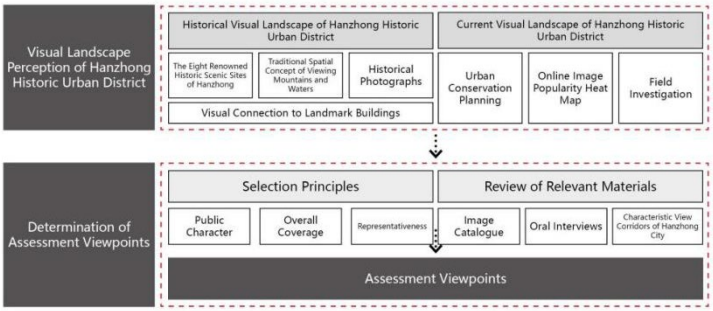


Figure 2.Viewpoint Selection

2.4.2. Viewshed Analysis

After the evaluation points were determined, a parametric viewshed analysis was conducted using the Rhino–Grasshopper platform. A "view cone" was employed to simulate the observer's visual field, and the Viewshed and Weaverbird plugins were used to identify high-rise buildings that are actually visible from these viewpoints. Buildings that exceed height limits but do not fall within any viewshed were excluded.

This process generated the final set of high-rise buildings that needed to be included in the visual perception experiments, ensuring that the study focuses only on structures that produce visual impacts from real human viewpoints(Fig. 3).



Figure 3.Calculation of a schematic diagram of high-rise buildings in view by Grasshopper

2.5. EYE-TRACKING EXPERIMENT AND SD QUESTIONNAIRE SURVEY

2.5.1. Experimental Materials and Sample Selection

The experimental samples were selected based on the viewshed analysis results from the second stage. Several representative views were chosen from the 41 evaluation points. The static stimuli used in the eye-tracking experiment consisted of current images of the historic urban areas of Hanzhong. All photographs were taken horizontally at a height of approximately 1.5 meters, with a focal length of 35 mm to approximate human vision. The final experimental material included 12 static view images, comprising composite scenes containing both high-rise buildings and heritage landscapes, as well as their corresponding comparison views.

2.5.2. Experimental Design and Implementation

Eye-tracking data were collected using a Tobii Pro device from 24 participants—14 heritage-related university students and 10 Hanzhong residents. Each image was displayed for 8 seconds, and the recorded metrics included Fixation Frequency (FF), Proportion of Fixation Duration (PFD), Average Fixation Duration (AFD), First Visit Time (FV), and Heat Maps (HM).

After the eye-tracking session, participants completed a Semantic Differential (SD) questionnaire, with items grouped into A/C-type heritage viewpoints and B-type street-space viewpoints. All indicators were rated on a seven-point scale to quantify perceptions of harmony, historical atmosphere, and the visual impact of high-rise buildings, providing essential support for subsequent analysis.

2.6. RESULT INTERPRETATION AND STRATEGY

Based on viewshed analysis and integrated perception results, a visual impact classification system was established, considering building visibility, visual prominence, and subjective perception from eye-tracking and SD data. High-rise impacts on heritage landscapes are categorized into three levels:

Level I: Acceptable Change — buildings are visible but cause minimal disturbance; Level II: Moderate Disturbance — partial skyline disruption or localized visual pressure; Level III: Significant Disturbance — strong conflicts in height or color that shift visual focus and weaken heritage integrity.

Corresponding mitigation strategies were proposed: Level III requires height adjustment, massing redesign, or façade modification; Level II can be improved through color correction, vegetation screening, or skyline recomposition. Level I represents visually compatible cases that may inform future urban design on form and proportion control.

3. Result&Conclusion

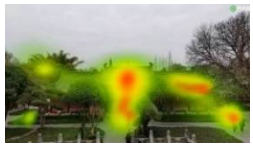
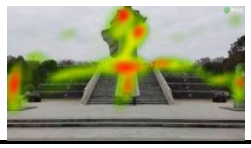
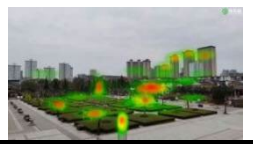
3.1. RESULTS OF URBAN MODEL CONSTRUCTION AND VIEWSHED ANALYSIS

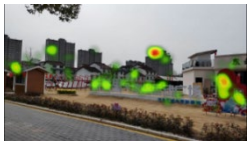
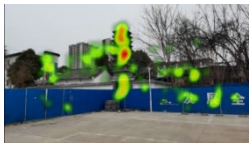
To ensure representativeness and scientific validity, this study integrated historical and contemporary viewpoints from Hanzhong's historic urban area. Grasshopper's viewshed analysis was then used to compute visual coverage for each evaluation point. By overlaying viewsheds with the distribution of over-height buildings, only high-rise buildings actually visible from key viewing directions were retained, while those outside all viewsheds were excluded. A total of 92 visually prominent buildings were identified as the final research sample, providing a clear spatial scope and object basis for subsequent eye-tracking and perception evaluations.

3.2. EYE-TRACKING EXPERIMENT & SD RESULTS

The representative eye-tracking heatmaps and corresponding gaze metrics are presented in Tables 1 and 2.

Table 1.Heatmaps Distribution of Representative Evaluation Points

Evaluation Point	Eye-Tracking Heatmaps		
A-Type Evaluation Points			
	A-1	A-5	A-8
			

B-Type Evaluation Points	A-10 	A-11 	A-14 
C-Type Evaluation Points	B-1 	B-5 	B-8 
	C-2	C-4	C-13

To compare the visual salience of different evaluation points, four eye-tracking metrics—Fixation Frequency, Proportion of Fixation Duration, Average Fixation Duration, and First Visit Time—were standardized using Z-scores to derive their relative levels across all viewpoints. The first three metrics were treated as positive indicators of visual salience, whereas First Visit Time was reverse-coded (shorter times indicating higher salience). The standardized values of the four indicators were then summed with equal weights to produce a composite Visual Salience Index (VSI). A higher VSI indicates that the heritage more readily attracts visual attention, whereas a lower VSI suggests weaker visual interference(See Table 2.). The calculation formula is as follows:

$$VSI_i = (Z_{FF,i} + Z_{PFD,i} + Z_{AFD,i} + (-Z_{FV,i}))$$

Table 2.Eye-Tracking Metrics, Z-Scores, and the Visual Salience Index (VSI)

Index	Fixation Frequency	Proportion of Fixation Duration	Average Fixation Duration/s	First Visit Time/s	Corresponding Evaluation Point	Visual Salience Index
1	0.037	0.421	0.477133	0.2976	A-10	0.6375
2	0.054	0.152	0.1823333	0.3215	A-5	0.0668
3	0.023	0.301	0.2985	0.138333	A-11	0.4842
4	0.047	0.38	0.2085	0.219666	A-8	0.4158
5	0.057	0.443	0.176166	0.2005	B-1	0.4757
6	0.046	0.485	0.214	0.400333	B-8	0.3447
7	0.0501	0.358	0.211333	0.331833	B-5	0.2876
8	0.00195	0.243	0.506333	1.255333	A-14	−0.5041
9	0.0501	0.263	0.187333	0.3995	C-13	0.1009
10	0.046	0.421	0.217667	0.11983	C-2	0.5648
14	0.046	0.543	0.2196667	0.5385	C-4	0.2702
18	0.0301	0.31	0.2012	0.1886	A-1	0.3527

SD perception scores (SDI) were calculated by averaging all SD indicators for each

viewpoint. The results show a clear contrast: Level I viewpoints (A-1, A-5, A-8, A-10, A-11, B-8) all exhibit positive SDI values, indicating higher visual harmony, while Level II–III viewpoints (A-14, B-5, C-2, C-4, B-1, C-13) show negative SDI, reflecting stronger perceived disturbance. These patterns closely match the eye-tracking results, confirming the reliability of the visual impact classification(Fig. 4).At the broader scale, 21 buildings were classified as having a significant impact, 43 as mild, and 28 as negligible across the full viewshed, consistent with the Level I–III distinction.

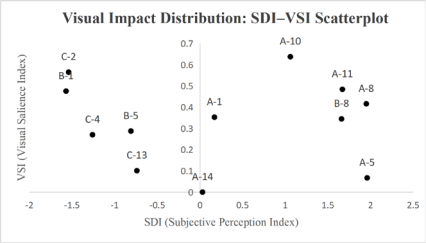


Figure 4. SDI–VSI Scatterplot for Visual Impact Classification

To further interpret the mechanisms underlying the three-level classification, representative viewpoints were examined by jointly comparing VSI values, SDI scores, and façade-related SD indicators(overall visual harmony, historical atmosphere, façade color contrast, material compatibility, visual prominence of high-rise buildings within the scene, and façade scale–proportion compatibility). Based on the VSI–SDI results across viewpoints, three patterns clarify the mechanism:

- Positive SDI with both low and high VSI (e.g., A-5, A-8, A-10) indicates that visual salience does not necessarily cause disturbance when façade color, material, and scale are compatible.
- High VSI with negative SDI (e.g., B-1, C-2) shows that strong salience amplifies negative evaluation when façade attributes conflict with the heritage context.
- Low VSI with negative SDI (e.g., B-5, C-13) suggests that even weakly salient buildings can impair overall perception through background incompatibility.

Overall, the results indicate that visual salience (VSI) and perceptual evaluation (SDI) represent two distinct yet interacting cognitive stages. Building height primarily determines whether high-rise elements become visually noticeable, but salience alone does not directly translate into perceived disturbance. When façade color, material compatibility, scale–proportion relationships, and positional prominence are well controlled, even visually salient high-rise buildings can be cognitively integrated into the scene, resulting in positive SDI values. Conversely, when façade-related attributes are incompatible with the heritage context, increased visual salience amplifies negative evaluation, leading to stronger perceived disturbance. In this sense, height establishes visibility, while façade design mediates whether visibility becomes evaluative conflict.

4. Discussion

This study, through integrated eye-tracking and SD evaluations, clarifies the multi-dimensional and stage-based nature of visual impacts caused by high-rise buildings in historic urban areas. The results show that height alone is not the decisive factor: several over-height buildings caused little disturbance, while some mid- and low-rise structures produced strong visual conflicts due to color, signage composition, or façade materials. This indicates that height primarily determines visibility conditions, while façade characteristics and spatial positioning mediate whether visibility becomes perceived disturbance.In this framework, eye-tracking captures attention allocation, while SD indicators explain perceptual interpretation.

At the practical level, the proposed workflow offers a shift from conventional two-dimensional control to three-dimensional visual management. By combining viewshed analysis with perceptual data, it enables the identification of "actually visible" impacts during urban renewal, providing a more refined basis for regulating color, façade details, material reflectivity, and the presentation of buildings in key viewpoints. This enhances the operability of heritage conservation, from design review to micro-scale interventions.

Limitations remain. The sample size is modest, and differences between participant groups were not examined in detail. In addition, screen-based eye-tracking experiments cannot fully reproduce embodied visual experience during real movement through urban space. Future research may incorporate VR/AR-based dynamic path experiments and expand participant samples to strengthen validity. More importantly, further studies should focus on modeling and quantifying façade-related attributes—such as building massing, proportion, color contrast, material reflectivity, window-to-wall ratio, and façade rhythm—and exploring their relationships with visual salience and perceptual disturbance. Such efforts would enable the development of parametric, data-driven tools for evaluating new and existing buildings in heritage contexts and facilitate the integration of visual impact assessment into urban design guidelines and management practices.

Attribution Statement

ChatGPT (OpenAI, 2026) was used to (1) suggest the initial structure of the manuscript and (2) to improve flow.

References

Ashrafi, B., Kloos, M., & Neugebauer, C. (2021). Heritage Impact Assessment, beyond an Assessment Tool: A comparative analysis of urban development impact on visual integrity in four UNESCO World Heritage Properties. *Journal of Cultural Heritage*, 47, 199-207. https://doi.org/10.1016/j.culher.2020.08.002

Kloos, M., Seyedashrafi, B., & Tebart, P. (2019). Heritage Impact Assessment–Heumarkt Neu construction project and development of the World Heritage property Historic Centre of Vienna. Aachen, Alemanha, Michael Kloos Planning and Heritage Consultancy [Versão electrónica]. Acesso em 10-07-2022, from: https://www. bmkoes. gv. at/dam/jcr: 6853b892-55ee-4d1b-903ad9ca565ede74/HIA_en. pdf], 250.

Li, N., Zhang, S., Xia, L., & Wu, Y. (2022). Investigating the visual behavior characteristics of architectural heritage using eye-tracking.
Buildings, 12(7), 1058.
https://doi.org/10.3390/buildings12071058

Li, Y., Huang, J., Liang, J., Zhang, Y., & Chen, Y. (2022). Visual attractiveness and perceptual factors of commercial streets in heritage sites: A case study of Longtou Road in Gulangyu.
Journal of Western Habitat Environment, (2), 114–121.

Liu, F., Kang, J., Wu, Y., Yang, D., & Meng, Q. (2022). What do we visually focus on in a World Heritage Site? A case study in the Historic Centre of Prague.
Humanities and Social Sciences Communications, 9(1).
https://doi.org/10.1057/s41599-022-01411-1

Liu Zixuan, Glory & Gong Hang. (2025). Visual quality analysis of ground floor space based on visual calculation and eye movement. Journal of Hebei University of Engineering (Natural Science Edition), 42 (05), 104-112.

Mundher, R., Al-Sharaa, A., Al-Helli, M., Gao, H., & Abu Bakar, S. (2022). Visual quality assessment of historical street scenes: A case study of the first “Real” street established in Baghdad.
Heritage, 5(4), 3680-3704.
https://doi.org/10.3390/heritage5040192

Qi, T., Wang, Y., & Wang, W. (2013).A review of international visual landscape research.
Progress in Geography, 32(6), 975–983.

Serra, J., Iñarra, S., Torres, A., & Llopis, J. (2021). Analysis of facade solutions as an alternative to demolition for architectures with visual impact in historical urban scenes.
Journal of Cultural Heritage, 52, 84-92.
https://doi.org/10.1016/j.culher.2021.09.005

Shen, Y., Wu, L., Ma, J., & Zhou, X. (2020).The “grid” above the old city: A digital-technology-based height control system for historic urban areas—A case study of Taiyuan Old City.
Architectural Journal, (6), 80–85.
https://doi.org/10.19819/j.cnki.ISSN0529-1399.202006016

World Heritage Centre. (2002, June 29). Report for the state of conservation of World Heritage (No. 2616). UNESCO World Heritage Centre. Retrieved November 20, 2021, from https://whc.unesco.org/en/soc/2616

World Heritage Centre. (2003, July 5). Report for the state of conservation of World Heritage (No. 2755). UNESCO World Heritage Centre. Retrieved November 20, 2021, from https://whc.unesco.org/en/soc/2755

Wen, J., & Feng, Y. (2022).Application of landscape visual impact assessment in urban regeneration: Insights from England.
Urban and Rural Planning, (4), 78–92.

Yang, B., Hu, M., & Fu, B. (2022).Forty years of historic city conservation: Review and outlook.
World Architecture, (12), 16–19.
https://doi.org/10.16414/j.wa.2022.12.002.

Zhang, S. (2017).Sustainable conservation of urban historic environments.
International Urban Planning, 32(2), 1–5.

Low-Cost Mobile LiDAR and RTK GNSS Integration for Georeferenced HBIM of Cultural Heritage

A Scan-to-BIM Workflow Based on Multi-Sensor Data Fusion for Accuracy Enhancement

Handan Aş Çemrek[1] and Ümit Işıkdağ[2]
[1]*Istanbul Health and Technology University*
[2]*Mimar Sinan Fine Arts University*
[1]*handan.as@istun.edu.tr, 0009-0007-7369-8576*
[2]*umit.isikdag@msgsu.edu.tr, 0000-0002-2660-0106*

Abstract. This study discusses a scan-to-BIM workflow based on the integration of low-cost mobile LiDAR and RTK GNSS, focusing on cultural heritage, and compares the performance of two different portable scanning systems. An 18th-century Sibyan School was selected as a case study, managed by Mimar Sinan Fine Arts University (MSGSU), Istanbul. The study employed a viDoc RTK-enabled iPad Pro (photogrammetry + LiDAR) and a Realsee Galois M2 tripod-mounted panoramic 3D LiDAR scanner. Point cloud data obtained from both systems were georeferenced with RTK GNSS control points, and their location accuracies were comparatively assessed. The generated point cloud datasets were combined using a multi-sensor data fusion approach and fed into an HBIM-based modeling process. Component geometries were analyzed for detail, data gaps, and noise. The findings indicate that the RTK-enabled viDoc method offers advantages in terms of location accuracy and level of detail, while the Realsee Galois M2 stands out for its rapid deployment, indoor coverage, and practical field use. In conclusion, the study suggests that low-cost mobile systems should be considered as part of a multi-sensor hybrid cultural heritage documentation and digital twin production strategy rather than completely replacing terrestrial laser scanning and suggests a repeatable workflow for similar projects.

Keywords. Mobile LiDAR, RTK GNSS, HBIM, Scan-to-BIM, Data Fusion, Cultural Heritage, Digital Twin, Sibyan School, MSGSU

1. Introduction

Digital documentation of cultural heritage structures is critical for protecting them from disasters and destruction, and for their preservation for future generations. Digital twin and Historic Building Information Modeling (HBIM) approaches provide three-dimensional, semantic, and high-fidelity representations of heritage, providing decision

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 71-80.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

support in restoration, maintenance, and virtual exhibition processes (Murphy et al., 2009). ICOMOS (2020) and UNESCO (2021) promote HBIM models with high data accuracy and compatibility with open standards such as IFC/CityGML.

While traditional Terrestrial Laser Scanning (TLS) offers high accuracy, it is limited in terms of cost and flexibility (Kuçak et al., 2023). In contrast, mobile LiDAR sensors (e.g., iPad Pro) and SLAM-based systems provide low-cost and rapid data collection (Teppati Losè et al., 2022). However, absolute location accuracy decreases without GNSS support (Ubaidillah et al., 2023). Integration of GNSS receivers such as the Pix4D viDoc RTK Rover can improve the georeference accuracy of mobile LiDAR scans, providing positional accuracies close to those of traditional systems (Zollini & Marconi, 2025). Consumer-grade Matterport systems, with a dimensional accuracy of approximately 1% (±10 cm at 10 m), offer time and cost advantages in studies that do not require high precision (Matterport, 2025b; Fiorini et al., 2023; Hariffin et al., 2023; Wang et al., 2025). Mobile SLAM-based systems (iPad with viDoc RTK, Matterport Pro3, Realsee Galois, XGRIDS Lixel L2 Pro, etc.) introduce a new paradigm in terms of portability and speed (Pix4D, 2022; Realsee, 2025b). Literature shows that these low-cost mobile systems provide sufficient detail and accuracy in most scenarios. This study examines a scan-to-BIM workflow based on the integration of mobile LiDAR and RTK GNSS data at the cultural heritage scale. The aim is to georeference point cloud data collected with low-cost devices to the centimeter level to produce HBIM models and increase documentation accuracy. In this context, a repeatable and scalable method for historical structures is proposed by combining mobile scanning and high-precision positioning techniques. The following sections summarize related studies and technologies and evaluate the XGRIDS Lixel L2 Pro. A comparative study of the same sample structure using viDoc RTK + iPad Pro and Realsee Galois M2 is then presented, and data quality and HBIM integration are discussed.

2. Literature Review Mobile 3D Scanning Technologies

2.1. DEVELOPMENT OF MOBILE SCANNING METHODS

Traditional fixed Terrestrial Laser Scanners (TLS) have been reliable in detailed measurement of cultural heritage, but their high cost, heavy equipment, and multi-stage recording processes require time and expertise (Kuçak et al., 2023). These limitations have led to the rapid development of mobile laser scanning approaches that offer more flexible solutions in the field. Handheld/shoulder-type SLAM scanners and smartphone-based LiDAR systems reduce scanning time by allowing the operator to move freely, enabling measurements even in hard-to-reach areas.

This transformation has accelerated with the integration of LiDAR sensors into iPhone and iPad models after 2020; these sensors can generate 3D point cloud data at close range. Teppati Losè et al. (2022) demonstrated that the iPad Pro LiDAR sensor can perform rapid scans with a few centimeters of accuracy in cultural heritage measurement. However, drift accumulation due to IMU-based positioning can lead to deviations in large or long-term scans. To address this issue, solutions such as the viDoc RTK mounted on a smartphone significantly reduce drift by stamping absolute coordinates onto each frame with an RTK-GNSS receiver, providing global positioning accuracy of 1–3 cm under favourable conditions. Llabani & Kopliku

(2024) reported that an Ottoman bridge could be modeled with a precision of ±3 cm using a viDoc-enabled iPhone and that an archaeological site showed a difference of <1 cm from the TLS reference; this demonstrates that mobile scanning can compete with traditional instruments.

Another advantage of mobile scanning is the speed of data collection. The ability to scan while moving continuously allows for the documentation of larger areas at the same time. However, smartphone-based mobile scanning has limitations in terms of accuracy and noise: periodic position corrections are recommended for large areas and loop closure strategies in indoor environments. It is emphasized that mobile LiDAR points can be sparser and noisier than those of professional scanners, resulting in better results when combined with photogrammetry (Jeftha & Shoko, 2024).

2.2. INTEGRATED LIDAR CAMERAS (MATTERPORT VS REALSEE)

The Realsee Galois offers advantages such as long battery life (up to 8 hours), fast scanning time (93 m² scanned in ±15 minutes), and immunity to sunlight (capable of scanning outdoor historical structures). Most importantly, the scanning application allows previewing and merging data in the field; this provides a more flexible and controllable workflow compared to Matterport, which only enables cloud processing (Realsee, n.d.; Matterport, 2025c). The most striking aspect of Realsee is its open data outputs and ease of integration. E57/PLY point cloud files, OBJ/FBX mesh models, 2D floor plans (PNG), and 360° panoramas are automatically generated from data uploaded to the cloud; these outputs can be directly imported into software such as Autodesk Revit, AutoCAD, or CloudCompare. The business model is designed to allow device owners to upload and process unlimited scans free of charge while maintaining data ownership by charging only low-cost fees for web hosting/sharing (RobotShop, n.d.). In conclusion, the Realsee Galois M2 offers significant advantages over Matterport in terms of speed, ease of use, and open data outputs; Although its measurement accuracy lags that of classical TLS, it constitutes an accessible documentation alternative for a wide range of users (Realsee, n.d.; Matterport, n.d.).

2.3. SLAM BASED SCANNERS (XGRIDS LIXEL L2 PRO)

Although not field-tested in this study, the XGRIDS Lixel L2 Pro is a SLAM-based handheld LiDAR system developed for professional mapping and large-scale 3D scanning. Thanks to LiDAR–IMU–camera sensor fusion, it generates real-time point cloud data while the operator is on the move. Its 32-channel rotating LiDAR unit achieves a range of up to 300 m and positional accuracy of 2–3 cm. Its lightweight, ergonomic, and modular design (RTK kit, extension rod, vehicle/drone mount) adapts to various scanning scenarios, from single monuments to large historical urban areas (Epotronic, n.d.). The LixelStudio software within the XGRIDS ecosystem combines and filters raw data to produce LAS, E57, or mesh outputs. Lixel CyberColor Studio, based on 3D Gaussian Splatting, enables rapid export of photorealistic 3D models to web, Unity/Unreal, and AR/VR environments (Epotronic, n.d.). This allows the workflow from mobile scanning to a virtual navigable model to be largely automated.

While L2 Pro's direct cultural heritage applications are limited, similar SLAM-based systems demonstrate the technology's potential. Giuliani et al. (2024)

documented the Pisa city walls with mobile LiDAR (GeoSLAM Horizon) and RTK GNSS to produce an HBIM model with ±3–5 cm accuracy. Tanduo et al. (2023) achieved ±3 cm accuracy in the basement of Castello del Valentino. These examples demonstrate the potential of systems like L2 Pro to accelerate scan-to-HBIM processes by providing sufficiently accurate data in significantly less time than TLS, provided appropriate data fusion and control points are implemented; however, instrument cost and computational requirements remain significant constraints.

3. Scanning Comparison with ViDoc and Realsee

This section presents a comparative analysis of scans taken with viDoc RTK + iPad Pro and Realsee Galois M2 devices on the 18th-century Sibyan School and Fountain. The case study site is managed by Mimar Sinan Fine Arts University (MSGSU). Field access and on-site data acquisition were conducted with institutional authorization. The two-story structure, constructed in alternating stone and brick, has recently undergone restoration, and detailed documentation is crucial (Kültür Envanteri, n.d.).

3.1. DATA COLLECTION

Scanning was conducted using viDoc RTK hardware and Pix4Dcatch software, using the iPad Pro's LiDAR sensor, with the operator walking around the structure. Simultaneous RTK GNSS measurements (with NTRIP corrections) achieved centimeter-level accuracy, and the data was linked to the ITRF-based TUREF (Treccani et al., 2024). Thanks to the ViDoc RTK Rover's precise geotagging of each image frame, hundreds of photos and millions of LiDAR points obtained during the approximately one-hour scan were directly georeferenced. Additional photogrammetry shots were taken with a DSLR camera to document facades and interiors in one go, without interruption. Gaps identified during the instant preview during the scan were corrected with additional passes.

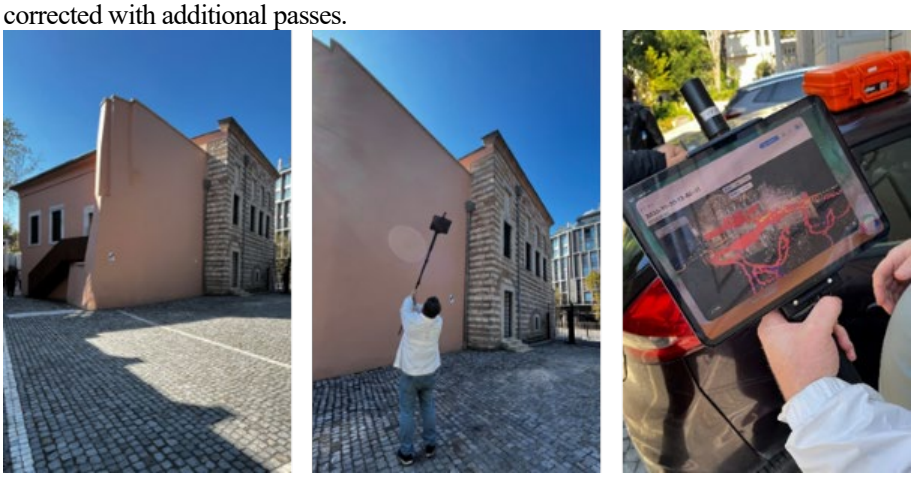


Figure 1. ViDoc – Pix4Dcatch Data Collection

The second scan was performed using a Realsee Galois M2, a static 360° multi-position scan. Sixty-seven tripod positions were identified to cover all facades and

accessible interior spaces of the building. At each location, the device generated a color preview point cloud after internal pre-processing, and the interface provided feedback to the operator on the adequacy of overlap during movement between setups. Once all locations were complete, adjacent scans were automatically aligned using common points to create a unified point cloud. Total field time, including setup and location changes, was approximately two hours. During the sunny afternoon shooting, the panoramic camera's HDR feature ensured balanced exposures. Thanks to the LiDAR system's protection from sunlight, no significant degradation in data quality was observed.



Figure 2. Realsee Galois M2 Data Collection

Neither data collection method physically marked the structure. The viDoc + iPad approach produced direct georeferenced data in a short time thanks to free navigation, while the Realsee method provided high-resolution static scans at each location in exchange for longer field time. The viDoc scan recorded data directly into the national coordinate system using RTK GNSS. Because Realsee does not have an onboard GNSS, three RTK GPS control points near the structure were measured and used to convert the Realsee point cloud to the ITRF-based coordinate system. This conversion process allows both datasets to be compared within the same reference system.

3.2. POINT CLOUD PROCESSING

Point cloud datasets obtained from different sensors were subjected to rough alignment, noise filtering, density optimization, and georeferencing using open-source software such as CloudCompare (Çemrek et al., 2025). DSLR/drone photogrammetry data was integrated with the LiDAR cloud to enhance both geometric accuracy and visual realism, and color and texture were added (Ubaidillah et al., 2023).

For processing the viDoc data, RGB images collected from the iPad Pro and LiDAR depth data were transferred to Pix4Dmatic. Thanks to the precise position/direction information provided by viDoc, photogrammetric alignment was completed quickly, and LiDAR depth information was integrated into the solution to strengthen the model on surfaces where photogrammetry alone was challenging. In

Pix4Dmatic, using the "fused" point cloud option, the RGB-based dense point cloud and LiDAR points were combined into a single cloud, noise filtered, and exported in .las format (Pix4D, 2022). This process resulted in both a colored point cloud and a triangular surface mesh for visualization.

Despite the device's pre-merging, the final solution for processing the Realsee data was performed using the manufacturer's cloud-based Realsee Console platform. The platform automatically saved the multi-position scans and generated a full-resolution combined point cloud. After processing, the .e57 (colored point cloud), .obj (3D surface model), and 360° images/floor plan were downloaded. Although the Realsee point cloud was sparser than the viDoc cloud (approximately 1–2 cm point spacing), it was observed that both scans successfully represented the building geometry, demonstrating a high degree of overlap when overlaid in the same coordinate system.

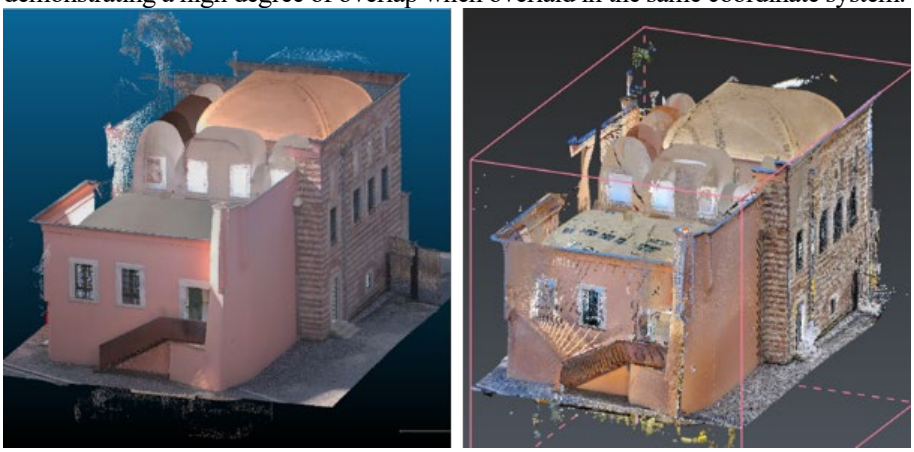


Figure 3. Left side viDoc data displayed in CloudCompare .las format, Right side Realsee data displayed in Cyclone 3DR Viewer .e57 format

3.3. VERIFICATION AND OUTPUT

When compared to RTK GNSS points measured in the field, viDoc RTK supported iPad scanning provided high absolute accuracy for a mobile system, staying within ±2 cm at critical points (Pix4D, 2022; Zollini & Marconi, 2025; Llabani & Abazaj, 2024). In Realsee Galois M2, similar reference accuracy was achieved by aligning according to three RTK points instead of in-device positioning. In local precision, ±1–2 cm noise was observed on smooth walls in the viDoc + photogrammetry cloud, while the Realsee cloud provided a more homogeneous and smooth geometry thanks to the calibrated LiDAR sensor (Realsee, n.d.; Treccani et al., 2024). However, while point density decreased in shadowed areas in the Realsee scan, the ability of the operator to physically approach the blind spots mitigated this problem in the viDoc method. As a result, both methods provided accuracy at the centimeter level; viDoc is advantageous in absolute position, while Realsee is advantageous in surface continuity and smoothness.

Thanks to high-resolution photographs, the viDoc + iPad method captured millimetric details such as marble inscriptions and stone joints with an average point

spacing of 4 mm. While Realsee's LiDAR data is limited to 1–2 cm with single-point precision, the continuity of the façade geometry is complete thanks to the approximately 44 million points generated from 67 scans. While the viDoc point cloud exhibits dense distribution at close range and sparse distribution at a distance, the integration of the photogrammetry cloud filled in the gaps by increasing the combined cloud to 108 million points. The colored point cloud data of both methods are sufficient to read material distinctions, but the viDoc model exhibits minor color seaming errors due to multiple photographs, while the Realsee model, thanks to HDR panoramas, produces more uniform tones and more controlled shadow/glow effects (Pix4D, 2022; Realsee, n.d.).

At the end of the viDoc + Pix4D workflow, a point cloud in LAS format of approximately 3.5 GB was generated. In addition, photogrammetric OBJ mesh and project files were obtained. The .e57 point cloud downloaded from the Realsee Console is approximately 3 GB in size and comes with HDR panoramas (JPEG) and a textured OBJ mesh. The fact that both methods offer industry-standard formats ensures compatibility issues during subsequent analysis and HBIM modeling stages. Realsee's automatic ground detection and scaled 2D floor plan generation offer a significant additional advantage (Realsee, n.d.), while a similar plan view was obtained by manually sectioning the point cloud on the viDoc side. These generated datasets were validated with TLS reference scanning and converted into an integrated HBIM data package containing plan-section-views, material inventory, and AR/VR integration files (Giuliani et al., 2024).

3.4. HBIM INTEGRATION

Point cloud datasets obtained from both mobile scanning systems are fully configurable into HBIM models using a similar workflow. The geo-referenced LAS point cloud generated with ViDoc + Pix4D was automatically aligned with the Turkish national coordinate system (ITRF-based), converted to .rcp format with Autodesk ReCap Pro, and imported into Autodesk Revit. The Realsee .e57 data, initially with local coordinates, was merged with CloudCompare to 67 scans. Then, the data was aligned with control points, exported to LAS format, and similarly converted to RCP using ReCap before being imported into Revit. This allowed the two data sets to be superimposed within a single project. Using the imported point clouds in the Revit environment as reference, it is possible to perform geometric and parametric modeling at the LOD 200–300 level, classify architectural components according to the IFC standard, and add material information and typological features to the model's semantic database (Rocha et al., 2020; Çemrek et al., 2025). This geo-referenced study offers the potential to transform the generated HBIM model into a digital twin readable by GIS systems, with its real location and environmental context. From a GeoBIM perspective, GIS-readable digital twins typically depend on open standards (IFC, CityGML, LADM), yet interoperability remains contingent on consistent georeferencing because BIM commonly uses local coordinates while geoinformation is defined in geographic coordinate frameworks, making cross-domain conversion unavoidable (Abdul Rahman, et al., 2026). In modeling studies, viDoc data provides detailed visual reference with its photo-based structure, while Realsee data, with its homogeneous colors from HDR panoramas, can be used for textural interpretation. The high accuracy

achieved enables damage and deformation analyses; exporting the generated model in IFC format enables compliance with the open data standard and can be easily transferred to the VR environment and used in interactive presentations.

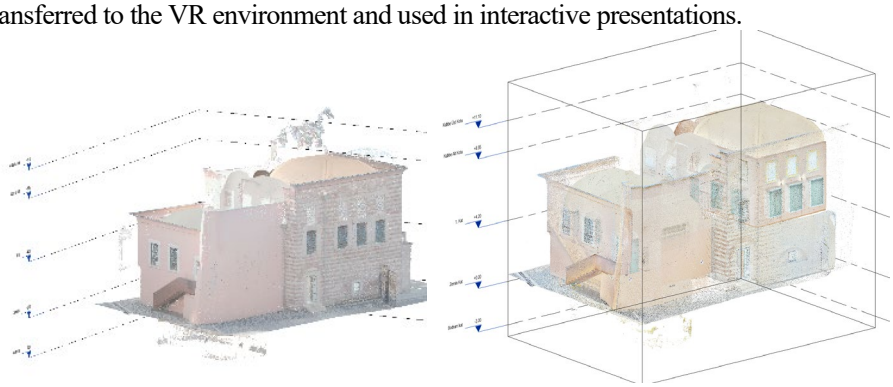


Figure 4. Comparison of ViDoc and Realsee Point Cloud in Revit

In conclusion, the Scan-to-BIM workflow has been successfully implemented with both viDoc and Realsee data, and both low-cost methods have been tested to produce HBIM outputs of sufficient quality for conservation projects.

4. Conclusion and Future Work

This study evaluated the usability of mobile SLAM-based systems in cultural heritage documentation using data acquired with the viDoc RTK-enabled iPad Pro and the Realsee Galois M2. While both systems offer low-cost and portable solutions, they also have distinct strengths. viDoc, with its centimeter-level absolute positioning accuracy thanks to its RTK GNSS integration and high detail with photogrammetry support, is particularly suitable for restoration and deformation monitoring applications requiring precise measurements. In contrast, its user-controlled workflow requires technical expertise and longer data processing times. Realsee Galois M2, with its automatic registration, cloud-based processing, and direct E57/OBJ output, stands out for its ease of use and speed. It offers a suitable tool for rapid documentation, virtual tours, and presentations to large audiences. However, its lack of absolute positioning and limited accuracy may necessitate calibration with external control points or viDoc-like systems in projects where measurement precision is critical. Overall, both devices can contribute to the widespread adoption of cultural heritage documentation by offering more accessible alternatives to traditional terrestrial laser scanning. The literature also demonstrates that low-cost mobile scanning combined with augmented reality applications can increase visitor engagement (Akbaylar et al., 2024).

This study demonstrated that a multi-sensor Scan-to-BIM workflow built with viDoc and Realsee data offers a scalable, repeatable framework suitable for HBIM production in cultural heritage structures. With rapid data collection in the field, minimal equipment, and relatively low operational costs, geo-referenced, information-rich models are produced, which can be used for restoration planning, due diligence, virtual exhibition, and training purposes.

The research also plans to apply XGRIDS Lixel L2 Pro to the same cultural heritage

site to compare its quantitative and qualitative performance with viDoc and Realsee. Furthermore, by integrating AI-based automatic point cloud processing and semantic segmentation methods into the workflow, the HBIM production process is aimed at partially automating and improving output quality. The goal is to contribute to the development of tools and methods that will make the documentation of cultural heritage faster, more accurate and accessible to a wider range of users.

Acknowledgements

We would like to thank the professional surveying and scanning service providers for their support during the on-site data acquisition. This study was supported by Mimar Sinan Fine Arts University (MSGSU) Scientific Research Projects (BAP) under Project No: 36, titled "3D Mobil Lidar İle Yüksek Geo-Hassasiyetli Geo YBM Oluşturulması.".

References

A. Llabani* and N. Kopliku, Application of handheld mobile terrestrial laser scanning for cultural heritage documentation., in: *ISC2024*. URL https://www.scipedia.com/public/Llabani*_Kopliku_2024a

Abdul Rahman, A., Zamzuri, A., As Cemrek, H., & Isikdag, U. (2026). BIM for Geoinformation in Malaysia and Turkiye-Current Status. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 1–5. https://doi.org/10.5194/isprs-archives-XLVIII-4-W17-2025-1-2026

Akbaylar Hayreter, İ., Güleç Özer, D., & As Cemrek, H. (2024). Enhancing cultural heritage digitalization and visitor engagement through lidar scanning and gamification. In *The 29th Annual Conference for Computer-Aided Architectural Design Research in Asia (CAADRIA).* The Association for Computer-Aided Architectural Design Research in Asia.

Çemrek, H. A., Sönmez, S., & Işıkdağ, Ü. (2025) 'Mobile 3D LiDAR-based Scan2BIM Process: A comparison of point cloud processing tools'*, Confluence – Proceedings of the 43rd Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2025),* Vol. 1, pp. 503–512. https://papers.cumincad.org/data/works/att/ecaade2025_589.pdf

Giuliani, F., Gaglio, F., Martino, M., & De Falco, A. (2024). A HBIM pipeline for the conservation of large-scale architectural heritage: the city Walls of Pisa. *Heritage Science*, 12(1), 35. https://doi.org/10.1186/s40494-024-01141-4

Epotronic. (n.d.). *XGRIDS – High-End Handheld Scanner & Professional 3D Modeling Software.* Retrieved November 12, 2025, from https://epotronic.com/eng/manufacturers/xgrids/

Fiorini, G., Tini, M. A., Montelli, F., & Bitelli, G. (2023). Comparison of two technologies in 3D surveying of real estate assets and cultural heritage. *Proc. MetroArchaeo*, 435-440.

Llabani, A., & Abazaj, F. (2024). A COMPARATIVE ANALYSIS BETWEEN PERSONAL AND TERRESTRIAL LASER SCANNING FOR THE DOCUMENTATION OF HERITAGE SITES. *Stavební obzor-Civil Engineering Journal*, 33(2), 199-214.

Hariffin, N. A. B., Razali, M. H. B., Luh, L. C., Sulaiman, S. A., & Hashim, M. A. B. M. (2023, October). The suitability of Matterport for building parcel dimension survey. In *2023 IEEE 13th International Conference on System Engineering and Technology (ICSET)* (pp. 239-244). IEEE.

Jeftha, K. J., & Shoko, M. (2024). Mobile phone based laser scanning as a low-cost alternative for multidisciplinary data collection. *South African Journal of Science*, 120(11/12).

Kuçak, R. A., Erol, S., & Alkan, R. M. (2023). iPad Pro LiDAR sensörünün profesyonel bir yersel lazer tarayıcı ile karşılaştırmalı performans analizi. *Geomatik*, 8(1), 35-41.
Kültür Envanteri. (n.d.). *Zevkî Kadın Sıbyan Mektebi*. https://kulturenvanteri.com/yer/zevki-kadin-sibyan-mektebi/.
Matterport. (2025a, August 1). *FAQ: Matterport E57 File.* Matterport Support. https://support.matterport.com/s/article/FAQ-Matterport-E57-File
Matterport. (2025b, August 14). *Measurement accuracy in Matterport Spaces.* Matterport Support. https://support.matterport.com/s/article/How-accurate-are-dimensions-in-Matterport-Spaces
Matterport. (2025c, October 1). *Matterport BIM File*. Matterport Support. https://support.matterport.com/s/article/Scan-to-BIM
Matterport. (n.d.). *BIM Files.* Retrieved November 10, 2025, from https://matterport.com/bim
Murphy, M., McGovern, E., & Pavia, S. (2009). Historic building information modelling (HBIM). *Structural Survey*, 27(4), 311–327. https://doi.org/10.1108/02630800910985108
Pix4D. (2022, March 2). *Joint LiDAR and photogrammetry workflow for 3D bridge model.* https://www.pix4d.com/blog/rtk-rover-bridge-restoration
Realsee. (2025a, June 19). *Exploring the future of 3D scanning with Louis Pont and Aerial Perspectives Australia.* https://home.realsee.ai/en/article/exploring-the-future-of-3d-scanning-with-louis-pont-and-aerial-perspectives-australia
Realsee. (2025b, June 19). *German startup Winobo uses Realsee digital twins to revamp half-timbered houses & monuments*. https://home.realsee.ai/en/article/german-startup-winobo-uses-realsee-digital-twins-to-revamp-half-timbered-houses-monuments
Realsee. (n.d.). *Galois – Specifications*. https://home.realsee.ai/en/specs-galois
RobotShop. (n.d*.). Realsee Technology Galois M2 3D LiDAR Camera Kit*. Retrieved November 10, 2025, from https://www.robotshop.com/products/realsee-technology-galois-m2-3d-lidar-camera-kit-us
Rocha, G., Mateus, L., Fernández, J., & Ferreira, V. (2020). A scan-to-BIM methodology applied to heritage buildings. *Heritage*, 3(1), 47–67. https://doi.org/10.3390/heritage3010004
Shults, R., Levin, E., Habibi, R., Shenoy, S., Honcheruk, O., Hart, T., & An, Z. (2019). Capability of matterport 3d camera for industrial archaeology sites inventory. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 1059-1064.
Tanduo, B., Teppati Losè, L., & Chiabrando, F. (2023). Documentation of complex environments in cultural heritage sites. A SLAM-based survey in the Castello del Valentino basement. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 489-496.
Teppati Losè, L., Spreafico, A., Chiabrando, F., & Tonolo, F.G. (2022). Apple LiDAR Sensor for 3D Surveying: Tests and Results in the Cultural Heritage Domain. *Remote Sensing*, 14(17), 4157. https://doi.org/10.3390/rs14174157
Treccani, D., Adami, A., Brunelli, V., & Fregonese, L. (2024). Mobile mapping system for historic built heritage and GIS integration: a challenging case study. *Applied Geomatics,* 16(1), 293–312.
Ubaidillah, M.R., Cahyadi, M.N., & Suhandri, H.F. (2023). Utilization of iPad Pro LiDAR and Low-Cost GNSS Ublox F9R for Data Collection of Road Assets in the Form of 3D Models. *IOP Conference Series: Earth and Environmental Science*, 1127(1), 012013. https://doi.org/10.1088/1755-1315/1127/1/012013
Zollini, S., & Marconi, L. (2025). Evaluation of Positioning Accuracy Using Smartphone RGB and LiDAR Sensors with the viDoc RTK Rover. *Sensors*, 25(13), 3867.
Wang, M., Liu, Y., Wang, H., Dong, H., & El Saddik, A. (2025). Empirical Studies of Large Scale Environment Scanning by Consumer Electronics. *IEEE Consumer Electronics Magazine.*

Perception-Driven 3D Printing

A Study on Adaptive Toolpath Generation for Complex Surfaces

Yujiao Li[1], Hao Wu[2], Jiapan Sun[3] and Philip F. Yuan[4]
[1,2,3,4]*College of Architecture and Urban Planning, Tongji University*
[1]*lyj2000@tongji.edu.cn, 0009-0009-9605-4575*
[2]*2032186@tongii.edu.cn, 0000-0002-0503-1696*
[3]*2510252@tongji.edu.cn, 0009-0005-6319-277X*
[4]*philipyuan007@tongji.edu.cn, 0000-0002-2871-377X*

Abstract. To address the limited adaptability of architectural 3D printing on complex surfaces, this study proposes an integrated workflow that couples computer vision with process-parameter modelling. The research evaluates 3D reconstruction performance by comparing voxel-based and feature-based visual SLAM algorithms across various scanning strategies, identifying ORB-SLAM3 as the most robust method for real-time environmental sensing. Quantitative models relating nozzle speed to layer height (V*-H*) and curvature (V*-R*) were established to ensure consistent filament width regardless of substrate irregularities. These findings were successfully implemented into an adaptive toolpath generation framework, validated through variable-height compensation and conformal printing experiments. The results demonstrate that the proposed workflow effectively regulates printing parameters in response to environmental variations, overcoming the constraints of traditional one-directional information pipelines to achieve high-quality, adaptive deposition in complex scenarios.

Keywords. Complex Surfaces, Computer Vision, 3D Reconstruction, Process-parameter Modelling, Adaptive Toolpath Generation

1. Research Background

3D printing fabricates objects directly from digital files through a layer-by-layer process and has gained growing relevance in architecture. Despite decades of development, most architectural 3D printing systems still follow a linear design–code generation–execution workflow. This assumes controlled environments and standardised materials, making the systems poorly suited to the irregular, changing conditions typical of on-site or non-standard construction tasks (Capunaman et al., 2025). Real-world substrates are frequently non-planar, curved, or partially damaged—as seen in urban repair, heritage reproduction, in-situ fabrication or conformal printing—and inadequate environmental capture severely limits precise toolpath

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 81-90.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

planning and adaptive control.

In response to the limited perceptual capability of current 3D-printing systems in complex environments, recent studies increasingly employ computer vision to enhance environmental understanding, covering 3D reconstruction, toolpath generation, and process correction. Nicholas et al. (2020) used real-time 3D scanning to reconstruct unknown geometries, encoded the models as neural-network inputs, and generated toolpaths from predicted panel performance. Naboni et al. (2022) integrated sensors into a concrete-printing system in the SENS-ENV experiment to scan construction sites and design adaptive paths through generative algorithms. Other studies customise prints manually from reconstructed surfaces; for example, Ribeiro et al. (2024) used photogrammetry to model rock terrains and produce bespoke concrete elements. Thomsen et al. (2025) combined computer vision with photogrammetric modelling for damage detection, diagnosis, and repair-path generation, forming a perception-driven repair workflow. Çapunaman's work progresses from a computer-vision-based digital-twin framework for unknown surfaces (2022) to large-scale concrete printing with real-time deposition monitoring and adaptive control (2025).

While existing research has advanced vision-based 3D printing through environmental reconstruction and path generation, these workflows often lack a direct link between sensing data and real-time process regulation. This study addresses this gap by establishing a quantitative coupling between substrate geometry and extrusion parameters. Our approach utilises high-precision 3D reconstruction to drive adaptive models, specifically V*-H* and V*-R*. These models dynamically adjust nozzle speed relative to layer height and curvature, ensuring consistent filament width on complex surfaces. This integrated workflow enables adaptive toolpath generation that responds to physical irregularities through active process self-correction.

2. Methodology

The workflow proposed in this study comprises three main components: an investigation of environmental perception methods, the construction of process-parameter models, and experiments on path-generation algorithms.

2.1. ENVIRONMENTAL PERCEPTION METHOD

As illustrated in Figure 1, the experimental setup consists of hardware equipment and an artificially created substrate. The hardware system is organised into three modules: a UR10 robotic arm, a perception module, and a material-delivery module. The motion programme for the UR10 is generated using FURobot based on Grasshopper. Within the perception module, a Realsense D455 depth camera is mounted at the end of the printhead, with the captured images transmitted to a PC via USB. The material-delivery module consists of a dispenser syringe containing the material; compressed air is used to drive the material through a pipe to the printhead. The rotation of the motor, which actuates the screw extruder, is regulated by analogue output signals from a control unit, enabling stable material extrusion. Prior to conducting the formal experiments, hand–eye calibration was performed to obtain the intrinsic and extrinsic parameters of the camera and to determine accurately its spatial relationship relative to the robotic arm.

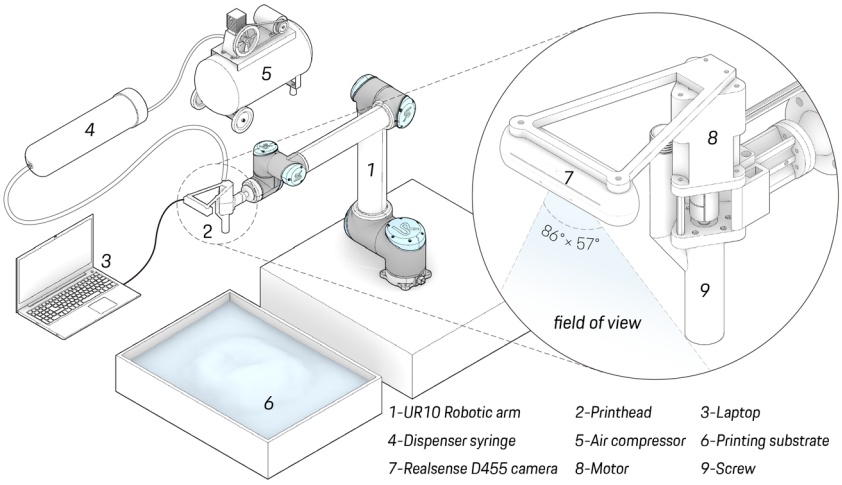


Figure 1. Hardware Equipment Diagram

This part of the study uses depth-camera scanning to generate 3D environmental models and to identify effective reconstruction strategies and parameters. Following Çapunaman et al. (2025), perception strategies were defined by their degrees of freedom, and four types—(a), (b), (c), and (d)—were adopted (Figure 2). Strategy (a) is a static scan, while (b) and (c) use 3-DoF motion with the camera positioned 600 mm above the surface. Strategy (d) employs 6-DoF multi-angle scanning along a spherical trajectory. For strategies (b)-(d), the number of scanning poses formed a second variable, ranging from 2×2 to 6×6. Figure 2, using strategy (d) as an example, shows the corresponding camera trajectories.

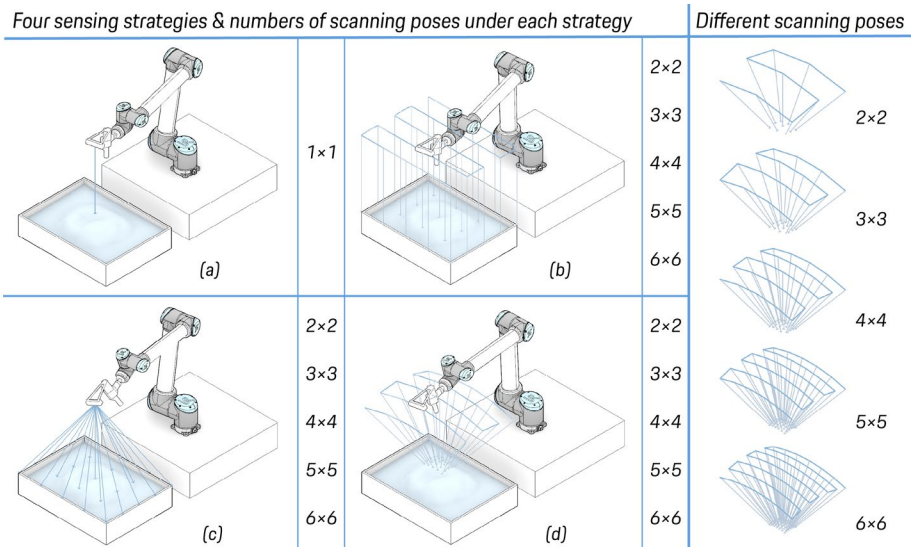


Figure 2. Diagram of the four sensing strategies and the different numbers of camera poses

Beyond the two variables associated with the scanning strategies, this experiment introduces the choice of 3D reconstruction algorithm as a third variable. Two representative categories of reconstruction methods were selected for comparison: the voxel-based Open3D approach, which fuses depth data (Zhou, 2018), and the feature-based visual SLAM method ORB-SLAM3 (Campos, 2021). As illustrated in Figure 3, the Open3D pipeline begins with the acquisition of multi-view RGB-D images, followed by the conversion of depth maps into point clouds. Local and global registration are then applied to refine the relative poses between views, and the aligned point clouds are subsequently integrated within a TSDF voxel framework to produce a dense reconstruction. In contrast, the ORB-SLAM3 workflow first extracts ORB feature points from the RGB-D input and estimates camera motion continuously through feature matching. During local mapping, a sparse point-cloud map is incrementally constructed. With the aid of loop detection and global optimisation, the method is able to generate temporally consistent camera trajectories and spatial structures over long sequences.

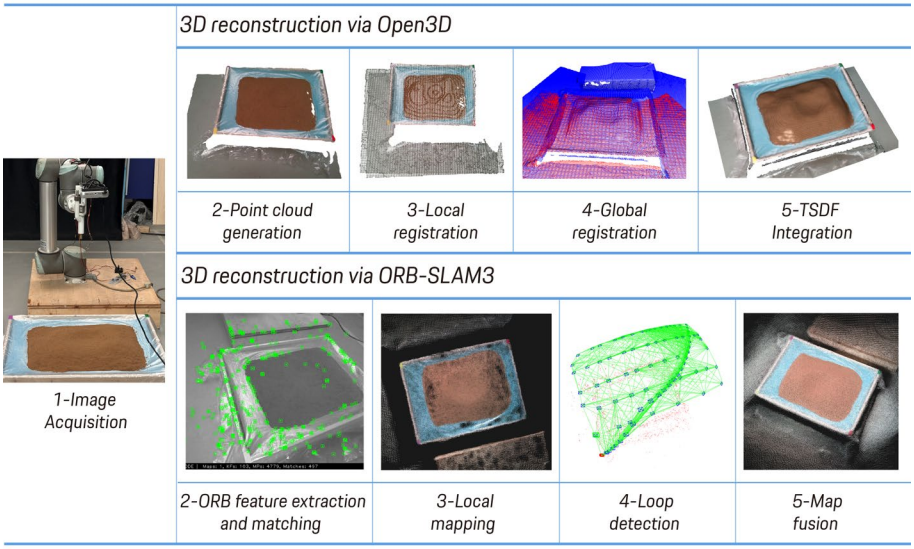


Figure 3. Diagram of the reconstruction workflows based on Open3D and ORB-SLAM3

Following the completion of scanning and model reconstruction, the accuracy of the reconstructed models was evaluated with reference to established methods in prior research. Specifically, a digital dial gauge was mounted on the end-effector of the robotic arm, and its TCP was determined. An orthogonal measurement grid with approximately 30 mm spacing was then established on the surface of the physical object. The robotic arm probed each point along the surface normal defined by the digital model, and the gauge readings were recorded and compared with the corresponding nominal offsets in the digital model. The deviation between the two was calculated as the reconstruction error.

As shown in Figure 4, reconstruction errors were recorded under the four perception strategies and varying numbers of scanning poses using the Open3D

reconstruction framework. After compiling all data, the error range was set to −8.0 mm to +8.0 mm, and the average, maximum, and minimum absolute errors were computed for each strategy. The results indicate that strategy (b) achieved the highest overall reconstruction quality, with consistently low average error. Strategies (c) and (d) exhibited poorer registration accuracy when the number of scanning poses was small, particularly for (c), where the presence of irrelevant surrounding objects led to occasional extreme errors. Increasing the number of poses reduced the average error for both strategies, although (d) continued to show relatively large extreme deviations. Strategy (a), constrained by highly limited viewpoints, produced the least satisfactory reconstruction performance. Furthermore, for strategies (b) and (d), once the number of scanning poses exceeded 4×4, additional scanning time and images contributed little further improvement to reconstruction accuracy.

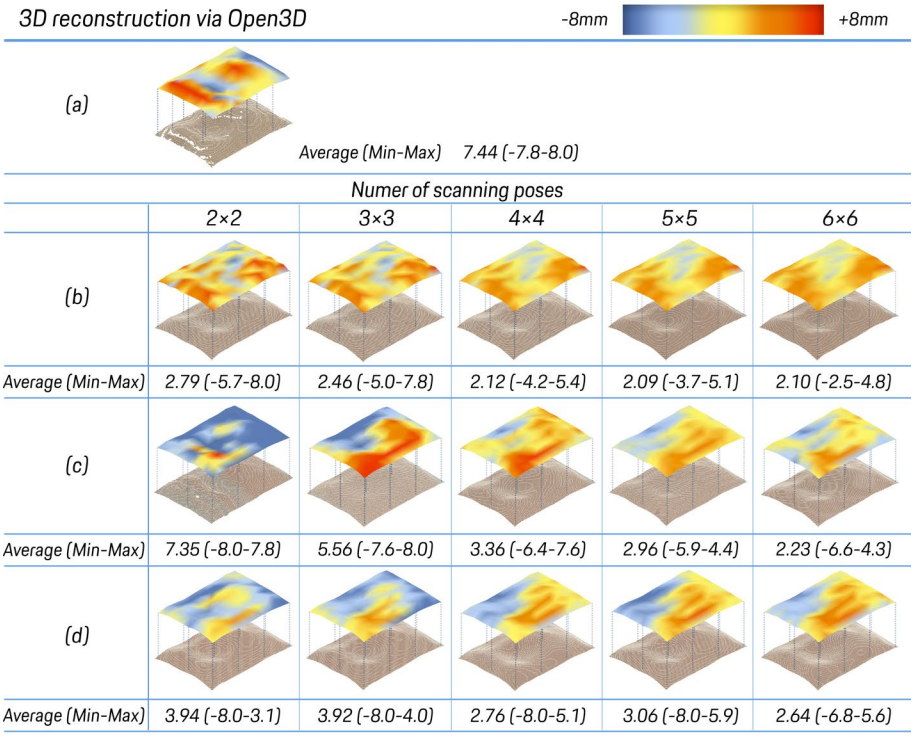


Figure 4. Error evaluation of the model reconstructed using the Open3D method

Figure 5 presents the reconstruction errors associated with the ORB-SLAM3-based perception strategies. Regardless of the specific strategy employed, this method consistently produced high-quality reconstructions. Strategies (c) and (d) exhibited larger local errors when the scanning duration—and thus the number of captured images—was limited; however, increasing the number of poses to 4×4 substantially reduced both the magnitude and the range of errors. Strategy (b), by contrast, maintained stable and reliable reconstruction accuracy across all pose configurations.

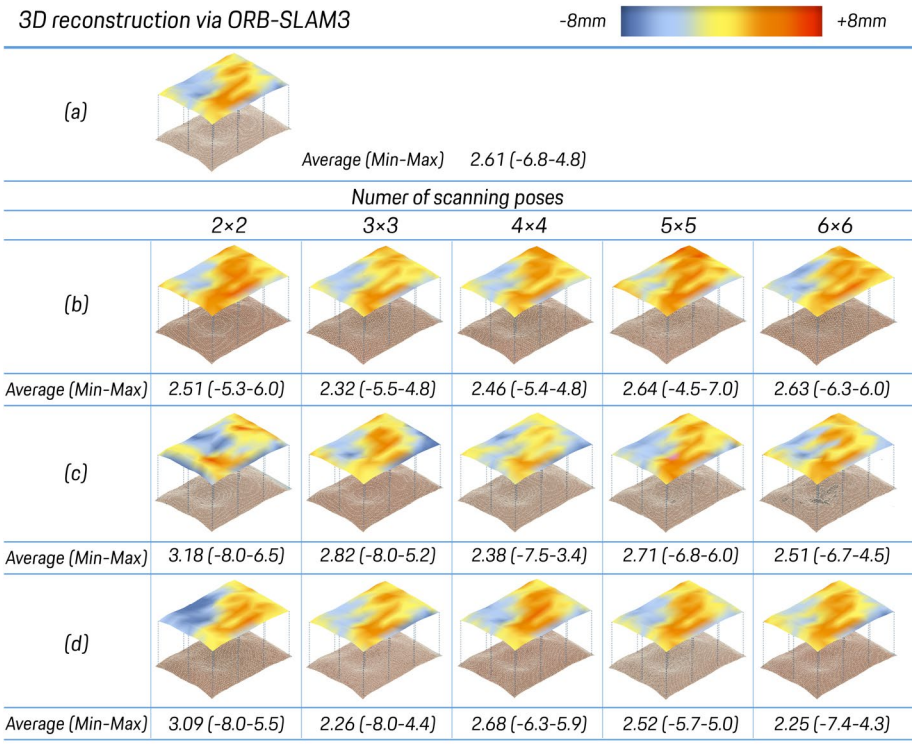


Figure 5. Error evaluation of the model reconstructed using the ORB-SLAM3 method

Synthesising the results across perception strategies, numbers of scanning poses, and reconstruction methods, an optimal perception approach can be identified for the purposes of this study. Overall, strategy (b) emerges as the most effective for image acquisition, likely owing to the relatively planar nature of the scanned object in this experiment. Increasing the number of scanning poses improves reconstruction accuracy to a certain extent, but beyond a threshold—where additional time and images yield diminishing returns—the precision no longer improves significantly. With respect to the reconstruction algorithms, visual SLAM methods such as ORB-SLAM3 demonstrate strong accuracy and robustness, producing low-error models across all scanning strategies, while also offering the advantage of real-time performance.

2.2. CONSTRUCTION OF PROCESS-PARAMETER MODELS

3D printing involves multiple process parameters—including nozzle speed, nozzle diameter, nozzle height, and material flow rate—all of which influence filament geometry and structural stability. This part of the study focuses on two scenarios commonly encountered when printing on complex surfaces, with the aim of establishing models that describe the relationships between key variables and process parameters. The material used in the experiments is clay, and the experimental methods and procedures are equally applicable to other materials.

The first scenario concerns printing a relatively planar object on an uneven substrate. Near the substrate, an adaptive toolpath is required, while the upper layers gradually revert to the original geometry, resulting in a variable-height printing path. This variation necessitates corresponding adjustments to nozzle speed or material flow to compensate for surface irregularities. In this context, maintaining a consistent filament width across layers is essential. Layer height therefore serves as the primary independent variable, with nozzle speed and material flow acting as associated control parameters. Preliminary tests were conducted to determine suitable parameter ranges, after which the material flow rate was fixed and a two-variable model—defined by nozzle speed and layer height—was established. As shown in Figure 6, a 10 mm nozzle was used, with nozzle speed V* varied between 4-12 mm/s and layer height H* between 3-9 mm, producing 20 parameter combinations. Each combination was used to print a three-layer, 100 mm linear filament. Cross-sectional profiles and widths were recorded, using the average width of the second and third layers to avoid the base-pressure influence present in the first layer. Results show that filament width increases as H* decreases or is V* reduced. An isoline map was subsequently generated from the parameter matrix, with each contour representing the relationship between V* and H* for a given target filament width. This map can be directly applied to toolpath planning for compensating depressions on uneven surfaces.

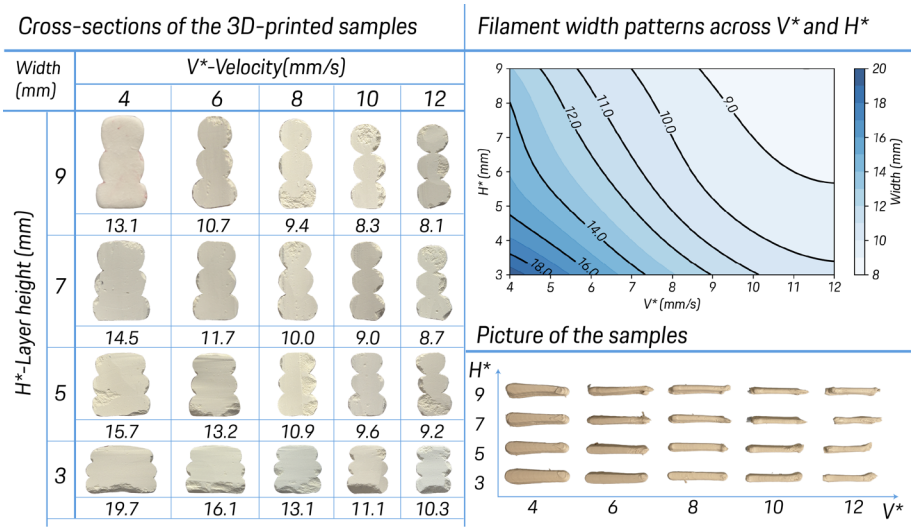


Figure 6. Diagram of the process-parameter model construction for V and H**

The second application scenario focuses on conformal printing on complex surfaces. Owing to significant variations in path curvature, steep turns often cause filament dragging. Allowing nozzle speed to adapt to curvature can therefore improve print quality in such regions while maintaining uniform filament width overall. As shown in Figure 7, nozzle speed V*(four levels, 4-10 mm/s) and curvature radius R*(six levels, 20-70 mm) were selected as variables, generating 24 parameter combinations. Using a 10 mm nozzle, three-layer circular samples

were printed at a constant layer height of 6 mm, and their cross-sectional profiles and widths were recorded. The results indicate that when V*<10mm/s, filament width increases as nozzle speed decreases and curvature decreases (i.e., as the radius increases). Although still narrower than straight-line filaments printed under the same conditions, the widths gradually converge as curvature decreases, with faster convergence at lower speeds. Based on these observations, an isoline map was produced, enabling the corresponding combinations of V* and R* to be applied directly to velocity planning for complex toolpaths.

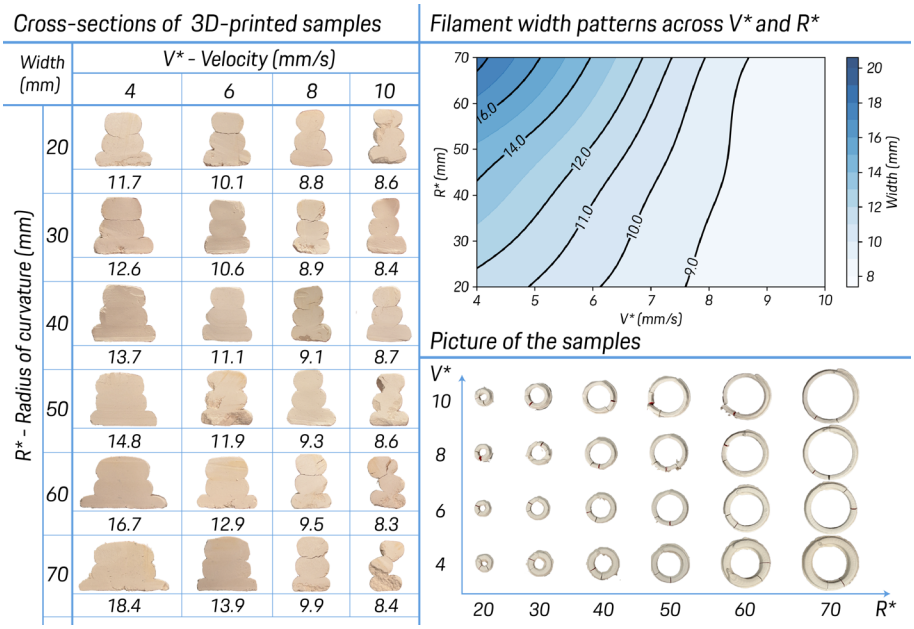


Figure 7. Diagram of the process-parameter model construction for V and R**

2.3. PATH-GENERATION ALGORITHM EXPERIMENTS

This section evaluates the applicability of the environmental perception method and the process-parameter models through two scenario-based experiments. As shown in Figure 8, the first experiment addresses variable-height compensation for ground undulation. The sand substrate contained two sculpted depressions beneath an otherwise level surface. A planar target path was generated above it and then locally adjusted through a parametric method: layer heights were increased over the depressions and gradually restored in upper layers, producing a smoothly varying height profile along the path. During toolpath generation, the V*-H* relationship corresponding to a target filament width of w=11mm was applied to the continuously changing layer heights, yielding the required printing speed for each layer. As illustrated by the comparison between sections (a) and (b), despite notable differences in layer height, the resulting filament widths closely match the intended 11 mm target, demonstrating the effectiveness of the model.

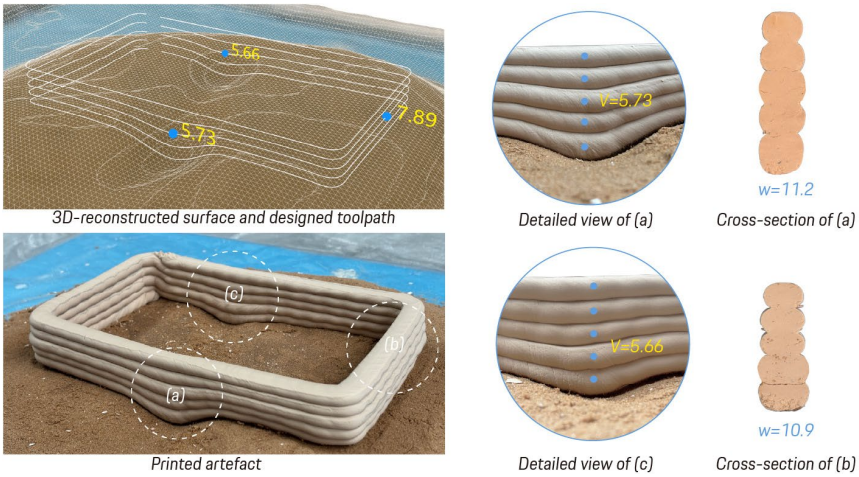


Figure 8. Variable-height compensation experiment for ground undulation

Figure 9 presents the conformal 3D printing experiment, focusing on the adjustment of printing speed in response to local curvature. An undulating surface was first shaped in sand, and an adaptive toolpath was generated on the reconstructed model using a differential growth algorithm. The V*-R* relationship model corresponding to a target filament width of w=10 mm was then applied to the curvature at each point along the path to determine the appropriate nozzle speed. Four representative points were selected and the corresponding speeds were indicated in yellow. As shown in the cross-sections (a)-(d), the filament width remained consistently close to the intended 10 mm across all conditions, demonstrating the effectiveness of the model.

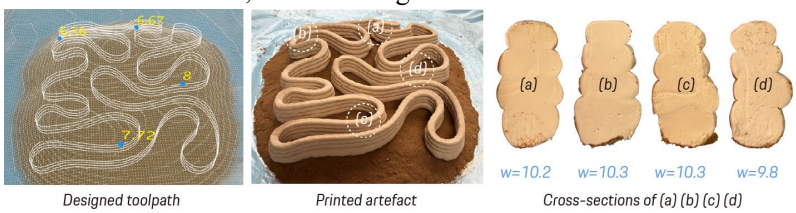


Figure 9. Conformal printing experiment under varying curvature

3. Conclusion

This study demonstrates an integrated workflow coupling computer vision with process-parameter modelling to achieve adaptive 3D printing in complex environments. However, limitations regarding fixed nozzle orientation and material dynamics persist. Future developments will focus on multi-axis toolpaths for improved deposition and implementing real-time closed-loop control to overcome data latency and sensor occlusion. Beyond architecture, this framework offers significant potential for sectors like manufacturing and robotics. Integrating advanced AI and computer vision for autonomous parameter tuning will further enhance system robustness, transforming passive sensing into an intelligent, self-correcting fabrication process.

Acknowledgement

This study was funded by the National Key R&D Programme of China (2023YFC3806904).

Attribution Statement

ChatGPT (OpenAI, 2024) was used to (1) translate texts from Chinese to English. (2) to improve flow.

References

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., & Tardós, J. D. (2021). Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE transactions on robotics*, 37(6), 1874-1890. https://doi.org/10.1109/TRO.2021.3075644

Capunaman, Ö. B., Dong, W., & Gürsoy, B. (2022). A vision-based sensing framework for adaptive robotic tooling of indefinite surfaces. *Construction Robotics*, 6(3), 259-275. https://doi.org/10.1007/s41693-022-00081-4

Çapunaman, Ö. B., Farrokhsiar, P., Bilén, S. G., Duarte, J. P., & Gürsoy, B. (2025). Vision-based sensing and digital twin technologies in conformal 3D concrete printing: exploring operational accuracy, adaptability, and scalability, and investigating monitoring capabilities in large-scale applications. *Construction Robotics*, 9(1), 4. https://doi.org/10.1007/s41693-024-00145-7

Naboni, R., Breseghello, L., & Sanin, S. (2022, September). Environment-aware 3d concrete printing through robot-vision. In *40th Conference on Education and Research in Computer Aided Architectural Design in Europe*:*Co-creating the Future,* eCAADe 2022 (pp. 409-418). Education and research in Computer Aided Architectural Design in Europe (eCAADe).

Nicholas, P., Rossi, G., Williams, E., Bennett, M., & Schork, T. (2020). Integrating real-time multi-resolution scanning and machine learning for Conformal Robotic 3D Printing in Architecture. *International Journal of Architectural Computing*, 18(4), 371-384. https://doi.org/10.1177/1478077120948203

OpenAI. (2024). ChatGPT Model: GPT-4 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/

Ribeiro, J., Morais, A., Silva, J. M., Brandão, F. J., Figueiredo, B., & Cruz, P. J. (2024). Robotic 3DCP fabrication of custom-fit slabs for irregular pontoons. *Architectural Intelligence*, 3(1), 14. https://doi.org/10.1007/s44223-024-00056-1

Thomsen, M. R., Nicholas, P., Chiujdea, R. S., Nielsen, S. D., Sonne, K., & Eppinger, C. (2025). Additive Manufacturing for Repair: Continual Construction Through Bio-Based Materials. *3D Printing and Additive Manufacturing*, 12(2), 112-121. https://doi.org/10.1089/3dp.2023.0344

Zhou, Q. Y., Park, J., & Koltun, V. (2018). Open3D: A modern library for 3D data processing. *arxiv preprint arxiv*:1801.09847. https://doi.org/10.48550/arXiv.1801.09847

XR and Spatial Cognition

Simulating Interactive Architecture

A Behaviour-Based Interactive Design Platform

Chao Yan[1], Te Li[2], Ruyi Yang[3], Ziyi Chen[4], Yanzi Jin[5], Jiaxin Yao[6], Wenqin Cai[7], Aize Han[8] and Yuxuan Deng[9]
[1,2,4,5,6,7]*Tongji University.*
[3]*Harvard University.*
[8]*Tianjin University.*
[9]*Shanghai University.*
[1]*yanchao@tongji.edu.cn, 0000-0003-1446-4861*
[2]*2330265@tongji.edu.cn, 0009-0000-2396-7060*
[3]*ruyi_yang@gsd.harvard.edu, 0009-0009-5248-1751*
[4]*2430083@tongji.edu.cn, 0009-0005-8747-2493*
[5]*2430288@tongji.edu.cn, 0009-0000-9349-1352*
[6]*2540751120@qq.com, 0009-0009-6052-9313*
[7]*caiwenqin103@163.com, 0009-0001-3479-289X*
[8]*han13113137137@163.com, 0009-0002-4820-0792*
[9]*18798625092@163.com, 0009-0006-4597-1996*

Abstract. Interactive architecture has played a significant role in the evolution of architectural discourse and practice. Advances in artificial intelligence and smart sensing now enable dynamic interactions between occupants and spaces. In the conceptual design phase, however, simulating formal transformations and testing human interactions remains challenging. This study addresses these challenges by integrating hardware and software devices to develop a behaviour-based MR design tool for interactive architecture. BodyMR is a virtual simulation design platform that integrates depth cameras and mixed-reality devices with the BodyAnalytics behaviour computation plugin in Grasshopper. The platform uses body tracking algorithms to analyse the motion data captured by the depth camera in design experiments, generating human posture information and transmitting it to Grasshopper to trigger model transformations. The updated model is then sent to mixed-reality devices, influencing participants' behaviours and thereby establishing an interactive feedback loop. This paper presents the application results of the design platform through the projects of three design groups in design studios, which have effectively validated the efficacy and scalability of the platform. In conclusion, to address the challenges of interactive space design, a simulation design platform was developed, providing an innovative, low-cost, and rapid-feedback approach for interactive architecture design.

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 93-102.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Keywords. Interactive Design, Adaptable Environment, Behaviour Computation, Mixed Reality, Computer Vision, Human-in-the-loop Interaction

1. Introduction

Since the mid-20th century, interactive architecture has played a significant role in the evolution of architectural discourse and practice. Responsive architecture is anticipated to represent the next stage in the evolution of the discipline. With the advancement of artificial intelligence and intelligent sensing technologies, the dynamic interaction between humans and architectural spaces has gradually transformed from a speculative concept into a tangible reality. Intelligent and adaptive environments have thus become a central theme in contemporary architectural design (Makanadar, 2024).

However, responsive architecture faces numerous challenges during the conceptual design stage. Due to its inherently dynamic and interactive nature, it is difficult to explore and refine responsive architectural designs using traditional static media such as drawings and physical models. Before construction, there is a lack of effective means to experimentally validate interaction behaviours. The design and testing of responsive architecture therefore suffer from a shortage of practical tools and methodologies, making it urgent to develop a universal simulation and experimental platform to address these issues.

In this context, this study explores the interaction between humans and the architectural environment and develops a virtual simulation design platform for behavioural analysis and spatial interaction modelling. By employing mixed reality technologies to visualize architectural dynamic responses, the platform constructs a feedback loop between users and the environment. This system provides a foundational tool for design validation and experimental testing in responsive architecture.

2. Literature Review

The concept of responsive architecture originated in the mid-20th century from architects' pursuit of spaces that adapt to human activities and environmental changes. Cedric Price's Fun Palace marked the intellectual genesis of this idea and inspired subsequent discussions on adaptable and variable architecture. The Japanese Metabolists and the British group ARCHIGRAM envisioned architecture as an organic, transformable entity responsive to external stimuli (Koolhaas et al., 2011; Simon, 2005). Within this framework, architecture could dynamically change according to users' needs. Based on the effective acquisition of user states and the inherent variability of architectural systems, the architect's role shifted from shaping form to designing rules (Frazer, 1995). Consequently, research focus moved from building form to optimizing architectural behaviour (Lee et al., 2021).

To date, numerous practical explorations of responsive and interactive architecture have emerged. Philippe Beesley's Living Architecture Systems Group (2025) integrates mechanical apparatuses and sensor technologies to construct architectural environments with life-like vitality, enabling dynamic human–space interaction.

Participants engage with these living architectures within spatial fields, and in turn, their behaviours in time and space are reshaped through this interaction. Similarly, Carlo Ratti's Digital Water Pavilion (Zaragoza Expo 2008) employed programmable water curtains to create transformable interfaces. The pavilion's spatial divisions, wall configurations, and entry points shifted dynamically in response to crowd density, while the changing patterns and frequencies of the falling water influenced participants' spatial distribution and movement (Carlo Ratti Associati, 2008). These projects demonstrate the potential of responsive architecture for broader application.

Based on the cases, it can be concluded that the key to designing responsive architecture lies in establishing a feedback loop between human behaviour and architectural transformation. However, existing studies mainly focus on exploring the potential for dynamic architectural transformations through computer-assisted planar configurations, yet it has seldom delved into the simulation of interactive feedback mechanisms (Tsoumpri, 2023). Designing interactive architecture through feedback loop exploration often requires the construction of small-scale mock-ups, which demand significant time and financial resources. Consequently, there is currently a lack of experimental simulation platforms specifically tailored for the design and prototyping of responsive architecture.

3. BodyAnalytics & BodyMR: A Virtual Simulation Interactive Design System

3.1. RESPONSIVE DESIGN FEEDBACK LOOP

To construct the Virtual Simulation Interactive Design System for responsive architecture, this study integrates computer vision and mixed reality technologies. It develops two key components: the BodyAnalytics plugin for human behaviour analysis and the BodyMR data communication software. Together, these modules unify all hardware, software, and algorithmic tools involved in the interactive design process, forming a comprehensive framework for simulating and analysing responsive architectural interactions.

First, to identify individuals within the scene, this study employs a body tracking algorithm, a technique that has been widely applied in environmental behaviour research (Yan et al., 2025). To further enhance the accuracy and depth of behavioural analysis, a ZED 2i depth camera was utilized as the primary hardware device. This system operates a three-dimensional human pose estimation algorithm capable of precisely capturing the spatiotemporal coordinates of the subject's body joints. To translate these spatiotemporal joint-coordinate sequences into meaningful behavioural indicators that architects can apply in interactive design, the research team developed the BodyAnalytics plugin based on the Rhino Grasshopper parametric design platform. By processing behavioural data through this plugin, architects can construct spatial interaction rules and generate corresponding dynamic spatial response patterns informed by human behaviour.

Second, to test whether spatial interaction patterns are correctly triggered and achieve the intended design outcomes during the iterative design process, and to feed the dynamic changes of space back into user behaviour, this study incorporates the

Meta Quest 3 mixed reality device to present spatial interaction effects. Mixed reality has been widely applied in architectural design, not only to assist construction but also to support participatory and collaborative design processes. Previous studies have highlighted its role in reducing operational difficulty and cognitive load for collaborative designers (Silcock et al., 2021). However, current research lacks approaches that utilize mixed reality for passive, participatory design, in which designers can directly experience and test spatial interaction effects during the iterative process. Such an approach allows the impact of space on user behaviour to be fed back into design evaluation, thereby establishing a space–behaviour feedback loop.

This study models spatial interaction dynamics through the collection and computation of human behavioural data and presents them in real time using mixed reality. Participants adjust their behaviour in response to the dynamic changes of the space, and these behavioural adjustments, in turn, influence subsequent spatial iterations. As illustrated in Figure 1, this process establishes an iterative feedback design loop for responsive architecture, providing a human-in-the-loop simulation method that supports the iterative refinement of architectural design.

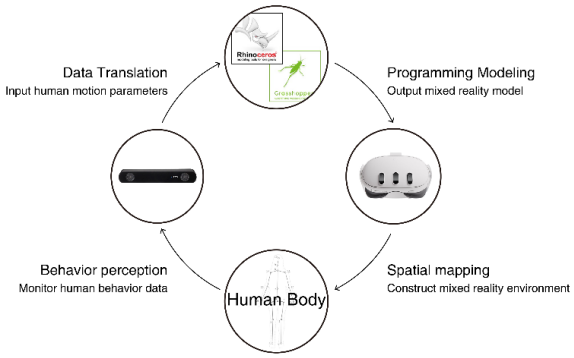


Figure 1. Feedback Design Loop for Simulating Responsive Architecture

3.2. SYSTEM ARCHITECTURE

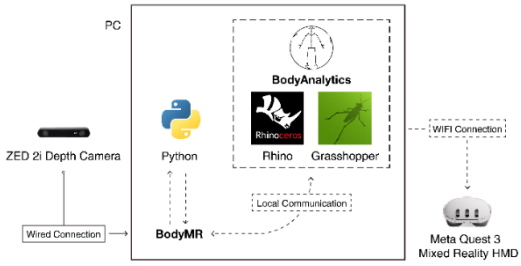


Figure 2. The Architecture of the Simulation Interactive Design System

To integrate software and hardware components within the feedback design loop and manage data transmission, the BodyMR data communication server was developed, which integrates all information flows within the interactive design simulation system

(Figure 2). BodyMR receives the video stream from the ZED depth camera, executes a human pose estimation algorithm to extract 3D joint coordinates, and transmits the data to Rhino Grasshopper for analysis. Within Grasshopper, spatial models generated through rule mapping are subsequently sent back through BodyMR to the Meta Quest, completing the core system architecture. Additionally, the server also includes pre-experiment data registration functions: human model calibration, which scales and adjusts pose outputs for accurate alignment, and world coordinate alignment, which synchronizes virtual objects with the physical environment.

In addition, the BodyMR server provides an extended Python interface, which allows for advanced data processing before the information is transmitted to Grasshopper. Through this module, users can perform expanded analyses on the raw data streams and pass the processed results into Grasshopper for further use. In the design case presented later in this study, this interface is employed to integrate multimodal data and perform motion prediction, thereby enhancing the system's capability for complex interaction modelling.

3.3. BEHAVIOURAL COMPUTATION

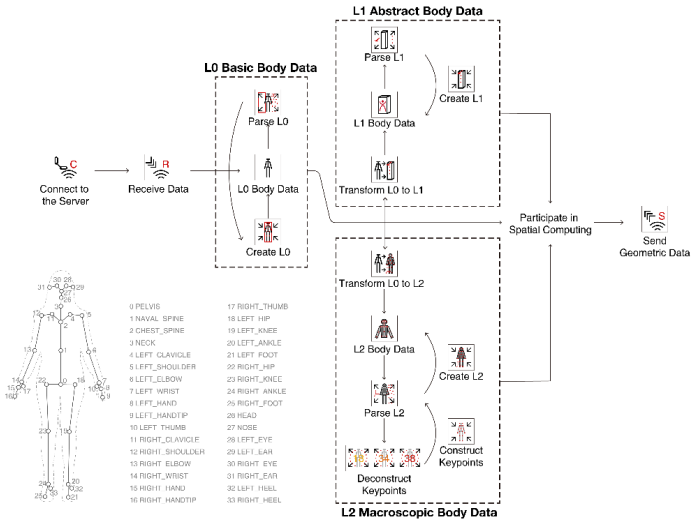


Figure 3. The Architecture of the BodyAnalytics Behavioural Computation Plugin and the 34-Keypoint Human Body Model

In the interactive design feedback loop, behavioural computation serves as the core module that maps human actions to spatial form transformations. To realize this function, the BodyAnalytics behavioural computation plugin was developed. Within this plugin, in order to address the diverse computational requirements arising from different design objectives, bodily information is structured into multiple hierarchical levels (Figure 3). This multi-layered information provides architects with a multidimensional computational foundation, ranging from overall movement patterns to localized postural adjustments, thereby supporting spatial interaction design across various scales and levels of complexity (Zhang, 2021).

- L0 (Basic Body Data): This level contains the raw data directly output from the depth camera's developer toolkit.
- L1 (Abstract Body Data): At this level, the body is represented as a cuboid with attributes such as velocity and direction. Parsing L1 body data yields indicators including Position, Velocity, Centre Axis, Bounding Box, Dimensions, Bounding Column, Radius, and Orientation.
- L2 (Macroscopic Body Data): Here, the body is modelled as a set of 3D joints and connecting lines. From L1 data, L2 parameters such as Key Points, Local Position per Joint, Local Orientation per Joint, and Global Root Orientation are derived. Depending on the human pose estimation model used, the key points set can include 18, 34, or 38 points, allowing architects to select the most suitable configuration for their design scenario. By applying the Deconstruct Key Points function, spatiotemporal data for specific joints can be extracted. For instance, in the 34-keypoint pose estimation model shown in Figure 3.

In the future, behavioural analysis will be extended to L3 (Microscopic Body Data), which includes the recognition of users' facial expressions and emotional states, thereby enabling more sophisticated and immersive spatial interaction logics.

4. Responsive Architectural Project for Interactive Design

Using the virtual simulation design platform, the research team conducted multiple design explorations in responsive architectural space. This paper presents three design studio projects to demonstrate the platform's scalability and adaptability. The Interactive Micro Office project illustrates the fundamental framework of interactive design. The Yoga Space project incorporates electroencephalography (EEG) data to explore multimodal, human-centred approaches to interactive architectural design. Finally, the Dormitory Space project integrates a Long Short-Term Memory (LSTM) neural network for behavioural prediction, achieving a more intelligent and predictive form of interactive architecture.

4.1. INTERACTIVE OFFICE SPACE DESIGN PROJECT

A student team designed an immersive office interaction system integrating intelligent environmental response mechanisms. The system aims to create a dynamically interactive workspace tailored to individual user states. Inspired by Mimosa pudica, the system emulates dynamic responses to external stimuli. The office system features movable wall structures composed of hundreds of independently rotating micro-panels surrounding the user. Through posture recognition and work-state evaluation algorithms, the system assesses the user's level of attention and fatigue to calculate their concentration state. Based on this analysis, the spatial openness is adjusted in real time, achieving an intelligent balance between efficiency and comfort in the workspace.

To establish the specific interaction patterns for the intelligent office space, it was necessary to determine the relationship between users' work performance indicators and spatial configurations. Previous studies have shown that the relationship between spatial openness and behavioural concentration is nonlinear (Mehta, 2012). Through

the interactive design platform, the design team collected L2 body parameters from participants, calculating the activity frequency of facial joint plane normal vectors as an indicator of concentration. By monitoring participants' activity states over time within design prototypes of varying openness, a quantitative relationship between spatial openness and concentration was derived. Similarly, based on human posture data, the experimental platform was used to compute and analyse combinations of L2 joint angle parameters, enabling quantitative modelling of attention and fatigue levels. By integrating these three indicators, the design team established a data-informed strategy for the dynamic adaptation of the interactive office environment (Figure 4).

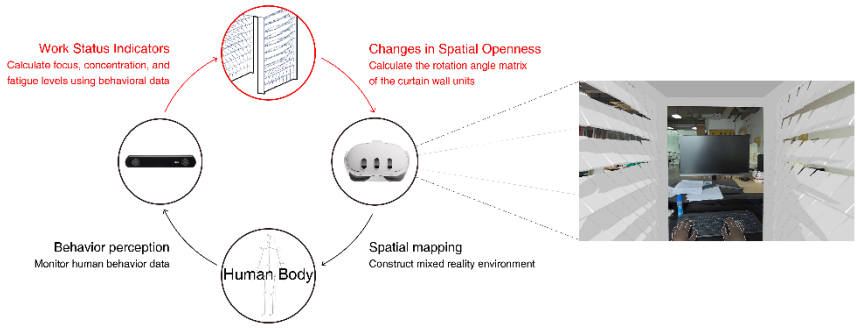


Figure 4. Feedback Design Loop of the Interactive Office Space Design Project

This design project demonstrates the application of the virtual simulation design platform as a real-time responsive framework, providing an empirical basis for the iterative design of dynamic interactive spaces through quantitative experimentation. In doing so, it establishes a foundational tool that supports the design-as-research paradigm within the field of responsive architectural design.

4.2. YOGA SPACE DESIGN PROJECT WITH MULTIMODAL DATA INTEGRATION

A group of students developed a mixed reality yoga space as a therapeutic environment, investigating how spatial design can enhance behavioural and psychological effects such as emotional relief and flow. The design features three concentric spatial layers enclosing the yoga practitioner, each defined by distributed tetrahedral geometries. The edge lengths, enclosure distances, and degrees of spatial encapsulation of these geometries dynamically change in response to the practitioner's movements and emotional states. To evaluate the effects of these spatial adjustments, the design team incorporated multimodal data into the experimental analysis. L1 body parameters derived from BodyAnalytics were mapped onto spatial forms, while EEG were collected using an EMOTIV EPOC X EEG Headset to assess the user's level of mental relaxation. This feedback continuously modulates spatial parameters, enabling the system to generate configurations that optimize the therapeutic effects of yoga.

During the experiment, the proportion of alpha wave signals at occipital electrodes (O1, O2, Oz) was used as an indicator of relaxation (Sanei et al., 2007). For each electrode, the α ratio was calculated, and the arithmetic mean of the three values was taken as the overall occipital α proportion. Before the experiment, participants wore

both the EPOC EEG headset and the mixed reality device, maintaining a resting seated posture for three minutes, followed by the performance of a designated yoga sequence (Figure 5). During this process, the Wallacei multi-objective optimization plugin was employed to automatically adjust spatial parameters with the goal of maximizing the occipital α ratio. Through iterative optimization, conducted with 10 participants each completing two experimental sessions, the system identified the optimal spatial configuration and element distribution for promoting relaxation and enhancing the experiential quality of the yoga environment.

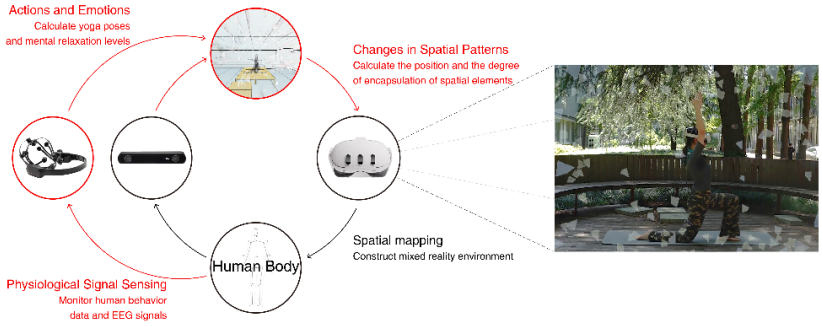


Figure 5. Feedback Design Loop of the Yoga Space Design Project with Multimodal Data Integration

Through multimodal data interaction, this project demonstrates the hardware extensibility of the virtual simulation design platform, facilitating the integration of diverse data streams to comprehensively assess users' physical movements and underlying psychological states. This approach enables more effective and human-centred spatial interactions.

4.3. PREDICTIVE DORMITORY SPACE DESIGN PROJECT BASED ON LSTM

A group of students explored an intelligent, adaptable dormitory space, investigating the potential of responsive architecture to enhance comfort and liveability in small-scale residential environments. The spatial configurations not only accommodate multiple functional needs but also guide users toward improved daily habits. To achieve this goal, the system detects the user's current actions, caches them in a memory sequence, and intelligently predicts the user's next behaviour based on the sequence, enabling the space to respond dynamically. In the design, 0.6 m square panel units fold and transform to create multiple functional spatial configurations.

Firstly, 107 questionnaires were collected to summarize behavioural patterns within the dormitory space, including activities such as resting, using a phone, socializing, folding clothes, sleeping and others. Each reported action was assigned a unique identifier. Through feature engineering, these actions could be identified in real time based on combinations of multiple L2 joint angle parameters from body data. The questionnaire also included statistics on action sequences, which were synthesized into several common behavioural patterns, forming a dormitory action sequence dataset. This dataset was used to train an LSTM network, capable of processing sequential

information and returning predictive results (Hochreiter and Schmidhuber, 1997). By leveraging the Python interface, the LSTM model can be loaded to predict the next user action online based on cached sequences (Figure 6). Spatial geometric transformations were defined for each current action-predicted action combination, enabling the simulation of an intelligent, predictive interactive dormitory space.

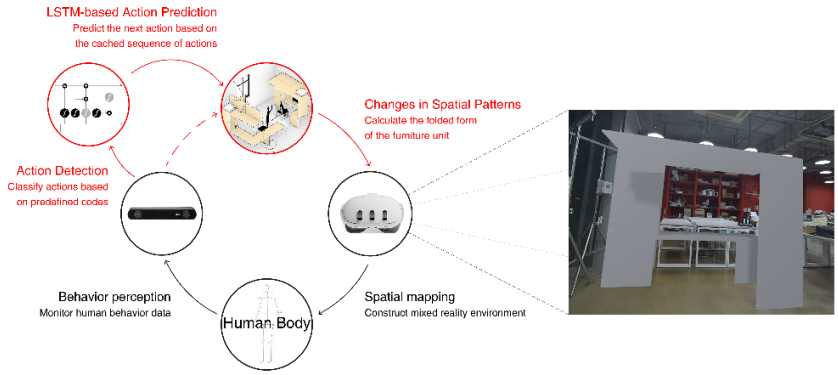


Figure 6. Feedback Design Loop of the Dormitory Space Design Project with Predictive Interaction

Through predictive interactive design, this project demonstrates the computational extensibility of the virtual simulation design platform, enabling the integration of machine learning algorithms and other data-analytic methods to create more intelligent and adaptive spatial interaction logics.

5. Conclusion and Discussion

In conclusion, this study represents a paradigm shift in the design methodology of responsive architecture. By leveraging intelligent algorithms and advanced simulation tools, it enables embodied exploration of complex human–architecture interactions.

Methodologically, this work contributes a framework for integrating behavioural data into architectural design. The platform supports real-time virtual experimentation, allowing designers to explore human-space interactions through embodied experience and derive insights for developing interactive spatial feedback mechanisms.

Systemically, this research introduces an integrated platform composed of the custom behavioural analysis tool BodyAnalytics and the data communication server BodyMR. Together, they establish a coherent feedback loop between behavioural tracking and architectural form generation, enabling iterative design of responsive spaces. The system's hardware-algorithm integration ensures flexibility, scalability, and practical applicability in responsive architecture.

The system still faces certain limitations. For instance, with regard to platform latency, Grasshopper receives behavioural input data at a default frequency of 250 ms, while the speed of spatial model generation is currently affected by the complexity of both geometry and interaction logic, which can be mitigated by simplifying mesh counts and other optimization strategies. Moreover, the platform's current applications have been limited to design studio contexts. Future development should focus on

enhancing system stability and improving support for more complex scenarios, thereby providing a truly practical and deployable design tool.

Attribution statement:

ChatGPT (OpenAI, 2024) was used to (1) assist in translating the manuscript and (2) correct spelling errors.

References

Carlo Ratti Associati. (2008). *Digital Water Pavilion.* Carlo Ratti Associati. Retrieved October 28, 2025, from https://carlorattiassociati.com/project/digital-water-pavilion/.

Frazer, J. (1995). *An evolutionary architecture: Themes VII.* Architectural Association Publications.

Hidayat, A., & Gegana, G. A. (2025). Virtual skill development centre based on neuroarchitecture approach for young people. In *30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia: Architectural Informatics, CAADRIA 2025, Volume 2* (pp. 501–510). The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation,* 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Koolhaas, R., Obrist, H. U., Ota, K., & Westcott, J. (2011). *Project Japan: Metabolism talks.* Taschen.

Lee, J. H., Ostwald, M. J., & Kim, M. J. (2021). Characterizing smart environments as interactive and collective platforms: A review of the key behaviours of responsive architecture. *Sensors,* 21(10), 3417. https://doi.org/10.3390/s21103417.

Living Architecture Systems Group. (2025). *Can architecture integrate living functions?*. Retrieved October 28, 2025, from https://livingarchitecturesystems.com/about/.

Makanadar, A. (2024). Neuro-adaptive architecture: Buildings and city design that respond to human emotions, cognitive states. *Research in Globalization,* 8, 100222. https://doi.org/10.1016/j.resglo.2024.100222.

Mehta, R., Zhu, R. (J.), & Cheema, A. (2012). Is noise always bad? Exploring the effects of ambient noise on creative cognition. *Journal of Consumer Research,* 39(4), 484–799. https://doi.org/10.1086/665048.

Sanei, S., & Chambers, J. A. (2007). *EEG signal processing.* John Wiley & Sons, Ltd.

Silcock, D., Schnabel, M. A., Moleta, T., & Brown, A. (2021). Participatory AR - A parametric design instrument. In *26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia: Projections, CAADRIA 2021, Volume 2* (pp. 295-304). Association for Computer-Aided Architectural Design Research in Asia (CAADRIA). https://doi.org/10.52842/conf.caadria.2021.2.295.

Simon, S. (2005). *Archigram: Architecture without architects.* MIT Press.

Tsoumpri, D. (2023). Architectural space as an open, adaptable system: A design experiment. *International Journal of Architectural Computing,* 21(4), 602-621. https://doi.org/10.1177/14780771231162382.

Yan, C., Li, T., Xing, M., Jin, Y., Yao, J., Lu, X., & Fang, Y. (2025). The relationship between illumination and social distance in indoor corridors: a behaviour-based experimental study of peri-personal space & inter-personal space. *Architectural Science Review,* 1–17. https://doi.org/10.1080/00038628.2025.2562021

Zhang, L. (2021). Urban ergonomics: An emerging field of cross-disciplinary studies. *World Architecture,* (3), 8–9.

Streets as Shared Canvases

An In-Situ Mixed Reality Platform for Citizen-Led Urban Restorative Design

Yuhao He [1], Tian Tian L**o** [2], Fenggui Rao [3]
[1,2,3] *The Hong Kong Polytechnic University.*
[1] *yuhao.he@connect.polyu.hk, 0009-0005-7640-492X*
[2] *tt-sky.lo@polyu.edu.hk, 0000-0002-1992-0777*
[3] *robbie.rao@connect.polyu.hk, 0009-0000-8990-1515*

Abstract. Urban streets are vital public interfaces where spatial form and everyday movement converge to shape human well-being. As health-oriented urban design increasingly values restorativeness, current evaluation methods remain largely detached from lived street experience. Photographic audits and lab-based virtual walkthroughs capture static form but overlook embodied and dynamic dimensions. To address this gap, this study develops an In-Situ Mixed Reality (ISMR) platform that frames streets as shared canvases for citizen-led restorative design. The platform enables on-site mixed reality (MR) assessments of perceived restorativeness. The MR environment situates perception, interaction, and design through large language models and context-aware visual synthesis. Field deployment revealed that localized social, micro-spatial, and sensory affordances, such as resting amenities, spatial maintenance, and micro greenery, exert a strong influence on perceived restorativeness. These findings indicate that restorative potential emerges not from monumental spatial form in isolation but through the cooperative interplay between large-scale morphology and assorted human-scale affordances of street life. By foregrounding perception as design intelligence, the study suggests new pathways for embedded human–AI collaboration in shaping emotionally responsive and citizen-authored urban environments.

Keywords. Urban micro interventions, In-Situ Mixed Reality (ISMR), AI-assisted co-creation, restorativeness

1. Introduction

Urban streets constitute a primary interface of human perception and social life. Yet, as urban streets expose citizens to cumulative sensory stressors, such as traffic congestion and noise pollution, the question of how streets can foster urban health has become a pressing concern. Restorativeness, defined as an environment's capacity to support psychological recovery, offers a valuable theoretical lens for addressing this question. However, current approaches to assessing street restorativeness, such as

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 103-112.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

photo audits, virtual walkthroughs, or post-hoc questionnaires, underplay embodiment and temporality. This disjunction between evaluative representation and lived perception has limited theoretical insight and guideline executability.

To address these limitations, the study introduces an In-Situ Mixed Reality (ISMR) platform that integrates perception capture, spatial analysis, and participatory creation into a unified workflow. The platform pursues three aims: (1) to enable ecologically valid assessment of perceived restorativeness via the ISMR platform; (2) to develop a human–AI co-creation framework that interprets citizen input into micro interventions; and (3) to examine how and which urban elements influence restorative perception. Through this workflow, the study contributes to the following domains. First, it pioneers an ISMR platform for situated evaluation, expanding restorativeness research from lab-based representation to field-based embodiment. Second, it establishes a human–AI co-creation framework that leverages large language models (LLMs) and context-aware visual synthesis, enabling a form of crowdsourced design authorship. Third, it proposes a microscale restorativeness mechanism that elucidates how localized sensory and situational affordances collectively shape restorative experience.

2. Literature review

2.1. RESTORATIVENESS IN URBAN STREETS

Grounded in Attention Restoration Theory (ART) (Kaplan, 1995) and Stress Recovery Theory (SRT) (Ulrich et al., 1991), early research on restorativeness predominantly examined natural environments. As research attention has increasingly turned toward built environments, urban streets have been recognized as potential restorative settings that shape psychological well-being. A growing body of evidence demonstrates that streets' naturalness, captured by metrics such as green view index and vegetation types, is a dominant predictor of perceived restorativeness (Baniadam et al., 2025). Spatial morphology likewise plays a critical role: street canyons with a moderate width–height ratio (Zhang et al., 2025) and adequate sky visibility (Park & Ergan, 2025) tend to elicit stress recovery. As illustrated in Figure 1, most studies have followed nature-dominant and morphological pathways, linking visual attributes to psychological outcomes through photographic and Virtual Reality (VR) representations.

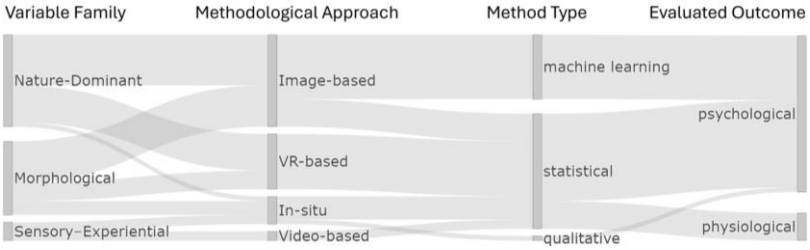


Figure 1. Mapping of urban street restorativeness studies (2015-2025)

However, this visually centred paradigm has produced what might be analytically rich but experientially thin knowledge. Few studies have examined how sensory-experiential factors mediate restorative perception in the lived street environment.

While computational approaches, such as Multiscale Geographically Weighted Regression (MGWR) or Graph Neural Networks (GNNs), have significantly improved predictive accuracy, they still rely on these visually derived inputs. Consequently, current understandings of urban street restorativeness remain rooted in macro and visual representation rather than embodied experience, highlighting the need to capture how people sense street space in real time.

2.2. IMMERSIVE AND SITUATED EVALUATION METHODS

To better approximate real-world experience, urban environmental research has adopted immersive paradigms to enhance ecological validity beyond static imagery. VR-based approaches enable parametric manipulation under controlled conditions, facilitating psychophysiological measurements and causal inference (Chen et al., 2024). Nevertheless, most immersive environments remain predominantly visual, with auditory cues serving as the primary supplement to enhance realism (Dong et al., 2024). Consequently, ex-situ VR studies are constrained by the immaturity of multisensory simulation technologies and the absence of natural human-environment interaction.

Correspondingly, recent work has shifted toward in-situ experiments, leveraging mobile and wearable sensing to collect synchronized biodata and environmental data during natural walking (Lei et al., 2026). These methods capture affective dynamics within real-world contexts but still follow a linear workflow of data collection followed by retrospective analysis. As a result, psychological traces are rarely interpreted in real time through context-aware human–AI systems. Moreover, existing frameworks seldom close the loop between perceptual experience and design feedback, limiting their relevance for participatory urban decision-making. This gap highlights the need for in-situ mixed-reality (MR) methods bridging perception, meaning, and action.

2.3. PARTICIPATORY URBAN PLATFORMS

While traditional immersive methods position users as subjects of measurement, participatory urban paradigms reconceptualize citizens as active agents of design. Rooted in participatory design and increasingly informed by citizen science, these approaches seek to democratise urban decision-making by embedding citizens directly in the creative process. Based on the reviewed literature, participatory urban platforms can be broadly categorised into four types according to their interaction media: screen- (Yao et al., 2025), VR- (Ehab et al., 2025), Augmented Reality (AR)- (Boos et al., 2023), and Tangible User Interfaces (TUI)-based (Hesselbarth et al., 2025).

Despite these diverse forms, current participatory platforms still exhibit a bias toward macro-scale planning rather than human-centred spatial experience. Most systems prioritize technical performance and usability-centric evaluation over their design utility. Moreover, their workflows frequently rely on structured, ex-situ workshops and expert facilitation, constraining in-situ spontaneity and long-term public engagement. Consequently, participatory urban design remains episodic instead of evolving into a continuous civic practice. To bridge this gap, platforms that embed co-creation within everyday urban encounters are needed.

To sum up, the field still lacks an integrated framework that (a) collects real-time perceptual data during natural walking, (b) analyses contextual dynamics through

human–AI interpretation, and (c) translates participatory citizen input into actionable urban design feedback for continuous, situated co-creation.

3. Methods

3.1. CONCEPTUAL FRAMEWORK

Building on the above gaps, this study proposes an In-Situ Mixed Reality (ISMR) platform, integrating ecological measurement, MR interaction, and AI-assisted co-creation. This study adopts Actor–Network Theory (ANT) to frame the ISMR platform as a heterogeneous network of human and non-human actors. Applying this lens, the platform is conceptualized as a dynamic ecology where humans, embodied AI, and urban streets co-construct the restorative experience (Figure 2).

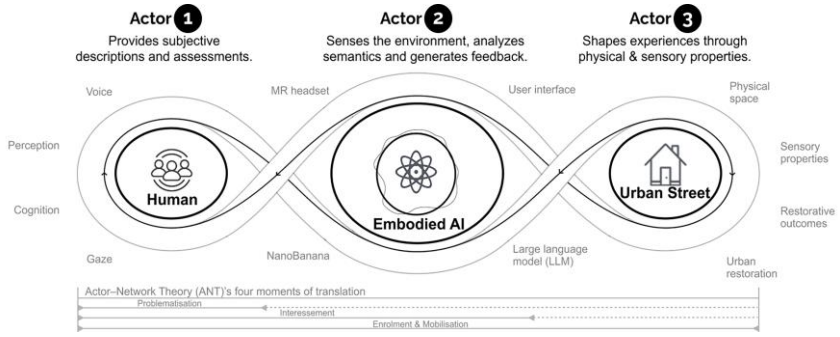


Figure 2. Conceptual diagram of the human–AI–environment network

In this human-AI-environment network, the human actor provides perceptual descriptions and subjective assessments; the embodied AI actor interprets and translates these expressions into semantic and visual representations; and the urban space actor shapes experience through its physical and sensory affordances, mediating environmental cues that influence perception. Following ANT's moments of translation, the ISMR platform orchestrates these interactions into a living system where design knowledge and restorative meaning emerge from distributed agency.

3.2. ISMR PLATFORM BUILDING

The ISMR platform was developed in Unity and designed for MR glasses to enable unobtrusive in-situ interaction. The User Interface emphasises a layered dialogue between perception and interpretation (Figure 3a). This configuration transforms embodied perception into structured input. Building on this platform, a Human–AI Co-creation Pipeline was developed, operating in two modes: Capture Mode, where users record and describe their immediate surroundings, and Explore Mode, where they refine proposals created by others (Figure 3b). Users capture a scene or select an earlier proposal and voice their perceived restorativeness and desired improvements. To minimise interruption during on-foot MR use, restorativeness was captured via a brief single-item overall rating (1–7; not restorative–highly restorative) rather than a long multi-item questionnaire. Their input is processed through an LLM–NanoBanana

pipeline: the LLM interprets semantic intent, while NanoBanana generates context-aware visual updates conditioned on the scene. Generated overlays were screened for scene fidelity and scale plausibility; failed outputs were re-generated with tightened prompts, otherwise logged and excluded. This workflow enables both human–AI and peer co-creation, turning each MR encounter into an incremental, citizen-authored restorative micro-intervention.

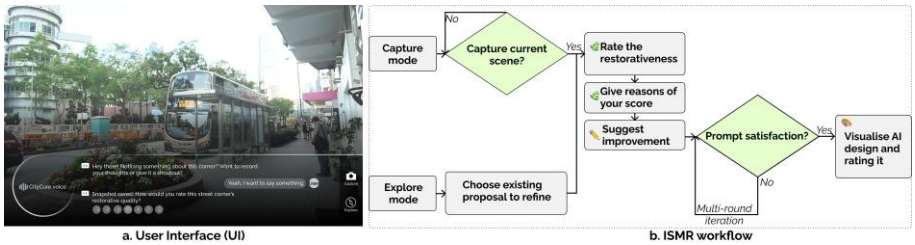


Figure 3. User interface and ISMR workflow

3.3. FIELD EXPERIMENT AND DATA ANALYSIS

A field experiment was conducted along Chatham Road South in Hong Kong, within a 400-meter segment of a prototypical high-density urban street, using the ISMR platform (Figure 4). The site offered two contrasting pedestrian conditions: one side comprised a conventional sidewalk lined with mixed commercial–residential frontage, while the opposite side incorporated a sequence of embedded micro-plazas within the pedestrian corridor. Participants were randomly assigned to walk on one side to ensure balanced exposure across conditions.



Figure 4. Study area along Chatham Road South: satellite map (Left) and Nolli map (Right)

The pilot experiment was implemented on Meta Quest 3 headsets, which served as a substitute for lightweight MR glasses given current hardware limitations. Twenty participants (10 male, 10 female; aged 20–35) were recruited through on-campus poster advertisements. Each participant experienced a 20-minute natural walk while wearing the MR headset, during which they completed approximately ten Capture and Explore tasks, followed by the Single Ease Question (SEQ) and the System Usability Scale (SUS) questionnaire. The study protocol was approved by the institutional ethics committee and all participants provided informed consent.

During each in-situ session, the ISMR platform automatically captured multimodal data, including geospatial coordinates, subjective restorativeness scores, verbal

feedback, and records of co-creation interactions. To evaluate the ecological validity of the ISMR platform, usability measures were analysed to confirm system stability and comfort. Co-creation events were further modelled into a co-creation graph to trace the provenance of peer-to-peer interventions, with frequency-based clustering used to identify engagement hotspots. Finally, textual feedback was analysed through a hybrid LLM–human synthesis to identify the underlying mechanism driving perceived restorativeness. In parallel, a Restorative Mood Map was generated to visualize the overall distribution of perceived restorativeness across the study area.

4. Results

The ISMR platform demonstrated stable usability and fluency across field deployment. Participants reported high ease of operation across functional phases, as reflected in SEQ ratings (Figure 5), with the co-creation phase achieving the highest mean score (M = 6.0). In this deployment, NanoBanana generated context-aware overlays within 10 seconds, and outputs were screened for semantic consistency and scale plausibility; failed cases (e.g., scale distortions) were re-generated under tightened prompts or excluded. As NanoBanana is an off-the-shelf diffusion model with limited public disclosure of architecture and training data, we treat it as a black-box generator and report reliability at the pipeline level rather than claiming model-level accuracy. SUS testing yielded an average score of 73.9 and good internal consistency (Cronbach's α = 0.867), corresponding to "Good" usability under Bangor's benchmark. These results confirm that the prototype functioned reliably under complex in-situ street conditions. Building upon this validation, the following sections detail the human–AI co-creation dynamics that emerged during participatory restorative design, as well as the restorativeness mechanism derived from participants' perceptual and semantic data.

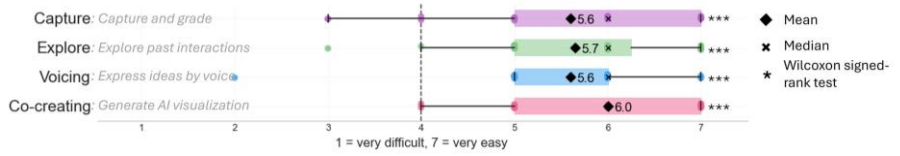


*Figure 5. Box-and-Whisker plot showing SEQ ratings (P < .001 ***)*

4.1. HUMAN–AI CO-CREATION DYNAMICS

A total of 196 valid co-creation events for restorative design were recorded across 20 participants. Quantitative comparison of restorativeness ratings before and after AI-assisted co-creation revealed a significant improvement in perceived restorativeness, with mean scores rising from 0.65 to 1.73 (Figure 6). Specifically, 70.62 % of samples showed higher perceived restorativeness after co-creation, 19.07% remained unchanged, and 10.31% decreased. These results suggest that iterative human–AI collaboration was associated with higher reported restorativeness, and that the co-creation process tended to increase citizens' ratings in this pilot. The magnitude of this shift suggests that the process is not merely cosmetic but perceptually meaningful.

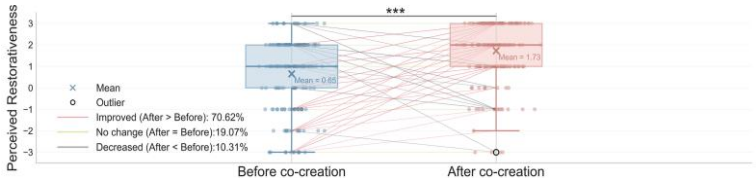


*Figure 6. Comparison of perceived restorativeness before and after co-creation (P < .001 ***)*

A dynamic network of iterative interventions mediated by the AI system emerged in this experiment (Figure 7). The radial diagram visualises provenance trajectories: each node represents a captured or generated scene image, and each link denotes an explicit reference relationship, i.e., the later proposal was created by selecting the earlier image as the reference. Nodes near the centre correspond to the originally captured reality, while those toward the periphery represent successive derivatives produced through citizen edits and AI-assisted reinterpretation. The longer a radial branch, or the greater the number of derivatives linked to a single anchor node, the stronger the collective interest and engagement elicited by that scene. As shown in Figure 7, the visualized trajectories reveal how the human–AI loop shaped collective design. Participants converged toward several recurrent restorative themes (e.g., micro greenery, resting amenities), which will be elaborated in Section 4.2, while the AI introduced serendipitous alternatives that broadened the design search space.



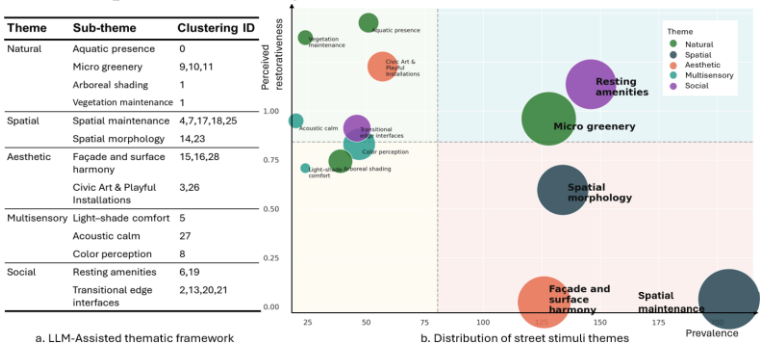
Figure 7. Radial co-creation network of human–AI iterations (left) and examples (right)

4.2. RESTORATIVENESS MECHANISM

To examine how citizens collectively perceive street restorativeness, sentence embeddings were generated using MiniLM-L12-v2 (Sentence Transformers), reduced to 20 dimensions via UMAP, and then clustered using HDBSCAN (chosen for its ability to handle variable-density clusters and label noise). We compared embedding models and clustering configurations and selected the setting that yielded the strongest internal validity and thematic coherence (clusters = 29; noise = 0.30; DBCV = 0.359; Silhouette = 0.583), outperforming BGE-M3 and BGE-Small-ZH-v1.5 (BAAI). Given the known sensitivity of unsupervised clustering, we treat cluster boundaries as exploratory and rely on subsequent expert interpretive coding in a hybrid LLM–human synthesis to consolidate results into five overarching themes and thirteen sub-themes (Figure 8a): 'Natural' features embody ecological vitality; 'Spatial' themes capture morphology and maintenance of built environment; 'Aesthetic' reflects façade articulation and visual expressiveness; 'Multisensory' concerns the balance of light,

colour, and sound; and 'Social' focuses on human affordances for rest, proximity, and interaction.

Building upon this framework, a bubble-plot model (Figure 8b) mapped the relationship between each sub-theme's prevalence and its perceived restorativeness, with bubble size representing the number of captured scenes. The vertical and horizontal mean lines divide the visualization into four quadrants, from which two major zones emerge. In the upper-right quadrant, resting amenities and micro greenery stand out as both widely perceived and strongly restorative, suggesting that small, localised social and biophilic affordances are commonly perceived as supportive of everyday recovery. In contrast, the lower-right quadrant is dominated by spatial maintenance and façade and surface harmony: these cues are often noticed but tend to co-occur with lower restorativeness ratings, consistent with participants' negative appraisals of neglect and visual discord. Spatial morphology appears in the same quadrant but as a weaker case, indicating a milder association in this dataset. Notably, high-prevalence themes are not necessarily restorative: maintenance and surface discord were frequently mentioned yet associated with lower ratings, suggesting a "deficit-salience" pattern where negative cues dominate citizen narratives.



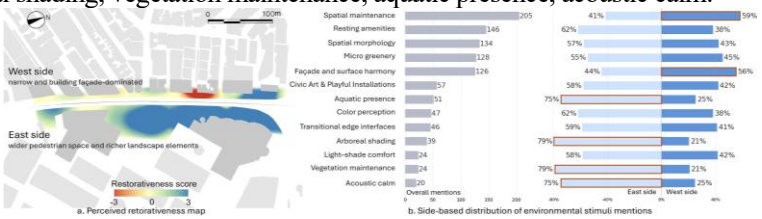
Theme	Sub-theme	Clustering ID
Natural	Aquatic presence	0
	Micro greenery	9,10,11
	Arboreal shading	1
	Vegetation maintenance	1
Spatial	Spatial maintenance	4,7,17,18,25
	Spatial morphology	14,23
Aesthetic	Façade and surface harmony	15,16,28
	Civic Art & Playful Installations	3,26
Multisensory	Light–shade comfort	5
	Acoustic calm	27
	Color perception	8
Social	Resting amenities	6,19
	Transitional edge interfaces	2,13,20,21

a. LLM-Assisted thematic framework

b. Distribution of street stimuli themes

Figure 8. LLM-assisted thematic framework and distribution of street restorativeness stimuli

The spatial distribution of perceived restorativeness (Figure 9a) reveals a noticeable asymmetry between the two pedestrian sides. Semantic analysis of participant expressions (Figure 9b) further suggests differing thematic emphases. The western sidewalk—narrow and façade-dominated—was associated with lower reported restorativeness and more mentions of spatial maintenance and façade/surface concerns. In contrast, the eastern sidewalk, with wider pedestrian space and richer landscape elements, was associated with higher reported restorativeness and more mentions of arboreal shading, vegetation maintenance, aquatic presence, acoustic calm.



a. Perceived retorativeness map

b. Side-based distribution of environmental stimuli mentions

Figure 9. Perceived restorativeness map and diverging thematic emphases by different pedestrians

5. Discussion

The ISMR platform demonstrated good ecological validity in the field deployment. Compared with lab-based evaluations, it captures in-situ, location-specific fluctuations in experience during real walking. Beyond usability, this supports a practical workflow between citizen feedback and AI-assisted visual proposals: participants describe what feels restorative (or not), and the system generates scene-conditioned alternatives that help them express and refine design ideas. Crucially, the AI's reinterpretation often introduces serendipity, generating imaginative yet locally grounded alternatives that expand users' design thinking.

In this pilot, morphology acted more like a structural constraint than a direct driver of restorativeness, while micro-scale affordances accounted for most perceived variation. Street-canyon form provides a relatively fixed spatial baseline, while day-to-day restorative relief appears to hinge on small, human-accessible affordances —for example shade, places to sit, permeable frontage, and façade upkeep. It is this tension between morphological stability and micro-scale variability that produces perceptual divergence within the same street segment of Chatham Road South. To translate this into a computational design method, future work will quantify morphology via street-canyon geometry (e.g., height-to-street width ratio, sky view factor) and micro-scale variability via editable affordance proxies (shade, seating, frontage permeability), enabling counterfactual micro-intervention simulation. The resulting executable designs bridge the long-standing gap between experiential feedback and spatial policy, enabling cities to listen to and respond to their inhabitants.

This study remains an early-stage validation. The modest sample size and narrow age range limit generalisability. Using a Meta Quest 3 also introduces device effects (weight and restricted field of view), which may add noise to ratings during a 20-minute walk. The precision of the Restorativeness Mood Map is challenged by GPS drift, while occasional scale distortions in AI-generated overlays call for stricter filtering and failure-case handling in the co-creation workflow. Future work will refine localisation and interaction flow, expand the dataset with richer multimodal environmental metadata, and replicate deployments with more diverse participants across multiple street contexts.

6. Conclusion

This paper presents an in-situ mixed-reality method for citizen-led restorative street design. Through the ISMR platform, the research demonstrates how restorativeness can be sensed, interpreted, and re-created within the very spaces where it is experienced. By enabling citizens to articulate perceptions, witness AI-generated visualizations, and refine collective visions, the platform transforms restorativeness into a collectively authored design practice. A future emerged, where urban design becomes less about planning from above and more about listening from within. In this future, the city becomes an evolving interface that learns from its inhabitants, heals through their collective imagination, and, ultimately, restores the humanity that builds it.

Acknowledgements

This study was approved by the Institutional Review Board of The Hong Kong Polytechnic University.

Attribution statement

ChatGPT (ChatGPT, 2025) was used to refine writing flow and grammar.

References

Baniadam P., Mars F., Normand J.-M., Requena-Ruiz I., & Siret D. (2025). Mental restorative effects of streetscape vegetation: Assessment of various design strategies using virtual environments. Urban Forestry & Urban Greening, 112, 128938. https://doi.org/10.1016/j.ufug.2025.128938

Boos, U. C., Reichenbacher, T., Kiefer, P., & Sailer, C. (2023). An augmented reality study for public participation in urban planning. Journal of Location Based Services, 17(1), 48–77. https://doi.org/10.1080/17489725.2022.2086309

ChatGPT. (2025). ChatGPT Model. [Large Language Model]. https://chatgpt.com/

Chen, D., Yin, J., Yu, C.-P., Sun, S., Gabel, C., & Spengler, J. D. (2024). Physiological and psychological responses to transitions between urban built and natural environments using the cave automated virtual environment. Landscape and Urban Planning, 241, 104919.

Dong, W., Dai, D., Shen, P., Zhang, R., & Liu, M. (2024). How public urban space enhance restoration benefits through combined multisensory effects: A systematic review. Land, 13(12), 2018.

Ehab, A., Aladawi, A., & Burnett, G. (2025). Exploring AI-integrated VR systems: A methodological approach to inclusive digital urban design. Urban Science, 9(6), Article 6. https://doi.org/10.3390/urbansci9060196

Hesselbarth, L., Scholz, T., Dratzidis, L. T., Dänekas, B., & Malaka, R. (2025). Tangible windmills: Exploring tangible user interfaces for urban infrastructure planning. Proceedings of the Mensch Und Computer 2025, 726–731. https://doi.org/10.1145/3743049.3748550

Kaplan S. (1995). The restorative benefits of nature: Toward an integrative framework. Journal of environmental psychology, 15(3), 169–182.

Lei, B., Liu, P., Fujiwara, K., Frei, M., Miller, C., Chua, Y. X., & Biljecki, F. (2026). Multidimensional analysis of human outdoor comfort: Integrating just-in-time adaptive interventions (JITAIs) in urban digital twins. Cities, 168, 106443.

Park, K., & Ergan, S. (2025). Quantification of the impact of street design features on restorative quality in urban settings. Sustainable Cities and Society, 121, 106216. https://doi.org/10.1016/j.scs.2025.106216

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. Journal of Environmental Psychology, 11(3), 201–230. https://doi.org/10.1016/S0272-4944(05)80184-7

Yao, Y., Li, Z., & Ma, S. (2025). Exploring gamification's role in shaping socially sustainable urban spaces: A case study of gensen in SOLANA, beijing. Buildings, 15(12), Article 12. https://doi.org/10.3390/buildings15121969

Zhang, N., Zhao, L., Guan, G., & Liu, C. (2025). Investigating the impact of street width-to-height ratios on stress recovery. Indoor and Built Environment, 34(5), 993–1004. https://doi.org/10.1177/1420326X251327543

From Automation to Interaction

A Representation Framework for Integrating Occupant Control in Intelligent Environments

Ilayda Memis[1], Yuval Kahlon[2], Frank Petzold[3], Davide Schaumann[4], and Christiane M. Herr[5]
[1,3]*Technical University of Munich.*
[2]*Institute of Science Tokyo.*
[4]*Technion – Israel Institute of Technology.*
[5]*Southern University of Science and Technology.*
[1]*ilayda.memis@tum.de, 0000-0003-1063-1327*
[2]*kahlon.y.af4e@m.isct.ac.jp, 0000-0002-3914-7945*
[3]*petzold@tum.de, 0000-0001-8974-0926*
[4]*d.schaumann@technion.ac.il, 0000-0001-9899-2818*
[5]*cmherr@sustech.edu.cn, 0000-0002-1381-8288*

Abstract. Intelligent environments (IEs) integrate sensing, automation, and actuation technologies to support human activity. However, the growing complexity of these systems has created a need for clear conceptual frameworks on how control and decision-making are shared between humans and buildings. This paper introduces a representation framework designed to classify and visualise different levels of human involvement within IEs. The proposed framework identifies system components and information flows, while illustrating varying degrees of automation and providing a structured means to compare and analyse human–building interactions. By testing this approach in a case study, this framework aims to facilitate interdisciplinary coordination and support the design of more transparent, inclusive, and human-centred IEs.

Keywords. Intelligent environments, human-building interaction, levels of automation, occupant control

1. Introduction

The field of intelligent environments (IEs) is broad and multifaceted, encompassing multiple disciplines and attracting interest from diverse groups of people. However, each discipline approaches IEs from its own perspective, resulting in a fragmented understanding of human–system interaction and the varying degrees of automation involved (Luna-Navarro et al., 2020, p. 2).

As IEs become more adaptive and data-driven, the complexity of interactions

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 113-122.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

between humans and buildings continues to grow. Understanding these interactions requires not only identifying potential levels of automation but also clarifying how decision-making and execution are distributed between human users and intelligent systems. For designers, finding an appropriate balance between automation and human control presents a significant challenge. While highly automated systems are often assumed to improve building efficiency, reduced human agency can lead occupants to override or interfere with automated operations, ultimately decreasing performance and satisfaction (Koyaz et al., 2022, p. 30). Therefore, it is crucial for designers to understand and provide occupants with the exact level of control needed of the system.

Existing classifications often focus on automation technologies or functional capabilities, with less emphasis on the human dimension of control. To bridge this gap, this paper introduces a human-oriented perspective to the field of IE, focusing on the dynamics of human control within automated systems. Existing levels of automation classification were adapted to IEs, based on current research and case studies, to incorporate multi-user scenarios and recent developments in machine learning. Further, a representation framework is proposed that illustrates key system components, information flows, and varying levels of human involvement. This framework enables researchers and designers to situate their work within the broader continuum of automation, facilitating clearer understanding and dialogue across disciplinary boundaries.

2. Theoretical Framework for Shared Control in Intelligent Environments

2.1. REINTERPRETING AN EXISTING TAXONOMY FOR IE

This paper focuses on the representation of existing automation systems in the context of IEs. We consider, as a starting point, the level of automation taxonomy presented by Vagia et al. (2016, p. 196), due to its increasing level of automation in human-computer systems. In the classification system, there are eight levels of automation, ranging from a completely manual system (level 1) to a completely autonomous one (level 8). The authors did not specify a use case for the levels of automation; instead, the taxonomy was presented as suitable for a general HCI system (Vagia et al., 2016, p. 192). In this paper, this taxonomy is taken as a basis and adapted specifically to IEs.

Figure 1 illustrates a classification of automation levels and their differences applied to an IE context. This original classification describes interactions between a single user and the system, often reflecting scenarios where a building manager makes decisions on behalf of all occupants. Therefore, in this case, "human" denotes the building manager. Level 1 represents a manual system, where no information exchange occurs. As this case is not very relevant to the IE context, it was faded out in the diagram. Similarly, the levels after 6 were also faded out, as they represent a fully automated system where occupants are only responsible for fixing errors in case of problems. Instead, we consider only levels that consider a balance between building automation and human control.

For levels 2-6, the building is responsible for analysing the environment and offering different options for possible actions to improve the situation. In level 2, where people have the highest level of control, the system offers these options, and the

building manager is responsible for both decision-making and the execution of those decisions. The execution of tasks by people can be classified as human adaptability, which encompasses various examples incorporating spatial, social, and operational adaptations (Schaumann, 2024, p. 9).

In level 3, people still make decisions based on offered options, but decision execution is delegated to the building systems, mainly mechanical actuators. In level 4, the system calculates different options and selects the most suitable one. However, people are still included in the control mechanism before the execution phase. The system communicates the decision to the people and requests their approval. The execution starts after the decision is approved. In level 5, people are involved only after the whole process is finished. After the decision execution phase, people are notified about the decision. Level 6 is similar to 5, but the decision is communicated to people only if requested. The arrows represent human-building communication. As seen from levels 3 and 4, there is a two-way communication between the people and the system; unlike at other levels, where the interaction only involves informing people about the options or changes. With the help of this diagram, the differences between various levels of automation can be clearly categorised.

A ventilation system in an office building could be used as an example to explain this concept. If sensors send the information to the controller about CO_2 levels in an environment and the locations where people work, for level 2, it can inform the building manager which windows can be opened for how long to make the air fresher. For level 3, the manager selects one of these options, and the ventilation system is activated mechanically. For level 4, the system requests the manager's approval to activate ventilation for a specified duration, and it is activated only after the manager's approval is obtained. For level 5, ventilation has already begun, and the manager is notified afterwards. For level 6, everything works automatically, and the manager wants to verify that it is functioning properly by checking the information through the interface.

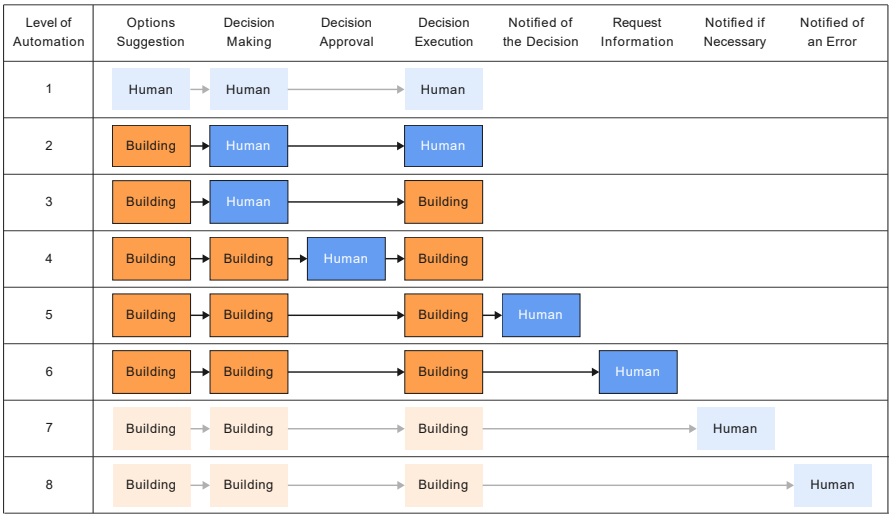


Figure 1: Human involvement for different levels of automation (adapted from Vagia et al., 2016).

2.2. EXTENDING THE FRAMEWORK WITH HUMAN-IN-THE-LOOP CAPABILITIES

Conventional IEs for shared spaces, such as offices and libraries, provide occupants with limited control. Building management decisions are either made by the system or the building manager. Therefore, systems only allow the interaction between the building manager and the interface (Boychenko, 2019). Even though these automated systems take the responsibility of controlling the environment away from people and reduce their burden, research shows that people feel uncomfortable when they have no control over their surroundings (Koyaz et al., 2022, p. 37).

On the other hand, new research suggests that other types of systems can be created, enabling people to have more power over the systems they inhabit (Yuill & Rogers, 2012, pp. 1-2). At the same time, recent developments in machine learning allow these algorithms to be updated in the long term based on user feedback (Mofidi & Akbari, 2020, p. 4), which was not included in the original taxonomy of levels of automation. Therefore, the classification system described in Section 2.1 was extended to include multi-user scenarios and a feedback-based learning algorithm.

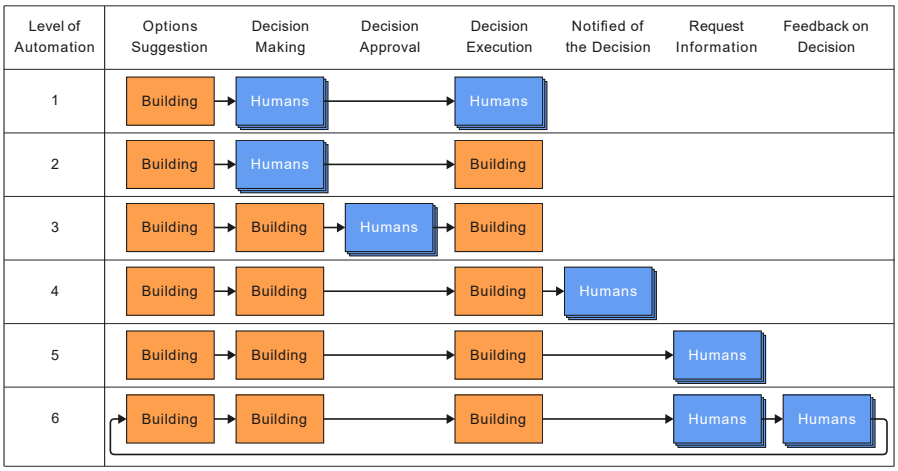


Figure 2: Extended representation of human involvement in IEs.

As seen in Figure 2, which is an adaptation of Figure 1, the role of human involvement within the flow of information remains relatively unchanged. However, within the multi-user scenarios, the interacting person does not have to be the building manager; it can be any occupant who experiences the changes in the environment. Therefore, in this figure, humans are represented as 'stacked' icons. Additionally, a new step in the flow of information is added, called feedback on decision. It represents the incorporation of a feedback-based learning algorithm to reflect current advancements in the processing systems. Since this is a novel form of control, it is represented as a new level 6. Similar to levels 4 and 5, for level 6, people are informed after the decision has been executed. However, this time, they have the chance to provide feedback about their discomfort. It is not direct control, but the impacts can be observed in the long term, as highlighted with the additional arrow.

For levels 1-2, there are parallels between single-user and multi-user scenarios in terms of interaction mechanisms. Both of them involve changes or decisions made by one person. However, in the case of multi-user IEs, options should be made available to everyone simultaneously. Since the interface is typically positioned in a central location in the room, interactions are carried out in a public manner. Thus, introducing additional factors that influence how these technologies are used or underused. These factors include familiarity between co-users (e.g., family, coworkers, or acquaintances), hierarchy between them (e.g., parent-child, employer-employee), or proximity of the users to the device (Shavitt et al., 2021, pp. 4-9), making the design of multi-user systems more complex than single-user ones (Augusto et al., 2013, p. 13).

For level 5, the transformation of the single-user systems into multi-user ones is also straightforward, as it primarily serves as an information display mechanism. It can be a screen located within the space that is accessible for everyone (Lassen et al., 2021), or, regarding more personal interfaces, a phone app (Winkler et al., 2016) or a toolkit (Luna-Navarro et al., 2021) that allows users to view the details of environmental calculations. And in the case of level 6, it is used to provide feedback on their comfort.

3. Representation Framework for Human-Building Interactions

We present a novel representation framework to illustrate the extended version of the levels of automation classification system presented in Section 2.2. The goal is to make this complex system understandable and clearly illustrate its elements, processes, and information flow. The representation system categorises the components of IEs into six groups: environmental, social, sensing, interaction, processing, and actuation. To ensure broad applicability across contexts, the environmental group is represented by a generic environmental quality element, which can be replaced or extended as needed. The social group comprises three actors: the building manager and two occupants (occupant 1 and occupant 2). To preserve continuity with this conventional model while accommodating more contemporary, participatory systems, the manager's role is kept in the representation. The inclusion of two occupants enables representation of shared environments, with the number adjustable to reflect specific case studies.

For the sensing group, various sensors were included to illustrate the different types of equipment used to measure the surroundings in a given space, including both environmental and human-related sensors. The different types of sensors were not specified in detail, since the focus of this paper is to represent different levels of human control. For the interaction group, one interface was added to the system, as the only interaction taking place is between the building manager and the system. For processing, the controller element is responsible for processing the data from sensors, suggesting options, and, at higher levels of autonomy, making decisions. For the actuation, architectural elements were selected that have an impact on the space in terms of environmental quality and human comfort. When the actuators change the environment, both environmental quality and people are impacted, and this is true for each level of automation. Therefore, the system can be considered a loop rather than a linear process. However, in this figure, the lines represent the flow of information, and the final actuator impact involves a physical exchange rather than an informational one. Therefore, to provide a clearer representation, the final impact is excluded, and the process is depicted as linear.

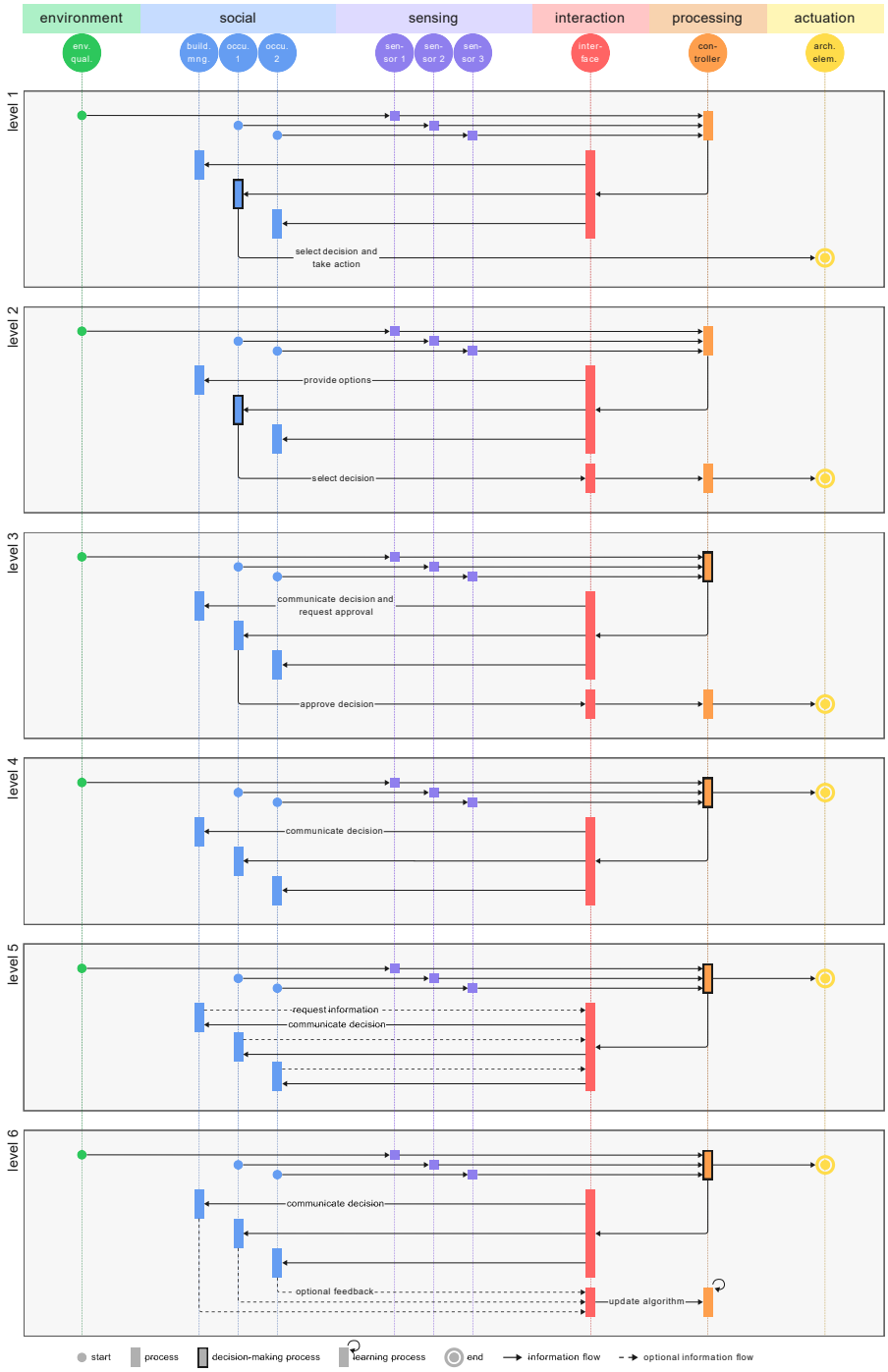


Figure 3: Levels of automation representation for multi-user decision making.

Each level begins with sensors collecting data on environmental quality and the occupants. This data is sent to the controller, and different options for actions are generated. After this point, the levels start to differentiate. In level 1, the options are sent to the interface, which visualises these options for the users. One of them selects the appropriate decision and takes action. The black outline of the occupant 1 rectangle indicates the decision-making capacity within the system. After selecting the appropriate option, occupant 1 also executes this decision. The advantage of this level is that, since the person making the change is visible, others can clearly see why that person is making the change, ensuring they are completely informed about the situation (Zhong et al., 2021). In level 2, after the options are provided to the users, one of them selects the decision on the interface, and the choice is sent to the controller. Afterwards, the system mechanically changes the relevant architectural element. Since the information is displayed on a single interface within the environment, the person making a decision must go to the interface location and click on the selected option. While they are doing that, the action remains visible to others, and occupants will still be informed about the situation (Van De Werff et al., 2017, pp. 211-212).

In level 3, the options are generated, and one of them is selected by the controller. Since it has the decision-making capacity, it is highlighted with an outline for this level. The decision is communicated through the interface, and approval is requested from any one of the users. After they approve, the system initiates the actuators. Similarly, one person approaches the interface and approves the decision, making this action visible to others. In level 4, the system makes the final decision and executes it. Afterwards, the users are notified about the selected decision and the action taken. From this level on, the interface does not have to be publicly located in the space; it can also be designed as a personal device. Since both decision-making and execution are already handled by the system, and occupants do not have any direct control, it is not crucial for occupants to see each other's interactions. At level 5, the selected decision is communicated only if requested by people. At the final level (6), the decision is communicated to people, and they are allowed to give optional feedback (Lassen et al., 2021, p. 6). This feedback is connected to the learning algorithm within the controller, but does not have a direct impact on actuation. Therefore, it does not exert direct control over the system. However, the feedback does improve the algorithm in the long term, particularly with respect to assumptions of comfort.

4. Representation of a Case Study

As a real-life example, Figure 4 illustrates a complex case study that incorporates multiple levels of automation simultaneously (Gupta et al., 2016, p. 143). Their goal was to create a system where people feel comfortable while reducing energy usage at the same time. That's why they divided the space into different zones and incorporated a feedback mechanism into the system that takes the multi-user situation into consideration. The feedback provided by humans is associated with the relevant zone in the building. It helps the system to configure itself in the long term in a more precise way. Further, the roles of manager/occupant are distinctly separated, and the latter can only view information about the IE and give feedback about their comfort levels. On

the other hand, managers have the decision-making power (Gupta et al., 2016, p. 146).

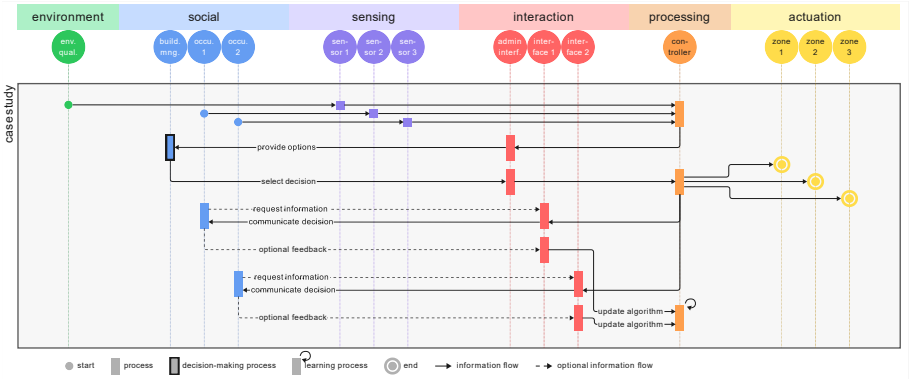


Figure 4: Representation of a multi-user case study.

Figure 4 summarizes this case study. Since this study is focused on a multi-user, multi-zone interaction system, both the interaction and actuation sections are extended. Each user interacts with the system through their mobile devices; therefore, the flow of information for each person is separated, and interactions are individual. The building manager holds the decision-making power in the system, having a level 3 automation for single-user systems. Therefore, their rectangle is outlined, and their interaction is represented at the beginning of the flow. As shown in the diagram, they select a decision from their own interface to be executed immediately. In comparison to the previous diagram, different zones in the actuation section are included, as it was a crucial part of the setup of the case study. Note that the relationship between the physical location occupants and their corresponding zone is not represented.

After this decision is executed, the interaction with the occupants comes into light - occupants can view the changes made to the IE through the app. If they wish, they can also give feedback about how comfortable they feel (Gupta et al., 2016, p. 143). This feedback is connected to a separate rectangle in the diagram, which represents the learning algorithm, as indicated by the loop sign. Feedback does not alter the actuation outcome for this loop of events; however, it will be incorporated into the algorithm in the long term. Therefore, the occupants have level 6 automation for multi-user systems.

As seen in this case study, this representation framework helps us clearly visualize the involved actors, their decision-making capabilities within the system, the stages at which they are involved, and their impact on the execution. It also highlights the coexistence of different levels of automation for various actors within a single system.

5. Discussion & Conclusion

Building upon an existing classification of levels of automation (Vagia et al., 2016, p. 196) and incorporating current research, a representation framework that provides a foundation for discussing IEs was developed, which is especially useful for shared spaces. This representation highlights the varying degrees of human involvement within each level.

First, the original level of the automation classification scheme was investigated in the context of IEs. The original classification is based on interactions between an individual and the system. It depicts that the building manager has both the power and responsibility to make decisions related to environmental quality, comfort, and other aspects, while other occupants passively participate. Such partial control for occupants may result in decreased trust in the system, reduced user satisfaction, and even in overriding or not using the system (Kaasinen et al., 2012, p. 7).

Analysis of recent case studies reveals emerging approaches that integrate multiple occupants into the control process (Gupta et al., 2016; Lassen et al., 2021; Van De Werff et al., 2017; Zhong et al., 2021). To address these developments, the level of automation classification system has been adapted to accommodate more realistic multi-user scenarios and incorporate advancements in machine learning algorithms. In this version, occupants are no longer mere factors in the system; they become active users and are given the power to make decisions.

To illustrate this extended version of the level of automation classification, a visual representation was developed, particularly for shared IEs. This representation highlights the crucial elements within each system and visualizes the flow of information. By juxtaposing different levels of automation side by side, similarities and differences between levels are highlighted. Delegating task execution to the building or adding an extra step of communication before executing decisions changes the entire level of automation in a system. It also highlights that the difference between each level of automation is subtle yet substantial.

The analysed examples are clustered around levels 1-2 or levels 5-6. They suggest that, for levels 1-2, if the control is given to people, the actions should be visible from everywhere (Van De Werff et al., 2017; Zhong et al., 2021). If the level of automation is higher, where they simply view the information (level 5) or provide feedback for long-term calculations (level 6), the information display system can be public (Lassen et al., 2021) or personalised (Gupta et al., 2016). The inclusion of occupants in the system increases people's satisfaction without creating conflict (Aryal et al., 2021). Yet, levels 3-4, which hold potential for more nuanced, context-aware control strategies, remain underexplored, possibly due to their reliance on notification mechanisms that may interrupt daily routines (Aryal et al., 2021). Research on adjustable autonomy systems based on personal preferences also points towards interesting application possibilities (Ball & Callaghan, 2012).

In summary, our framework offers a structured means of understanding and communicating the diverse ways human control can be integrated into IEs. It supports researchers and designers in localizing systems across the automation spectrum and exploring alternative configurations based on available options. By clarifying how control is distributed between users and systems, and communicated to occupants, the framework contributes to the development of more adaptive and inclusive IEs, while expanding the view on automation towards more human-centred systems.

References

Aryal, A., Becerik-Gerber, B., Lucas, G. M., & Roll, S. C. (2021). Intelligent Agents to Improve Thermal Satisfaction by Controlling Personal Comfort Systems Under Different

Levels of Automation. *IEEE Internet of Things Journal, 8*(8), 7089–7100. https://doi.org/10.1109/JIOT.2020.3038378

Augusto, J. C., Callaghan, V., Cook, D., Kameas, A., & Satoh, I. (2013). Intelligent Environments: A manifesto. *Human-Centric Computing and Information Sciences, 3*(1), 12. https://doi.org/10.1186/2192-1962-3-12

Ball, M., & Callaghan, V. (2012). Explorations of Autonomy: An Investigation of Adjustable Autonomy in Intelligent Environments. *2012 Eighth International Conference on Intelligent Environments*, 114–121. https://doi.org/10.1109/IE.2012.62

Gupta, S. K., Atkinson, S., O'Boyle, I., Drogo, J., Kar, K., Mishra, S., & Wen, J. T. (2016). BEES: Real-time occupant feedback and environmental learning framework for collaborative thermal management in multi-zone, multi-occupant buildings. *Energy and Buildings, 125*, 142–152. https://doi.org/10.1016/j.enbuild.2016.04.084

Kaasinen, E., Kymäläinen, T., Niemelä, M., Olsson, T., Kanerva, M., & Ikonen, V. (2012). A User-Centric View of Intelligent Environments: User Expectations, User Experience and User Role in Building Intelligent Environments. *Computers, 2*(1), 1–33. https://doi.org/10.3390/computers2010001

Koyaz, M., Prieto, A., Ünlü, A., & Knaack, U. (2022). Towards a Human Centred Approach for Adaptive Façades: An Overview of User Experiences in Work Environments. *Journal of Facade Design and Engineering, 10*(1), 29–54. https://doi.org/10.47982/jfde.2022.1.02

Lassen, N., Josefsen, T., & Goia, F. (2021). Design and in-field testing of a multi-level system for continuous subjective occupant feedback on indoor climate. *Building and Environment, 189*, 107535. https://doi.org/10.1016/j.buildenv.2020.107535

Luna-Navarro, A., Fidler, P., Law, A., Torres, S., & Overend, M. (2021). Building Impulse Toolkit (BIT): A novel IoT system for capturing the influence of façades on occupant perception and occupant-façade interaction. *Building and Environment, 193*, 107656. https://doi.org/10.1016/j.buildenv.2021.107656

Luna-Navarro, A., Loonen, R., Juaristi, M., Monge-Barrio, A., Attia, S., & Overend, M. (2020). Occupant-Facade interaction: A review and classification scheme. *Building and Environment, 177*, 106880. https://doi.org/10.1016/j.buildenv.2020.106880

Schaumann, D. (2024). Human Behavior Adaptability in Responsive Buildings: An Exploratory Study in Workplace Settings. *Buildings, 14*(6), 1830. https://doi.org/10.3390/buildings14061830

Shavitt, C., Kuzminykh, A., Ridel, I., & Cauchard, J. R. (2021). Naturally Together: A Systematic Approach for Multi-User Interaction With Natural Interfaces. *Proceedings of the ACM on Human-Computer Int., 5*(CSCW2), 1–31. https://doi.org/10.1145/3476090

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics, 53*, 190–202. https://doi.org/10.1016/j.apergo.2015.09.013

Van De Werff, T., Niemantsverdriet, K., Van Essen, H., & Eggen, B. (2017). Evaluating Interface Characteristics for Shared Lighting Systems in the Office Environment. *Proceedings of the 2017 Conference on Designing Interactive Systems*, 209–220. https://doi.org/10.1145/3064663.3064749

Winkler, D. A., Beltran, A., Esfahani, N. P., Maglio, P. P., & Cerpa, A. E. (2016). FORCES: Feedback and control for occupants to refine comfort and energy savings. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1188–1199. https://doi.org/10.1145/2971648.2971700

Computing on the Physical Model

Interactive Visual Feedback of Quantitative Criteria in Design through Physical Models

Artemis Pantopoulou[1] and Dimitris Papanikolaou[2]
[1,2]*National Technical University of Athens*
artemispantopoulou@gmail.com, 0009-0000-4937-8053
dimitri.c.pap@gmail.com, 0000-0002-6832-463X

Abstract. This paper explores how hybrid physical–digital interfaces embed performance feedback into early architectural design. Bridging hands-on form-finding with human–computer interaction, it examines how designers shift between intuition and analysis. It introduces a hybrid workflow that connects a physical model to structural and environmental simulations, enabling spatialized feedback during direct manipulation in the conceptual phase. Through a case study with architecture students, the research investigates how this system shapes design reasoning by integrating live data into a tangible environment. Results indicate that spatializing performance criteria fosters a fluid negotiation between qualitative intent and quantitative assessment. The work demonstrates how hybrid interfaces can support situated, performance-aware decision-making without disrupting intuitive workflows.

Keywords. tangible interface, hybrid, computational, form finding, simulation

1. Introduction

In early architectural design, intuition and experimentation often outweigh precision. Designers sketch and model to explore form and experience, yet as performance-based criteria, structural, environmental, and material, gain importance, the challenge arises to integrate analysis without constraining creativity. Tools enabling quantitative evaluation are rarely intuitive, while those that encourage exploration, such as physical models, often lack measurable feedback (Figure 1). This divide separates qualitative freedom from quantitative rigor.

For example, in a competition involving the second author, the team explored how varying roof slopes affected shading and structure, but the lack of immediate feedback fragmented the process. Such cases highlight the need for hybrid interfaces that bridge physical intuition with analytical insight.

This paper investigates how interactive visual feedback of quantitative criteria can inform conceptual design through physical models. A hybrid workflow was developed,

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 123-132.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

using computer vision to extract geometry from a physical model, feed it into a parametric simulation, and project the results back as real-time visualization, connecting tactile manipulation with digital analysis.

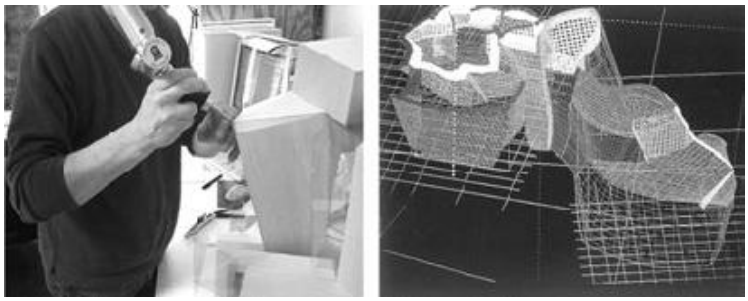


Figure 1. Models of the Guggenheim Museum in Bilbao. Physical model (left) and instant digital representation using CATIA software (right).

Focusing on positional data linking physical and digital components, the system was evaluated through self-experimentation, informal discussion, and a user study. It examines how designers interpret and act upon quantitative feedback while maintaining creative flow. The research addresses a key challenge in computational design: enabling engagement with complex data without interrupting the material and spatial thinking central to architecture. When data becomes tangible, visible and manipulable it can drive design reasoning.

The case study involved students using a vision-based system that connected a physical canopy model to structural simulation in Rhino. As users adjusted elements, real-time feedback informed their decisions, revealing how hybrid models merge intuition with performance awareness. The paper situates this within broader debates on computation in design, demonstrating how hybrid interfaces shift performance feedback from post-analysis to an active part of form-making

2. Background

2.1. FORM FINDING THROUGH PHYSICAL MODELS

Architectural design has always balanced the abstract and the material. Physical models historically served as instruments of experimentation, enabling architects to reason through geometry, gravity, and force. Long before digital computation, such analog simulations offered embodied understanding through direct manipulation.

Antoni Gaudí's hanging chain models for the Sagrada Família exemplify this approach: by inverting catenary forms, he derived stable compression geometries that merged form and structure("Front Matter" 2020; Boller and D'Acunto 2021). Frei Otto continued this lineage at the Institute for Lightweight Structures, using soap films and tensile membranes to study minimal surfaces, letting matter "find" its own equilibrium(Graefe 2018) (Hooke 1676; Burkhardt 2020) . Similarly, Heinz Isler's inverted fabric models and Sergio Musmeci's sculptural bridge geometries demonstrated that material behavior could guide design logic (Boller and D'Acunto

2021; Chilton 2020; Abel and John C. Chilton 2011,"3.Kilian, A., & Ochsendorf, J. (2005)., Negroponte 1975).

While these analog methods offered intuitive insights, they were limited in precision, repeatability, and quantitative evaluation. Designers could observe deformations but not measure performance numerically. Yet, they established a mindset in which design was simultaneously spatial, structural, and performative.

With the advent of digital tools, new possibilities for variation and simulation emerged, but at the cost of material immediacy. Parametric modeling and algorithmic workflows often abstracted design away from embodied feedback, relegating physical models to representational rather than generative roles. This historical shift frames the current need to reconcile digital precision with tactile intuition.

2.2. FORM FINDING THROUGH HYBRID DIGITAL-PHYSICAL MODELS

In recent decades, architecture has intersected with human–computer interaction (HCI) and tangible computing, fields that study how physical engagement shapes understanding. Tangible User Interfaces (TUIs) make digital information physically manipulable, reuniting "atoms and bits" as proposed by Ishii and Ullmer's Tangible Bits framework(H Ishii et al. 2004, Hiroshi Ishii et al. 2012).

Projects such as MIT Media Lab's inFORM introduced shape-changing surfaces capable of representing data through physical movement, while SandScape allowed users to sculpt terrain and view environmental simulations in real time. Similarly, research at ETH Zurich has explored robotic and augmented interfaces that respond to material or spatial input, merging physical modeling with live computation (Follmer et al. 2013, Leithinger et al., n.d., Alexander et al. 2018, Coelho and Zigelbaum 2011).

Despite advances in software such as Karamba3D, Kangaroo Physics, and Ladybug Tools, performance-based workflows remain primarily screen-bound and technically demanding. Hybrid interfaces offer an alternative: they turn the physical model into a feedback surface where force, form, and data coexist, making performance directly perceptible and actionable. (Apolinarska et al. 2024).Recent systems using fiducial tracking, projection mapping, and actuation show that designers can grasp and manipulate performance data as tangible phenomena rather than abstract parameters. Such systems serve as cognitive scaffolds, helping uusers reason spatially through perception–action loops (Nejur and Akbarzadeh 2021, Efficient Computation for 3D Graphic Statics" 2021; Lee, Post, and Ishii 2011, Mueller, Fivet, and Ochsendorf 2015).

Building on this discourse, our research introduces a hybrid workflow connecting physical manipulation with real-time simulation. The contribution lies in:

- A hybrid workflow that connects physical modeling with real-time simulation, enabling designers to interactively visualize performance metrics related to the geometric positioning of a physical model's components, and thus support design exploration during the conceptual phase.
- A documented design experiment that tests this workflow in an educational context, combining self-use, peer discussion, and user feedback
- Observations on how spatially embedded feedback influences the way designers

interpret and respond to structural and environmental criteria

3. Methods

This study is based on a fictional yet plausible design scenario, used to frame the development of a hybrid interface connecting a physical model to real-time computational feedback. The methodology combines physical prototyping, parametric simulation, and observational testing. The system was iteratively tested by the first author and later evaluated through a short user study with architecture students, focusing on early-stage design tasks involving structural and environmental performance.

3.1. INITIAL FRAMING: THE FICTIONAL DESIGN PROBLEM

Two students, Designer A and Designer B, are tasked with developing a canopy structure on sloped terrain. Their goal is to understand how variations in support placement, canopy inclination, and span affect deformation and solar exposure. Using a small-scale model, they quickly discover that minor changes cascade through the system, complicating their ability to evaluate structural and environmental interdependencies. To address this, a hybrid interface was created to help them visualize and manage these relationships.

3.2. SYSTEM DESIGN

The system (Figure 2, left) is a hybrid interface that connects physical modeling with real-time digital simulation, allowing designers to receive immediate performance feedback while manipulating a tangible model. It is composed of several physical components that represent parts of a shell structure. The central element of the model is tagged with a fiducial marker, a high-contrast printed symbol used for computer vision tracking (Figure 2, right). These markers, following the reacTIVision protocol (Lastovetska, 2025), enable the detection of position and orientation in physical space.

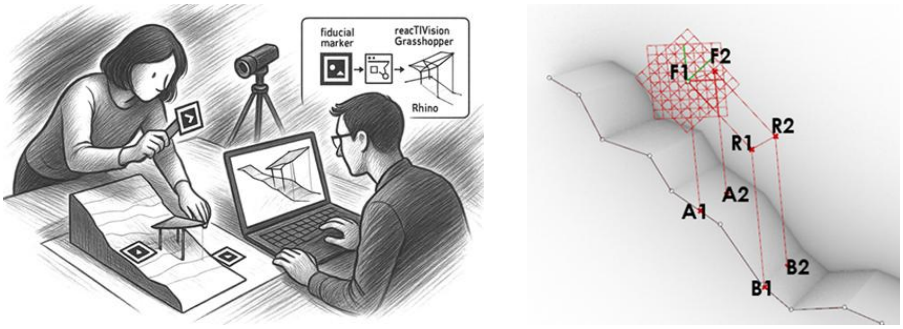


Figure 2. Left: Case Study sketch representation (AI-generated image using ChatGPT (OpenAI, 2025). Right: Naming points , F1 is the anchor point of the canopy which is tagged with the fiducial marker

A standard HD webcam is positioned above the model to capture the physical

layout. The video feed is processed in real time by the reacTIVision software (Bencina, Kaltenbrunner, and Fabra, n.d.), which identifies fiducial IDs and extracts their 2D coordinates and rotational angle. This spatial data is transmitted via the TUIO protocol, an open framework for transmitting touch and object data from physical interfaces to digital applications.

The tracking is deliberately minimal: only the position (X, Y) and orientation (angle) of each physical component is recorded. This simplicity serves two purposes: to keep the interface lightweight and fast, and to preserve the intuitive nature of physical manipulation without requiring complex sensor integration.

The tracked data is received in Grasshopper, (Figure 3) a parametric design environment within Rhino3D, where each tagged element is represented as a corresponding digital geometry in 3D space. Through this live linkage, any movement or rotation of a physical element results in an immediate update of the digital model.

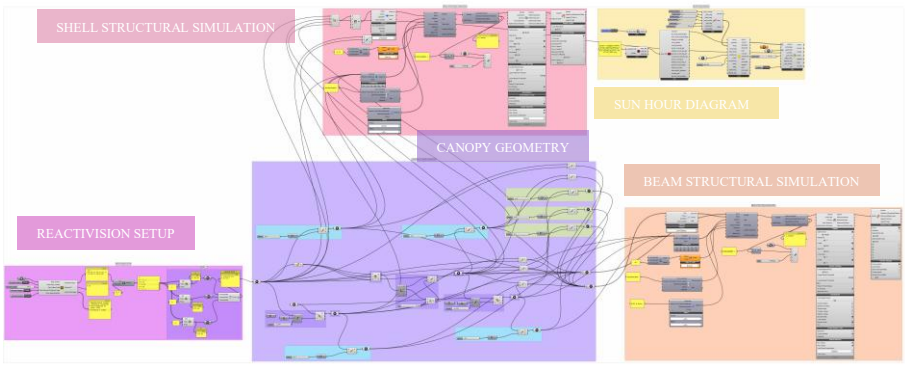


Figure 3. Grasshopper Fiducial Marker Setup

This updated geometry forms the input for structural and environmental simulations. Structural performance is evaluated using Karamba3D, which calculates principal stresses and displacement vectors based on simplified support conditions and load assumptions. Environmental feedback is generated through Ladybug Tools, specifically using solar radiation analysis to visualize solar exposure across the shell surface. Both toolsets output color-mapped visualizations, scalar annotations, and vector fields that are displayed on a connected screen adjacent to the physical model.

The entire workflow is designed to support iterative, hands-on experimentation. Designers can physically reposition components to explore spatial configurations and immediately observe how these changes impact performance metrics. In this way, the system maintains the intuitive and spatial logic of physical modeling while embedding simulation as an active, real-time feedback layer. The goal is the integration of structural and environmental awareness into the early, conceptual stages of design.

3.3. EXPERIMENT

Before engaging participants, the first author conducted self-guided sessions to observe how real-time feedback influenced design thinking (Figure 4). Decisions once driven

by form or composition began responding to displacement and solar patterns.

The bending-moment diagram (My) illustrates zones of positive and negative

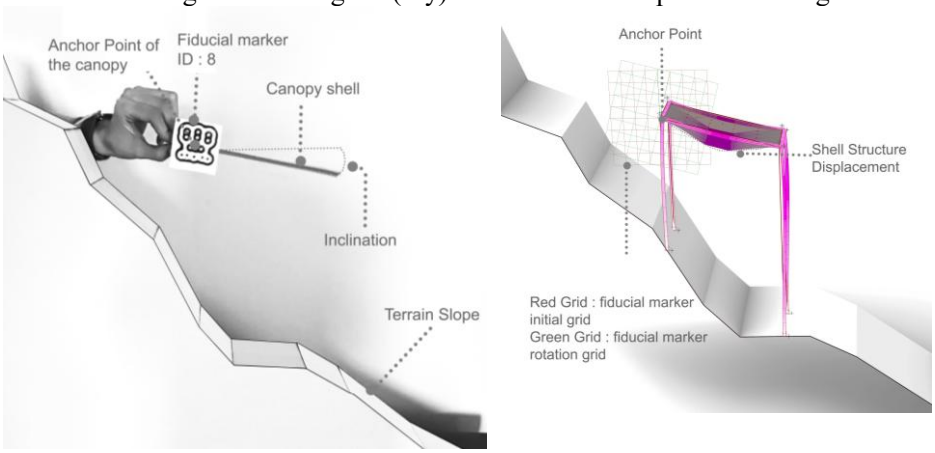


Figure 4. Left: physical model, right: digital twin, displacement

curvature along the canopy members, indicating where bending stresses peak under loading. The shear-force diagram (Vz) highlights the distribution of vertical shear along the structure, revealing critical sections where load transfer changes abruptly. The sun-hours diagram maps annual solar exposure across the canopy surface, identifying areas of high and low insolation that influence shading performance and thermal behaviour (Figure 5). Adjustments to one support prompted compensations elsewhere, insufficient shading led to asymmetrical experiments. The process became a rhythmic feedback loop of gesture, response, reflection, promoting predictive intuition as the designer began to anticipate outcomes before acting. Over time, the model evolved from representational object to active reasoning medium, merging spatial and performative awareness.

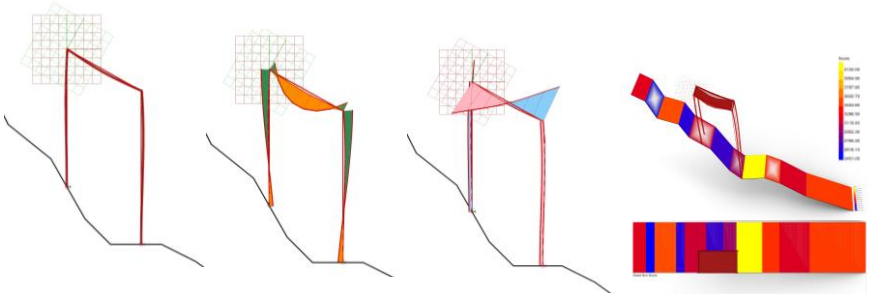


Figure 5. Structural Simulation using Karamba: a: Canopy form tests, b: Bending moment diagram for y axis (My), c: Shear force diagrams (shear force Vz) d: Direct Sun Hours simulation (Ladybug)

3.4. USER STUDY

A user study with ten architecture students further explored how the system mediated design reasoning. Each participant was asked to design a small canopy minimizing

deformation and solar exposure within 30 minutes. The physical setup consisted of a rectangular cardboard piece representing the canopy surface, with a fiducial marker attached to one edge. This marker, corresponding to point F1 in the digital model, (Figure 3) was recognized by the computer-vision system and had to remain oriented directly toward the camera to be detected. For this reason, participants could only move the cardboard horizontally and vertically, up, down, left, and right, altering the canopy's position and height. Each movement of the F1 point triggered an automatic update in the Grasshopper model, which recalculated the geometry of the entire canopy in real time. Participants could also rotate the cardboard between −90° and +90° around the horizontal (0°) axis, allowing them to observe how changes in inclination affected structural deformation and solar exposure. Rotations along other axes were restricted to ensure the marker remained parallel to the camera, meaning the canopy could only pivot within the XY plane. All manipulations took place directly above the physical topographic model, ensuring precise spatial correspondence with the digital twin and accurate simulation of structural and environmental conditions.

The terrain used in the experiment was intentionally varied in slope, enabling a richer understanding of how topographic inclination influenced both the sun-hour direction and the displacement behaviour of the canopy. Because the canopy geometry adapted dynamically to the marker's position, different slopes produced noticeably distinct simulation outcomes.

Some participants further experimented by modifying the canopy's span. In such cases, they fabricated new cardboard pieces at different lengths, reattached the fiducial marker at the same reference point, and made a minor adjustment to the Grasshopper setup to match the new shell dimensions. This allowed them to test alternative configurations both structurally and in terms of environmental comfort.

After a brief tutorial, participants interacted with the system while screen activity and verbal feedback were recorded. Observations focused on how visualized performance guided decisions and how users balanced intuition with analysis. Post-session interviews assessed perceived benefits and challenges (Figure 6). The data revealed patterns of embodied reasoning and how immediate feedback shaped iterative exploration.

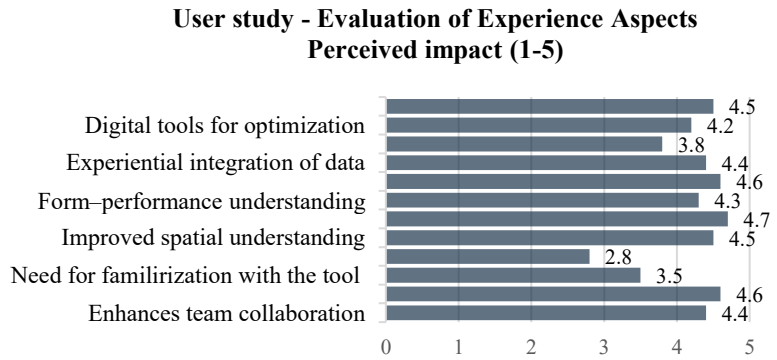


Figure 6. Results of the user study evaluating key aspects of the hybrid interface experience.

4. Results

The hybrid interface and its use by architecture students revealed clear patterns in how designers interpreted and acted upon performance feedback. Although modest in scope, the system exposed how physical manipulation and computational analysis can coexist in conceptual design.

All participants actively explored multiple iterations, responding directly to visualized structural data. Initial gestures based on composition gradually shifted toward performance-driven reasoning. One student noted they "had not realized how much a small shift in support placement would change deformation," adjusting the F1 point accordingly. Another sought to "balance the canopy," distributing stress more evenly by observing color gradients. Participants described the experience as "thinking through the model" or even "a conversation with the model," suggesting a sense of reciprocal interaction.

None relied on numerical data,instead, they reasoned visually and spatially, reading patterns and testing hypotheses through direct manipulation. This embodied computation aligns with theories of material reasoning, where design emerges from perception–action cycles. The interface appeared to reduce cognitive load by externalizing complexity: performance consequences became visible, freeing users from mental prediction and allowing them to iterate intuitively.

These observations support the hypothesis that a hybrid interface, even a simple one, can mediate the interplay between qualitative and quantitative reasoning, enabling designers to engage with performance criteria without sacrificing creative fluency.

4.1. LIMITATIONS

The study also revealed several limitations. While feedback was clear in terms of direction (e.g., "more stress here"), it lacked granularity, and the deliberately minimal set of tracked parameters constrained the range of observable differences. Participants at times struggled to distinguish small variations between iterations, and some expressed a desire to "zoom in" or "compare versions." Additionally, physical manipulation was intentionally restricted to planar movement and rotation to ensure robust computer-vision tracking, and the need to view feedback on a separate screen rather than directly on the physical model introduced a small but perceptible disconnect.

This study suggests that spatialized, embodied feedback can support performance awareness even in novice users and during early conceptual phases. The use of tools like Karamba3D or Ladybug is not new, what is novel here is how their output is delivered, interpreted, and acted upon in the context of physical manipulation.

5. Discussion

Our results suggest that even a minimal system can shift designers' attention from purely visual and compositional concerns toward performance-aware thinking. Beyond its specific implementation, this paper uses the proposed workflow to reflect on the role of hybrid interfaces in early-stage architectural design. By linking tangible interaction with real-time analysis, the system allows designers to respond to

quantitative criteria such as structural behavior or environmental exposure without leaving an intuitive design mode. Importantly, the simplicity of the physical model is a deliberate methodological choice: the system functions as a research probe to examine how designers reason when performance feedback becomes spatially embedded. What emerges is a hybrid cognitive environment in which analysis is not a separate evaluative step but part of the evolving design space. This perspective reframes hybrid physical–digital systems as supports for situated, informed, and iterative decision-making.

5.1. NEXT STEPS

Several directions could extend this research. A clear next step is the integration of a computer vision system capable of recognizing a wider range of physical components, such as arbitrary cardboard shapes and automatically generating their digital counterparts. Building on this, future development could include an algorithm for detecting topological relationships, allowing the system to understand how components are physically connected, not just positioned, and reflect those connections in the digital model.

Beyond geometry and topology, future work could explore the layering of additional performance criteria, acoustic behavior, lighting quality, or airflow, and how designers negotiate trade-offs between them in real time.

The system also lends itself to collaborative contexts such as studios, juries, or co-design workshops. Its tangible nature supports shared attention, making it suitable for multi-user interaction. Future studies could examine how collective decision-making unfolds when performance data is embedded in the physical model.

Lastly, in terms of pedagogy, this research reinforces that computational design does not need to be code-heavy or abstract. Hybrid tools that link physical making with live analysis can lower barriers to entry and foster intuitive understanding of performance, especially for students with limited scripting experience. Such systems may open up new ways of thinking, where design exploration becomes an active dialogue between materials, space, and data.

In summary, the study presented here offers a set of prompts for further inquiry into how hybrid interfaces might reshape our relationship with data in architecture. By reconfiguring not what we know, but how we come to know it, such systems invite a reconsideration of the design act itself: as a site where both intuition and analysis can converge in creative and embodied dialogue.

References

Abel, J. F., and John C. Chilton. 2011. *"HEINZ ISLER – 50 YEARS OF 'NEW SHAPES FOR SHELLS': PREFACE."* In . 131-134.

Alexander, Jason, Anne Roudaut, Jürgen Steimle, Kasper Hornbæk, Miguel Bruns Alonso, Sean Follmer, and Timothy Merritt. 2018. "*Grand Challenges in Shape-Changing Interface Research." In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems,* 1–14. Montreal QC Canada: ACM. https://doi.org/10.1145/3173574.3173873.

Apolinarska, Aleksandra Anna, Gonzalo Casas, Luis Salamanca, and Matthias Kohler. 2024. *"ARA - Grasshopper Plugin for AI-Augmented Inverse Design." In Scalable Disruptors,*

edited by Philipp Eversmann, Christoph Gengnagel, Julian Lienhard, Mette Ramsgaard Thomsen, and Jan Wurm , 231–40. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-68275-9_19.

Boller, G., and P. D'Acunto. 2021. *"Structural Design via Form Finding: Comparing Frei Otto, Heinz Isler and Sergio Musmeci." In History of Construction Cultures, by João Mascarenhas-Mateus, Ana Paula Pires, Manuel Marques Caiado, and Ivo Veiga, 1st ed* ., 431–38. London: CRC Press. https://doi.org/10.1201/9781003173434-168.

Burkhardt, Berthold. 2020. *"Soap-Film and Soap-Bubble Models." In PHYSICAL MODELS*, 569–85. John Wiley & Sons, Ltd. https://doi.org/10.1002/9783433609613.ch19.

Chilton, John. 2020. *"Heinz Isler and His Use of Physical Models." In PHYSICAL MODELS,* 613–37. John Wiley & Sons, Ltd. https://doi.org/10.1002/9783433609613.ch21.

Coelho, Marcelo, and Jamie Zigelbaum. 2011. *"Shape-Changing Interfaces." Personal and Ubiquitous Computing* 15 (2): 161–73. https://doi.org/10.1007/s00779-010-0311-y.

Follmer, Sean, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. 2013. *"inFORM: Dynamic Physical Affordances and Constraints through Shape and Object Actuation." In Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, 417–26. St. Andrews Scotland, United Kingdom: ACM. https://doi.org/10.1145/2501988.2502032.

Front Matter. 2020. *In PHYSICAL MODELS, i–xxxii. John Wiley & Sons, Ltd*. https://doi.org/10.1002/9783433609613.fmatter.

Graefe, Rainer /. 2018. *"Frei Otto – Architecture and Civil Engineering." Journal-International Association for Shell and Spatial Structures, March.* https://doi.org/10.20898/J.IASS.2018.195.906.

Hooke, Robert. 1676. *A Description of Helioscopes and Some Other Instruments Made by Robert Hooke, Fellow of the Royal Society.* https://name.umdl.umich.edu/A44317.0001.001.

Ishii, H, C Ratti, B Piper, Y Wang, A Biderman, and E Ben-Joseph. 2004. *"Bringing Clay and Sand into Digital Design — Continuous Tangible User Interfaces." BT Technology Journal* 22 (4): 287–99. https://doi.org/10.1023/B:BTTJ.0000047607.16164.16.

Ishii, Hiroshi, Dávid Lakatos, Leonardo Bonanni, and Jean-Baptiste Labrune. 2012. *"Radical Atoms: Beyond Tangible Bits, toward Transformable Materials."* Interactions 19 (1): 38–51. https://doi.org/10.1145/2065327.2065337.

Kilian, A., & Ochsendorf, J. (2005). *Particle-Spring Systems for Structural Form Finding. Journal of the International Association for Shell and Spatial Structures* , 46(2), 77-84. | PDF | Force | Damping." n.d. Scribd. Accessed May 30, 2025.

Lee, Jinha, Rehmi Post, and Hiroshi Ishii. 2011. *"ZeroN: Mid-Air Tangible Interaction Enabled by Computer Controlled Magnetic Levitation." In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, 327–36. Santa Barbara California USA: ACM. https://doi.org/10.1145/2047196.2047239.

Leithinger, Daniel, David Lakatos, Anthony DeVincenzi, Matthew Blackshaw, and Hiroshi Ishii. n.d. *"Direct and Gestural Interaction with Relief: A 2.5D Shape Display."*

Mueller, Caitlin, Corentin Fivet, and John Ochsendorf. 2015. *"Graphic Statics and Interactive Optimization for Engineering Education."* In , 1–14. https://doi.org/10.1061/9780784479117.223.

Negroponte, Nicholas. 1975. *Soft Architecture Machines. Cambridge, Mass: The MIT Press.*

Nejur, Andrei, and Masoud Akbarzadeh. 2021. *"PolyFrame, Efficient Computation for 3D Graphic Statics." Computer-Aided Design* 134 (May):103003. https://doi.org/10.1016/j.cad.2021.103003.

OpenAI. 2025. ChatGPT (Version 4.5). Accessed June 10, 2025.*"I acknowledge using ChatGPT (OpenAI, version 4.5, 2025) to generate image and to improve clarity and grammatical accuracy."*

Reimagining Injustice Heritage through Virtual Reality

Designing Immersive Experiences of Memory and Space

Zi-Ru Chen[1], Kai-Hsiang Liang[2], Yu-Jie Liu[3], Wen-Xin Chen[4] and Yu-Tzu Tseng[5]
[1,3,4] *Southern Taiwan University of Science and Technology*
[2,5] *National University of Kaohsiung*
[1] *zrchen@stust.edu.tw, 0000-0001-8086-8368*
[2] *KLiang@nuk.edu.tw, 0000-0002-2085-4279*
[3] *luj619957@stust.edu.tw, 0009-0007-9120-1674*
[4] *mb3j0107@stust.edu.tw, 0009-0007-3933-2917*
[5] *m1141416@mail.nuk.edu.tw, 0009-0002-6323-2477*

Abstract. Injustice sites challenge conventional interpretation by demanding embodied, affective engagement. This study proposed a VR-based framework that translates spatial observation into immersive first-person narratives to re-present historical memory. Using the Jing-Mei White Terror Memorial Park as the field site, we integrated LiDAR/photogrammetry, narrative analysis, and system design to build a 360° experience structured by three modules- spatial guidance, situational immersion, and narrative linking. The model and prototype demonstrated how architectural evidence, temporal pacing, and interactive cues converge into a coherent spatial narrative that elicits moral emotions and deepens understanding of oppression and justice. The approach operationalizes a Space—Narrative—Emotion translation model linking spatial anchors to target affective states. Our contribution lies in connecting architectural analysis with historiographic interpretation and offering an extensible pathway for human-rights education and digital-heritage practice.

Keywords. Injustice sites, virtual reality (VR), immersive narration, spatial narrative, human-rights education

1. Introduction

The concept of "injustice heritage" gained global attention in the early 21st century, particularly after former Auschwitz-Birkenau concentration camps in Germany were inscribed as UNESCO World Heritage sites (Macdonald, 2009; Young, 2009). Especially in regions undergoing transitional justice, these sites not only preserve painful memories of state violence and political suppression but also provide important opportunities for civic education and public reflection (Bell & de-Shalit, 2011; Mitchell, 2018). In Taiwan, the preservation and reinterpretation of injustice heritage

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 133-142.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

have become central to promoting social dialogue and advancing human rights awareness (Wang et al., 2022). While traditional guided tours and educational activities remain essential, they are often constrained in their ability to deliver immersive and participatory experiences. Emerging immersive technologies, particularly virtual reality (VR), present new possibilities by allowing visitors to engage with reconstructed spaces, narratives, and historical atmospheres in a more embodied and experiential manner.

1.1 SPATIAL EXPERIENCE AND PROMOTION OF INJUSTICE HERITAGE

The concept of injustice heritage gained global attention in the early twenty-first century when Konzentrationslager Auschwitz-Birkenau were inscribed as UNESCO World Heritage sites (Macdonald, 2009; Young, 2009). Such sites, also referred to as "dark," "difficult," or "negative" heritage, represent places marked by state violence and collective trauma (Dai et al., 2021). In Taiwan, injustice heritage specifically denotes locations associated with human rights violations during the authoritarian period (Wang et al., 2022). Preserving these sites not only safeguards historical memory but also fosters democratic values and human rights awareness. However, preservation efforts face multiple challenges, including the collection of oral histories and archival data, property disputes, and physical deterioration (Lennon & Foley, 2010; Wang et al., 2022). Beyond conservation, political, social, and psychological factors further complicate the process (Young, 1993; Lin, 2010; Tsao, 2017).

Public engagement often relies on guided tours and exhibitions, which have led to growing interest in "dark tourism" (Tsao, 2017; Lennon & Foley, 2010; Huang, 2023). Human rights museums also play an important educational role by transforming sites of trauma into spaces for civic learning and social justice (Carter & Orange, 2012; Mitchell, 2018). Recent creative works, such as Detention and The Man Who Couldn't Leave Tsao, demonstrate how immersive media can enhance public reflection on freedom and human rights. These examples suggest that design-based and experiential approaches can expand the interpretive and educational potential of injustice heritage.

1.2 THE APPLICATION AND DEVELOPMENT OF IMMERSIVE TECHNOLOGIES

The integration of digital media has transformed museum and heritage interpretation, enabling visitors to access contextual information beyond physical environments and fostering interactive, multisensory learning experiences. As exhibition tools evolve from traditional guided tours to mobile and multimedia systems, interaction design becomes essential in ensuring intuitive navigation and user engagement (Kristof & Satran, 1995; Zhao, 2001).

"Immersion" refers to a user's psychological engagement within a digital environment, achieved through technologies such as VR, AR, and MR. Rooted in Csikszentmihalyi's theory of "flow" (1975, 1990), immersive experiences involve spatial presence, narrative engagement, and emotional resonance (Calleja, 2011). Beyond psychological flow, Slater (2009) further defines the VR experience through the concept of "presence," which is constituted by Place Illusion (PI)—the sensation of

"being there"—and Plausibility Illusion (Psi)—the belief that the virtual events are actually occurring. In heritage contexts, these illusions are critical for transforming historical data into an embodied, lived experience. Such technologies are increasingly applied in education, health, and cultural sectors, offering multidimensional and participatory learning opportunities (Redação, 2019).

In heritage preservation, the Québec Declaration (2008) highlights the integration of tangible and intangible heritage through digital methods (Brouwer, 2021). Immersive technologies enhance interactivity and accessibility while reconstructing historical narratives beyond physical limits. Global examples include the Metropolitan Museum of Art's 3D collection scans (Erlick, 2017), the digital restoration of Notre-Dame Cathedral (Amaclio Productions, 2022), and the Dachau Concentration Camp AR project The Liberation (KZ-Gedenkstätte Dachau, 2020).

In summary, immersive technologies have transformed interpretive and educational practices, serving as vital media for re-presenting cultural memory and social history. Applying such technologies to the education and exhibition of injustice heritage can further human rights understanding and foster emotional resonance among diverse publics.

2. Problem and Objective

While injustice heritage sites carry profound historical and emotional meanings, conventional interpretation methods—such as guided tours and exhibitions—often struggle to engage visitors in deeper, participatory learning experiences. The challenge lies in how to transform these historically charged spaces into environments that encourage embodied reflection and emotional connection through digital media.

Therefore, this study aimed to explore how virtual reality (VR)–based immersive design can propose the interpretation of injustice heritage sites. The research specifically sought to: (1) conduct spatial observation and analysis of a selected injustice heritage site to identify its architectural and historical characteristics; and (2) design a preliminary VR experience flow that integrates these spatial features into a coherent framework for human rights education. User evaluation will be conducted in future work to examine experiential effectiveness and perceived presence.

3. Methodology and Steps

This study adopted a qualitative and design-oriented research approach, with an emphasis on the translational process that moved from spatial observation to the design of narrative structures. Virtual reality (VR) was employed as the medium for developing immersive experiences at sites of injustice. The research centered on the conceptual development of the VR experience flow and system design. The overall process comprised three principal phases:

1. Spatial observation and documentation
2. Narrative analysis and design translation
3. VR experience flow and system design

Ultimately, the study proposed a preliminary model of a "VR Experience Design

Process for Sites of Injustice," demonstrating how spatial observation and narrative analysis were integrated within immersive design, and providing a foundational framework for future human rights education and digital heritage interpretation.

3.1 SPATIAL OBSERVATION AND DOCUMENTATION

This study designated the Jing Mei White Terror Memorial Park as a representative site of injustice and the primary field site. Through on-site survey, photographic documentation, and spatial surveying/mapping, we analyzed the architectural fabric, spatial hierarchies, and symbolic elements, and examined their relationships to historical events, political power, and collective memory. The procedure comprised:

- **Step 1: field recording.** Site study was conducted at the site using an infrared rangefinder, 2D still and video capture, 360° panoramic imagery and video, and LiDAR/photogrammetry to record in detail the architectural components, material textures, and environmental features, thereby establishing a precise spatial data baseline.
- **Step 2: spatial surveying and mapping.** Based on the field data, we built planimetric and volumetric spatial models (e.g., cells, courtyards, interrogation rooms), analyzed spatial hierarchies, circulation patterns, and zoning structures, and synthesized diachronic transformations, morphological characteristics, and mnemonic imagery of the site. Particular attention was given to diachronic spatial change: what was observed and surveyed reflected alterations over time and shifts in use. The study sought to reconstruct how the spaces appeared at the time when human rights were violated, including their materials, colors, furnishings, and related attributes. In addition to removing contemporary exhibition elements for historical consistency, we also introduced narrative-driven virtual assets—such as character avatars and contextual objects (props)—to support key events and interactions in the experience flow. See Figure 1.

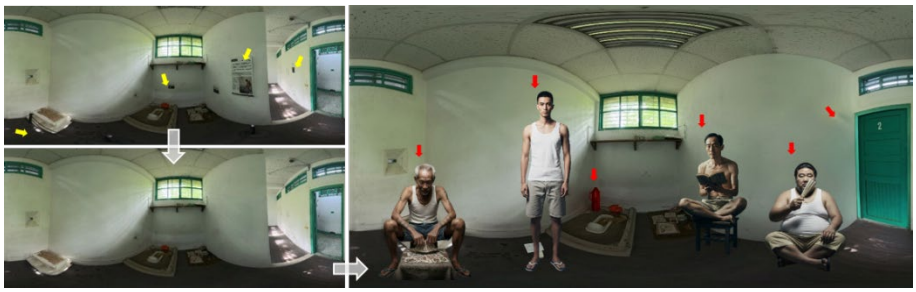


Figure 1. Cell interior (360°): on-site capture and reconstructed VR scene with exhibition elements removed and context props added to support narrative interaction

3.2 NARRATIVE ANALYSIS AND DESIGN TRANSLATION

This phase focused on translating findings from spatial observation into a narrative structure and examined how space functioned as a medium for historical narration and the re-articulation of memory. By aligning temporal nodes, spatial

paths, and event imagery, the study developed a translational model of "Space—Narrative—Emotion," which served as the conceptual foundation for the immersive design. The phase comprised:

- **Step 3: historical data collection.** The research compiled archival records, oral histories, images, and literature related to the site to reconstruct its political, social, and cultural background and to trace the processes through which victims' human rights were deprived. In addition, we participated in and recorded guided tours at the Jing-Mei White Terror Memorial Park to document the guides' narration comprehensively. See Figure 2. We then interviewed a survivor formerly detained at the site and drew on his oral testimony and publications to reconstruct conditions at the time. Finally, the study synthesized these historical sources with the on-site findings as the basis for spatial narrative translation.



Figure 2. Guided-tour documentation at the Jing-Mei White Terror Memorial Park

- **Step 4: narrative scriptwriting.** Historical materials and field-investigation data were translated into a spatial narrative script that defined the principal plotline, temporal pacing, and transition nodes, thereby establishing an initial framework for a guided narrative and ensuring logical coherence and emotional continuity. Figure 3 illustrates the Story Experience Flow: after wearing the VR headset, participants entered a 360° virtual environment that opened with a 1970s survivor's diary as the narrative cue and led to an entry-selection interface, where they chose one of six survivor roles as the first-person perspective.

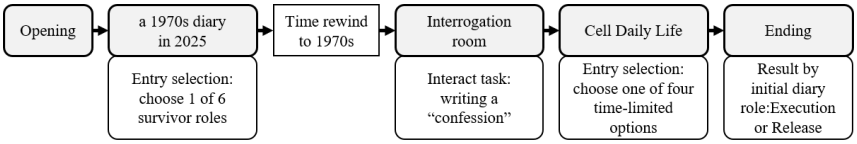


Figure 3. Story experience flow of VR interaction in injustice sites

The story then rewound to the interrogation room and introduced three survivor recollections to build context and affect. Participants subsequently encountered a verdict interaction; after writing a "confession," they proceeded to trial and punishment. The experience then shifted to its core—the "Cell Daily Life" sequence—implemented using authentic spaces of the Jing-Mei White Terror Memorial Park, where soundscapes, object interactions, and rhythmic control accumulated tension. The narrative culminated in two endings—Execution

(shooting) or Release (freedom)—predetermined by the initial role selection. Each role was derived from a real case; regardless of branch, both paths converged on the Character Archetype and a closing synthesis, completing narrative linking.

The script operationalized the Space-Narrative-Emotion model by using spatial anchors (e.g., interrogation room, cells, release courtyard) as carriers for narrative nodes and, through voice guidance, interactive events, and first-person viewpoint, mapping each node to target affective states (fear, anxiety, helplessness, pre-verdict ambivalence, post-release complexity). The divergence and reconvergence of the dual endings closed the story loop and demonstrated the results of the model in immersive human-rights education for eliciting moral emotions and deepening memory comprehension.

3.3 VR EXPERIENCE FLOW AND SYSTEM DESIGN

Building on the preceding analyses, this phase focused on the holistic design of the VR experience and articulated three core modules: (1) Spatial Guidance—virtually reconstructing the site and its historical context; (2) Situational Immersion—representing the historical atmosphere through visuals, sound, and narrative pacing; and (3) Narrative Linking—organizing interrelated narrative nodes into a continuous historical experience. The phase comprised:

- **Step 5: sensory and guidance design.** Based on narrative nodes and spatial features, the experience was configured as a 360-degree immersion. Visually, participants viewed the overall space as it appeared historically, with minimal textual and pictorial overlays. Aurally, background music was reduced to emphasize dialogue (e.g., judge and prison guard voices) and diegetic sound effects (e.g., footsteps with shackles), with directional audio aligned to spatial context to heighten immersion and emotional tension. Plot progression was grounded in period-appropriate encounters (court personnel, guards, cellmates) and supported by subtle navigation cues- lighting, signage, and audio prompts- to guide users through historical scenes in sequence.
- **Step 6: VR system construction.** Scenes were captured with 360-degree cameras and modeled in 3ds Max, then integrated in Unity Engine to build the 3D virtual environments. Narrative content—text, 2D images and video, 3D images and video, 3D models, sound effects, voice-over, and ambient tracks—was embedded at interaction nodes to ensure consistency between spatial experience and historical narration, while preserving extensibility for future updates.

3.4 SUMMARY

Through an integrated three-stage process of spatial observation, narrative analysis and system design, this study proposed a space-design-based mode of injustice sites that moved from the physical site to its virtual interpretation. The methodological contribution lay in: (1) linking, from a design-research perspective, architectural analysis of injustice sites with the interpretation of historical narratives; and (2) constructing a VR-mediated framework for immersive experiences at such sites, thereby providing a practical foundation for future human-rights education and the

transmission of cultural memory.

4. Results

The research presents design outcomes derived from the integrated three-stage process of spatial observation—narrative analysis—system design, including the prototype experience flow (Figure 3) and the experience framework (Figure 4). Focused on immersive narration, the model reconstructs pathways of historical memory through spatial guidance and interaction design, outlining an application framework for human-rights education and digital-heritage interpretation.

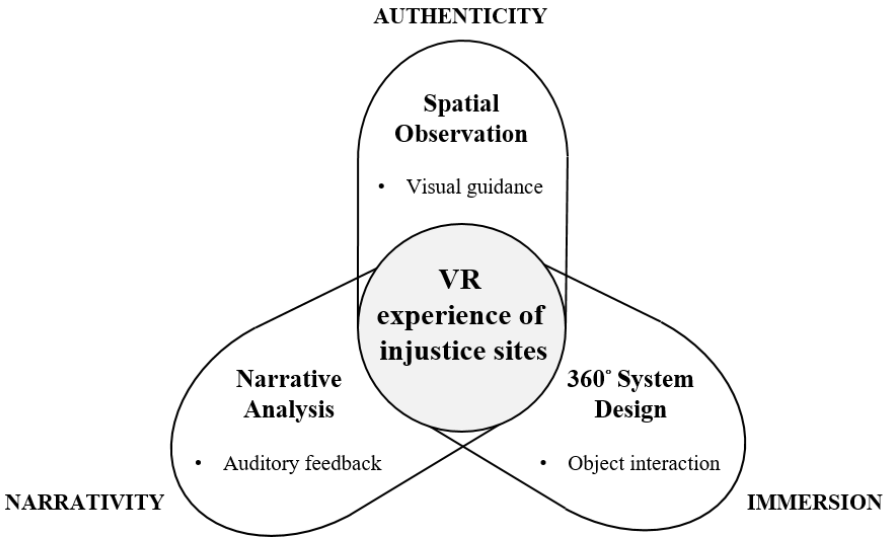


Figure 4. A model for the VR experience at injustice sites

4.1 DESIGN CONCEPT

The virtual interpretation of sites of injustice was framed by the dual aims of re-presentation and reflection. VR functioned not only as a tool to re-present historical scenes but also as a medium to guide users to perceive, understand, and reflect on the past. The design emphasized three principles:

- **Authenticity:** Reconstructing spatial structures and atmospheres on the basis of in-situ data.
- **Narrativity:** Using space as a narrative carrier; re-articulating events through node transitions and situational chaining.
- **Immersion:** Mobilizing multisensory cues (sound, light, materiality, pacing) to intensify affective engagement and sense of presence (Slater, 2009).

4.2 EXPERIENCE FRAMEWORK

The VR experience consolidated prior findings into three principal modules:

- **Module 1: Spatial Guidance.** This module reconstructed the site's spatial configuration and supported understanding of spatial hierarchy through guided prompts and lighting cues. Through concrete virtual reconstructions (e.g., cells, corridors, interrogation rooms), users formed an overall spatial sense of the site.
- **Module 2: Situational Immersion.** The narrative script of the VR story was translated into spatial situations. Historical events became interactive triggers, such as sonic reverberations, projected archival imagery, or shifts in ambient atmosphere—through which users experienced temporal progression, affective tension, and the presence of history.
- **Module 3: Narrative Linking.** Interactive relations wove nodes into a continuous storyline. Each node coupled a historical event—spatial image—interactive action, forming a pathway from understanding to reflection and concluding in a contemplative space that invited users to reconsider the meanings of memory and justice.

4.2 SYSTEM IMPLEMENTATION AND DEMONSTRATION

The system was implemented using the Unity engine as the primary development platform, with scenes reconstructed from LiDAR and photogrammetry-based 3D models. Core interaction modes included:

- **Visual guidance:** guiding movement via changes in lighting and focal cues;
- **Auditory feedback:** shaping atmospheres through environmental sound and narrative voice-over;
- **Object interaction:** triggering displays of historical materials (text, images, audio) via clicking and gaze.

For demonstration, the cells and interrogation room of the Jing-Mei Memorial Park were recreated to evoke detention and interrogation scenarios of political victims. The experience concluded with a "Wall of Memory" as a reflective node, encouraging users to pause and look back within the virtual environment.

5. Conclusion

The VR experience design presented in this study demonstrated that virtual reality can serve as a medium linking historical interpretation and affective experience. By integrating spatial observation, narrative translation, and interaction design, the research proposed a prototype for immersive interpretation at sites of injustice. The model's value lay in: (1) showcasing the integrative potential of historical space and narrative structure; (2) offering innovative directions for digital human-rights education; and (3) establishing an extensible framework for subsequent research on interactive heritage.

First, at the spatial level, the application of LiDAR and photogrammetry enabled high-precision reconstruction of architectural form and atmosphere, making authenticity the basis for historical perception. Second, at the narrative level, the study

integrated historical events, spatial circulation, and user experience to formulate a time-paced, affect-laden spatial narrative structure. This approach surpassed the constraints of text-based sources: history was no longer passively read but re-experienced through interaction.

Moreover, the research illuminated VR's potential in human-rights education. When an injustice site was transformed into an explorable, experiential, and immersive space, viewers entered history from a first-person perspective, sensing the conditions of oppression and silencing. Such engagement was not mere simulation but an ethical learning process that fostered self-understanding of justice and memory within the virtual domain.

In conclusion, by employing VR as a design medium for interpreting sites of injustice, this study constructed an immersive experience flow organized around spatial observation—narrative analysis—system design. The findings confirmed that, when history is translated into a perceptible and interactive spatial narrative, users achieve a deeper understanding of past traumas and memories and rearticulate contemporary accounts of transitional justice through virtual experience. The study thus offers a new trajectory for cultural-heritage digitization and opens a design-centered pathway for interpretive practice in human-rights education.

Acknowledgements

We thank the interviewees for sharing their time and testimonies. This work was supported by the National Science and Technology Council, Taiwan (Grant NSTC 113-2222-E-218-002-).

Attribution Statement

ChatGPT (OpenAI, 2025) was used to (1) suggest the initial structure of the manuscript and (2) translate Chinese-language drafts into academic English. All content, data, and references were reviewed and verified by the authors.

References

Amaclio Productions. (2022). *Trailer – Eternelle Notre-Dame.* YouTube. Retrieved February 20, 2024, from https://www.youtube.com/watch?v=vGC89oPLtao

Bell, D. A., & de-Shalit, A. (2011). *The spirit of cities: Why the identity of a city matters in a global age.* Princeton University Press.

Brouwer, I. (2021). *Simulations & virtual reality, digital twins and augmented reality supporting automation and production solutions.* Retrieved February 20, 2024, from https://www.linkedin.com/pulse/simulations-virtual-reality-digital-twins-augmented-2-ivo-brouwer

Calleja, G. (2011). *In-game: From immersion to incorporation.* MIT Press.

Carter, J., & Orange, J. (2012). Contentious terrain: Defining a human rights museology. *Museum Management and Curatorship,* 27(2), 111–127.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience.* Harper Perennial.

Csikszentmihalyi, M., & Csikszentmihalyi, I. (1975). *Beyond boredom and anxiety.* Jossey-Bass.

Dai, W.-F., Chuang, Y.-C., & Yeh, Q.-X. (2021). Memorial sites and human rights in Tainan II. Tainan, Taiwan: Tainan City Government, Bureau of Culture.

Erlick, N. (2017). *Museums are turning to virtual reality, apps, and interactive experiences to keep tech-savvy visitors engaged.* The Verge. Retrieved February 20, 2024, from https://www.theverge.com/2017/5/6/15563922/museums-vr-ar-apps-digital-technology

Huang, S.-M. (2023). Prisons, empire, and dark tourism: Shu-mei Huang on East Asian colonial carceral landscapes. Humanity·Island. Retrieved from https://humanityisland.nccu.edu.tw/shu-mei-huang/.

Kristof, R., & Satran, A. (1995). *Interactivity by design: Creating and communicating with new media.* Adobe Press.

KZ-Gedenkstätte Dachau. (2020). *Die Befreiung / The Liberation.* Retrieved February 20, 2024, from https://interaktiv.br.de/die-befreiung/en/

Lennon, J., & Foley, M. (2010). *Dark tourism.* Cengage Learning EMEA.

Lin, H.-W. (2010). 'Remembrance' and 'Forgetting' of Difficult Pasts - Holocaust: Reflections on the Issue of Tragic Sites as Traumatic Cultural Heritage. Journal of Cultural Heritage Preservation, 13, 5–18.

Macdonald, S. (2009). *Difficult heritage: Negotiating the Nazi past in Nuremberg and beyond.* Routledge.

Mitchell, P. (2018). Concentration camp buildings as artifacts: The archaeology of buildings in the Mauthausen–Gusen complex. *International Journal of Historical Archaeology,* 22(3), 553–573.

OpenAI. (2025). ChatGPT 5 Thinking, March 2025 version. OpenAI. https://chat.openai.com/.

Québec Declaration on the Preservation of the Spirit of Place: Adopted at Québec, Canada, October 4th 2008. (2008). *International Journal of Cultural Property,* 15(4), 393–396. https://doi.org/10.1017/S0940739108080430

Redação, P. (2019). *Projeto usa realidade virtual para te conectar com a Mona Lisa.* Folha de Alphaville. Retrieved February 20, 2024, from https://www.folhadealphaville.com.br/mundo/projeto-usa-realidade-virtual-para-te-conectar-com-a-mona-lisa.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), 3549-3557.

Tsao, C.-R. (2017). Heritage for freedom: Stories of Taiwan's 228 and White Terror memorial sites. Taipei, Taiwan: Taiwan Youyi.

Wang, S.-J., Li, Y.-X., Chang, Y.-N., Chang, W.-X., & Su, X.-W. (2022). Injustice sites: Spatial practices of transitional justice. Taipei, Taiwan: Transitional Justice Commission; Youguan Creative Strategy Co., Ltd.

Young, J. E. (1993). *The texture of memory: Holocaust memorials and meaning.* Yale University Press.

Young, K. (2009). *Auschwitz-Birkenau: The challenges of heritage management following the Cold War. In Places of pain and shame: Dealing with "difficult heritage"* (pp. 50–67). Routledge.

Zhao, Y.-L. (2001). A Survey on the Media Uses and Motivations of the Visually Impaired in Taiwan. Mass Communication Research, 66, 61–96.

Perceptual Environments for Regulating Interaction Balance in Cross-Disciplinary Teams

A spatial mediation approach for multi-user collaborative dynamics

Jia-Rong Li[1], Teng-Wen Chang[2]
[1,2]*National Yunlin University of Science and Technology*
[1] *lotus.softlab@gmail.com, 0009-0002-5446-3353*
[2]*tengwen@softlab.tw, 0000-0001-9503-2766*

Abstract. Smart and responsive environments have largely focused on single-user adaptation and physical comfort, offering limited support for regulating interaction dynamics in collaborative settings. In cross-disciplinary collaboration, interaction imbalance often emerges at the group level, manifesting as discourse dominance, participation withdrawal, and collective silence. This paper presents a spatially embedded perceptual environment that functions as a perceptual mediator for regulating group-level interaction imbalance in multi-user, cross-disciplinary collaboration. The system captures multimodal social signals, including speaking rhythm, turn-taking patterns, posture changes, and collective silence, and interprets them through a three-layer perceptual logic integrating sensing, interaction state inference, and non-intrusive spatial feedback. The system was deployed in interdisciplinary design sessions involving mixed design engineering teams. Results indicate that spatial feedback was associated with reduced participation dominance, increased turn-taking frequency, and shorter periods of collective silence, supporting smoother interaction rhythms without disrupting creative flow. These findings suggest the potential of perceptual environments to support real-time regulation of group interaction dynamics in cross-disciplinary collaboration.

Keywords. interactive environments, spatially embedded sensing, interaction imbalance, multimodal social signals, real-time spatial feedback, Research through Design

1. Introduction

Research on smart environments can be traced to Weiser (1991) vision of ubiquitous computing, which proposed that computation should recede into everyday environments, sensing context and responding seamlessly to users. Building on this idea, responsive architecture research has emphasized context awareness as a core capability of intelligent spaces, suggesting that built environments may adapt to occupants' evolving needs (Nguyen et al., 2022).

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 143-152.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Despite these developments, much smart environment research remains oriented toward control and comfort optimization. Existing systems typically automate environmental adjustments based on individual behavior patterns or optimize parameters such as lighting and temperature through IoT and BIM technologies. While such approaches demonstrate the feasibility of sensing individual behavior (Chang et al., 2023; Chang et al., 2020), they primarily address physical environment regulation, leaving collective social dynamics insufficiently perceived at the spatial level. Consequently, interaction imbalance—manifested as asymmetries in discourse power, cognition, and authority (Gray et al., 2022) continues to be addressed mainly through interpersonal or facilitative interventions.

This study explores how interactive environments might perceive and respond to multimodal social signals to support interaction fluency in cross-disciplinary collaboration. It aims to prototype a spatially embedded sensing framework in which physical space functions as a perceptual mediator, detecting and regulating interaction rhythms through quantifiable metrics and adaptive spatial feedback. Accordingly, the study addresses the following research questions:

RQ1: How can spatial agents perceive multimodal social signals indicative of interaction imbalance in cross-disciplinary collaboration, and which quantifiable metrics support reliable detection?

RQ2: How can physical space regulate interaction imbalance through spatial feedback without disrupting creative processes, and what measurable interaction changes are associated with such interventions?

2. Literature Review

2.1. CROSS-DISCIPLINARY COLLABORATION AND INTERACTION IMBALANCE

Cross-disciplinary collaboration frequently exhibits interaction imbalance, manifesting as persistent asymmetries in discourse power, participation, and authority. Existing interventions, including AI-supported collaboration tools that promote semantic alignment (Li et al., 2024), primarily operate at organizational or interpersonal levels and tend to emphasize coordination efficiency rather than the real-time regulation of interaction balance during collaborative work.

Although recent studies have explored multimodal sensing approaches to capturing behavioral cues such as location and posture (Chang et al., 2020), and have conceptualized space as a perceptual agent (Chang et al., 2023), these approaches have not been applied to detect and regulate interaction imbalance in real-time, group-based collaborative settings. imbalance in real-time collaborative settings.

Consequently, interaction imbalance in cross-disciplinary collaboration continues to be managed largely through facilitation or organizational mechanisms, leaving the role of physical space as a non-intrusive, in-situ mediator of interaction imbalance insufficiently explored.

2.2. INTELLIGENT AND RESPONSIVE ENVIRONMENTS: FROM CONTROL TO UNDERSTANDING

From the perspective of intelligent environments, interaction imbalance has rarely been addressed as a spatial perception problem. Cross-disciplinary collaboration has often been described as exhibiting interaction imbalance, manifested as asymmetries in discourse power and participation. Prior studies suggest that unequal voice distribution may influence whose expertise becomes visible and how collective decisions are shaped, potentially placing fewer dominant participants at a disadvantage (Gray et al., 2022). Recent AI-supported collaboration tools have sought to improve semantic alignment across disciplines (Li et al., 2024); however, such approaches tend to operate at organizational or interpersonal levels and appear less concerned with regulating interaction balance as collaboration unfolds in real time.

Design research indicates that multimodal sensing can be used to capture behavioral cues such as location and posture (Chang et al., 2020), and some studies have conceptualized space as a perceptual agent within fabrication and co-making contexts (Chang et al., 2023). Nevertheless, in group-based cross-disciplinary collaboration, interaction imbalance is still largely addressed through facilitation or organizational mechanisms, while the role of physical space as an in-situ mediator remains underexplored. Research on intelligent environments further suggests that many systems remain control-oriented, focusing on comfort or efficiency optimization rather than group-level interaction imbalance (Augusto, 2022; Lee et al., 2021).

2.3. PERCEPTUAL MEDIATION AND NON-INTRUSIVE SPATIAL INTERVENTION

Recent research has suggested reconsidering collaborative space not merely as a passive container, but as an active mediator that may sense, interpret, and respond to interaction imbalance. From a distributed cognition perspective, collaborative activity is shaped through the interplay of individuals, artifacts, and environments that mediate attention, coordination, and shared understanding (Sidra & Mason, 2025) . Within AI-augmented design contexts, spatial tools and ambient interfaces have been discussed as mechanisms that distribute cognitive effort across human–artifact systems, potentially supporting emergent collaboration during creative work.

Within this framing, spatial intervention is often understood as a form of cognitive mediation rather than direct behavioral control. Research on interactive and responsive environments emphasizes non-intrusive strategies, such as spatial feedback and soft fascination, which may guide attention while minimizing disruption to creative flow (McDonnell & Strayer, 2024). However, existing studies frequently address sensing, interpretation, and actuation as separate components, with limited examination of how these elements might be integrated to mediate group interaction imbalance in real time (Lee et al., 2021). This gap points to the relevance of exploring perceptual mediation as a design principle for collaborative environments.

3. Methodology

3.1. RESEARCH THROUGH DESIGN FRAMEWORK

This study adopts Research through Design (RtD) as its core methodology, positioning design practice as a mode of inquiry through which knowledge is generated by

constructing and reflecting on artifacts (Busciantella-Ricci & Scataglini, 2024). RtD is particularly suitable for this research because perceptual environments for multi-user collaboration remain an underexplored design space that requires iterative exploration rather than hypothesis-driven validation. As illustrated in Figure 1, the RtD process is structured into two interrelated stages: Formulation and Construction. The Formulation stage synthesizes insights from the literature to articulate the research questions, focusing on how interaction imbalance can be perceived at the spatial level (RQ1) and how such imbalance can be regulated through spatial feedback (RQ2). This stage establishes a three-layer perceiving–interpreting–acting framework that guides the subsequent design exploration.

The Construction stage operationalizes this framework through iterative prototyping in team-based collaborative scenarios. Each iteration follows a reflective loop comprising five steps: (1) Perceive, capturing multimodal interaction cues such as speech rhythm, gesture activity, and stillness; (2) Interpret, identifying recurring patterns of interaction imbalance; (3) Feedback, delivering non-intrusive spatial cues; (4) Measure, examining interaction changes through system logs and qualitative observation; and (5) Refine, adjusting sensing and interpretation rules based on observed dynamics. Through this iterative process, the RtD approach enables the articulation of transferable design knowledge concerning how physical space can mediate collaborative interaction.

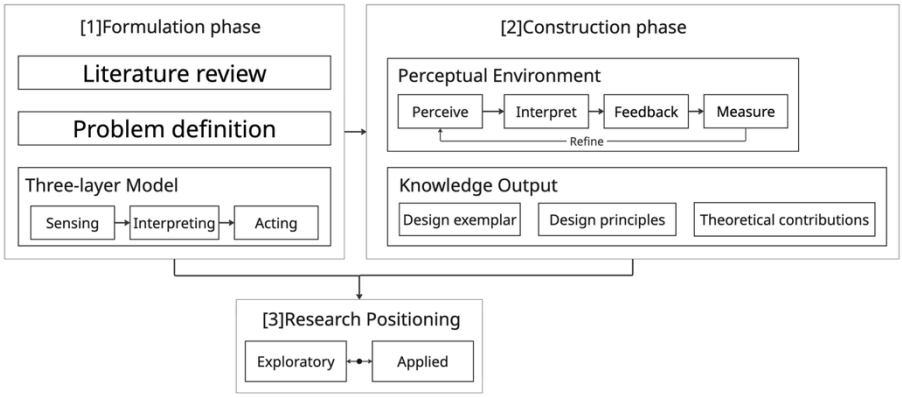


Figure 1 Research through Design Process Framework

3.2. PERCEPTUAL LOGIC AND SYSTEM CONSTRUCTION

This study constructs a perceptual environment that enables physical space to sense, interpret, and respond to interaction imbalance in cross-disciplinary collaboration. The system operates at the group interaction level to address emergent multi-user phenomena such as discourse dominance, participation withdrawal, and collective silence. As illustrated in Figure 2, the system follows a three-layer perceptual logic consisting of Perception, Interpretation, and Feedback, with iterative refinement across sessions. At the Perception layer, multimodal interaction cues are approximated through spatially embedded audio–visual sensing. Visual sensing was implemented using an Intel RealSense D435 depth camera to capture participant presence, posture

change, and spatial positioning. Human presence and movement cues were detected using a YOLO-based model implemented in PyTorch via the Ultralytics framework, with OpenCV used for image processing and spatial tracking.

Audio sensing was implemented using a multi-channel microphone array, with PyAudio and NumPy supporting real-time audio capture and signal processing. Speaking activity and collective silence were estimated through audio energy analysis and direction-of-arrival (DOA) algorithms. Multimodal cues from audio and visual streams were fused to approximate group-level interaction rhythms rather than stable individual identification. As the sensing system functioned at a prototype level, ambiguous detections were supplemented through manual annotation based on synchronized video recordings.

At the Interpretation layer, rule-based thresholds were used to translate multimodal cues into four recurring interaction imbalance patterns: reduced turn-taking, prolonged dominance, participation withdrawal, and group-level stillness. Based on detected patterns, the Feedback layer delivered ambient, non-intrusive spatial responses, such as illumination modulation and localized cues, following principles of soft fascination. Through this logic, the environment functions as a perceptual mediator supporting balanced participation and smoother interaction in collaborative settings.

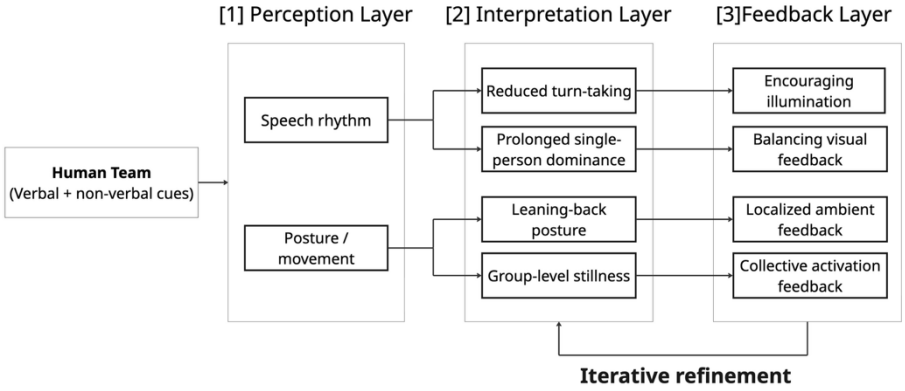


Figure 2 Three-layer system architecture with learning loop

3.3. EVALUATION METHODS

To evaluate the effectiveness of the perceptual environment, this study adopts a dual evaluation framework that combines qualitative observation with system-derived metrics, enabling interaction imbalance to be examined as both socially situated phenomena and measurable patterns captured through spatial sensing. Qualitative analysis draws on video recordings and post-session participant interviews to examine the social meaning of detected interaction imbalance patterns, participants' experiences of spatial feedback, and communication frictions arising from cross-disciplinary collaboration. Video data provide contextual grounding for interpreting group-level interaction imbalance and non-verbal behavior.

Quantitative evaluation focuses on changes associated with each imbalance pattern, as summarized in Table 1. Reduced turn-taking is examined through shifts in speaking

distribution, dominance through changes in speaking time ratios, participation withdrawal through posture recovery and verbal re-engagement, and group-level stillness through the resumption of gesture activity and conversational flow. By triangulating system-recorded indicators with qualitative observations, this framework validates both the interpretive logic of the system and the effectiveness of its spatial feedback interventions.

Table 1: Imbalance detection rules and adaptive feedback strategies

Layer	Cue (sensed behavior)	Trigger (interaction pattern)	Feedback mechanism
Speech rhythm	Reduced speaking rate; extended silence	Reduced turn-taking	Encouraging illumination (soft ambient increase)
Speech rhythm (dominance)	Prolonged uninterrupted speech by one participant	Single-person dominance	Balancing visual feedback (subtle brightness redistribution)
Posture / movement	Sustained backward-leaning; increased distance	Participation withdrawal	Localized ambient feedback (brief, low-intensity seat vibration)
Posture / movement (group)	Group-level stillness; reduced micro-movement	Interaction stagnation	Collective activation feedback (brief ambient sound cue)

4. Result

4.1. OBSERVED INTERACTION PATTERNS

This section examines how interaction imbalance manifested during collaborative sessions and evaluates the system's ability to perceive such patterns in a multi-user setting. Data were collected from approximately 9 hours of cross-disciplinary design discussions involving three mixed design–engineering teams (N = 13). Interaction indicators, including speaking activity, turn-taking frequency, and collective silence, were recorded at one-second resolution. As automatic speaker attribution was intermittently unstable, manual annotation was used to verify speaker identity in ambiguous segments.

Across three iterative prototyping rounds, the system's capacity to approximate interaction imbalance patterns improved, with detection accuracy increasing from 72% to 89%. These improvements reflect refinements in speech rhythm interpretation and participant attribution, which reduced false activations during natural pauses, rather than indicating fully automated detection. Consistent interaction tendencies were observed across teams. Early discussion phases frequently showed pronounced speaking imbalance, with dominant participants accounting for approximately 75% of total discourse time. Qualitative observations suggested that dominant speakers were often unaware of this imbalance, while quieter members gradually reduced participation attempts, contributing to interaction inertia. Periods of collective silence commonly emerged during conceptual disagreement and were accompanied by non-verbal discomfort cues such as gaze avoidance and backward-leaning postures. Engineering-background participants appeared more likely to withdraw during abstract

or terminology-heavy discussions, as reflected in reduced speaking frequency and increased physical distancing.

Figure 3 summarizes three representative interaction patterns derived from observational analysis and annotated logs: speech dominance, dominance with progressive withdrawal, and balanced interaction. These patterns represent recurrent tendencies rather than complete interaction sequences. In contrast, sessions exhibiting more balanced interaction were characterized by reciprocal exchanges, smoother turn-taking, and responsive body language. Notably, these states did not correspond to equal speaking time, suggesting that interaction fluency may relate more to regulated participation rhythms than enforced equality. Together, these findings establish an empirical baseline for examining how spatial feedback influenced subsequent interaction dynamics.

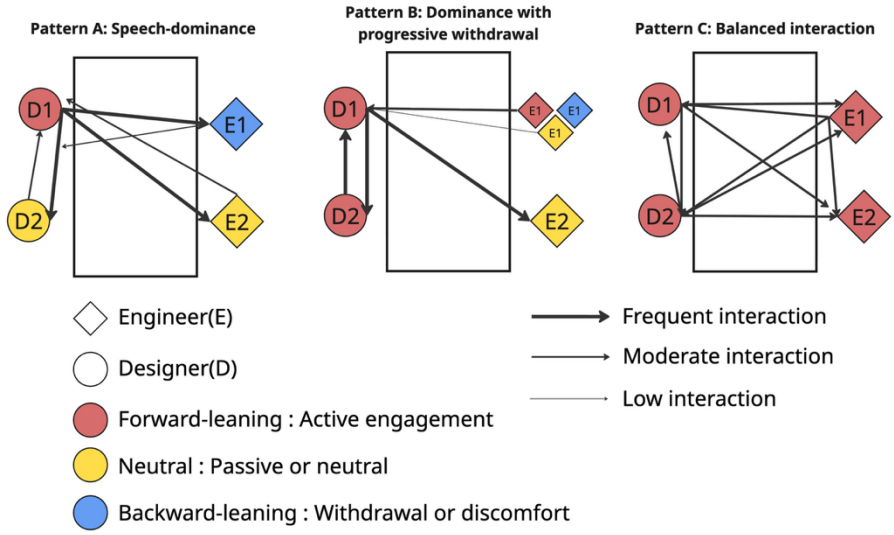


Figure 3 Representative interaction patterns observed across cross-disciplinary teams

4.2. CHANGES IN INTERACTION BALANCE FOLLOWING SPATIAL FEEDBACK

This section examines changes in interaction balance following the introduction of spatial feedback, using pre- and post-intervention indicators supported by targeted qualitative observations. Across all three teams (N = 13), spatial feedback was associated with a redistribution of speaking participation. Prior to intervention, discussions were frequently characterized by single-participant dominance, with dominant speaking time averaging 75 ± 8% of total discourse. Following feedback, this proportion decreased to 58 ± 6%, indicating a consistent downward trend in extreme dominance across teams rather than an enforced equalization of participation.

Turn-taking behavior also exhibited noticeable shifts. The frequency of turn exchanges increased from an average of 2.1 to 6.8 per ten minutes, suggesting smoother conversational transitions. Participants who had previously shown

withdrawal tendencies appeared to re-engage, with speaking instances increasing from 1.3 to 4.7 per ten minutes. At the group level, mean collective silence duration decreased from 6.8 ± 2.1 seconds to 3.2 ± 1.4 seconds, representing a substantial reduction in interaction stagnation.

Qualitative observations helped contextualize these quantitative trends. In teams with pronounced imbalance, localized spatial feedback, such as subtle seat vibration, appeared to support re-engagement through self-regulation rather than explicit intervention. In teams with relatively balanced interaction, ambient lighting and visual cues were observed to help sustain turn-taking rhythms and collective focus without disrupting creative flow. Changes in speaking distribution were further reflected in shifts in the Gini coefficient (from 0.68 to 0.52), suggesting movement toward more equitable participation while preserving situationally appropriate asymmetry. Taken together, these observations indicate that spatial feedback tended to regulate interaction extremes and restore conversational fluency, supporting the role of perceptual environments as subtle, in-situ mediators of collaborative interaction rhythms.

4.3. CROSS-TEAM CONSISTENCY AND ADAPTIVE SYSTEM BEHAVIOR

This section examines whether the effects of spatial feedback exhibited consistent patterns across teams and how the perceptual environment responded to differing collaboration rhythms over iterative sessions. Rather than framing adaptation as system optimization, the analysis focuses on how spatial mediation appeared to align with team-specific interaction dynamics. Across the three cross-disciplinary teams (N = 13), spatial feedback was consistently associated with directional changes in interaction patterns, including reduced dominance, increased turn-taking, and shorter periods of collective silence. While the magnitude of these changes varied, teams exhibiting higher initial interaction imbalance tended to show more pronounced redistribution of participation, whereas teams with relatively fluid interaction patterns demonstrated subtler yet stable shifts. These observations suggest that spatial feedback functioned in a regulative manner, supporting interaction fluency without enforcing uniform participation.

Differences in interaction rhythm across teams further highlighted the relevance of adaptive interpretation. Engineering-oriented teams commonly displayed longer reflective pauses that did not necessarily indicate interaction breakdown, whereas design-oriented teams showed faster conversational pacing. Over iterative sessions, interpretive parameters related to silence duration and dominance were gradually adjusted to better reflect these differences, contributing to improved alignment between perceived interaction states and observed group behavior. Correspondingly, pattern detection accuracy increased from 72% to 89% across iterations.

Session records indicated that interpretive adjustments stabilized after the third iteration, and participant feedback suggested that later-stage spatial feedback was perceived as less intrusive and more contextually appropriate. Taken together, these findings suggest that the effectiveness of spatial mediation may depend on its capacity to adapt interpretive logic to team-specific interaction rhythms, supporting cross-team consistency while avoiding the imposition of uniform interaction norms.

5. Discussion and Conclusion

This study explores how physical space may participate in regulating interaction imbalance in cross-disciplinary collaboration through perceptual mediation. Rather than treating imbalance as an individual or communicative failure, the findings suggest that interaction imbalance can be understood as a spatially perceivable condition that emerges through collective rhythms of speech, posture, and silence. From this perspective, interactive environments may support collaboration not by enforcing equal participation, but by moderating interaction extremes and sustaining conversational fluency.

Across the observed sessions, spatial feedback was associated with consistent shifts in interaction patterns, including reduced dominance, increased turn-taking, and shorter periods of collective silence. These changes did not indicate uniform outcomes across teams; instead, they reflected differences in disciplinary communication styles and collaboration rhythms. Through iterative refinement, the perceptual environment adjusted its interpretive thresholds, aligning more closely with team-specific patterns. This suggests that spatial mediation may operate more effectively when it remains adaptive and context-sensitive, functioning as a responsive participant rather than a controlling system.

From a design research perspective, this work contributes a reframing of interactive environments as perceptual mediators that integrate sensing, interpretation, and ambient feedback into a coherent spatial logic. The emphasis is not on sensing accuracy or automation performance, but on how multimodal social signals may be translated into non-intrusive cues that support self-regulation within collaborative groups. Methodologically, the research-through-design approach enabled the examination of these dynamics in situ, combining system logs with qualitative observation and participant reflection.

Overall, this study positions perceptual mediation as a promising design approach for collaborative environments, extending discussions of intelligent and responsive spaces toward social and group-level interaction dynamics. Future work may explore longer-term deployments, different collaborative contexts, and how participants appropriate spatial cues over time to support collective sense-making.

Attribution Statement

ChatGPT (OpenAI, 2025) was used to enhance the flow and clarity of the abstract and to correct spelling and grammar errors. All AI-assisted outputs were thoroughly reviewed and verified by the authors.

References

OpenAI. (2025). ChatGPT Model: GPT-4 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/

Chang, T.-W., Huang, H.-Y., Hong, C.-C., Datta, S., & Nakapan, W. (2023). SENS+: A Co-Existing Fabrication System for a Smart DFA Environment Based on Energy Fusion Information. Sensors, 23(6), 2890. https://doi.org/https://doi.org/10.3390/s23062890

Chang, T.-W., Huang, H.-Y., Hung, C.-W., Datta, S., & McMinn, T. (2020). A network sensor fusion approach for a behaviour-based smart energy environment for Co-making spaces. Sensors, 20(19), 5507. https://doi.org/https://doi.org/10.3390/s20195507

Gray, B., Purdy, J., & Ansari, S. (2022). Confronting power asymmetries in partnerships to address grand challenges. Organization Theory, 3(2), 26317877221098765. https://doi.org/https://doi.org/10.1177/26317877221098765

Li, J.-R., Chang, T.-W., & Chang, C.-C. (2024). CocoBot: Developing a consent communication tool based on co-evolutionary model. Proceedings of the 29th International Conference of the Association for ComputerAided Architectural Design Research in Asia (CAADRIA) 2024, 49-58. https://doi.org/https://doi.org/10.52842/conf.caadria.2024.2.049

Nguyen, B. V. D., Han, J., & Vande Moere, A. (2022). Towards Responsive Architecture that Mediates Place: Recommendations on How and When an Autonomously Moving Robotic Wall Should Adapt a Spatial Layout. Proc. ACM Hum. Comput. Interact., 6(CSCW2), 1-27. https://doi.org/https://doi.org/10.1145/3555568

Weiser, M. (1991). The Computer for the 21 st Century. Scientific american, 265(3), 94-105. https://doi.org/https://www.jstor.org/stable/24938718

Lee, J. H., Ostwald, M. J., & Kim, M. J. (2021). Characterizing smart environments as interactive and collective platforms: A review of the key behaviors of responsive architecture. Sensors, 21(10), 3417. https://doi.org/https://doi.org/10.3390/s21103417

Augusto, J. C. (2022). Contexts and context-awareness revisited from an intelligent environments perspective. Applied Artificial Intelligence, 36(1), 2008644. https://doi.org/https://doi.org/10.1080/08839514.2021.2008644

McDonnell, A. S., & Strayer, D. L. (2024). Immersion in nature enhances neural indices of executive attention. Scientific Reports, 14(1), 1845. https://doi.org/https://doi.org/10.1038/s41598-024-52205-1

Sidra, S and Mason, C: 2025, Generative AI in human-AI collaboration: validation of the collaborative AI literacy and collaborative AI metacognition scales for effective use, International Journal of Human–Computer Interaction: 1-25.

Busciantella-Ricci, D., & Scataglini, S. (2024). Research through co-design. Design Science, 10, e3. https://doi.org/10.1017/dsj.2023.35

From Interaction to Synergy

Human–AI–Robot Collaborative Design in Immersive Mixed-Reality Environments

Zhiyong Dong[1], Peter Búš[2], Qi Wang[3], Maowei Jiang[4]
[1,2,3,4]*Tsinghua University.*
[1]*dongzy23@mails.tsinghua.edu.cn, 0000-0003-0958-2986*
[2]*peter_bus@sz.tsinghua.edu.cn, 0000-0002-1730-4559*
[3]*wangqi24@mails.tsinghua.edu.cn, 0009-0001-6327-5402*
[4]*jmw24@sz.tsinghua.edu.cn, 0000-0002-4267-5570*

Abstract. Architectural assembly often involves dynamic, uncertain, and spatially complex conditions, yet traditional robotic systems rely on fixed programming and screen-based interfaces that separate human intuition from robotic execution. This study investigates a unified workflow for human–AI–robot collaborative design in immersive mixed-reality environments, aiming to bridge the gap between generative design intelligence and precise robotic assembly. We propose an integrated system where augmented reality enables designers to manipulate virtual components in real space, conditional Generative Adversarial Networks (cGANs) generate context-aware modular assembly proposals, and robotic arms execute precise physical assembly through real-time bidirectional communication. A paired-dataset approach and image-based representation allow the cGAN to learn modular relationships, while a digital twin ensures continuous feedback and low-latency interaction. Experimental results demonstrate that the system supports both manual and AI-assisted workflows, achieving accurate alignment of modular components and enabling iterative co-creation. The approach captures designers' tacit knowledge and integrates it into the generative model, creating a synergistic loop between human spatial reasoning, AI prediction, and robotic actuation. Findings suggest that immersive mixed-reality interfaces combined with data-driven reasoning can transform architectural robotics from reactive control to adaptive, collaborative, and co-creative design-to-production processes.

Keywords. Human–Robot Collaboration, Generative Design, Augmented Reality Interaction, Conditional GAN, Mixed Reality Robotics, Adaptive Construction Workflow, Digital Twin

1. Introduction

Architectural robotics has long relied on fixed programming, predefined motion sequences, and screen-based interfaces that separate robotic execution from human

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 153-162.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

spatial reasoning. Such workflows are poorly suited for the dynamic and uncertain conditions of architectural assembly, where desi gners depend on embodied intuition and situational judgment. In recent years, extended reality (XR, including AR, VR and MR) technologies have demonstrated strong potential to address this challenge by enabling designers to manipulate virtual components directly within physical space, providing intuitive three-dimensional feedback and more fluid control of robotic systems. Yet despite these advances, most AR-robotic frameworks remain limited to direct manipulation: humans act, robots follow, and the interaction stops short of allowing robots—or the computational system behind them—to actively reason or contribute to the design process.

Concurrently, advances in artificial intelligence, particularly in generative models such as conditional Generative Adversarial Networks (cGANs), have shown remarkable capability in learning design logic, producing structural variations, and adapting to complex modular systems. These models can capture latent spatial patterns and generate feasible next-step configurations, offering a new form of machine reasoning that complements human intuition. However, their integration into real-time, on-site construction workflows remains underdeveloped. The separation between AI-based design generation and embodied robotic execution limits the potential for genuine synergy. This research addresses these gaps by establishing a unified human–AI–robot workflow in an immersive mixed environment, where AR supports intuitive human input, cGANs provide data-driven generative reasoning, and robotics ensures precise physical execution. By enabling these three agents to operate collaboratively, the study aims to move architectural robotics from reactive control toward coordinated, adaptive, and co-creative design-to-production processes.

2. Background and Related Work

2.1. IMMERSIVE EXPERIENCE AND AUGMENTED REALITY TECHNOLOGY FOR ROBOTIC CONTROL

Augmented Reality (AR) can be emerged as a viable alternative for improving robotic control systems, enhancing user interaction by overlaying virtual information onto the physical workspace, providing real-time visualization of robotic trajectories, operation status, and task execution. Huang and Spaw (2023) explored the integration of AR with robotic tube bending, improving fabrication precision, streamlines complex bending processes, and facilitating human-robot collaboration in assembly tasks. Zhao et al. (2024) integrated AR to enable intuitive interaction, robotic trajectory planning, and flexible workspace configuration for creative design and prototyping. An AR-assisted framework is developed for masonry structures, enabling intuitive gesture-based design and robotic assembly through an interactive interface, enhancing designer participation in the construction process (Song et al. 2023). Song and Hahm (2023) explored AR-based robotic bricklaying, utilizing information transmission to remotely visualize robotic arm motion simulation. A mobile AR application is used for human-robot collaboration in timber assembly, supporting real-time task distribution and synchronization (Alexi et al. 2024). Mitterberger et al. (2022) integrated AR interfaces with a robotic spraying system, enabling designers and skilled workers to intuitively engage with in-situ plastering through controller-based interaction and real-time audio-

visual guidance. These approaches focus on pre-setting design displays and guiding the construction process. However, further testing is needed for true real-time control and path planning within a three-dimensional mixed space, enabling the human user to interact with the robot intuitively and immediately.

2.2. AI-DRIVEN ENHANCEMENT AND EMBODIED ROBOTICS

In recent years, embodied intelligence has emerged as a cutting-edge field, emphasizing the robot's ability to autonomously perceive, interact with, and make decisions within its physical environment, overcoming the limitations of traditional reliance on external programming. By integrating perception, learning, and motion control, embodied intelligent robots can respond to environmental changes in real time, enhancing their adaptability in complex tasks (Attalla et al. 2023).

The technological core of embodied intelligence lies in multimodal perception and self-optimization capabilities. Advances in vision systems, particularly with the introduction of deep learning technologies, have equipped robots with powerful environmental perception capabilities. For instance, the application of convolutional neural networks (He et al. 2016) in object detection has significantly improved the ability of robots to identify and localize complex targets. Additionally, digital twin technology enables robots to simulate and optimize their action strategies in virtual environments, reducing the error rates in real-world operations. This adaptability is particularly critical in architectural design and production. On the one hand, the uncertainties and environmental changes in sites demand that robots possess rapid adjustment capabilities. On the other hand, tasks involving complex geometries and modular assembly require robots to efficiently identify components and plan optimal paths. With the development of artificial intelligence, generative neural networks have been applied in architectural and urban design for project generation (Dong 2022; Zheng and Yuan 2021), offering potential intelligent solutions for the design-to-production process of robotics.

3. Methodology

3.1. MAIN WORKFLOW

To achieve more efficient AR-driven robotic control, the system integrates human spatial manipulation, AI-based generative reasoning, and robotic execution into a unified data loop, as shown in Figure 1. The workflow begins with the user arranging virtual components in the mixed-reality environment, where gestures and hand-tracking define the intended spatial configuration. This interaction is captured and translated into structured data, which is then processed by the cGAN model to generate the next-step assembly proposal based on modular logic. The resulting design—whether human-defined, AI-suggested, or collaboratively refined—is transmitted to the robot through a bidirectional communication channel, enabling precise physical assembly while continuously updating the digital twin. This closed, synergistic cycle supports adaptive decision-making and real-time co-creation between human intuition, AI prediction, and robotic actuation.

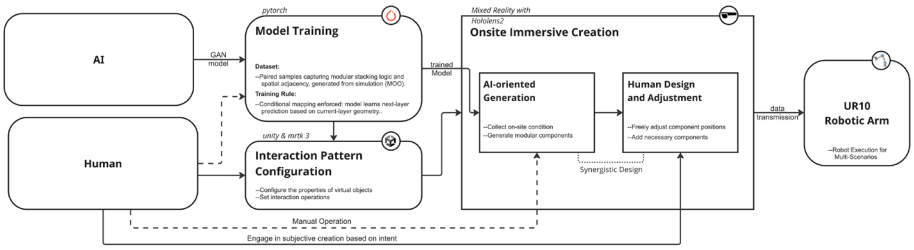


Figure 1. Synergistic Design Workflow for Robotic Assembly

3.2. MIXED-REALITY IMMERSIVE OPERATION CONFIGURATION

The immersive operation environment is built using the Mixed Reality Toolkit 3 (MRTK3), an open-source development framework released by Microsoft (2023) for creating mixed-reality and augmented-reality applications in Unity. MRTK3 provides a modular set of components that support interaction, input handling, and spatial visualization, offering a consistent development experience for mixed-reality interfaces.

Within this setup, virtual components are configured as interactable objects that can be grasped, moved, and rotated directly in mid-air. Their physical properties, such as rigid-body behavior and gravity settings, are defined to maintain stable manipulation. Using MRTK3's interaction system, we enable multiple input modes, including motion-controller interaction, hand-tracking, and direct hand manipulation. The system captures the user's gestures in real time, allowing components to be repositioned through air-hold gestures for direct grabbing or via hand-ray interactions for remote manipulation. Paired with a head-mounted display (HMD), this configuration creates an immersive operation environment that supports intuitive spatial manipulation and real-time design adjustments.

3.3. DATA-DRIVEN GENERATIVE REASONING

To accommodate the diverse and highly flexible assembly configurations in this experiment, we adopt an Image-to-Image conditional Generative Adversarial Network (cGAN) (Isola et al. 2017) as the core model for data-driven reasoning. A standard GAN consists of two main components: a generator and a discriminator. During training, the generator synthesizes candidate images from latent representations, while the discriminator evaluates these outputs by distinguishing them from real samples. Through this adversarial process, the generator progressively learns to produce outputs that become increasingly indistinguishable from the ground truth, achieving a stable balance between the two networks.

The introduction of conditional variables by Mirza and Osindero (2014) extended traditional GANs into conditional GANs, enabling mappings not only from random noise z but also from an input condition x to generate an output y. In our workflow, the conditional input x is extracted directly from the current assembly state, guiding the generator to produce structured and context-specific predictions. This configuration enhances the model's ability to generate modular assembly proposals that respond to

user-defined spatial layouts, forming the basis of the AI-driven reasoning process, as shown in Figure 2.

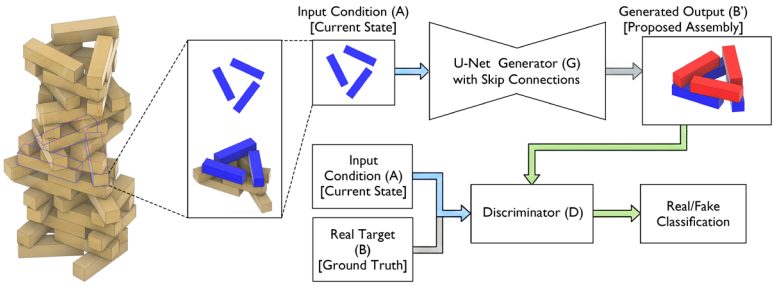


Figure 2. cGAN Framework and Data Processing

3.4. BIDIRECTIONAL COMMUNICATION AND ROBOTIC EXECUTION

Throughout the experiment, robust connectivity among the HMD, robot, and host computer was essential to maintain an ultra-low latency real-time data loop (Figure 3). All devices were connected via TCP/IP on the same local network. The first link, between the HMD and host computer, served two purposes: enabling a digital twin, where a virtual UR10 arm in the HMD simulates robot movements and transmits its state back to the host in real time for live visualization and JSON recording; and supporting human–AI interaction, allowing the trained cGAN to generate design proposals based on the current assembly state, write results to JSON, and transmit sequential data back to the HMD for collaborative operation, supporting synergistic collaboration between human and AI.

The second link connects the HMD to the physical robotic arm. Movements of the virtual arm within the immersive environment are directly translated into joint angles for the real robot, which are then sent as URScript commands to execute physical motion. The transformation from end-effector position to the six joint angles of the UR10 constitutes the inverse kinematics (IK) computation. IK is implemented within the Unity environment by configuring the virtual arm. The process begins by mirroring the robot geometry according to the Denavit-Hartenberg (DH) convention, which defines link lengths, joint axes, and offsets for kinematic calculations, establishing a mapping between joint space and Cartesian space. Inverse kinematics is then applied to determine joint angles for a given target pose, defined either through the AR interface or a predefined trajectory. Analytical methods are used when a closed-form solution exists, while numerical methods iteratively minimize the error between the desired and computed poses when analytical solutions are not feasible, ensuring convergence to a valid joint configuration.

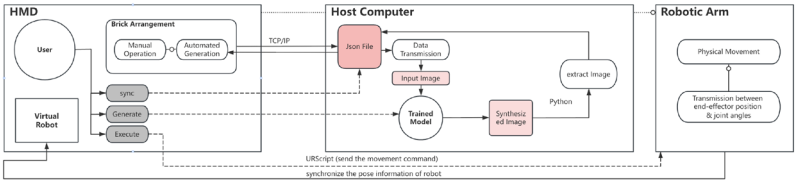


Figure 3. Bidirectional Communication among HMD, Host Computer and Robotic Arm

4. Experiment

4.1. INTERACTION MODE SETUP TO CONTROL ROBOTS

The immersive interface, developed with the MRTK package, enables interaction with digital objects in the real-world environment and collaboration with the trained agent for design completion. By merging virtual and physical spaces, users can manipulate virtual objects through intuitive gestures, simulating both modular component generation/adjustment and robotic motion control. In the mixed-reality environment (Figure 4), the base interface includes a button bar, a virtual UR10 arm, a display window for the physical robot's pose, and the corresponding real brick placements.

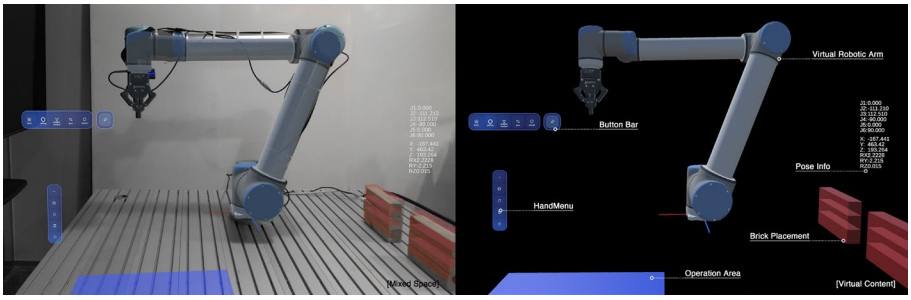


Figure 4. Initial Interface in the Immersive Mixed-Reality Space

To enhance functionality, a button bar was implemented with toggle options for user-follow and fixed modes (Figure 5), includeing 5 menus: Brick, Setting, Gripper Control, Sync, and Reload. In the Brick submenu, the Add button lets users place and manipulate virtual bricks, the Gravity button toggles gravity, the Generate button sends signals to the host computer for the cGAN to produce new assembly proposals displayed in the immersive space, the Erase button clears bricks, and the Execution button commands the physical robot to assemble them. The Setting submenu includes Start/Stop Transmission buttons to control the real-time connection and a Base Display button with a slider for robotic base angle calibration. The Gripper submenu manages opening/closing and end-effector rotation via sliders and angle buttons. The Sync button sends the current brick state to the host computer as a JSON file. Additionally, a HandMenu follows the user's hand for quick access to key functions.

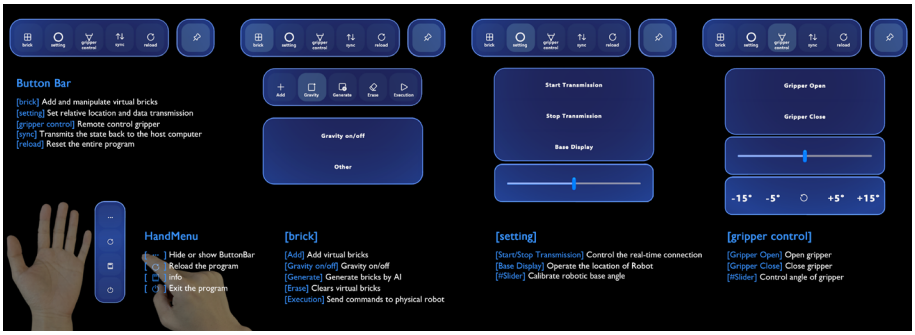


Figure 5. Button Bar and HandMenu for Immersive Operation

4.2. DATA-DRIVEN APPROACH FOR MODULAR GENERATION

4.2.1. Paired-Dataset Construction

To achieve stable modular structure generation, the structural problem was reformulated into an image-based representation, enabling the model to learn spatial relationships visually. Within the simulation workflow, a variety of assembly configurations were produced and subsequently converted into standardized cases aligned with the modular design logic. Each layer of the modular structure was encoded using quantitative stability metrics, including the number of intersection surfaces between upper and lower components (Nc,i), the total intersecting area (Sc), and the number of intersections among upper-layer elements (Ic). Safety zones were also defined for each component to prevent overlap within the same layer (Figure 6).

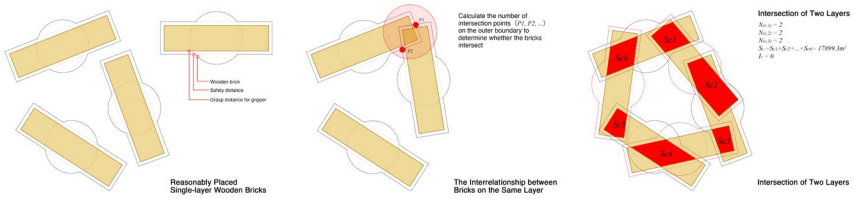


Figure 6. Modular Design Logic and Generation Rules for Paired Dataset Construction

The projects were converted into a paired dataset: the bottom layer labeled in blue formed dataset A, and combining layers A and B created dataset B. This pairing enables the network to learn the transformation from lower to upper layers, generating upper-layer configurations from given lower-layer arrangements. Initially, 105 pairs were generated via multi-objective optimization, then expanded through mirroring and rotation to 4,200 pairs, with 3,800 for training and 400 for testing (Figure 7).

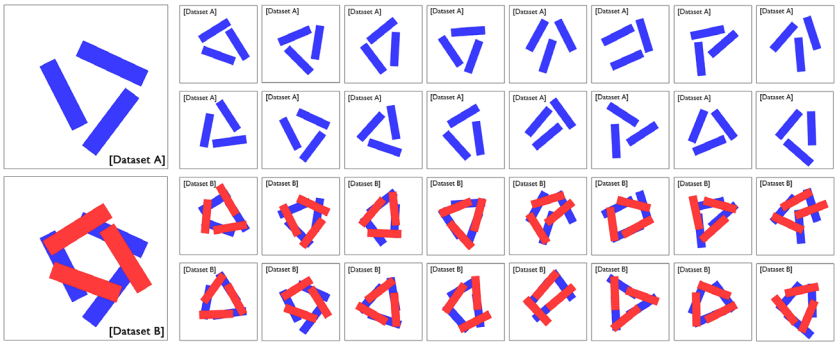


Figure 7. Part of Sample Paired Training Dataset after Augmentation

4.2.2. Training and Testing

The GAN model was trained using the constructed dataset. All input images from the training dataset were resized to 512×512 pixels to match the network's input requirements, facilitating the learning of modular assembly logic. The training process

is an adversarial interaction between the generator and discriminator: the generator synthesizes candidate upper-layer images based on the bottom-layer input, while the discriminator evaluates whether the outputs resemble real examples. Loss values of both networks were recorded throughout training (Figure 8). At the equilibrium point, representing the optimal solution of the minimax game, the discriminator assigns a probability of 0.5 to the generator's outputs. In practice, a higher discriminator loss paired with a lower generator loss suggests that the generator is effectively learning the target mapping. As training progressed, the loss curves of both the generator and discriminator converged and reached a stable state, indicating effective learning.

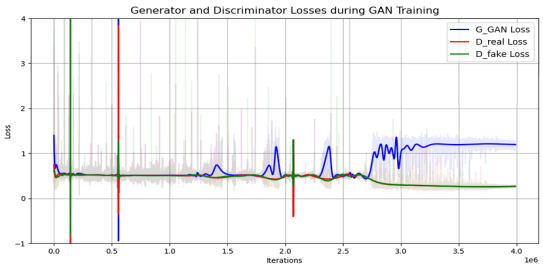


Figure 8. Generator and Discriminator Loss Values during Training

4.2.3. Post-Processing of Generated Data

When the "Generate" button is triggered, data from the immersive environment is sent to the host computer and stored as a JSON file. The top three bricks are extracted and rendered as an image, which is input to the trained cGAN to generate the next-layer arrangement. The output image is processed with OpenCV to detect red pixel regions representing upper-layer components. Their centroids are calculated as component centers, and components are organized according to their primary rotation directions (Figure 9). Through iterative generation and processing, the system derives the number, positions, and rotation angles of components per layer, and transmits this data back to the immersive space for integration into the synergistic design workflow.

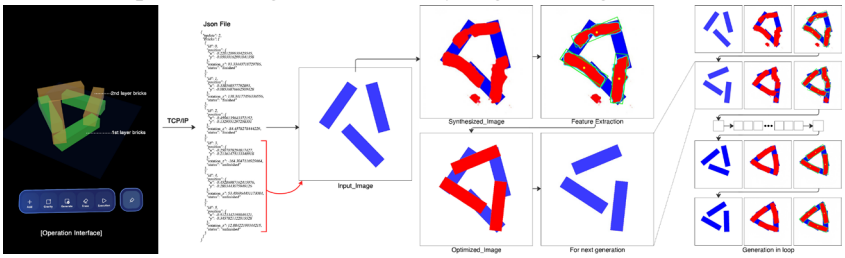


Figure 9. Generator and Discriminator Loss Values during Training

4.3. RESULTS AND DISCUSSION

Using the completed program, modular assembly can be carried out either through direct guidance of the UR10 robotic arm or via AI-assisted planning. In practice, virtual components are positioned within the immersive space, and the system commands the UR10 to transport the corresponding physical components from the magazine to their

target locations. Component IDs and positions are simultaneously recorded on the host computer in a JSON file. Placed virtual objects are visually updated with reduced transparency and highlighted in green. For AI-assisted operation, the current assembly state is transmitted to the trained cGAN model, which generates a new placement proposal based on the top-layer components and renders it directly in the immersive space. The robotic arm then executes the assembly according to this plan. This interactive loop between human input and AI reasoning enables designers to iteratively explore, refine, and validate modular assembly strategies.

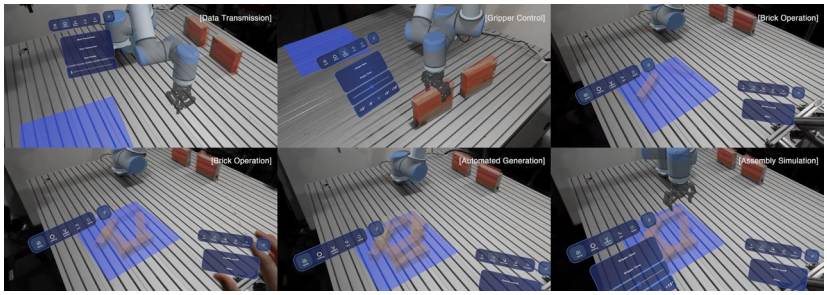


Figure 10. Onsite Human–Robotic Synergistic Assembly in the Mixed-reality Space

The workflow enables effective collaboration between human intuition, AI generativity, and robotic precision, achieving accurate component alignment in both manual and AI-assisted modes with stable performance across repeated tasks. Although the current prototype targets simple structures, the framework is scalable to larger datasets and more complex configurations. AR-based interaction captures designers' tacit knowledge, which can be fed back into the AI model for continuous learning, highlighting mixed-reality environments as both control interfaces and knowledge-exchange systems for adaptive, co-creative design-to-production workflows.

5. Conclusion

This study presents a unified framework that shifts architectural robotics from one-way interaction to genuine human–AI–robot synergy. By integrating immersive mixed-reality manipulation, data-driven generative reasoning, and real-time robotic execution within a continuous feedback loop, the system allows designers to operate fluidly across physical and virtual spaces while enabling AI models to contribute directly to decision-making. The cGAN-based reasoning engine learns modular logic from paired datasets and offers context-aware assembly proposals, while the digital-twin architecture preserves low-latency coordination among human actions, AI predictions, and robotic actuation. Experimental results show that this collaborative workflow supports both manual and AI-assisted assembly, strengthens spatial understanding, and improves the accuracy and adaptability of modular construction tasks. Immersive interfaces paired with embodied intelligence can elevate architectural robotics from reactive execution tools to proactive, co-creative partners, opening new possibilities for adaptive, iterative, and human-centered design-to-production processes.

Acknowledgements

The research is supported by Cross-disciplinary Research and Innovation Fund provided by Shenzhen International Graduate School, Tsinghua University, grant n. JC2024008 under the project titled "Intelligent design-to-built-production: Reusing Demolition Waste in Robotic Fabrication and Circular Construction".

References

Alexi, Eleni Vasiliki, Joseph Clair Kenny, Lidia Atanasova, Gonzalo Casas, Kathrin Dörfler, and Daniela Mitterberger. 2024. "Cooperative Augmented Assembly (CAA): Augmented Reality for on-Site Cooperative Robotic Fabrication." *Construction Robotics* 8 (2): 28.

Attalla, Ahmed, Omar Attalla, Ahmed Moussa, Duaa Shafique, Saleh Ba Raean, and Tarek Hegazy. 2023. "Construction Robotics: Review of Intelligent Features." *International Journal of Intelligent Robotics and Applications* 7 (3): 535–55.

Dong, Zhiyong. 2022. "Perceiving Fabric Immersed in Time, an Exploration of Urban Cognitive Capabilities of Neural Networks." In *POST-CARBON - Proceedings of the 27th CAADRIA Conference*. Sydney.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Deep Residual Learning for Image Recognition." In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–78. doi:10.1109/CVPR.2016.90.

Huang, Lee-Su; Spaw. 2023. "Augmented Reality Assisted Robotic: Tube Bending." In *Proceedings of the 43rd Annual Conference for the Association for Computer Aided Design in Architecture (ACADIA)*, 340-349.

Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. "Image-to-Image Translation with Conditional Adversarial Networks." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5967–76.

Microsoft MRTK team. 2023. "Mixed Reality Toolkit 3." https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-overview/.

Mirza, Mehdi, and Simon Osindero. 2014. "Conditional Generative Adversarial Nets." arXiv. doi:10.48550/arXiv.1411.1784.

Mitterberger, Daniela, Selen Ercan Jenny, Lauren Vasey, Ena Lloret-Fritschi, Petrus Aejmelaeus-Lindström, Fabio Gramazio, and Matthias Kohler. 2022. "Interactive Robotic Plastering: Augmented Interactive Design and Fabrication for On-Site Robotic Plastering." In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–18. CHI '22. New York, NY, USA: Association for Computing Machinery.

Song, Yang, Asterios Agkathidis, and Richard Koeck. 2023. "Augmented Bricks an Onsite AR Immersive Design to Fabrication Framework for Masonry Structures." In *Hybrid Intelligence*, edited by Philip F. Yuan, Hua Chai, Chao Yan, Keke Li, and Tongyue Sun, 385–95. Singapore: Springer Nature.

Song, Yang and Hahm. 2023. "Augmented Robotic Bricklaying." In *Immanuel Koh, Dagmar Reinhardt, Mohammed Makki, Mona Khakhar, Nic Bao (Eds.), HUMAN-CENTRIC - Proceedings of the 28th CAADRIA Conference, Ahmedabad, 18-24 March 2023, Pp. 323–332*. CUMINCAD.

Zhao, Yi, Lynn Masuda, Lian Loke, and Dagmar Reinhardt. 2024. "Designing a Dynamically Configurable Digital Twin for Human-Robot Collaboration Tasks: A Case of Working Environment Configuration for the Robotic Lab." In *CAADRIA 2024: Accelerated Design*, 381–90. Singapore. doi:10.52842/conf.caadria.2024.3.381.

Zheng, Hao, and Philip F. Yuan. 2021. "A Generative Architectural and Urban Design Method through Artificial Neural Networks." *Building and Environment* 205: 108178.

A Vision-Perception-Based Adaptive Building Façade

Research and Prototype Validation

Di Zhang[1], Yunsong Han[2,], Yuxiao Wang[3] and Feixue Chen[4]
[1,2,4]*Harbin Institute of Technology*
[3]*Tsinghua University*
[1]*22s034017@stu.hit.edu.cn, 0009-0007-3401-3829*
[2]*hanyunsong@hit.edu.cn, 0000-0003-3158-6123*
[3]*yuxiao-w23@mails.tsinghua.edu.cn, 0000-0001-8215-6540*
[4]*1190500409@stu.hit.edu.cn, 0009-0002-0435-831X*

Abstract. Growing demands for daylight quality and visual comfort underscore the need for occupant-responsive adaptive façades, yet maintaining real-time balance between glare control and daylight access remains challenging. This study presents a vision-driven control framework that integrates human visual perception, simulation-based decision making, and physical actuation. The prototype façade, composed of origami-inspired triangular modules, operates through zonal adjustments guided by real-time visual field recognition and glare simulation feedback. A deep-learning perception module detects head pose and gaze direction, while DGP is derived directly from HDR images captured inside the scale model, ensuring that all control decisions are based on real measured glare rather than simulated values. Control weights derived from the gaze-projected target zones inform rotation through a rule-based algorithm implemented on a dual-Arduino platform. Field experiments conducted in a south-facing office in Harbin, China, under winter clear-sky conditions showed that the system maintained DGP$\leqslant$0.35 in the majority of control cycles and improved daylight uniformity by 11.6% relative to a static reference. These results demonstrate that coupling computer-vision occupant sensing with simulation-informed façade actuation effectively enhances visual comfort and daylight performance, providing a scalable pathway toward human-centered, perceptually intelligent adaptive façades.

Keywords. Adaptive façade, Visual comfort, Computer vision, Daylight glare probability, Deep learning, Human-building interaction

1. Introduction

Balancing natural daylight and visual comfort remains a critical yet unresolved challenge in contemporary building design (Hosseini et al., 2019). Excessive daylight exposure often results in glare and visual fatigue, while insufficient daylight increases

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 163-172.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

dependence on artificial lighting and, consequently, energy consumption. Achieving an optimal balance between these competing demands is therefore essential for both occupant well-being and sustainable building operation (Tabadkani et al., 2021).

Adaptive façades have emerged as an effective approach to regulate daylight transmission and solar gain. Most existing systems rely on environment-driven predictive control, adjusting façade geometry based on environmental parameters such as solar altitude or outdoor illuminance. However, these methods often neglect real-time occupant perception and cannot dynamically adapt to subjective variations in comfort or task requirements (Zabin et al., 2022). The lack of responsiveness to human perceptual feedback limits their effectiveness in real-world applications.

To address this gap, this research introduces computer vision – based occupant sensing into adaptive façade control. By employing non-contact visual field recognition, the system perceives real-time user gaze and integrates it with daylight simulations to quantify visual comfort using Daylight Glare Probability (DGP). This integration allows façade systems to respond to both environmental and human factors simultaneously.

In this case study, the adaptive façade prototype is controlled using the occupant's real-time head orientation and measured glare conditions. The system applies a practical rule-based control strategy calibrated through experiment, rather than relying on predictive simulation models.

2. Related Work

2.1. ADAPTIVE FAÇADE CONTROL STRATEGIES

Adaptive façade systems are widely recognized as effective solutions for enhancing indoor environmental quality and reducing building energy consumption (Tabadkani et al., 2021). Existing control strategies can generally be categorized into predictive simulation-based approaches and real-time feedback approaches. Predictive strategies utilize building performance simulation tools, such as Radiance, EnergyPlus, and Grasshopper with Ladybug and Honeybee, to optimize façade configurations according to solar geometry, illuminance levels, and thermal parameters (Lee, D et al., 2021). Although these methods enable precise optimization under predefined boundary conditions, they often overlook the dynamic nature of human perception, which leads to static assumptions about occupant behaviour and visual comfort thresholds.

In contrast, real-time feedback control depends on direct occupant input, including subjective comfort voting, manual blind adjustments, and sensor-based illuminance thresholds, to govern façade movement. While these systems can respond to actual user discomfort, they typically lack the integration of physical simulation models that can represent the complex interactions among daylight distribution, glare formation, and occupant adaptation (Wen et al., 2024). As a result, their control decisions are often reactive rather than predictive. Recent studies have increasingly explored hybrid control frameworks that merge predictive simulations with perceptual feedback, thereby allowing adaptive façades to operate based on both environmental and occupant-driven data (Gordo-Gregorio et al., 2025). Nevertheless, such human-in-the-loop approaches remain underdeveloped, particularly with respect to non-intrusive

sensing, data interoperability, and validation under real-world conditions.

2.2. DEEP LEARNING IN HUMAN SENSING

Recent developments in human sensing technologies have greatly enhanced the ability of building systems to understand occupant behaviour and comfort (Ji et al., 2023). Traditional methods often rely on self-reported feedback, wearable sensors, or environmental detectors such as motion or light sensors. Although these approaches provide reliable quantitative data, they can be intrusive, costly, and limited in spatial or temporal coverage, making continuous application in real settings difficult. To overcome these limitations, advances in computer vision, acoustic sensing, and multimodal fusion now enable non-intrusive monitoring of occupants' postures, gestures, and visual attention, offering richer information for modelling human–environment interaction (Li et al., 2023).

In recent years, deep learning has become a powerful tool for robust and real-time human sensing (Hempel et al., 2022). Convolutional and transformer-based architectures achieve high accuracy in head pose estimation, gaze tracking, facial expression analysis, and activity recognition under diverse indoor conditions. Models such as 6DRepNet, trained on large annotated datasets, such as AFLW2000-3D, allow passive inference of attention, emotion, and fatigue without the need for wearable devices. Beyond vision, emerging work combines acoustic, thermal, and Wi-Fi sensing to capture speech, heat signatures, and movement patterns (Abade et al., 2018). Despite these advances, most applications remain in controlled environments, and integrating such multimodal sensing with simulation-informed environmental control remains a key direction toward human-centered and perceptually intelligent buildings.

3. Methodology

In this study, the adaptive façade control workflow was implemented specifically for the Harbin office case, where a scale-model façade prototype was operated in a closed loop using real occupant-derived visual information. Real-time head pose and gaze direction, obtained through camera-based deep learning, were used to capture the user's field of view during the experiment. For each gaze instance, High Dynamic Range (HDR) images recorded inside the model provided the measured DGP, which served as the basis for determining the façade's rotation state. The actuation module then adjusted the physical façade units accordingly and recorded new HDR imagery for subsequent evaluation. This workflow reflects the practical control approach used in this particular case study rather than a generalized façade control framework.

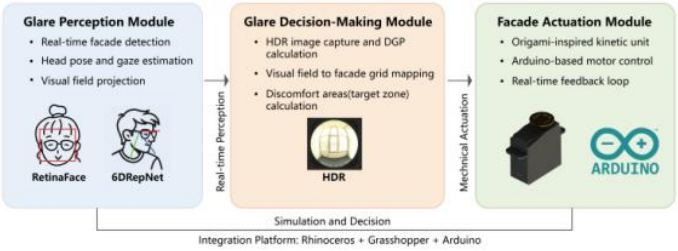


Figure 1. Overall system architecture of vision-driven adaptive façade control

3.1. RESEARCH CASE & BOUNDARY CONDITIONS

Field measurements were conducted in a south-facing office located on the fourth floor of a building in Harbin, China. The test space measured 5 m in depth, 4 m in width, and 3.6 m in floor-to-floor height. To ensure reproducibility, all evaluations were conducted through on-site measurements in a controlled real-environment setup. The scale models were embedded within the actual office space, driven by real occupant head-pose sequences, and assessed using measured HDR imagery and illuminance data, as shown in Figure 2.

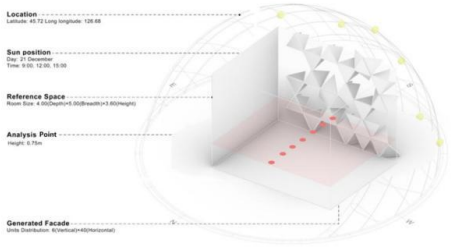


Figure 2. Schematic of the test space and the embedded scale models

3.2. PROTOTYPE & CONTROL STRATEGY

The proposed adaptive façade functions through a closed-loop process encompassing visual perception, glare assessment, and façade actuation. The three corresponding modules, glare perception, glare decision-making, and façade actuation, are responsible for data acquisition, state evaluation, and physical execution.

3.2.1. Physical prototype

The adaptive façade prototype consists of equilateral, origami-inspired triangular units as shown in Figure 3. Each unit achieves bidirectional rotation through a central pulling-disk mechanism with front and back cable paths. Units are arranged in a grid and clustered under shared actuators to balance high kinematic flexibility with a low actuator count. The supporting frame is constructed from aluminium profiles and acrylic plates, while centralized servo mounts and limit stops facilitate rapid assembly, calibration, and maintenance at the scale-model level.

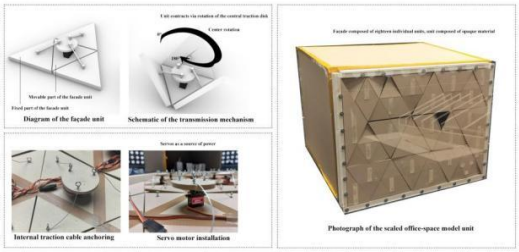


Figure 3. Design and assembly of the adaptive façade prototype and scaled office-space model

3.2.2. Glare perception module - Real-time visual field recognition

An indoor camera captures the occupant's face, where RetinaFace combined with MobileNet-0.25 performs lightweight face detection, and 6DRepNet estimates the head's yaw, pitch, and roll (Deng et al., 2020). The resulting gaze vector is then mapped onto the façade plane through a pinhole camera model, defining the user's visual-field cone.

To enhance robustness under high-contrast luminance, the model was fine-tuned using a custom dataset of approximately 18,000 images, combining samples from AFLW2000-3D and 300-VW datasets (Shen et al., 2015, Zhu et al., 2016). The module outputs the occupant's gaze direction and its corresponding façade projection for downstream glare evaluation.

3.2.3. Glare decision-making module - Simulation-based comfort assessment

The glare decision-making module integrates perceptual information from the visual field identified by the perception module with glare evaluation derived from measured HDR imagery to determine the optimal rotation states of the façade units. For each detected head field of view, HDR images captured within the corresponding gaze cone are processed to compute the DGP, using a comfort threshold of DGP=0.35 to distinguish acceptable from discomfort conditions (Wienold & Christoffersen, 2006). The projected gaze cone defines Target Zones (TZs) where visual discomfort is most likely to occur. Each façade cell is then assigned a dynamic control weight based on its distance from the gaze axis, its relative contribution to measured glare, and a seasonal weighting factor that balances daylighting and glare suppression under varying conditions. A rule-based solver computes the target rotation angle for each unit and applies spatial smoothing to avoid abrupt transitions and maintain geometric continuity. Through this process, the module converts real, perception-derived measurements into actionable control parameters that drive the façade's real-time adaptation.

3.2.4. Façade actuation module - Rule-based physical execution

The façade actuation module physically executes the control commands generated by the decision-making process through a clustered servo array managed by dual Arduino microcontrollers. Target rotation angles, computed in Grasshopper with Python, are exported as CSV files and transmitted via serial-LoRa communication to the two synchronized Arduino boards for parallel multi-channel control. The firmware converts the target angles into PWM signals using a non-linear mapping function and dynamically adjusts the step size according to the real-time DGP deviation. Tunable parameters govern the responsiveness and damping behaviour, ensuring both stability and sensitivity. This closed-loop framework enables real-time, perception-informed façade adjustment, reducing implementation complexity while achieving effective glare mitigation and maintaining balanced daylight performance.

3.3. HARDWARE–SOFTWARE INTEGRATION

The control system consists of two Arduino ATMEGA328P boards, a 16-channel

PWM servo driver with an isolated power supply, and a LoRa 433 MHz serial repeater for wireless communication. A dual-supply isolation scheme was adopted to improve system stability: the servos were powered by an MW LRS-100-5 converter (5V, 18A), while the Arduinos were powered independently through Type-C interfaces to reduce digital interference. All boards and peripheral modules were solder-mounted on a prototyping backplane and enclosed within a reinforced housing, which improved mechanical rigidity and minimized vibration during testing.

The Grasshopper–Python workflow computed the target rotation angles for each façade unit, exported the results as CSV files, and transmitted the data via serial–LoRa communication to the dual Arduino boards configured for 18-channel control. The embedded firmware performed non-linear mapping between PWM signals and rotation angles and employed ID-based servo addressing to coordinate multi-actuator movements. During the testing phase, HDR images were captured for each detected head field of view identified by the perception module. These HDR datasets, corresponding to actual occupant gaze directions, were then used to calculate DGP and evaluate the façade's adaptive control performance under real daylighting conditions, as shown in Table 1.

Table 1. Pseudocode for gaze to DGP to angle computation

Control logic of adaptive façade system
INPUT: Illuminance (lux), DGP
OUTPUT: Facade transparency level, rotation command ▫ i
BEGIN
Measure Illuminance, DGP
IF Illuminance ≥ 3000
→ Reduce transparency significantly near work zones
→ Lower indoor light to prevent discomfort
ELSE IF 750 ≤ Illuminance < 2000
→ Apply minor adjustments
→ Maintain comfort and efficiency
ELSE IF Illuminance < 300
→ Increase transparency
→ Enhance brightness and reduce artificial lighting
ELSE IF DGP ≥ 0.45
→ Strongly reduce transparency in glare zones
→ Ensure comfort and safety
ELSE IF 0.35 ≤ DGP < 0.40
→ Moderately reduce transparency
→ Balance daylight and glare control
END IF
END

3.4. EXPERIMENTAL DESIGN

Two scale models, a vision-driven adaptive façade model and a non-façade control model, were constructed and installed in the same south-facing office room in Harbin. The models were positioned adjacent to the window to ensure comparable daylight exposure, and sufficient spacing was maintained to prevent mutual shading. In the experimental model, a fisheye camera mounted on a two-axis gimbal was used to replicate the occupant's real viewing perspective. The gimbal was programmed to follow the ten most frequent head-pose directions identified from recorded office activity videos at three-minute intervals, representing typical gaze behaviour throughout the day. These head-pose sequences were used to control the adaptive façade at three representative time points (09:00, 12:00, and 15:00) corresponding to distinct solar positions and daylight conditions. Both models were operated simultaneously on December 22, a clear winter day, from 09:00 to 15:00, as shown in

Table 2. During each two-minute control cycle, the adaptive façade updated its configuration according to the latest head-pose input, after which illuminance measurements and HDR images were synchronously collected for quantitative glare evaluation.

Table 2. Experimental conditions and timeline

Parameter	Specification	Description
Location	Harbin, China	Real south-facing office room on the 4th floor used for field testing under cold-climate daylight conditions.
Date and Time	December 22, 09:00-15:00	Representative clear winter day capturing variations in solar altitude and azimuth.
Test Models	(1) Vision-driven adaptive facade Both scale models installed side-by-side near the model (2) Non-facade	Both scale models installed side-by-side near the window, spacing maintained to avoid mutual shading.
Control Input	Occupant head-pose sequences	Top-10 most frequent directions extracted from office activity videos at 3-minute intervals.
Sensing Setup	Fisheye camera + two-axis gimbal	Simulated occupant's real viewing perspective and synchronized facade response.
Control Cycle	2-minute interval	Each cycle included facade adjustment, illuminance recording, and HDR image capture
Measurement Metrics	Illuminance, HDR imagery, DGP	Used for quantitative glare evaluation and comparison between adaptive and control models.

3.5. DATA ACQUISITION & METRICS

Illuminance and glare were evaluated in both scale models using synchronized photometric and visual sensing systems. Illuminance was measured at seven points on the scaled work plane, forming a 70 mm × 70 mm sampling grid corresponding to full-scale office conditions. To minimize random noise, each measurement was averaged over three consecutive readings. For DGP assessment, each model incorporated a USB fisheye camera mounted on a two-axis gimbal. A custom Python routine synchronously captured low dynamic range (LDR) sequences, merged them into HDR composites, and computed DGP values for each control cycle.

4. Results

4.1. HOURLY ANALYSIS RESULTS

The hourly variation of DGP and average illuminance was analyzed for both the dynamic façade and the static reference models are shown in Figure 4. The dynamic system maintained DGP values below the comfort threshold of 0.35 for most cycles, while the static model exceeded 0.9 during morning hours. This indicates that the vision-driven control strategy effectively reduced glare within the occupant's field of view. At 09:00, when glare was most critical due to low solar altitude, the dynamic façade achieved a substantially lower DGP level, demonstrating its effectiveness in mitigating both direct and diffuse glare under winter conditions.

In terms of illuminance, the dynamic façade exhibited smoother variations and reduced peak levels, from approximately 10,000 lx to 8,000 lx, with a peak reduction of approximately 2,000 lx during midday compared to the static case. These results confirm that the adaptive façade effectively balances daylight transmission and glare

prevention, achieving a comfortable luminous environment throughout the day.

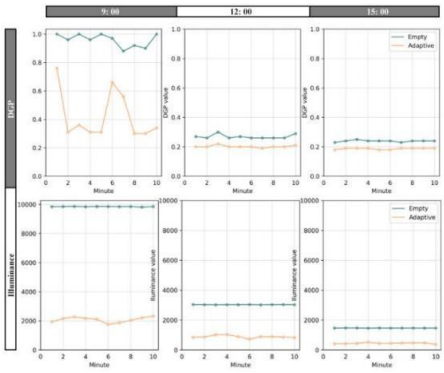


Figure 4. Hourly variation of DGP and average illuminance for the dynamic and static façade mode, adapted from the author's master's thesis research on architecture adaptive façade system for the office building unit based on glare perception

4.2. GLARE OCCURRENCE AND ILLUMINANCE DISTRIBUTION

The experiment maintained consistent occupant behavior and visual field conditions, ensuring reliable comparison between the dynamic façade and the static reference. Using a comfort threshold of DGP=0.35, glare levels below this value were considered imperceptible. Under this criterion, the dynamic façade consistently outperformed the static configuration, effectively reducing uncomfortable glare within the user's field of view. The proportion of intolerable glare was reduced from 33.33% in the static reference to 13.33% under the adaptive condition (-20 percentage points), confirming a clear improvement in visual comfort. Illuminance distribution analysis also indicated greater uniformity: 63.3% of measurements fell within 300–750 lx and 36.7% within 750–2000 lx, verifying the effectiveness of the adaptive control strategy in achieving a balanced and comfortable daylight environment, as shown in Figure 5.

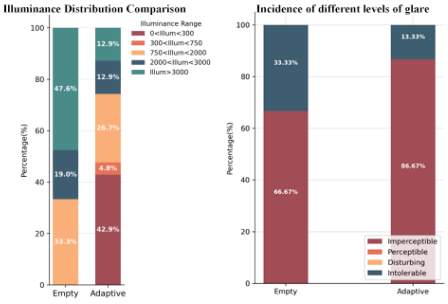


Figure 5. Distribution of illuminance ranges under dynamic and static façade conditions, adapted from the author's master's thesis research on architecture adaptive façade system for the office building unit based on glare perception

5. Conclusion

This study examines how real-time visual information can be used to operate an

adaptive façade in a controlled case study conducted in a south-facing office in Harbin. The prototype relied on measured head-pose-derived visual fields, HDR-based DGP evaluation, and physical actuation of façade modules during on-site testing. Across all experimental conditions, the dynamically controlled façade maintained DGP below 0.35 for most cycles.

Rather than proposing a generalized façade control framework, this work demonstrates that visually informed actuation can meaningfully improve glare mitigation and daylight distribution under real daylight conditions. Future work will extend this case study by testing the system at full scale, incorporating multiple occupants, and evaluating performance across different seasons, orientations, and climatic contexts.

6. Discussion

The results demonstrate that integrating real-time visual perception substantially enhances the adaptivity and human-centred responsiveness of façade control systems. Unlike conventional predictive or rule-based strategies that rely primarily on environmental parameters such as solar position or outdoor illuminance, the proposed approach directly incorporates occupants' visual perception into the control loop. By combining subjective visual feedback with objective glare evaluation, the system is able to respond dynamically to actual visual discomfort within the occupant's field of view rather than to assumed uniform comfort thresholds.

Compared with commonly adopted rule-based solar-angle or illuminance-threshold control methods, the vision-driven strategy introduces an additional layer of perceptual awareness. Façade responses are adjusted according to where occupants are looking and how glare is perceived within the visual cone, which helps explain the observed reduction in intolerable glare while maintaining a more balanced indoor illuminance distribution without excessive daylight obstruction. This perception-aware mechanism enables more nuanced and context-sensitive control than environment-driven approaches.

Scaling the system from a laboratory prototype to the field presents practical challenges. Large façade areas would require hierarchical zoning strategies to balance actuator density and control latency. In addition, long-term operation would need to address issues such as communication robustness, calibration drift, and mechanical wear through modular design and routine maintenance.

Multi-occupant environments introduce further complexity due to potentially conflicting gaze directions and visual preferences. While not implemented in this study, future extensions could aggregate multiple gaze inputs into zone-level comfort objectives using weighted or priority-based schemes. Overall, the findings indicate a shift from environment-driven façade control toward perception-centred adaptive systems, highlighting the potential of combining computer vision and building physics to achieve more responsive, occupant-aligned environmental control.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 52422801).

References

Abade, B., Perez Abreu, D., & Curado, M. (2018). A Non-Intrusive Approach for Indoor Occupancy Detection in Smart Environments. *Sensors*, 18(11), 3953. https://doi.org/10.3390/s18113953

Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr42600.2020.00525

Gordo-Gregorio, P., Alavi, H., Edwards, D. J., Forcada, N., & Guéna, F. (2025). An occupant-centric approach on digital twins for building management. *Building Research & Information*, 1–20. https://doi.org/10.1080/09613218.2025.2463676

Hempel, T., Abdelrahman, A. A., & Ayoub Al-Hamadi. (2022). 6d Rotation Representation For Unconstrained Head Pose Estimation. *2022 IEEE International Conference on Image Processing (ICIP)*. https://doi.org/10.1109/icip46576.2022.9897219

Hosseini, S. M., Mohammadi, M., & Guerra-Santin, O. (2019). Interactive kinetic façade: Improving visual comfort based on dynamic daylight and occupant's positions by 2D and 3D shape changes. *Building and Environment*, 165, 106396. https://doi.org/10.1016/j.buildenv.2019.106396

Ji, W., Lu Feng Yang, Liu, Z., & Feng, S. (2023). A Systematic Review of Sensing Technology in Human-Building Interaction Research. *Buildings*, 13(3), 691–691. https://doi.org/10.3390/buildings13030691

Lee, D., Cho, Y.-H., & Jo, J.-H. (2021). Assessment of control strategy of adaptive façades for heating, cooling, lighting energy conservation and glare prevention. *Energy and Buildings*, 235, 110739. https://doi.org/10.1016/j.enbuild.2021.110739

Li, Y., Chen, J., Ma, J., Wang, X., & Zhang, W. (2023). Gaze Estimation Based on Convolutional Structure and Sliding Window-Based Attention Mechanism. *Sensors*, 23(13), 6226–6226. https://doi.org/10.3390/s23136226

Shen, J., Stefanos Zafeiriou, Chrysos, G. G., Kossaifi, J., Georgios Tzimiropoulos, & Maja Pantic. (2015). The First Facial Landmark Tracking in-the-Wild Challenge: Benchmark and Results. International Conference on Computer Vision. https://doi.org/10.1109/iccvw.2015.132

Tabadkani, A., Roetzel, A., Li, H. X., & Tsangrassoulis, A. (2021). A review of occupant-centric control strategies for adaptive façades. *Automation in Construction*, 122, 103464. https://doi.org/10.1016/j.autcon.2020.103464

Wen, S., Xue, J., Hua, G., & Lai, D. (2024). Reconstructing fisheye luminance maps with a two-step network from a single low dynamic range image. *Automation in Construction*, 160, 105294. https://doi.org/10.1016/j.autcon.2024.105294

Wienold, J., & Christoffersen, J. (2006). Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras. Energy and Buildings, 38(7), 743–757. https://doi.org/10.1016/j.enbuild.2006.03.017

Zabin, A., González, V. A., Zou, Y., & Amor, R. (2022). Applications of machine learning to BIM: A systematic literature review. *Advanced Engineering Informatics*, 51, 101474. https://doi.org/10.1016/j.aei.2021.101474

Zhu, X., Lei, Z., Liu, X., Shi, H., & Li, S. Z. (2016). Face Alignment Across Large Poses: A 3D Solution. Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.2016.23

Zhang, D. (2024). Research on Architecture Adaptive Façade System for the Office Building Unit Based on Glare Perception. Master's thesis, Harbin Institute of Technology, Harbin, China. (in Chinese)

Embodied Navigation and HCI

Interactive Multi-Modal Sensory Open-Plan Workspaces

A Methodological Pilot Study in Virtual Open-Plan Offices

Yuqing Du[1], Anastasia Globa[2] and Thomas Parkinson[3]
[1,2,3]*The University of Sydney.*
[1]*yuqing.du@sydney.edu.au, 0000-0001-8196-0246*
[2]*anastasia.globa@sydney.edu.au, 0000-0002-4749-5675*
[3]*thomas.parkinson@sydney.edu.au,0000-0002-0088-8754*

Abstract. Designing indoor spaces that support comfort and productivity is a central goal in architecture, especially in open-plan offices where sights and sounds strongly influence how people feel and perform. While design has conventionally focused on visual modalities, Virtual Reality (VR) technology lets us economically create and test a wider range of sensory experiences during pre-occupancy stages. Our VR experiment allows participants to adjust their surroundings for greater sensory comfort and personal wellbeing. By collecting users' subjective feedback and objective biosensing features, we explore how empowering users to shape their own workspace in virtual environment can support comfort, and engagement. This work also validates a previously proposed data collection synchronisation workflow, and the findings indicate that several items of subjective ratings had marginal differences between Visual and Visual + Acoustic. Participants also provided how experimental settings can be improved. These insights inform the iterative advancement of integrating VR technologies with built environment studies.

Keywords. Virtual Reality, sensory comfort, multisensory modalities, open-plan workstations, biosensor

1. Introduction

The design and performance of indoor environments significantly influence human well-being and productivity, making comfort investigation in these spaces a key area of research (Latini et al., 2023). Focusing on office environments, design settings have a major impact on employees' daily comfort, and work efficiency (Du et al., 2025a; Yildirim et al., 2024). Open-plan offices in the research context are important due to their collaborative benefits, yet they often compromise individual comfort. Such layouts can foster communication but also increase noise and visual distraction, leading to lower performance (Abdalrahman & Galbrun, 2020). An analysis of satisfaction surveys and found that (a) acoustics were the most common source of dissatisfaction, and (b) occupants of open-plan offices were approximately twice as likely to be

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 175-184.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

dissatisfied compared to occupants in private offices (Parkinson et al., 2023).

Employees can rarely customise their environment, whereas multimodal sensory design can improve comfort in shared offices, yet architectural practice still focuses mainly on the visual and overlooks other senses (Du et al., 2025a; Heydarian et al., 2015). Recent research is shifting toward engaging multiple senses (Lyu et al., 2022; Yildirim et al., 2024). Evaluating multi-sensory adjustments in real offices poses practical challenges, so fully immersive Virtual Reality (VR) has emerged as a powerful platform for user-centred evaluation of pre-occupancy workspace designs (Du et al., 2025a; Heydarian et al., 2015). Beyond immersive visualisation, the technology enables real-time, user-driven testing of design options that would be difficult and costly to replicate physically (Turner et al., 2016). Integrating multi-modal sensory simulations in VR can change design features, such as interior materials, background noises, and sun exposure (Liu et al., 2023; Lyu et al., 2022), facilitating understanding of users' perceptual pleasantness. Furthermore, the sensory stimuli generated by VR can increase user engagement in the experimental environment more effectively than real-world simulations (Fathy et al., 2023; Lyu et al., 2022).

To understand the impact of sensory adjustments in fully immersive VR, a mixed approach was employed to include subjective evaluations, cognitive tests, and biosensors (Yildirim et al., 2024). Subjective perceptions are typically assessed through self-report surveys and interviews, which provide insights into user comfort, engagement, and overall experience (Globa et al., 2022; Latini et al., 2023). In parallel, diverse cognitive tests offer objective measures of task performance, attention, and cognitive function under different sensory conditions (Du et al., 2025b; Latini et al., 2023). To further complement these data, biosensors can capture real-time indicators of emotional arousal, VR experimental engagement, and cognitive load (Du et al., 2024; Heydarian et al., 2015). Many studies now employ the Lab Streaming Layer (LSL) protocol to facilitate synchronised collection of biosensing signals across different devices (Kothe et al., 2019). Two LSL workflows for synchronising multiple streams were proposed to reduce experimenters' disturbance in VR studies (Du et al., 2025a). The workflows can integrate VR experience with the measurement of physiological performance.

Despite advancements in VR-based sensory investigation in office environments, there are still opportunities to extend the research scope. First, fully immersive VR studies with multi-domain sensory adjustments have addressed personal workstation design in semi-outdoor spaces (Lyu et al., 2022) or private offices (Latini et al., 2021; Yildirim et al., 2024). Few have explored how individual sensory optimisations are perceived in shared workspaces (Du et al., 2025b; Yildirim et al., 2024). Second, the integration of VR headsets with multiple biosensors raises questions about wearing comfort and the potential for experimental disturbances (Shui et al., 2025). Prior studies found that different brands and wearing styles of VR headsets affect weight balance on the head (Du et al., 2024). However, no study has examined the weight pressure of combining VR headsets with biosensors. Although lighter and more wearable devices were developed to record biosensing metrics (e.g., Emotiv) (Duvinage et al., 2013), previous studies have only evaluated the practical challenges of sensor wearability without a VR headset (Shui et al., 2025). Additionally, many studies have conducted cognition test outside the virtual environment, few have explored how users perceive

the connections between their VR experiences and real-world tests (Latini et al., 2023).

To address these gaps, this study developed a fully immersive VR open-plan office environment with multi-modal sensory adjustments for individual workstations. Biosensing measurements and cognition tests were integrated in the experiment. A mixed-methods approach was adopted, combining a controlled experiment (quantitative) with post-experimental interviews (qualitative). Here we report only subjective and experiential findings from surveys and interviews; detailed analysis of the biosensing insights will be addressed in another paper. The quantitative part examined the first research question (RQ1): how do participants perceive the interactive VR prototype with adjustable sensory conditions and multi-domain sensory stimuli? The study compared two VR conditions: Visual Adjustment (V) and Visual + Acoustic Adjustment (VA). Building on insights from earlier studies where multi-sensory enhancements improved occupants' subjective experience in office environments (Yildirim et al., 2024), we hypothesise that VA will result in higher ratings than V. A semi-structured interview was employed to explore participants' experiences with the immersive VR environment and the experimental settings, responding to the second research question (RQ2): how do participants perceive the comfort of wearing a VR headset and multiple biosensors? and the third research question (RQ3): how do participants assess the relevance of real-world cognition tests after the immersive VR conditions?

2. Methodology

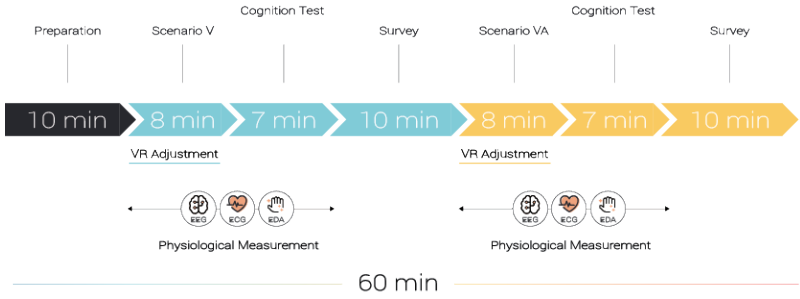


Figure 1. Experiment Workflow

This study implemented Path B of Lab Streaming Layer (LSL) workflows which were proposed for synchronising biosensing and behavioural signals (Du et al., 2025a). Path B uses graphical user interfaces, making deployment easier for standard experimental needs. Path B is suitable for rapid prototyping when common signals such as electroencephalography (EEG), galvanic skin response (GSR), and photoplethysmography (PPG) are adequate for assessing users' internal states (Chiossi, Ou, et al., 2023; Du et al., 2025a). Seven volunteers participated in a 60-minute user study in February 2025. During the preparation phase, participants signed consent forms. They briefly used both the *Oculus Quest 2* and *Pico 4* during a one-minute interaction tutorial of manipulating VR controllers, which included an interactive task adjusting screen brightness of a virtual monitor. After this trial, participants could select their preferred headset. This choice allowed them to consider the different comfort

features of each headset. Two VR scenarios include Visual Adjustment (V) and Visual + Acoustic Adjustment (VA). Each VR scenario comprised 1 minute to familiarise, 6 minutes to adjust, and 1-minute immersion in the optimised workspace. Following, a 7-minute cognition test in the real world and a 10-minute survey (see Figure 1). In Scenario V, participants adjusted only visual elements. In Scenario VA, they adjusted both visual and acoustic elements. Throughout, biosensing signals, EEG, PPG, and GSR, were synchronised via LSL to index arousal and stress. Surveys captured Sensory Pleasantness, Mood States, Adjustment Function, and VR Engagement. After both rounds, a 5-minute semi-structured interview gathered reflections on the overall experience, including sensory pleasantness and headset wearability.

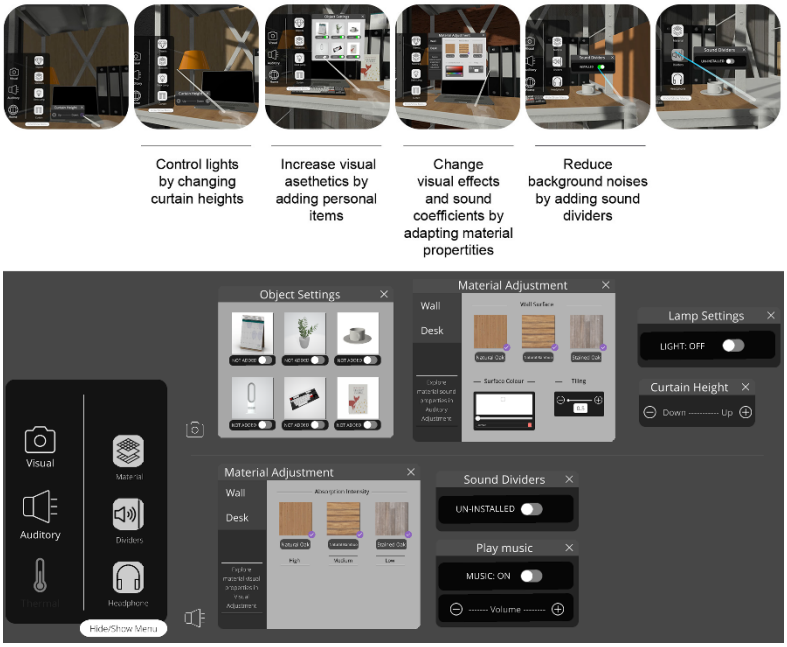


Figure 2. VR Interface for Sensory Adjustment.

2.1. SENSORY ADJUSTMENT

The fully immersive VR environment was developed using *Unity 2022.3.2* with an open-plan office asset package *London Office by* Oneiros. Both visual and acoustic adjustments were designed to reflect everyday strategies that individuals use to enhance sensory comfort at their own workstations (see Figure 2). Ambient audio combined typical office sounds with environmental noise, including employee conversations, printer noises, and telephone rings. In Scenario V, participants could set the curtain height to modulate daylight and toggle the desk lamp (see Figure 2). Wall materials offered three reflectance levels: natural oak (high), bamboo (medium), and stained oak (low), and desktops offered glass (high), wood grain laminate (medium), and rubber desk mat (low). Users can also add objects such as a calendar, a Dyson fan, a gaming keyboard, plants, a mug, and a book. In Scenario VA, besides visual adjustment,

changes to desk and wall materials were programmed to influence ambient noise by simulating sound absorption coefficients of 0.3, 0.6, and 0.9 for low, medium, and high background noise absorption. These values may overstate real-world differences, but they ensured that participants could clearly perceive how visual changes affected the acoustic environment. Beyond materials, users could add a sound divider for an additional 40 percent reduction in background noise. Participants could also play music, and adjust volume with a slider.

2.2. BIOSENSING MEASUREMENTS

This pilot study tested the proposed LSL workflow and suggested devices for real-time physiological data collection and less experimenter manipulations across multiple devices (Du et al., 2025a). The setup included VR headsets, the *Emotiv MN8* for measuring EEG, and the *Shimmer GSR+* sensor for measuring SCL and PPG for calculating heart rates. Details of the adopted device are in Du et al., 2025a. Biosensing signals were recorded continuously during V and VA.

2.3. COGNITION TESTS

Cognitive performance was assessed using two tasks delivered via Millisecond Inquisit 7: the Visual Simon Task and the Emotional Stroop Task. The Visual Simon Task tests response inhibition by requiring colour-based responses, capturing the Simon effect as the reaction time gap between congruent and incongruent trials. The Emotional Stroop Task measures attentional control by naming the colour of emotionally charged words; slower times or more errors indicate interference. Both tests were run outside VR.

2.4. SURVEY

After each cognition test, participants completed a survey on their workplace experience. Sensory Pleasantness (visual, acoustic, overall) used a seven-point Likert scale (very unpleasant to very pleasant). Mood used a nine-point bipolar scale with ten items spanning positive and negative affect (e.g., depressed–elated, anxious–relaxed) (Lyu et al., 2022; Yildirim et al., 2024). Adjustment Function used a seven-point scale to rate intuitiveness, efficiency, effectiveness, adaptability, satisfaction, and sustainability (Li et al., 2022; Ntoa et al., 2021). VR Engagement used a five-point scale to assess sensory involvement, movement, overall involvement, and enjoyment.

To compare participants' subjective ratings between V and VA, the Wilcoxon signed-rank test was employed for each survey dimension. All statistical analyses were performed using the *scipy* package (version 1.15.2) in the Python environment (version 3.12.7). This test is non-parametric, making it particularly suitable for within-subjects designs involving paired ratings, especially with small sample sizes ($N = 7$) and ordinal data such as Likert scale responses, where normality cannot be assumed. In addition to the inferential analysis, descriptive statistics were calculated for each condition to help illustrate the direction and magnitude of any observed differences. The non-parametric test is well suited for situations involving paired ordinal data or small sample sizes, as it makes no assumptions about data normality (Divine et al., 2013). Effect sizes (r) were calculated for each comparison to quantify the magnitude of observed differences. The r values are interpreted as small ($r \approx 0.1$), medium ($r \approx 0.3$), or large ($r \geq 0.5$) (Cohen,

2013). We treated results as statistically significant when $p < 0.05$ (Cohen, 2013; Divine et al., 2013). Given the small, exploratory design, we also flagged p-values between 0.05 and 0.10 as marginal, indicating possible effects that merit follow-up in larger samples (Cohen, 2013).

2.5. INTERVIEW

At the end of the experiment, participants completed a semi-structured interview to capture detailed feedback on adjustable sensory conditions, VR interface usability, and the impact of wearing multiple biosensors during immersive tasks. The interview questions in this study explored:

a. How do you think the multi-modal VR approach compared to experience the simulated adjustments in the real world?

b. How did wearing a VR headset and biosensors affect your sense of VR immersion?

c. Do you think the different VR scenario adjustments influenced your cognitive performance during the tests?

Sessions ran in English or Chinese. Recordings were transcribed and imported into Dovetail, a platform for multilingual processing and cross-transcript labelling.

3. Results

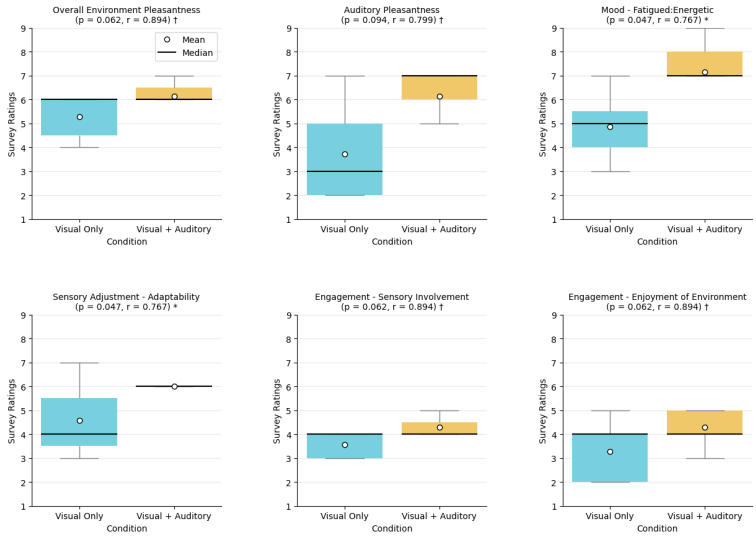


Figure 3. Survey items with significant and marginal rating differences between V and VA.

21 items were evaluated in Visual (V) and Visual + Acoustic (VA). Figure 3 shows boxplots with whiskers for minimum and maximum, a box for the interquartile range, a black line for the median, and a white circle for the mean. Two items differed significantly between conditions (Wilcoxon signed rank, $p < 0.05$, *), and three showed

marginal differences (p from 0.05 to 0.10, †).

Figure 3 shows marginal differences between V and VA for Overall Environment Pleasantness ($p = 0.062$, $r = 0.894$) and Auditory Pleasantness ($p = 0.094$, $r = 0.799$), with V yielding higher, more concentrated ratings, especially for overall environment pleasantness; the other Sensory Pleasantness items followed the same direction with higher means in VA. For Mood States, only the fatigued to energetic dimension differed significantly ($p = 0.047$, $r = 0.767$), with VA ratings higher and tighter, and both the minimum and maximum above VA; other mood items again showed higher means in VA. Within Adjustment Function, only Adaptability was significant ($p = 0.047$, $r = 0.767$), with VA both higher and highly consistent, whereas V was lower and more variable; other items showed higher means in VA. For VR Engagement, Sensory Involvement and Enjoyment of Environment were marginal ($p = 0.062$, $r = 0.894$ for both), each higher and more concentrated in VA.

3.1. INTERVIEW

According to interview question a, participants reported greater control and convenience in VR, with instant adjustments to lighting, sound, and materials. However, missing workplace cues, an inactive monitor, no visible colleagues, and limited screen interaction, reduced realism. They suggested a working monitor, and ambient social presence. Besides the virtual adjustment for controlling sensory experiences, most participants found the virtual environment with interactive sensory stimuli more engaging than a real office. However, they asked for visual guidance, and hand interaction with virtual objects.



Figure 4. Participant wearing Oculus Quest 2 (left); participant wearing PICO 4 (right).

Considering interview question b, only one participant chose *Oculus Quest 2*, citing minimal light leakage from the bottom edge of the headset. The remaining six participants preferred the *PICO 4* primarily due to its greater comfort and ergonomic advantages. Specifically, participants reported that *PICO 4's* rigid plastic head strap effectively distributed the device's weight and alleviated pressure compared to the soft belt design of the Oculus Quest 2 (see Figure 4). Participants said wearing the MN8 EEG and Shimmer GSR+ caused little loss of immersion. The main disruption was the technical setup: repeated signal checks and calibration before each session, especially when long hair or unstable sensor contact made detection harder. Two participants with longer hair were required to restart their sessions because of unstable biosensing signals. Several participants remarked that cognition tests increased their sense of involvement and provided a meaningful experimental structure.

In terms of question c, several participants described the out-of-VR cognition tests

are similar to game-like challenges. They suggested integrating work-related tasks directly within their personalised VR workstation

4. Discussion

To address RQ1, most of the survey results align with previous research. Participants expressed appreciation for the high degree of control and flexibility enabled by multi-modal VR adjustments. Similar to recent studies demonstrating the value of interactive design research in VR (Lyu et al., 2022; Yildirim et al., 2024), our results show that being able to dynamically manipulate lighting, materials, and soundscapes in VR provided better sensory engagement. Previous studies have demonstrated that multi-modal sensory adjustments in VR, particularly in semi-outdoor spaces (Lyu et al., 2022) or private offices (Latini et al., 2023; Yildirim et al., 2024), can produce significant improvements in pleasantness and mood states. In contrast, this pilot study indicates that the Visual + Acoustic (VA) condition showed only marginal increases in overall environmental and auditory pleasantness, with clearer trends for improved mood and adaptability, relative to the Visual-only (V) condition. These patterns tentatively support higher pleasantness, function and engagement under VA. Given the use of non-parametric analyses, the results are indicative only and not statistically robust. To clarify these effects, larger and more diverse samples are needed. Moreover, participants' qualitative feedback noted that the absence of several sensory elements reduces the sense of authenticity despite the successful simulation of visual and acoustic design adjustments. Their suggestion of new sensory stimuli, such as monitor display and virtual colleagues, will be considered in the next VR prototype.

To address RQ2, this study reveals a clear preference for the PICO 4 over Oculus Quest 2. This finding aligns with recent research: for longer sessions, users prefer hard-plastic head-strap headsets for comfort (Du et al., 2024). In terms of adding biosensors to the head, although previous research has raised concerns about the negative impact on user experience in VR (Setz et al., 2010; Shui et al., 2025), most participants in this study experienced minimal disruption. On the other hand, the study reveals that LSL-based workflow reduces direct experimenter intervention and operational disturbance (Du et al., 2025a; Kothe et al., 2019). By enabling simultaneous control of multiple biosensors with a single click, the unified management of sensors reduces setup friction during initial signal checking. However, as EEG signals of MN8 remain low quality which is sensitive to hair lengths and physical movement, future protocols should guide participants to move smoothly to maintain signal integrity without disrupting the humanistic experience. In addition, while many studies have focused on improving physical comfort and reducing device size for integrating measurement devices with VR headsets (Setz et al., 2010), our interview results indicate that highly stable biosensors should be integrated for seamless user experience.

To address RQ3, participants reported that finishing cognition tests in the real world reduced cybersickness and helped maintain engagement, supporting breaks needed between VR exposures (Lyu et al., 2022). Further, the 8-minute VR scenario and 10-minute survey were comparatively long, with many participants finishing sooner. Future work should test shorter exposures or add simple in-VR tasks to sustain engagement. Qualitative feedback showed some did not link out-of-VR test performance to in-VR sensory changes, more research should clarify whether the test

results are indicators that can reflect the influence under different sensory conditions.

Besides the findings responding to the RQs, we note several limitations and directions for improvement. The small sample reduces generalisability; future work should recruit larger, more cohorts based on gender or cultural diversity. Although an LSL workflow reduced researcher intervention, frequent calibration and signal checks broke immersion; more robust devices are needed. Limiting adjustments to two sensory modalities gives only a partial view of open plan perception. Next iterations should shape group comparisons of multisensory settings with interaction design, add modalities such as thermal/smell cues to improve comfort and engagement. While this pilot focused on sensory agency, participant interviews revealed that achieving humanistic realism requires broader environmental variables. Future iterations of this methodology should incorporate low-disturbance texture guidance, virtual colleagues, and spatial arrangement to simulate social density, and context-relevant tasks that more accurately reflect authentic workplace behaviour.

5. Conclusions

This study examined adjustable sensory conditions, headset and sensor wearability, and cognition testing in a fully immersive VR open workspace. Participants valued the control and flexibility of multi-modal adjustments, with Visual + Acoustic tending to score higher across most measures; however, with a small sample and non-parametric tests, only 5 of 21 items showed notable differences. Participants preferred PICO 4 to Oculus Quest. Biosensors had little impact on immersion, but setup and signal stability issues sometimes disrupted sessions, highlighting the need for more robust biosensing systems. Running cognitive tests outside VR reduced cybersickness and maintained engagement, though analysing task-evoked behaviour within VR may yield richer insights. Implementing a proposed workflow in fully immersive VR with multiple biosensors and lower experimenters' disturbance, this study contributes practical experience for integrating adjustable sensory settings, interaction logging, and biosensing capture. The quantitative and qualitative findings support the fully immersive VR with adjustable multi-modal sensory settings for improving open-plan workspace design and outline priorities for future work.

References

Abdalrahman, Z., & Galbrun, L. (2020). Audio-visual preferences, perception, and use of water features in open-plan offices. The Journal of the Acoustical Society of America, 147(3), 1661–1672. https://doi.org/10.1121/10.0000892,

Chiossi, F., Ou, C., Gerhardt, C., Putze, F., & Mayer, S. (2023). Designing and evaluating an adaptive virtual reality system using EEG frequencies to balance internal and external attention states. https://doi.org/10.1016/j.ijhcs.2024.103433

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Statistical Power Analysis for the Behavioral Sciences. https://doi.org/10.4324/9780203771587

Divine, G., Norton, H. J., Hunt, R., & Dienemann, J. (2013). A review of analysis and sample size calculation considerations for Wilcoxon Tests. Anesthesia & Analgesia, 117(3).

Du, Y., Globa, A., & Brambilla, A. (2024). Understanding VR-based investigations for perceptual experiences - A systematic literature review on VR-based indoor comfort empirical studies. SIGraDi 2024 - Biodigital Intelligent Systems Conference, (pp. 1977–1988). https://doi.org/10.52842/conf.sigradi.2024.1977

Du, Y., Globa, A., & Brambilla, A. (2025a). Developing a lab streaming layer workflow for coordinating physiological data collection in virtual reality: A case study on multisensory comfort in pre-occupancy office environments. ARCHITECTURAL INFORMATICS - Proceedings of the 30th CAADRIA Conference (pp. 571–580). https://doi.org/10.52842/conf.caadria.2025.4.571

Du, Y., Globa, A., & Brambilla, A. (2025b). Towards holistic multisensory fully immersive virtual reality prototypes in architectural design: An analytical hierarchy process-based systematic review. Architectural Engineering and Design Management, 1–25. https://doi.org/10.1080/17452007.2025.2549494

Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., & Dutoit, T. (2013). Performance of the Emotiv Epoc headset for P300-based applications.

Fathy, F., Mansour, Y., Sabry, H., Refat, M., & Wagdy, A. (2023). Virtual reality and machine learning for predicting visual attention in a daylit exhibition space: A proof of concept. Ain Shams Engineering Journal, 14(6), 102098.

Globa, A., Wang, R., Tokede, O., & Khoo, C. K. (2022). Pre-Occupancy evaluation of buildings in VR: development of the prototype and user studies. Architectural Science Review, 65(5), 407–419. https://doi.org/10.1080/00038628.2022.2123775

Han, K.-T. (2003). A reliable and valid self-rating measure of the restorative quality of natural environments. Landscape and Urban Planning, 64(4), 209–232.

Heydarian, A., Carneiro, J. P., Gerber, D., & Becerik-gerber, B. (2015). Immersive virtual environments, understanding the impact of design features and occupant choice upon lighting for building performance. Building and Environment, 89(2015), 217–228.

Kamińska, D., Smółka, K., & Zwoliński, G. (2021). Detection of mental stress through EEG signal in virtual reality environment. Electronics (Switzerland), 10(22).

Kothe, C., Medine, D., Boulay, C., Grivich, M., & Stenner, T. (2019). What is LSL? Labstreaminglayer. https://labstreaminglayer.readthedocs.io/info/intro.html

Latini, A., Di Loreto, S., Di Giuseppe, E., D'Orazio, M., & Di Perna, C. (2023). Crossed effect of acoustics on thermal comfort and productivity in workplaces: A case study in virtual reality. Journal of Architectural Engineering, 29(2), 1–10.

Li, W., Zhou, Y., Luo, S., & Dong, Y. (2022). Design factors to improve the consistency and sustainable user experience of responsive interface design. Sustainability, 14(15).

Liu, G., Zou, J., Qiao, M., Zhu, H., Yang, Y., Guan, H., & Hu, S. (2023). Stress recovery at home: Effects of the indoor visual and auditory stimuli in buildings. Building and Environment, 244, 110752. https://doi.org/10.1016/J.BUILDENV.2023.110752

Lyu, K., de Dear, R., Brambilla, A., & Globa, A. (2022). Restorative benefits of semi-outdoor environments at the workplace: Does the thermal realm matter? Building and Environment, 222(April), 109355. https://doi.org/10.1016/j.buildenv.2022.109355

Ntoa, S., Margetis, G., Antona, M., & Stephanidis, C. (2021). User experience evaluation in intelligent environments: A comprehensive framework. Technologies, 9(2).

Parkinson, T., Schiavon, S., Kim, J., & Betti, G. (2023). Common sources of occupant dissatisfaction with workspace environments in 600 office buildings. Buildings and Cities, 4(1), 17–35. https://doi.org/10.5334/bc.274

Shui, X., Xu, H., Tan, S., & Zhang, D. (2025). Depression recognition using daily wearable-derived physiological data. Sensors, 25(2). https://doi.org/10.3390/s25020567

Turner, C. J., Hutabarat, W., Oyekan, J., & Tiwari, A. (2016). Discrete event simulation and virtual reality use in industry: new opportunities and future trends. In IEEE Transactions on Human-Machine Systems (Vol. 46, Issue 6, pp. 882–894). IEEE.

Yildirim, M., Globa, A., Gocer, O., & Brambilla, A. (2024). Multisensory nature exposure in the workplace: Exploring the restorative benefits of smell experiences. Building and Environment, 262, 111841. https://doi.org/10.1016/J.BUILDENV.2024.111841

Merging Graph Theory and VR

Human-Centred Assessment of Fire Evacuation Cognition

Yixuan Wang[1], Marc Aurel Schnabel[2], Rong Fu[3] and Cheng Zhang[4]
[1,2,3,4]*Xi'an Jiaotong-Liverpool University.*
[1]*Yixuan.Wang20@student.xjtlu.edu.cn, 0009-0002-9752-6609*
[2]*MarcAurel.Schnabel@xjtlu.edu.cn, 0000-0002-2923-1609*
[3]*Rong.Fu21@student.xjtlu.edu.cn, 0009-0008-1082-419X*
[4]*Cheng.Zhang@xjtlu.edu.cn, 0000-0003-2465-1767*

Abstract. Conventional fire safety analysis focuses on distances and capacities but offers little insight into how occupants perceive and understand complex layouts under stress. This paper proposes a human-centred computational framework that links building spatial configuration with evacuation cognition. We introduce the Evacuation Cognitive Index (ECI), a composite metric that integrates Spatial Complexity (SC) from Graph Theory with Visual Load (VL) captured through eye tracking in immersive Virtual Reality (VR) evacuation scenarios. A BIM-based model provides the geometric and topological basis for GT analysis, while VR simulations record participants' gaze behaviour, route choice, and evacuation time in a school building under simulated fire conditions. Higher ECI values are associated with denser fixation patterns, increased decision uncertainty, and longer evacuation times, whereas low ECI routes support clearer spatial legibility. ECI thus functions as a cognition-informed design indicator that advances evacuation planning toward humanistic, perception-aware and performance-based building safety.

Keywords. Virtual Reality, Graph Theory, Eye-tracking, Fire Evacuation, Humanistic computation

1. Introduction

The spatial configuration of a building plays a decisive role in determining evacuation safety and efficiency during fire emergencies, as it governs accessibility, route selection, and overall movement performance. Although fire protection regulations such as GB50016-2014 (Ministry of Public Security of the People's Republic of China, 2014) specify quantitative requirements for evacuation distance and exit capacity, they provide limited guidance on qualitative aspects of spatial legibility and circulation clarity. Consequently, designers often rely on subjective judgment to address cognitive factors of wayfinding, such as intuitive navigation, spatial hierarchy, and environmental comprehension.

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 185-194.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

To quantitatively assess how building layouts influence evacuation performance, numerous studies have employed Graph Theory (GT) to represent architectural spaces as topological networks that represent connectivity, accessibility, and circulation efficiency (Zhao et al., 2023). While GT has effectively identified critical nodes and optimal egress routes, most conventional models evaluate evacuation performance mainly in terms of evacuation time. Edge attributes such as visibility, signage, and environmental hazards are typically excluded, limiting their capacity to capture dynamic interactions between human cognition and spatial configuration during emergencies (Donato, 2017).

Recent developments in VR have enabled the study of evacuation behaviour in immersive and controllable environments, providing empirical insights into visual attention, decision-making, and navigation (Bernardini et al., 2023). However, most existing VR studies have primarily focused on external guidance systems, such as exit signage or lighting cues, while the cognitive effects of spatial configuration remain insufficiently explored (Fu et al., 2024a).

This paper proposes a human-centred computational framework that integrates BIM, GT, immersive VR, and eye-tracking to link building spatial configuration with evacuation cognition. It introduces the ECI, combining Spatial Complexity (SC) and Visual Load (VL) into a unified indicator of cognitive difficulty. VR and eye-tracking experiments were employed to empirically validate that higher ECI values are associated with increased gaze effort, decision uncertainty, and longer evacuation times, thereby informing cognition-aware design. By integrating spatial, perceptual, and behavioural dimensions, the framework establishes a cognition-informed and performance-based approach to building safety design, advancing understanding of the interplay between spatial topology and human cognition.

2. Methodology

The methodology integrates BIM, GT, VR and eye-tracking to investigate the cognitive mechanisms of fire evacuation (Figure 1).

BIM provides geometric and semantic information to develop a GT-based spatial network for quantifying SC and connectivity. The derived topological parameters guide the identification of potential evacuation routes and the evaluation of their structural efficiency. These routes are further examined within an immersive VR platform that simulates dynamic fire conditions and environmental cues to reproduce realistic evacuation scenarios. Eye-tracking data collected during VR experiments capture participants' visual attention, enabling real-time measurement of participants' visual attention allocation and route-selection behaviour. A case study conducted on the fourth floor of an engineering building demonstrates the implementation of this framework, from BIM-based graph generation to VR-based behavioural analysis, enabling a comprehensive understanding of how spatial configuration and visual cognition jointly influence evacuation performance. The integrated results are synthesised through the ECI, which quantitatively represents cognitive load under different spatial conditions.

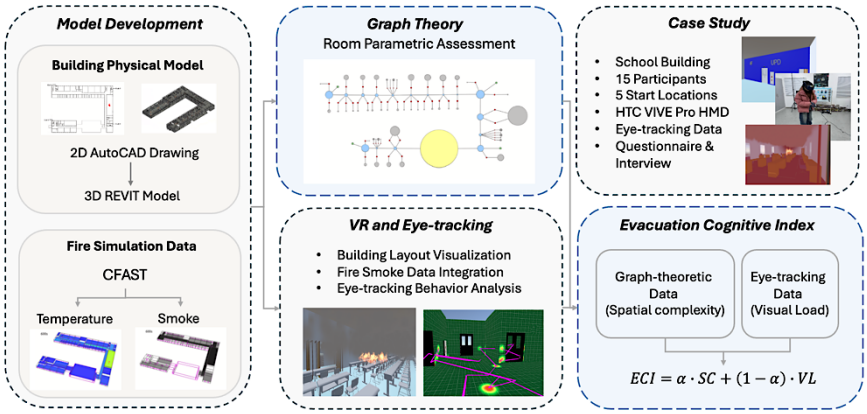


Figure 1. Methodology for cognition-informed fire evacuation analysis

2.1. MODEL DEVELOPMENT

The BIM model was generated from the floor plan to ensure the geometric accuracy of spatial components such as rooms, corridors, and vents, providing a reliable foundation for fire simulation. Fire dynamics were simulated in CFAST to compute the temporal evolution of temperature and smoke dispersion, generating realistic environmental data for behavioural experiments. The layer temperatures and smoke heights of each compartment were sampled at predefined intervals and compiled into structured datasets. These data were then integrated into the VR platform to enhance the visual and thermal realism of the evacuation simulations.

2.2. GRAPH-THEORETIC EVALUATION

The floor plan was abstracted into a graph to capture the building's spatial topology. Fire safety codes were formalised as graph-assembly rules to ensure that the GT network complied with evacuation design principles. To quantitatively assess the floor layout and evacuation performance, several GT parameters were derived from the network's structural and topological properties, including closeness centrality, circulation, connectivity, accessibility, and evacuation cost (Donato, 2017). Higher closeness centrality and connectivity indicate shorter path lengths and improved spatial integration, promoting efficient movement and reduced transition costs. Such configurations improve both routine circulation and emergency evacuation performance. However, excessive parameter distributions may cause congestion, overload critical nodes, and reduced route redundancy, thereby compromising system robustness. These parameters served as the basis for the SC component of the ECI. Table 1 summarises key GT parameters, their mathematical definitions, and their spatial implications. As illustrated in Figure 2, the building's schematic graph was constructed with vertices corresponding to spatial units, doors, transit points, and exits, positioned at their geometric centres and scaled by area. Edges represent the connectivity between nodes, weighted by walking distances. Nodes are colour-coded

to clarify the topological and functional organisation for quantitative analysis.

Table 1. Key GT parameters and their spatial implications

GT Parameter	Spatial Interpretation	Cognitive Implication
Closeness Centrality	The average of the shortest path length from the node to every other node in the graph.	↑ → accessibility ↑
Circulation	The average distance between all pairs of its nodes.	↑ → graph complexity ↑
Connectivity	The ratio of edges to nodes in a graph.	↑ → graph complexity ↑
Accessibility	The total number of nodes in the graph.	↑ → space and route complexity ↑
Evacuation Cost	The total walking distance from the initial point to the exit.	↑ → efficient routes ↓

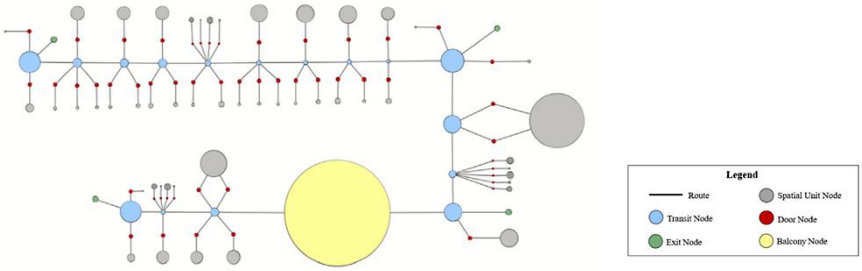


Figure 2. Schematic Graph

2.3. VR ENVIRONMENT AND EYE-TRACKING

The immersive virtual environment was developed in Unity3D using C# scripting, integrating the BIM model with fire simulation data from CFAST. The architectural geometry, fire safety facilities, and evacuation signage were modelled in Autodesk Revit and imported into Unity3D for real-time simulation. Temperature and smoke height outputs from CFAST were mapped to visual cues following established VR fire simulation practices (e.g., colour gradients, visibility reduction), consistent with prior evacuation cognition studies. Dynamic smoke and temperature fields were visualised to represent the temporal evolution of the fire, with colour transitions indicating heat intensity and a particle system rendering smoke dispersion. When local temperature exceeded 26 °C, the field of view shifted toward red hues, and darker tones represented higher thermal intensity.

Participants experienced the simulation through an HTC Vive Pro HMD equipped with Tobii eye-tracking. The system continuously recorded gaze coordinates, fixation counts, durations, and timestamps to capture visual attention and behavioural responses. Eye-tracking metrics, including fixation duration, fixation count, and Areas of Interest, were extracted for quantitative assessment. A customised 'UnityHeatmapShader' visualised gaze data as scan paths and heatmaps, with yellow spheres denoting gaze points, magenta lines representing scan paths, and colour

gradients from green to red indicating increasing attention intensity.

2.4. CASE STUDY

The case study was conducted on the fourth floor of a school building (4,856 m²), comprising 32 offices, 10 washrooms, 5 low-voltage rooms, and 4 exits. The fire ignition point was positioned in the middle room of the right-side corridor. The U-shaped layout posed potential fire safety risks in long corridors and dead-end rooms due to limited escape routes (Fu et al., 2024b). Therefore, five starting locations within these critical areas were selected to evaluate evacuation efficiency. The fire scenario was simulated using CFAST, with ignition in the centre of the computer room (Figure 3). Three workstation panels were defined as combustibles under normal environmental conditions. Simulation results were imported into Unity 3D to visualise dynamic smoke and temperature fields for VR experiments.

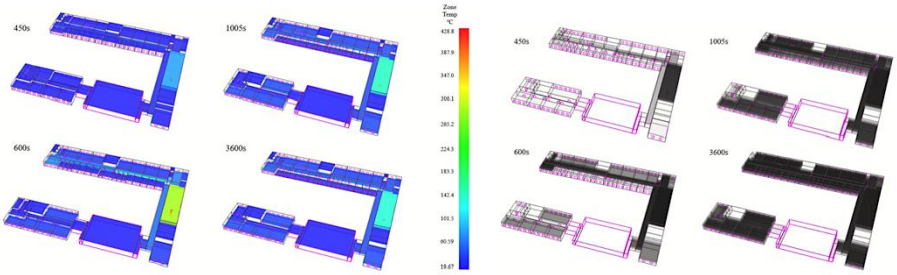


Figure 3. Left: Temperature change with timestamp. Right: Fire smoke distribution with timestamp.

Fifteen participants (10 female, 5 males; aged 20–31) with normal vision were randomly assigned to five groups, each starting from a designated room. After completing the pre-experiment questionnaire on demographics and prior VR experience and receiving brief navigation training, participants were instructed to evacuate the virtual environment via any available exit as efficiently as possible. Eye-tracking data, behavioural timing, and route selection were recorded. Complementary video recordings enabled classification of behaviour types and measurement of their frequency and duration. Following the task, participants completed a post-experiment questionnaire evaluating task difficulty, layout familiarity, exit visibility, and evacuation efficiency on a five-point Likert scale. Semi-structured interviews were subsequently conducted to investigate further participants' perceptions of the building layout and the reasons for their route selection during evacuation.

2.5. EVACUATION COGNITIVE INDEX (ECI)

The ECI integrates spatial and perceptual dimensions of evacuation cognition by combining SC and VL. SC represents the topological difficulty of evacuation routes, derived from normalised GT metrics (Lg, D, β, AC, Cost) and weighted via the entropy method to reflect each variable's information contribution. VL quantifies attentional effort in visual information processing. Both indices were normalised to [0–1] for comparability, with higher values indicating greater cognitive difficulty. The integrated

ECI was then computed as a weighted linear combination:

$$\mathrm{ECI} = \alpha \cdot \mathrm{SC} + (1 - \alpha) \cdot \mathrm{VL} \quad [1]$$

where α represents the relative contribution of spatial configuration to evacuation cognition, although the weighting parameter is partly heuristic, its selection was informed by both theoretical considerations in spatial cognition and empirical validation. Alternative weighting schemes (α = 0.3, 0.5, 0.7) and equal-weight baselines were tested, and correlation analyses revealed that α = 0.6 maximised the correlation between ECI and evacuation time. This indicates that integrating VL with SC significantly improves the index's explanatory power compared to using spatial metrics alone.

3. Results

3.1. SPATIAL ACCESSIBILITY AND ROUTE EVALUATION

This study evaluated 20 evacuation routes to assess spatial configuration and potential evacuation efficiency based on average path length, β-value, accessible space, and evacuation cost. Table 2 presents the average values calculated from 20 evacuation routes as a reference for 8 representative routes. Considerable variation among routes indicated that evacuation performance is highly dependent on the spatial configuration. Routes with lower evacuation costs generally exhibited shorter path lengths, lower topological complexity, and limited connectivity, suggesting that simplified spatial layouts with direct circulation and fewer decision points enhance evacuation efficiency.

Table 2. The characteristics of various evacuation routes

	Vertices	Edges	Density	Diameter	Lg	β	AC	Cost
422-Route1	77	78	0.027	15	6.77	1	31	174.38
436-Route1	111	112	0.018	19	8	1	46	208.85
441-Route1	44	45	0.048	10	4.77	1	17	137.25
441-Route3	11	10	0.182	5	2.56	0.91	2	54.38
461-Route2	4	3	0.25	3	1.67	0.75	1	11.37
461-Route4	93	93	0.022	16	6.65	1	38	147.22
482-Route1	6	5	0.333	4	2.07	0.83	1	9.57
482-Route4	118	119	0.017	20	8.24	1	48	221.77
Average	50	50	0.083	10	4.70	0.97	19.55	98.93

Four representative routes (436-Route4, 441-Route3, 461-Route2, and 482-Route1) exhibited optimal performance across all indicators. These routes featured fewer accessible subspaces along corridors, reducing SC and enhancing directional clarity. In VR experiments, 73.3% participants selected these routes, with an average evacuation time of 46.37s, SD = 5.12s, confirming their superior spatial legibility and navigational efficiency.

Conversely, high-ECI routes generally exhibited longer path lengths, higher accessibility, and increased decision points, leading to extended exploration and route verification (422-Route1, 436-Route1, 441-Route1, 461-Route4, and 482-Route4). Multiple accessible rooms along these paths increased topological complexity, leading to longer exploration behaviour and more frequent route changes during evacuation. Although high accessibility facilitates routine circulation, it becomes a cognitive distractor in emergencies by introducing additional decision points and visual uncertainty. Participants navigating these complex routes spent more time on exploration and route verification, ultimately reducing evacuation performance.

Behavioural data further confirmed this trend: 11.1% participants selected Route 3, characterised by the highest accessibility and complexity, even though it was not the shortest path. Their mean evacuation time (63.43s, SD=6.84s) was 36.8% longer than that of the optimal route. Overall, these findings reveal the duality of spatial accessibility: although it improves flexibility and usability under normal conditions, it can impede evacuation under stress by increasing cognitive demands, as evidenced by significant correlations between ECI and evacuation time ($r = 0.68$, $p = 0.01 < 0.05$), highlighting the necessity of balancing accessibility and legibility through optimised spatial hierarchy and clear circulation cues.

3.2. EVACUATION COGNITIVE INDEX ANALYSIS

Individuals' cognitive load during evacuation is influenced by both SC (e.g., path accessibility and topological configuration) and visual information-processing demands (e.g., signage, lighting, and obstacles). Evaluating these factors separately fails to capture overall cognitive stress and behavioural performance. To address this gap, ECI is used to integrate spatial and visual dimensions into a unified quantitative framework, enabling comparative analysis across building layouts and supporting evidence-based optimisation of evacuation design.

SC was calculated using normalised GT indicators, including average path length (Lg), network diameter (Diameter), complexity coefficient (β), path cost (Cost), number of edges (Edges), network density (Density), and angular connectivity (AC). Entropy-based weighting determined the relative contribution of each parameter (Table 3). Edges, Diameter, Cost, and AC increased SC, whereas Density had a negative effect, indicating that denser networks may reduce spatial differentiation and legibility.

Table 3. Entropy-Based Weighting for SC Calculation

	Lg	Diameter	β	Cost	Edges	Density	AC
Weight	0.159	0.269	0.056	0.202	0.122	0.064	0.129

VL was derived from eye-tracking data collected during VR evacuation experiments using three indicators: fixation count, total fixation duration, and mean fixation duration. According to cognitive resource allocation theory (Holmqvist et al., 2011), total fixation duration is a direct and reliable indicator of the overall cognitive resources allocated to a visual task. Therefore, the total fixation duration was assigned

the highest weight (0.4), with the other two indicators each receiving a weight of 0.3, reflecting their comparatively lower contribution to overall visual load. Both SC and VL were normalised to 0–1, with higher values indicating greater cognitive difficulty. The final ECI was calculated as [1] with $\alpha = 0.6$, emphasising the dominant role of SC in evacuation cognition (Franz and Wiener, 2008). Sensitivity analysis confirmed that this weighting maximised the correlation between ECI and evacuation time. The results of SC, VL, and ECI are summarised in Table 4.

Table 4. Results of SC, VL, and ECI

	SC	VL	ECI
422-Route4	0.464	0.503	0.480
422-Route3	0.468	0.992	0.678
436-Route3	0.523	0.828	0.645
436-Route4	0.375	0.420	0.393
441-Route3	0.403	0.599	0.481
461-Route2	0.269	0.300	0.281
482-Route1	0.292	0.283	0.289

Conversely, 436-Route3 and 422-Route3 exhibited the highest ECI values, accompanied by longer fixation durations and more frequent gaze shifts during evacuation. Introducing VL alone significantly improves SC's explanatory power, confirming that ECI captures evacuation cognition more comprehensively than topological indicators in isolation. These results demonstrate that ECI effectively quantifies the interaction between spatial topology and visual cognition, linking building layout with cognitive load and decision-making efficiency.

3.3. VR BEHAVIOURAL OBSERVATIONS

Eye-tracking data empirically validate the ECI analysis. Heatmaps and scan paths revealed that participants navigating high-ECI routes exhibited denser fixation and longer dwell times at corridor junctions and doors, indicating increased visual attention and decision uncertainty. For instance, one participant on 436-Route4 misidentified a blue wall near Exit 3 as an exit cue, resulting in unnecessary deviation and a higher VL value. Interviews confirmed that this misjudgement was driven by perceptual ambiguity between the blue wall at Exit 3 and the opposite corridor end, which terminated in a reflective window. Conversely, low-ECI routes produced more linear scan paths, fewer fixations, and smoother navigation, reflecting reduced cognitive load and clearer spatial comprehension. These visual patterns align with the quantitative relationship between higher ECI and longer evacuation time.

4. Discussion

This study demonstrates how computational tools can approximate the cognitive effort required for fire evacuation in complex buildings. By combining GT, VR, and eye-tracking within a BIM-based workflow, the ECI links spatial configuration with

evacuation cognition. ECI captures how SC and VL jointly influence behaviour. In the VR experiment, higher ECI values were associated with greater cognitive demand and longer evacuation times, whereas lower ECI values were associated with better spatial legibility and wayfinding efficiency.

The integration of GT, VR, and eye-tracking confirms that ECI provides a sensitive diagnostic measure for identifying cognitively demanding areas within architectural layouts. High-ECI zones are often located along corridors with numerous accessible subspaces or ambiguous junctions, exhibiting increased spatial and perceptual complexity. In contrast, simplified configurations with clear visual hierarchies promote efficient and confident navigation. These findings highlight the dual role of spatial accessibility: while beneficial for circulation flexibility in normal conditions, excessive accessibility may increase decision uncertainty and cognitive effort during emergencies.

From a design perspective, ECI enables evidence-based optimisation by quantitatively identifying high cognitive-load regions and informing design strategies to mitigate them. Designers can refine spatial hierarchy to reduce redundant connections, clarify circulation paths, and enhance visual differentiation through lighting, material colour, or contrast. Directional cues and strategically positioned signage further strengthen spatial legibility and reduce visual load. Through these interventions, ECI evolves from a diagnostic indicator into a predictive design tool that supports cognition-informed and performance-based building design.

Beyond practical implications, the ECI framework offers an initial integrated approach for combining spatial syntax, visual cognition, and behavioural modelling within a single cognitive evaluation scheme. When incorporated into simulation-based workflows, ECI allows architects to anticipate user behaviour, assess perceptual clarity, and optimise spatial configuration during early design phases. The framework thus bridges quantitative topology analysis with perceptual and behavioural evidence, offering a more holistic understanding of how humans interact with architectural space under stress.

However, this study has several limitations. The findings are based on a single school building with limited evacuation routes and a relatively small participant sample, constraining generalisability. The VR environment cannot fully reproduce the emotional and social conditions of real emergencies, such as higher stress, crowding and collective behaviour. Moreover, the weighting of spatial and visual components in ECI is informed by theory and sensitivity analysis. Still, it remains partly heuristic, and individual differences in visual exploration may influence VL measures. Future research should therefore involve larger and more diverse samples, additional building types, refinement of the weighting scheme and comparative studies between VR-based experiments and real or drill-based evacuations.

In the future, the ECI framework could be integrated into a Digital Twin platform to support cognition-aware and adaptive evacuation management. In this system, building layout and GT representations would form the spatial and topological layer, while VR and eye-tracking studies would provide parameters for a human-interaction layer. ECI could operate as a cognitive evaluation engine that quantifies SC and VL and is updated with sensor and localisation data. This would allow the Digital Twin to identify cognitively demanding regions, suggest adjustments to lighting, signage or

guidance strategies, and, in the future, provide more tailored evacuation support. Such an integration would extend building safety beyond static structural optimisation toward more dynamic, human-centred and cognition-aware evacuation systems.

5. Discussion

This study presented a human-centred computational framework that combines BIM, GT, VR, and eye-tracking to investigate the cognitive aspects of fire evacuation. The ECI was developed to couple SC and VL, providing a quantitative indication of the mental demands imposed by different spatial configurations and evacuation routes. In the VR evacuation study, higher ECI values were associated with increased gaze effort, greater decision uncertainty, and longer evacuation times, while lower ECI routes supported clearer spatial legibility and more efficient wayfinding.

The ECI framework offers an initial integrated approach for linking spatial syntax, visual cognition, and behavioural performance, supporting cognition-informed architectural decisions. In practice, ECI can function as a diagnostic indicator to highlight cognitively demanding zones and to guide adjustments to circulation structure, signage, and lighting. In the future, integrating ECI within Digital Twin environments could enable cognition-aware, adaptive evacuation planning and management. Overall, this research contributes to advancing building safety toward more human-centred, perception-aware and performance-based design.

References

Bernardini, G., Lovreglio, R., Quagliarini, E., & D'Orazio, M. (2023). Can active and passive wayfinding systems support fire evacuation in buildings? Insights from a virtual reality-based experiment. *Journal of Building Engineering,* 74, 106778. https://doi.org/10.1016/j.jobe.2023.106778

Donato, V. (2017). Towards validating the design process: integrating graph theory into BIM. *Architectural Engineering and Design Management*, 13(1), 22–38. https://doi.org/10.1080/17452007.2016.1208602

Franz, G., & Wiener, J. M. (2008). From space syntax to space semantics: A behaviorally and perceptually oriented methodology for the efficient description of the geometry and topology of environments. *Environment and Planning B: Planning and Design*, 35(4), 574–592. https://doi.org/10.1068/b33050

Fu, R., Zhang, C., Kan, R., Liu, Y., & Seo, H. (2024a). Virtual reality–based simulation for assessing building fire safety design. *SIMULATION*, 100(6), 581–594. https://doi.org/10.1177/00375497231225980

Fu, R., Zhang, C., Wang, Y., Yue, Y., & Di Sarno, L. (2024b). Building layout assessment on fire evacuation efficiency by integrating graph theory and virtual reality. *Journal of Building Engineering*, 98, 111498. https://doi.org/10.1016/j.jobe.2024.111498

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Ministry of Public Security of the People's Republic of China. (2014). *Code for Fire Protection Design of Buildings (GB 50016–2014, 2018 Edition)*. Beijing: China Planning Press.

Zhao, H., Yu, Z., & Zhu, J. (2023). A prediction model for the smoke spread path in high-rise building fires based on graph theory. *Fire*, 6(7), 258. https://doi.org/10.3390/fire6070258

Gamified Urban Design in Virtual Reality

Embodied Play as a Mode of Creative Participation

Mengdie Wang[1]
[1]*Nanyang Technological University.*
[1]*mengdie001@e.ntu.edu.sg, 0009-0008-7592-225X*

Abstract. Although virtual reality (VR) and gamification are widely adopted in participatory urban design, their integration is rarely examined in terms of how it mediates creative engagement and embodied design practice. This study investigates how embodied play, through the synthesis of bodily interaction and game elements can reshape participation into a creative design practice. VRban Designer is a gamified VR simulator developed for this research that integrates hand-tracking interaction with a progressive gameplay to support creative participation in urban design. Conducted on the Nanyang Technological University (NTU) campus, the study examines how embodied interaction and gamified progression influence participants' engagement and the spatial quality of their design outcomes. A mixed-method approach combining experiential questionnaires, behavioural observation, and spatial mapping reveals that gesture-based interaction enhances playfulness and expressiveness, while game progression sustains focus and motivation. Together, these elements enable participants to produce spatially coherent and meaningful design outcomes, demonstrating the potential of embodied play as a mode of participatory urban design.

Keywords. Virtual Reality, Embodied Interaction, Gamification; Participatory Urban Design, Hand Tracking

1. Introduction

Contemporary urban design increasingly views citizens as active contributors rather than passive recipients. Mueller et al. (2018) introduced citizen design science, positioning citizens as creators who collect data, envision alternatives, and propose designs that complement professional practice. As Sanders (2002) argues, understanding what people say, think, or do is not enough; participation must tap into their tacit and latent knowledge, their perceptions, emotions, and aspirations. This deeper engagement emerges when people are empowered not just to respond but to make, create, and experiment, revealing hidden needs and new possibilities.

Virtual Reality (VR) allows participants to experience design from an embodied, first-person perspective, understanding space through action rather than description.

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 195-204.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Gamification, in parallel, introduces game design elements in non-game context to structure motivation and sustain engagement over time (Deterding et al., 2011). Both paradigms hold potential for participatory urban design, but their integration has not yet been leveraged to unlock citizens' creative agency.

The gap lies not simply in combining VR and gamification, but in understanding how embodied play can transform participation into creative design practice. Existing studies often position VR as a visualization or consultation medium, while gamified approaches focus on external motivation: points, badges, or rankings. As a result, participation is often evaluated separately in terms of usability and enjoyment, rather than through their combined creative impact. What remains missing is a framework that shows how bodily interaction and playful challenge can jointly support the act of designing, where fun functions as a mechanism for situated creativity rather than a surface level incentive. Furthermore, empirical evaluations rarely extend beyond usability and enjoyment metrics, leaving open whether such engagement produces spatially coherent, contextually relevant design outcomes.

This study examines VRban Designer, a gamified VR simulator that integrates hand tracking within a progressive game structure to support creative participation in urban space design. The project is situated on the Nanyang Technological University (NTU) campus, focusing on the public space surrounding the upcoming JW3 Nanyang Gateway MRT station. A total of 54 participants engaged in the VR session to redesign the site. A mixed-method approach combined experiential and spatial analyses to evaluate participant engagement and design outcomes. Questionnaires, observations, and spatial logs together provided an integrated understanding of how participants experienced and enacted the gamified environment, linking embodied interaction to the creativity and coherence of their designs. Accordingly, the study addresses the following research questions:

- RQ1: How do embodied interaction (hand tracking) and gamification elements together create fun and meaningful engagement in immersive spatial design tasks?
- RQ2: What spatial design patterns emerge from playful, embodied engagement, and how do these patterns reflect creative and contextually meaningful participation?

By connecting the experiential and spatial dimensions of participation, this paper contributes in two ways. First, it conceptualizes embodied gamification as a participatory framework that transforms play into a mode of situated creativity. Second, it introduces an integrated evaluation model that links experiential evidence with spatial pattern analysis, demonstrating how playful engagement can lead to meaningful and coherent design outcomes. These contributions aim to advance participatory urban design toward an embodied, playful, and creative mode.

2. Related works

2.1. GAMES AND GAMIFICATION IN PARTICIPATORY URBAN DESIGN

Games are widely used to engage citizens in planning and design by framing urban space as a playground where participants can test ideas and imagine alternatives through structured play (Pienaru, 2018). Several mechanics are particularly relevant.

Simulation and sandbox mechanics treat urban elements as playable resources, allowing participants to place, modify, or remove features to actively create spaces (Bereitschaft, 2016). Role-play assigns players stakeholder identities, encouraging empathy, negotiation, and consensus-building by requiring participants to deliberate from diverse perspectives (Tan, 2017). Voting and competition provide structured ways to aggregate preferences, motivating creativity by ranking proposals and selecting favoured outcomes (Poplin, 2014). Socializing mechanics such as discussion channels, collaborative play, or shared editing encourage communication and co-creation, turning the planning process into a collective act (Gugerell et al., 2017).

Gamification suggests new potential of applying game elements into urban design participation. Common elements include points, badges, leaderboards, and quests that reward contributions and track progress (Koivisto & Hamari, 2019). These elements can also help visualize the impact of decisions, allowing participants to perceive trade-offs and consequences in real time. Avatars and storytelling create more engaging and personalized experiences, while team-based structures encourage collaboration and shared goals (Jahn et al., 2021). In urban planning and design, such elements have been applied in participatory projects that integrate challenges and point mechanisms (Poplin, 2012) and employ avatars, badges, and quests (Muehlhaus et al., 2023) to motivate decision-making and sustain engagement.

2.2. IMMERSIVE DESIGN AND EMBODIED INTERACTION

Immersive visualization enhances design cognition by allowing spatial understanding to emerge through perception and action rather than abstract representation (Marriott et al., 2018). Abstract data can be embedded spatially, viewpoints can be combined across egocentric and exocentric frames, and immersive settings foster exploratory behaviors (El Beheiry et al., 2019). VR extends perception beyond the audio-visual realm by incorporating haptic, olfactory, and gustatory stimuli, offering designers and stakeholders a multisensory means to assess spatial qualities (Apostolou & Liarokapis, 2022).

Embodied interaction underpins VR's potential in design participation, emphasizing engagement through bodily movement rather than abstract control inputs (Dourish, 2001). Such embodied engagement promotes intuitive understanding and reduces cognitive load by aligning digital actions with natural sensorimotor experiences (Lehrman, 2025). Within this paradigm, hand tracking introduces a gesture-based interaction modality that enables users to manipulate virtual elements directly, enhancing presence and immediacy (Buckingham, 2021; Johnson-Glenberg, 2018). Although hand tracking has been studied primarily in gaming and educational applications, its application to participatory and urban design remains limited.

3. Method

3.1. SYSTEM OVERVIEW

The gamified VR simulator was developed in Unity with Meta All-in-One SDK for hand tracking and tested on Meta Quest3 headset. The virtual scene reconstructs the NTU campus residential district adjacent to the upcoming JW3 MRT station.

The simulator uses a three-level gameplay framework that gradually scaffolds participants from exploration to open-ended design. Level 1 allows participants to explore and navigate the virtual site, collect badges at the station entrances, and select the next task site. Level 2 incorporates guided design challenges in the entrance area to teach core design toolkit functions such as placing assets, planting vegetation, editing pavement, and customizing placeholders. Level 3 supports open creative design, where participants redesign public spaces along their route. Actions are logged through a scoring system that rewards asset interventions and property edits and unlocks a hidden reward. The design toolkit includes 10 asset categories, each offering a set of corresponding assets that participants can select and apply to shape the space. Eight categories are accessed through the hand-mounted menu, while scene-embedded interfaces support large-scale surface edits. Through different interaction methods, participants can place and modify assets, iteratively adjusting properties such as style

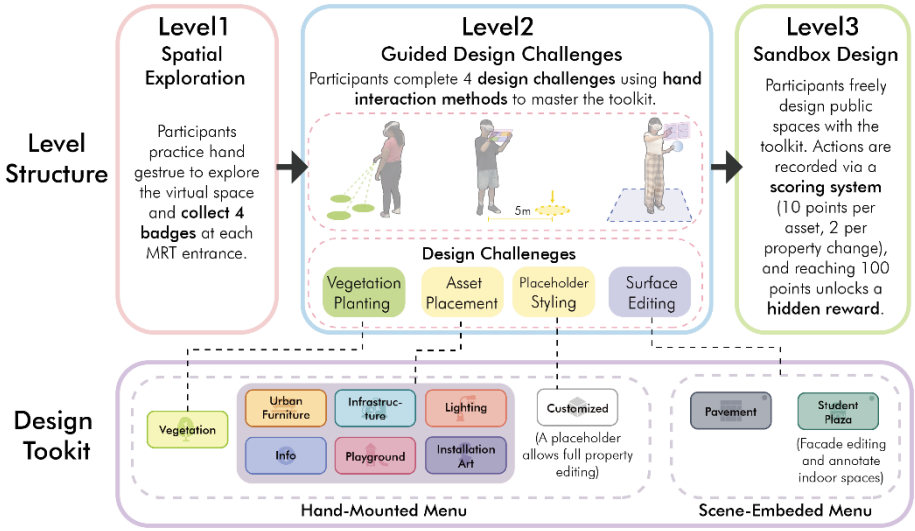


Figure 1. Gameplay Framework and Design Toolkit

options, scale, position, rotation, and materials, and add annotations (Figure 1).

3.2. PARTICIPANTS AND PROCEDURE

The study was approved by NTU's Institutional Review Board (IRB), and informed consent was obtained from all participants. A total of 54 participants (23 females, 31 males; aged 21-65) from the university community, mainly hall residents and frequent users of the station area took part in the experiment. After a short introduction and pre-test questionnaire to collect the demographic info (15 mins), participants completed all three levels of the VR design task (20-60 mins), followed by a post-test questionnaire and short interview (20 mins).

3.3. DATA COLLECTION

First, participants completed a post-test questionnaire consisting of 5-point Likert items

evaluating usability (SUS), presence, fun, gamification elements and game mechanics, and overall experience. Second, behavioural observation notes taken during sessions were complemented by post hoc qualitative review of video recordings to document bodily engagement, gesture patterns, persistence, and visible affective responses such as laughter or hesitation, providing evidence of embodied play. Third, design output logs were recorded by the simulator at real-time, capturing each placed asset's parameters and property changes, coordinates, and timestamp in JSON format for subsequent spatial analysis and scene reconstruction.

3.4. ANALYSIS FRAMEWORK

This analysis focuses on data collected from Level 3.

For RQ1: Performance data (scores, level duration, and intervention counts) were used to quantify engagement intensity. Questionnaire data based on 5-point Likert items were summarized using mean (SD), interquartile range (IQR) to describe central tendency and response consistency without assuming normality. Observational video logs were coded for bodily movement and affective expression, such as grabbing, stretching, crouching, or expressive laughter, to interpret how hand-tracking interaction elicited enjoyment and immersion.

For RQ2: spatial logs were processed and imported into QGIS to analyse the distribution and diversity of participant-generated layouts. The workflow included a 14 m grid tessellation to calculate asset density and compositional proportions, a Shannon diversity index (H′) was computed to quantify the richness and balance of asset types across the site, followed by k-means clustering (k = 7) to identify recurring design configurations. The resulting spatial patterns were compared with the reconstructed scenes to interpret underlying design intentions and assess overall design quality.

4. Results

4.1. PERFORMANCE OVERVIEW

Among 54 participants, one ended the session early due to motion sickness but completed sufficient design actions (score > 100), so the data were retained. The average Level 3 duration was 23.6 minutes (SD = 10), with most participants finishing within 17-27 minutes. Participants placed an average of 12 assets (SD = 4.39, IQR = 9-13), ranging from 3 to 28. The mean Level 3 score was 153.6 (SD = 54.55), spanning 38-316; one participant scored below 100, 31 between 100-150, and 22 above 150.

4.2. QUESTIONNAIRE RESULTS

The System Usability Scale (SUS) yielded a high mean score of 81.2 (SD = 11.3, IQR = 77.5-90), with no participant scoring below 50. Presence ratings were similarly strong, with all five items: spatial composition, self-location, sense of presence, spatial understanding, and design decision-making, averaging between 4.33 and 4.57 (SD = 0.60-0.81). The overall enjoyment item, "It is fun to play the game," received the highest score (M = 4.63, SD = 0.52).

Gamification elements were positively received (M = 3.61-4.46), particularly

"Level progression is clear and sets goals" (M = 4.46, SD = 0.66) and "Gamification elements help to master the toolkit" (M = 4.39, SD = 0.76). The mystery reward (n = 42) and score system (n = 34) were identified as the most meaningful achievements. Game mechanics also scored highly (M = 4.44–4.61), with enjoyment (M = 4.61, SD = 0.49) and intuitiveness (M = 4.56, SD = 0.57) rated highest.

4.3. BEHAVIOUR OBSERVATION

Participants' engagement during the experiment indicated that hand gestures and bodily movement frequently elicited visible joyfulness (Figure 2). When using their hands to grab, rotate, or reposition assets, participants displayed expressive movements and often responded with laughter or smiles to unexpected or exaggerated gesture effects.



Figure 2. Participants' bodily movement during the experiment

4.4. SPATIAL PATTERN ANALYSIS

4.4.1. Heatmap

The heatmap visualizes the spatial distribution of all placed assets across the Level 3 design area. As pavement and student-plaza elements are predefined in fixed scene locations, these categories were excluded from the analysis. The density of individual asset categories is shown in Figure 3.

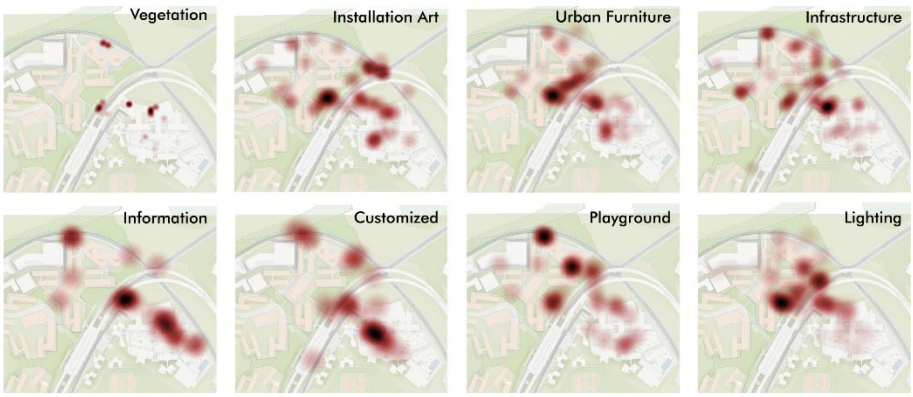


Figure 3. Heatmaps showing the spatial density of individual asset categories

Areas of high intensity emerge around the entrance zones leading to the two residential clusters and in the open public spaces adjacent to the residential buildings. Vegetation assets appear highly concentrated within their designated planting zones. Other categories display distinct spatial tendencies: urban furniture is primarily concentrated on open spaces; lighting assets are frequently placed along the sheltered pathway described by many participants as "very dark"; and information assets such as maps or directories tend to cluster near the main entrances of the residential areas. Infrastructure elements show a relatively dispersed distribution across the site, while installation art pieces are mostly located on the grass field, and playground assets tend to be positioned closer to the buildings, forming small activity nodes adjacent to residential spaces. Customized assets commonly appear beneath building overhangs, where participants experimented with their own interpretations of function and use.

4.4.2. Shannon Diversity (H') Index and k-Means Clustering (k = 7)

The H′ index map illustrates the local diversity of asset types within each grid cell. Shown in Figure 4, darker tones indicate areas with a more varied composition of design elements. The distribution closely corresponds to the heatmap, showing higher diversity around the entrances leading to the two residential clusters on both sides of the main road. Additionally, several public spaces adjacent to buildings exhibit greater compositional richness, suggesting more active and varied design interventions in these zones.

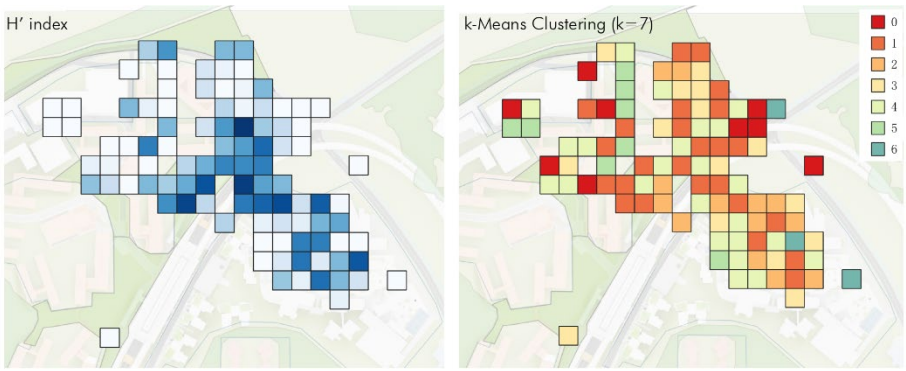


Figure 4. H′ index Map (left) and k-Means Clustering Map (right)

The k-means analysis identified seven distinct spatial composition types across all grid cells. Cluster 0 primarily contains installation art elements with minor infrastructure and low diversity (H′ = 0.17). Cluster 1 combines customized placeholders, lighting, and urban furniture, with moderate infrastructure presence (H′ = 1.17). Cluster 2 is vegetation-dominant, showing moderate compositional diversity (H′ = 1.04). Cluster 3 is largely infrastructure-oriented, with limited variety (H′ = 0.34). Cluster 4 integrates urban furniture and playground elements with secondary infrastructure (H′ = 0.80). Cluster 5 consists exclusively of customized and lighting elements (H′ = 0.00). Cluster 6 includes only information elements, also showing no diversity (H′ = 0.00).

4.4.3. Recurring design patterns

Several participants produced spatial configurations that showed clear similarities in asset composition. Figure 5 illustrates examples of these recurring arrangements. Participant 0531A, 0612B and 0618A created designed arrangements in which an art canopy, stylized benches, and ground lighting are combined to define a lounge space within the open public area. While Participants 0602A and 0603B designed a small leisure or waiting zone using swings and rocking horses near the bus stop.

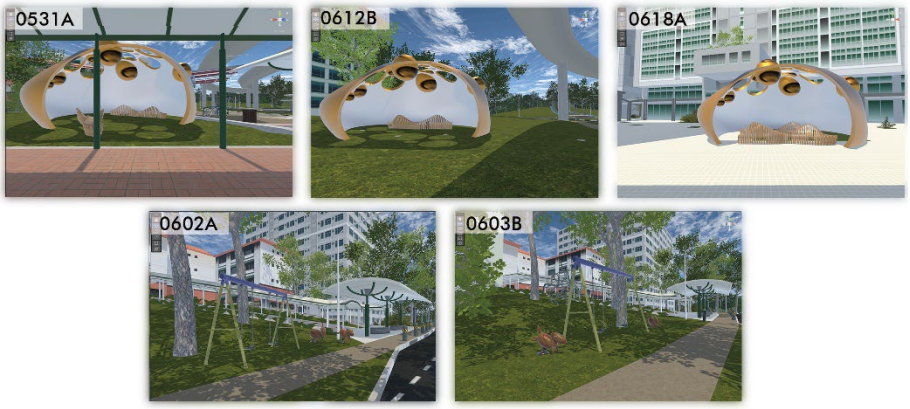


Figure 5. Similar design composition by different participants

5. Discussion

Through the study, participants demonstrated a high level of joy and engagement during the gamified VR design session. The questionnaire results indicate that most participants found the experience enjoyable, while observational evidence further revealed that hand gestures and bodily movement evoked a distinctive form of playful and expressive joy. This joy stems from bodily improvisation within technological constraints. When participants mimicked grabbing or reaching motions with their hands and bodies, their movements often became exaggerated or slightly awkward, generating moments of humour through imperfection. Hand tracking mirrors real motion but introduces subtle limitations, prompting users to compensate through expressive gestures. The pleasure therefore emerged from breaking habitual expectations of movement. Physical exaggeration transformed functional design tasks into small performances that felt spontaneous and liberating. Such movement produces expressive bodily play, a spontaneous, childlike form of enjoyment (Sutton-Smith, 2010). Rather than detracting from immersion, these clumsy gestures heighten awareness of one's body as an active agent. This experience aligns with the idea of playfulness as an attitude that values exploration over control (Sicart, 2014).

The embodied interaction transforms the design process into an affective, performative experience that merges functional control with playful experimentation. When combined with gamification elements such as mystery rewards and point accumulation, this embodied expressiveness is further amplified. The hidden-reward structure sustains curiosity and motivates participants to experiment with more tools,

while the points provide immediate feedback that validates their exploratory actions. Together, these elements form a feedback loop between bodily play and motivational design: curiosity encourages physical engagement, and physical engagement reinforces emotional investment. This synthesis of embodied humour and gamified reward thus explains how hand tracking and game mechanics jointly foster both fun and meaningful engagement in immersive spatial design.

Most participant designs aligned with real-world expectations and practical needs. As shown in the heatmap and H′ index maps, design activity concentrated around residential entrances. Clustering analysis further revealed functionally driven interventions: improving orientation, providing relaxation, and enhancing shading near gathering points. Meanwhile, many participants demonstrated aesthetic awareness through deliberate choices of vegetation, spatial composition, and distinctive assets, balancing pragmatic problem-solving with playful exploration.

Recurring spatial compositions also revealed a shared intuition or collective imagination. Similar spatial logics and asset configurations suggested that participants drew upon common mental models of desirable entrance spaces. This convergence underscores the interpretive coherence of their collective output and indicates the potential to extract generalizable design tendencies or reference patterns, which is valuable for informing early-stage urban design with public insights.

6. Conclusion and Future Work

This study demonstrates that embodied play can serve as a catalyst for meaningful participation in urban design. By integrating hand tracking interaction and gamified progression, VRban Designer reframed participation into an active, creative design practice. After brief onboarding, most participants operated the system as intended, though VR novices and elder adults may face a higher learning threshold, and extended headset wear and sustained hand tracking can increase fatigue. The findings indicate that fun is not only an affective outcome, but also a practical mechanism that sustains engagement and supports creative and contextually grounded design outcomes.

Future work will extend this framework to broader participant groups and more diverse planning contexts. Through refinement of gesture interaction and gamified mechanics, the system aims to bridge playful immersion with actionable urban insights, advancing participatory design as both an experiential and generative practice.

References

Apostolou, K., & Liarokapis, F. (2022). A Systematic Review: The Role of Multisensory Feedback in Virtual Reality. 2022 IEEE 2nd International Conference on Intelligent Reality (ICIR), 39–42. https://doi.org/10.1109/ICIR55739.2022.00024

Bereitschaft, B. (2016). Gods of the City? Reflecting on City Building Games as an Early Introduction to Urban Systems. Journal of Geography, 115(2), 51–60. https://doi.org/10.1080/00221341.2015.1070366

Buckingham, G. (2021). Hand Tracking for Immersive Virtual Reality: Opportunities and Challenges. Frontiers in Virtual Reality, 2, 728461. https://doi.org/10.3389/frvir.2021.728461

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." Proceedings of the 15th International Academic

MindTrek Conference: Envisioning Future Media Environments, 9–15. https://doi.org/10.1145/2181037.2181040
Dourish, P. (2001). Where the action is: The foundations of embodied interaction. MIT Press.
El Beheiry, M., Doutreligne, S., Caporal, C., Ostertag, C., Dahan, M., & Masson, J.-B. (2019). Virtual Reality: Beyond Visualization. Journal of Molecular Biology, 431(7), 1315–1321. https://doi.org/10.1016/j.jmb.2019.01.033
Gugerell, K., Jauschneg, M., Platzer, M., & Berger, M. (2017). Playful Participation with Urban Complexity – Evaluation of the Co-Located Serious Game Mobility Safari in Vienna. REAL CORP 2017 – PANTA RHEI – A World in Constant Motion. Proceedings of 22nd International Conference on Urban Planning, Regional Development and Information Society, 413–420.
Jahn, K., Kordyaka, B., Machulska, A., Eiler, T. J., Gruenewald, A., Klucken, T., Brueck, R., Gethmann, C. F., & Niehaves, B. (2021). Individualized gamification elements: The impact of avatar and feedback design on reuse intention. Computers in Human Behavior, 119, 106702. https://doi.org/10.1016/j.chb.2021.106702
Johnson-Glenberg, M. C. (2018). Immersive VR and Education: Embodied Design Principles That Include Gesture and Hand Controls. Frontiers in Robotics and AI, 5, 81. https://doi.org/10.3389/frobt.2018.00081
Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. International Journal of Information Management, 45, 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013
Lehrman, A. L. (2025). Embodied Learning Through Immersive Virtual Reality: Theoretical Perspectives for Art and Design Education. Behavioral Sciences, 15(7), 917. https://doi.org/10.3390/bs15070917
Marriott, K., Chen, J., Hlawatsch, M., Itoh, T., Nacenta, M. A., Reina, G., & Stuerzlinger, W. (2018). Immersive Analytics: Time to Reconsider the Value of 3D for Information Visualisation. In K. Marriott, F. Schreiber, T. Dwyer, K. Klein, N. H. Riche, T. Itoh, W. Stuerzlinger, & B. H. Thomas (Eds.), Immersive Analytics (Vol. 11190, pp. 25–55). Springer International Publishing. https://doi.org/10.1007/978-3-030-01388-2_2
Muehlhaus, S. L., Eghtebas, C., Seifert, N., Schubert, G., Petzold, F., & Klinker, G. (2023). Game.UP: Gamified Urban Planning Participation Enhancing Exploration, Motivation, and Interactions. International Journal of Human–Computer Interaction, 39(2), 331–347. https://doi.org/10.1080/10447318.2021.2012379
Mueller, J., Lu, H., Chirkin, A., Klein, B., & Schmitt, G. (2018). Citizen Design Science: A strategy for crowd-creative urban design. Cities, 72, 181–188. https://doi.org/10.1016/j.cities.2017.08.018
Pienaru, M.-I. (2018). The City as a Playground—Game tools for interactive planning. 679–686. https://doi.org/10.52842/conf.ecaade.2018.2.679
Poplin, A. (2012). Playful public participation in urban planning: A case study for online serious games. Computers, Environment and Urban Systems, 36(3), 195–206. https://doi.org/10.1016/j.compenvurbsys.2011.10.003
Poplin, A. (2014). Digital Serious Game for Urban Planning: "B3—Design Your Marketplace!" Environment and Planning B: Planning and Design, 41(3), 493–511. https://doi.org/10.1068/b39032
Sanders, E. B.-N. (2002). From user-centered to participatory design approaches. In J. Frascara (Ed.), Design and the Social Sciences (pp. 18–25). CRC Press / Taylor & Francis. https://doi.org/10.1201/9780203301302-ch8
Sicart, M. (2014). Play matters. The MIT Press.
Sutton-Smith, B. (Ed.). (2010). The ambiguity of play. Harvard University Press.
Tan, E. (2017). Play the city: Games informing the urban development. Jap Sam Books.

Pedagogical Strategy Combining VR and Traditional Drawings to Depen Architectural Case Study Understanding

Merging Virtual Immersion and Symbolic Encoding: Validating the Synergistic Effect of Mixed Media via Dual Coding Theory

Kai-Hsiang Liang[1]
[1]*National University of Kaohsiung*
[1]*KLiang@nuk.edu.tw, 0000-0002-2085-4279*

Abstract. This study addresses the critical challenge of using traditional 2D drawings to convey complex 3D architectural space to novice students. The primary research question is: Does a VR-and-paper hybrid strategy enhance learning outcomes and student preference more effectively than standalone VR or paper methods? We empirically investigated the efficacy of integrating Virtual Reality (VR) and paper drawings against single-medium approaches in architectural case study instruction. A rigorous 2 x 3 mixed-design experiment was conducted with 120 vocational high school students, randomly assigned to Paper, VR, or Hybrid groups, studying two contrasting architectural cases. Learning outcomes were assessed via a domain-specific knowledge test and student preferences via a standardized questionnaire, supplemented by qualitative interviews. Quantitatively, the Media Group exerted a significant main effect on learning outcomes ($F(2, 117) = 37.52$, $p < 0.001$), with a large effect size (partial$\eta^2 \approx 0.38$). Post-hoc analysis confirms that both VR and Hybrid groups significantly outperformed the Paper group, with the Hybrid group achieving the highest mean scores. The Hybrid group also reported the highest satisfaction, validating its synergistic effect. In conclusion, the hybrid model constitutes an optimal pedagogical framework for novice spatial comprehension by strategically leveraging the complementary activation of Dual Coding Theory's cognitive pathways: VR's intuitive spatial immersion and paper's systematic symbolic encoding.

Keywords. virtual reality, case study, hybrid learning, spatial comprehension, dual coding theory

1. Introduction

This section provides the background and context for the study, highlighting the limitations of traditional methods in architectural education and outlining the research focus.

Case study instruction has always been a critical component in architectural

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 205-214.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

education for bridging theory and practice. Traditionally, students analyse architecture by relying on two-dimensional paper-based graphics such as floor plans, elevations, sections, and photographs. However, this type of planar medium is limited in representing the three-dimensional complexity of architectural space, making it challenging for novices who must expend significant cognitive effort to mentally transform planar information into volumetric concepts. For instance, when dealing with intricate circulation or large-scale spaces, relying solely on drawings often prevents a comprehensive grasp of key design features. Consequently, traditional paper-based teaching has limitations, necessitating the introduction of new tools to aid spatial understanding.

Driven by the recognized limitations of single-modality instruction, this research is specifically motivated to empirically validate the cognitive mechanisms and efficacy of a "VR and Paper Hybrid" teaching strategy, thereby addressing the gap in evidence-based hybrid instructional design for complex spatial subjects. This study focuses on two main dependent variables: (1) Learning outcomes; and (2) Media preference. Based on these two indicators, the following primary hypotheses are proposed:

- The VR group will outperform the pure Paper group in learning outcomes;
- The Hybrid group will achieve the best performance, expecting a synergistic effect from combining VR and paper that exceeds the benefits of a single medium;
- Students will express higher interest and preference for the novel VR-introduced learning methods compared to traditional paper-based methods.

2. Literature Review

This section reviews the existing body of literature regarding the adoption of Virtual Reality in architectural education, discusses the recognized limitations of single-modality learning, and builds the theoretical foundation for adopting a hybrid approach.

2.1. THE ROLE OF VR IN ARCHITECTURAL EDUCATION

With advancements in digital technology, immersive media such as Virtual Reality (VR) have been introduced into architectural pedagogy to compensate for the shortcomings of traditional drawings. VR utilizes head-mounted displays to provide an immersive 3D environment, allowing students to feel as if they are physically present within the architectural space, directly perceiving true scale and spatial relationships. Some researchers found that using VR to aid design training significantly increased students' creative process and motivational engagement (Obeid & Demirkan, 2023). Learning with VR models enhances first-year architecture students' visual recognition skills regarding architectural construction and form relationships (Ceylan, 2020). Other studies reported that learning in a virtual heritage environment leads to deeper student engagement and understanding (Shanti & Al-Tarazi, 2023). Furthermore, VR helps narrow educational resource gaps and improves learning performance in architectural history courses for students in different regions (Li et al., 2025). Overall, substantial empirical evidence demonstrates VR's potential to enhance spatial understanding in architecture through its immersive and interactive characteristics.

2.2. LIMITATIONS OF UNIMODAL MEDIA AND THE NEED FOR A HYBRID APPROACH

However, introducing VR is not a panacea; standalone VR may increase cognitive load due to interface barriers or information overload, potentially counteracting its advantages if not accompanied by appropriate instructional design. Conversely, while traditional paper drawings lack immersion, they possess distinct advantages in precisely reading annotations, quickly browsing multiple drawings, and making comparisons. Thus, a key question arises: Is it possible to combine VR with traditional drawings to capitalize on the strengths of both?

Some scholars propose that integrating VR with other forms of representation, such as floor plans, can offer diversified sensory stimuli and information presentation, generating a synergistic effect that surpasses the learning outcomes of a single medium. This perspective is supported by the Dual Coding Theory in educational psychology, which posits that human cognition processes information through two channels: visual imagery and verbal symbols (Paivio, 1971). A hybrid strategy allows students to simultaneously utilize both cognitive pathways, leading to deeper knowledge encoding and retrieval. Recent literature also advocates for accelerating the development of phygital hybrid teaching models in the post-pandemic era; for example, some propose adopting integrated physical and virtual teaching in architectural design studios (Zairul et al., 2023), and other research observed that combining VR with traditional drawing accommodates diverse student learning preferences (Lu, 2022). In short, the integrated use of multiple media warrants further empirical validation for its potential to yield additional learning benefits.

3. Methodology

This section details the experimental framework, participant characteristics, materials developed, and the systematic procedures employed to test the formulated hypotheses.

3.1. EXPERIMENTAL DESIGN AND CASE SELECTION

This study employs a mixed design of 2 case types paired with 3 media groups. Case Type served as the within-subjects independent variable (all participants studied both cases), and Media Group was the between-subjects independent variable (participants were assigned to only one group). This design allows for examining the consistency of the media effect across different case types.

- Case A: Villa Savoye (Le Corbusier): Characterized by clear geometric form and spatial logic, it serves as a representative of low-complexity cases.
- Case B: Row House in Sumiyoshi (Tadao Ando): Known for its compact spatial composition and unique circulation strategy, it presents a higher difficulty level for interpreting two-dimensional drawings.

The research developed three learning materials: the VR Material (a first-person immersive environment with simple interactive annotations developed in the Unreal Engine), the Paper Material (a compilation of traditional 2D drawings and textual information), and the Hybrid Material (simultaneously providing the VR environment and the paper booklet, utilizing QR codes to enable rapid teleportation between the virtual and real-world corresponding spatial locations). Figure 1 provides an overview of our experimental environment and the actual learning scenarios for all groups (Paper, VR, and Hybrid), visually illustrating the operational differences in media setup.



Figure 1: Experimental Participant Scenarios for Architectural Case Study Learning

3.2. SAMPLE AND GROUP ASSIGNMENT

The participants were 120 second-year students from architectural departments at three vocational high schools across northern, central, and southern Taiwan. All participants had at least one year of architectural design course background, and none had prior in-depth study of the two selected cases. A stratified random assignment design was adopted to divide the 120 students into three groups: the VR Group, the Paper Group, and the Hybrid Group, with 40 students per group. Prior to the experiment, students completed the adjusted Architecture and Interior Spatial Ability Test (AISAT). To statistically control for potential pre-existing group differences in baseline spatial ability and thus enhance the internal validity of the study (Cho & Suh, 2022), these scores were subsequently used as a covariate in the Analysis of Covariance (ANCOVA).

3.3. EXPERIMENTAL PROCEDURE

The ensuing discussion details the precise sequence and implementation of the experimental design, encompassing the two-phase instructional unit arrangement, the presentation order of the architectural case studies, the specific media tools deployed by each group, and the synchronized timing for the post-tests and preference questionnaire administration. This systematic approach ensures methodological rigor and reproducibility.

- Pre-test Phase: Students completed a background questionnaire and the spatial ability test (AISAT).
- Instruction and Training: Students were instructed on how to use their assigned media.
- Case Learning and Analysis: Students independently explored and studied the two cases for a total of 60 minutes (individual work). Hybrid group students were free to switch between VR and paper drawings.
- Assessment and Questionnaire: All learning materials were collected, and students

completed the Architectural Case Study Learning Assessment (20 multiple-choice and short-answer questions, Cronbach's $\alpha = 0.82$) and a Media Experience and Preference Questionnaire (5-point Likert scale).

- Interview: A subset of students from each group was randomly selected for brief semi-structured interviews.

3.4. DATA ANALYSIS METHODS

The analytical approach is delineated herein, starting with descriptive statistics to summarize group performance. The subsequent analysis leverages ANCOVA for assessing learning outcomes while statistically controlling for baseline spatial ability, employs ANOVA for examining media preferences, and concludes with a thematic analysis of the qualitative interview data, achieving a robust integration of mixed-method findings.

- Quantitative Analysis: Descriptive statistics were first used to summarize the overall performance trends of each group in the learning assessment and media questionnaire. To statistically control for pre-existing group differences in baseline spatial ability, a two-way repeated measures Analysis of Covariance (ANCOVA) was employed to analyse the learning assessment scores. Media Group (between-subjects) and Case Type (within-subjects) served as factors, with the pre-test AISAT score incorporated as the covariate. One-way ANOVA was used for the total media preference scores. Tukey HSD post-hoc comparisons were employed when significance was achieved, and effect sizes (partialη^2) were calculated.
- Qualitative Analysis: Thematic analysis was employed for coding and interpretation of the observation notes and interview transcripts.

3.5. NOTE ON MANUSCRIPT PREPARATION

The large language model, ChatGPT (OpenAI, 2025), was employed solely to enhance the manuscript's readability, specifically for improving flow and correcting errors in spelling and grammar.

4. Results

The empirical findings are reported in this chapter, stemming from the quantitative data analysis and supplemented by the thematic interpretation of qualitative data gathered through student interviews.

4.1. QUANTITATIVE RESULTS: COMPARISON OF LEARNING OUTCOMES AND MEDIA PREFERENCE

This subsection details the statistical comparison of the three media groups (Paper, VR, and Hybrid) on the primary dependent variables: the Architectural Case Study Learning Assessment scores and the Media Experience and Preference Questionnaire results.

4.1.1. Architectural Case Study Learning Assessment Scores

Both the VR and Hybrid groups significantly outperformed the Paper group, with the Hybrid group achieving the highest mean score. The descriptive statistics, including the adjusted mean scores, are presented in Table 1a. As shown, the Hybrid group achieved the highest mean scores for both Case A (M=80.1) and Case B (M=85.5), visually suggesting its superior performance. The complete two-way repeated measures ANCOVA results, controlling for baseline spatial ability, are summarized in Table 1b. The two-way repeated measures ANOVA results showed a significant main effect for Media Group ($F(2, 117) = 37.52$, $p < 0.001$, partial$\eta^2 \approx 0.38$) and a significant main effect for Case Type ($F(1, 117) = 7.95$, $p < 0.01$). A significant interaction effect was also found ($F(2, 117) = 3.20$, $p < 0.05$). Post-hoc comparisons indicated that the Paper group's scores were significantly lower than both the VR and Hybrid groups across both cases. For the Row House in Sumiyoshi (the more complex case), the Hybrid group's advantage was most pronounced, with its score being marginally significantly higher than the pure VR group ($p \approx 0.04$). The pure VR approach, while immersive, likely presented a higher cognitive load, thereby limiting the ability of students to mentally map the complex, non-linear circulation. This confirms that the hybrid approach provides a crucial scaffolding mechanism when spatial complexity is high.

Table 1a: Descriptive Statistics of Adjusted Mean Scores

(Note: Mean Scores are adjusted based on the AISAT pre-test score as a covariate.)

Media Group	N	Case Type	Adjusted Mean Score	Standard Deviation
Paper	40	Case A (Villa Savoye)	65.2	10.5
	40	Case B (Row House in Sumiyoshi)	54.8	9.8
VR	40	Case A (Villa Savoye)	78.5	11.2
	40	Case B (Row House in Sumiyoshi)	72.3	12.0
Hybrid	40	Case A (Villa Savoye)	80.1	10.8
	40	Case B (Row House in Sumiyoshi)	85.5	10.3
Total	120	-	72.7	15.2

Table 1b: Summary of Two-Way Repeated Measures ANCOVA Results for Learning Outcomes (Controlling for AISAT Score)

Source of Variation	Sum of Squares	df	Mean Square	F	p	Partialη^2
Media Group	11256.0	2	5628.0	37.52	< 0.001	0.38
Case Type	795.0	1	795.0	7.95	< 0.01	0.06
Group x Case Type	320.0	2	160.0	3.20	< 0.05	0.03
Error (Between Subjects)	17550.0	117	150.0	-	-	-

4.1.2. Learning Media Preference Evaluation

Hybrid group students reported the highest satisfaction and enjoyment with the

learning method (M=22.4), followed by the VR group (M=19.7), and the Paper group the lowest (M=15.8). ANOVA results indicated a significant difference between groups ($p < 0.001$). Tukey comparisons showed that both the Hybrid and VR groups were significantly higher than the Paper group, and the Hybrid group was also significantly higher than the VR group ($p < 0.05$).

4.2. QUALITATIVE RESULTS: LEARNING PROCESS OBSERVATIONS

Key findings derived from the thematic analysis of field observations and student interview transcripts are presented here, which specifically illustrate the behavioural differences and subjective learning experiences across the three media groups. Further analysis of the semi-structured interviews provided qualitative evidence on cognitive strategy use, categorized and summarized in Table 2.

Table 2: Qualitative Findings and Cognitive Strategy Synthesis Across Media Groups

Theme	Paper Group - Observation/ Behavior	VR Group - Observation/ Behavior	Hybrid Group - Observation/ Behavior	Cognitive Implication
Spatial Comprehension	Expends significant cognitive effort on "mental transformation," requiring gestures to convert 2D plans into 3D.	Activates intuitive 3D perception, understanding scale and vertical relationships.	Student (M08): "Found many things I couldn't see in the plans... took off the headset to check the floor plan, clarify it, and then continued in VR."	VR: Visual Imagery Code; Hybrid: Complementary macro- and micro-viewpoints.
Information Encoding	Relies on transcribing notes and sketching to aid memory, highly systematic.	Relies entirely on intuition and memory due to headset inconvenience, worried about "not remembering details."	Pauses VR exploration, consults paper drawings, records key findings or questions, and then re-enters VR for verification.	Paper: Verbal Code; Hybrid: Efficient integration, achieving deep encoding and knowledge structuration.
Emotional Engagement & Load	Emotional response is flat, prone to reading fatigue.	High interest and engagement, but some students (V19) felt "stuffiness and dizziness," indicating cognitive and physiological load.	Maintains high enthusiasm and focus; switching effectively mitigates the physiological discomfort associated with exclusive VR use.	VR: Increases motivation; Hybrid: Manages cognitive and physiological load while maintaining high motivation

4.2.1. Immersive Exploration and Spatial Comprehension

Both the VR and Hybrid group students exhibited distinct immersive exploration behaviours, enabling an intuitive understanding of the building's vertical relationships and structural grid, effectively reducing the cognitive difficulty of spatial comprehension. Paper group students, conversely, expended greater mental effort, often using their hands to gesture in the air while attempting to convert planar drawings into three-dimensional images. A Hybrid group student (M08) shared: "VR revealed many things I couldn't see in the plans, but sometimes a question would pop up... I'd take off the headset and check the floor plan, clarify it, and then continue in VR. This felt more grounded than only looking at drawings". These students would check the plans after a period of VR exploration to confirm their location and the overall context, alternating between macroscopic plans and microscopic immersion to achieve a more solid grasp of the complex space.

4.2.2. Information Acquisition and Note-Taking Strategies

Paper group students predominantly relied on transcribing key points or sketching diagrams to aid memory. VR group students, due to the inconvenience of immediate note-taking while wearing the headset, relied entirely on intuition and memory, with some confessing they were "a little worried about not remembering the details afterward". The Hybrid group, however, benefited from both approaches: they alternated note-taking, pausing their VR exploration to consult the paper booklet, recording important findings or questions, and then re-entering VR for verification.

4.2.3. Emotional Engagement and Learning Interest

VR and Hybrid group students generally exhibited higher enthusiasm and engagement, maintaining a sense of novelty and focus. In contrast, the Paper group students' emotional response was notably flatter, with some showing signs of reading fatigue or zoning out during the latter half of the study period. Although a VR group student (V19) expressed reservations about the VR headset causing stuffiness and dizziness over time, the overall consensus confirmed that VR's immersive and interactive nature effectively boosts students' learning interest and participation.

5. Discussion and Synthesis of Data

This discussion systematically interprets the empirical results reported in the preceding chapter, examining the observed effects against the backdrop of relevant cognitive theories. The primary goal is to synthesize the quantitative and qualitative evidence to derive robust, practical implications for architectural pedagogy.

5.1. COGNITIVE MECHANISM OF HYBRID MEDIA SUCCESS

The dual success of the Hybrid group in both learning outcomes and media preference is robustly explained by the Dual Coding Theory (Paivio, 1971). As synthesized in Table 2, the Hybrid model strategically engages both cognitive pathways. VR facilitates the non-verbal 'Visual Imagery Code,' crucial for intuitive spatial comprehension, while the traditional paper booklet activates the 'Verbal Symbols

Code,' essential for the logical structurization and systematic encoding of knowledge. Qualitative data, detailed under the 'Cognitive Implication' column in Table 2, revealed that Hybrid group students efficiently switched between these two representational modes using VR to acquire experiential knowledge, and the paper drawings to structure and encode that knowledge. This strategy not only deepened comprehension and retention but also mitigated the cognitive load and physiological discomfort potentially associated with prolonged, exclusive use of VR.

5.2. INFLUENCE OF CASE TYPE ON MEDIA EFFICACY

A significant interaction was observed between the Media Group and Case Type. For the geometrically clear Villa Savoye, the immersive experience provided by pure VR offered sufficient assistance, with the additional benefit provided by the hybrid medium being minimal. However, in the Row House in Sumiyoshi, the advantage of the hybrid strategy was more pronounced, yielding a marginally significant difference over the pure VR group. Critically, this finding suggests that the paper component serves as a crucial cognitive stabilizer, preventing spatial disorientation and reducing load when architectural complexity challenges the efficacy of pure immersive media. This confirms that for highly abstract or complex spatial organization, the hybrid strategy is essential for comprehensive knowledge acquisition.

6. Conclusion

The principal academic contribution of this research lies in empirically substantiating an optimized pedagogical strategy for spatial comprehension, rigorously confirming the synergistic effect hypothesized under the Dual Coding Theory. The synthesis of qualitative findings (Table 2) alongside quantitative results provides a robust, evidence-based instructional framework. This study provides robust empirical validation for the superiority of a VR and paper-based hybrid instructional model over single-modality approaches in enhancing architectural case study learning for novice students. The quantitative data conclusively demonstrate that the Hybrid group achieved the highest mean scores in the case knowledge assessment and reported significantly higher media satisfaction compared to both standalone VR and Paper groups. This success underscores a crucial paradigm shift for architectural education: the imperative is not merely the adoption of digital tools, but the strategic leverage of the complementary cognitive functions afforded by mixed media.

Specifically, the hybrid model acts as an explicit implementation of the Dual Coding Theory, guiding educators to design materials that strategically prompt students to alternate between the non-verbal spatial immersion (VR) and the structured symbolic encoding (Paper). This validated hybrid framework offers a clear pathway for educators navigating the digital transformation, particularly for instructing complex architectural instances where traditional 2D representations are inherently inadequate. Furthermore, to ensure pedagogical sustainability, promoting VR instruction must address individual differences in student physiological adaptation. Appropriate guidance and alternatives, such as allowing scheduled short breaks or providing readily available supplementary paper materials, should be implemented. ChatGPT (OpenAI, 2025) was used to (1) to improve flow and (2) to correct errors in spelling and grammar.

References

Ceylan, S. (2020). Using Virtual Reality to Improve Visual Recognition Skills of First Year Architecture Students: A Comparative Study. In Proceedings of the 12th International Conference on Computer Supported Education (CSEDU) (Vol. 2, pp. 54–63). SciTePress.

Cho, J. Y., & Suh, J. (2022). The Architecture and Interior Design Domain-Specific Spatial Ability Test (AISAT): Its Validity and Reliability. Journal of Interior Design, 47(1), 7-25.

Li, T., Hu, H., Ma, H., Ma, J., & Li, Q. (2025). Using Virtual Reality to Enhance Learning Performance and Address Educational Resource Disparities in Architectural History Courses. Sustainability, 17(3), 866. DOI: 10.3390/su17030866

Lu, S.-T. (2022). Integrating VR interactive experiences with traditional imagery in architectural education to accommodate diverse learning preferences. Computers in Education, 13(4), 245–260.

Obeid, S., & Demirkan, H. (2023). The influence of virtual reality on design process creativity in basic design studios. Interactive Learning Environments, 31(4), 1841–1859.

OpenAI. (2025). ChatGPT Model: GPT-5 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/

Paivio, A. (1971). Imagery and Verbal Processes. New York: Holt, Rinehart and Winston.

Shanti, Z., & Al-Tarazi, D. (2023). Virtual Reality Technology in Architectural Theory Learning: An Experiment on the Module of History of Architecture. Sustainability, 15(23), 16394. DOI: 10.3390/su152316394

Zairul, M., Rasdi, R., & Hanafiah, D. (2023). Embracing hybrid post-pandemic teaching in architectural studios. Journal of Architectural Education Innovations, 12(1), 50–62. DOI: 10.1108/ARCH-11-2022-0251

Effects Of The Biophilic Design Density Of Transitional Spaces On Human Stress Recovery In A Virtual Reality Environment

Zhixian Li[1] and Sisi Zlatanova[2]
[1,2]*The University of New South Wales*
[1]*zhixian.li@unsw.edu.au, 0000-0003-2649-8137*
[2]*s.zlatanova@unsw.edu.au, 0000-0002-8766-0487*

Abstract. Previous research has demonstrated that biophilic design has a significant impact on human emotion and stress. However, most research has been limited to indoor and outdoor spaces, and there is still limited research on transitional spaces. The effects of different density biophilic designs on human stress have also not been adequately researched. Transitional spaces are areas located between indoor and outdoor environments, such as building entrance canopies and semi-open cafés. This study used virtual reality (VR) to conduct a between-subjects experiment with 40 participants to investigate the impact of biophilic design density levels on human stress recovery in transitional spaces. Four virtual reality environments, including one non-biophilic environment and three biophilic environments with different density levels, were created. After participants experienced the Trier Social Stress Test (TSST), they were randomly assigned to one of the virtual reality environments for stress recovery. The participants' physiological indicators, including heart rate, diastolic blood pressure, and systolic blood pressure, were collected during the experiment. Psychological indicators were measured using the State-Trait Anxiety Inventory (STAI). Generally, the experimental results indicated that the relationship between the density of biophilic environments and human stress recovery in transitional spaces follows a U-shaped dose-response relationship. A medium-density biophilic environment had the most effective stress recovery outcome, while both higher and lower densities resulted in weaker restorative effects.

Keywords. Biophilic Design, Stress Recovery, Transitional Space, Virtual Reality, Human Well-being, Physiological Indicators

1. Introduction

The built environment plays a critical role in shaping human perceptions, preferences and reactions (Li et al., 2025). Modern urban lifestyles often reduce contact with nature, contributing to increased stress-related conditions and poorer psychological health. Biophilic design has therefore emerged as a key approach for integrating natural systems into constructed environments. Rooted in the *Biophilia Hypothesis*, it emphasizes humans' innate connection to nature (Kellert & Wilson, 1993), and

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 215-224.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

incorporates elements such as daylight, trees, grass, vegetation, and natural landscapes into architectural design. Recent developments in this field highlight its potential to improve health, well-being, and productivity, with numerous studies demonstrating its benefits for enhancing perception, supporting creativity, reducing fatigue, and promoting stress recovery (Gillis & Gatersleben, 2015).

In indoor spaces, biophilic design effectively enhances individuals' psychological and behavioral performance. For instance, the introduction of natural elements into office environments has been shown to improve employee well-being, work performance, creativity, and overall health (Al Horr et al., 2016). In educational environments, natural elements help reduce student stress (Li et al., 2024). In commercial environments, biophilic elements have also been found to enhance customer pleasure and satisfaction. Beyond indoor environments, biophilic design plays an equally vital role in enhancing human well-being in outdoor spaces. For instance, hospital healing gardens incorporate biophilic design to create restorative experiences for patients, thereby promoting recovery. Moreover, in community environments, biophilic design not only encourages physical activity but also effectively alleviates stress and fosters social interaction (Peters et al., 2010).

However, existing research has primarily focused on traditional indoor and outdoor spaces, while limited attention has been given to the intermediate zones that connect them. According to Winnicott (1971), transitional spaces are zones situated between the internal and external realms, possessing characteristics of both. In the context of the built environment, transitional spaces generally refer to buffer and connective areas linking indoor and outdoor environments, such as entrance canopies, foyers, elevator lobbies, corridors, and stairwells, as shown in Figure 1 (Gehl, 2011; Yan et al., 2019).

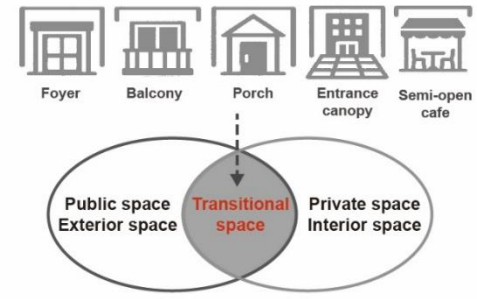


Figure 1. Definition of transitional space

Although numerous studies have demonstrated that biophilic design can promote emotional regulation and stress recovery, its effects within transitional spaces remain underexplored. Given that transitional spaces are frequently encountered in daily life and play an important role in shaping human environmental experience, understanding their restorative potential is essential (Li et al., 2022). Moreover, research examining how different densities of biophilic design influence human stress recovery in these spaces remains scarce. Therefore, this study aims to investigate the relationship between varying densities of biophilic design and human stress recovery within transitional spaces.

2. Research Methodology

2.1. EXPERIMENTAL ENVIRONMENT

The study opted to construct the experimental environment using Virtual Reality (VR) technology to achieve precise manipulation of the varying levels of biophilic design density. VR technology has been widely applied in environmental psychology research, as its primary advantage lies in its capacity to offer a highly immersive, realistic experience while maintaining a strictly controllable experimental setting (Gao & Mandryk, 2011). Compared to traditional methods, VR enables the precise isolation and control of key variables (e.g., spatial layout, lighting, and the density of biophilic elements), thereby ensuring the reproducibility and comparability of the experimental scenarios.

2.2. PHYSIOLOGICAL AND PSYCHOLOGICAL STRESS ASSESSMENT

This study employed a combined approach using physiological and psychological measurements to assess the participants' stress state and restoration more comprehensively. For the psychological assessment, the short-form of State-Trait Anxiety Inventory (STAI-S) was used to evaluate participants' anxiety and stress (Spielberger & Gonzalez-Reigosa, 1971; Marteau & Bekker, 1992). This is a widely used scale for measuring human stress. Furthermore, this study included three physiological indicators as objective criteria for stress recovery assessment: heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP). Multiple prior studies have demonstrated the efficacy of these indicators in measuring human stress recovery across different environments (Li et al., 2022).

3. Experimental Design

3.1. PARTICIPANTS

Forty healthy adults (>18 years) were recruited via campus announcements and social media. Participants with a history of mental illness or psychotropic medication use were excluded. All had normal or corrected-to-normal vision and no adverse reactions to VR. To minimize physiological confounding, participants avoided intense exercise, alcohol, and caffeine for 24 hours before testing. The study was approved by the UNSW Institutional Review Board (iRECS4558).

3.2. ENVIRONMENTAL SIMULATION

In this study, the VR environments were first constructed through 3D modeling in Rhinoceros, then imported into *Unity* software for immersive rendering and interactive programming. The final scenarios were presented to participants via an Oculus Quest 2 VR headset. This study employed a typical transitional space, a semi-open café as the base environment. Within this space, four experimental conditions were developed (Figure 2) to examine the effect of biophilic design density on stress recovery. These included one control scenario (non-biophilic space) and three experimental scenarios with low, medium, and high levels of biophilic density.

Scenario A (non-biophilic / control): Contained only standard café furnishings (tables, chairs, paved flooring) with no biophilic elements. **Scenario B** (low density): Added a small number of biophilic elements, including plant stands, sand landscaping, and a fountain. **Scenario C** (medium density): Added a moderate amount of biophilic elements, including plant stands, sand landscaping, a fountain, and rows of trees. **Scenario D** (high density): Added extensive biophilic elements, including plant stands, sand landscaping, a fountain, distant trees, and grass-covered ground.

A: non-biophilic



B: low density biophilic



C: medium density biophilic



D: high density biophilic



Figure 2. Four comparative experimental VR scenarios

3.3. EXPERIMENTAL PROCEDURE

The experimental procedure (Figure 3) lasted approximately 30 minutes. After reviewing instructions, providing consent, and completing demographic information, participants were equipped with HR and BP devices. EQ parameters (PM2.5, temperature, humidity, and CO_2) were continuously monitored.

Baseline Measurement. The experiment began with a 5-minute baseline measurement to acquire individual physiological indicators in a relaxed state. HR recording commenced at this point and continued until the end of the experiment. At the end of this phase, the baseline BP data (BP baseline) was recorded.

Stress Induction Phase. Then, participants performed the Trier Social Stress Test (TSST) to induce acute psychological and physiological stress (Kirschbaum, Pirke, & Hellhammer, 1993). This procedure involved two consecutive tasks: 5-minute speech and a 5-minute arithmetic task, with an audible alert for errors. Following the TSST, the pre-experimental BP data and pre-experimental STAI was measured immediately.

Stress Recovery Phase. Participants were randomly assigned to one of four VR scenarios for a 5-minute stress recovery. Previous research indicates that five minutes is sufficient for individuals to recover from acute stress (Suppakittpaisarn et al., 2023). Each participant experienced only one VR scenario to mitigate potential VR-induced discomfort (such as motion sickness or headaches), and performed the TSST only once to avoid potential residual stress effects.

Post-Recovery Measurement and Equipment Removal. The post-recovery BP (BP post) was immediately recorded, and participants completed the STAI-S questionnaire again (STAI post), after which the HR recording was terminated.

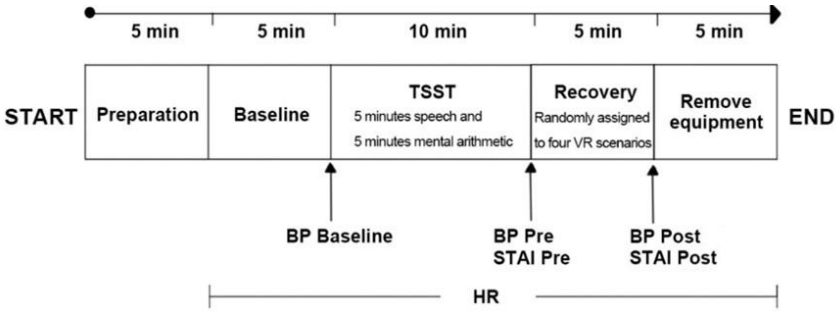


Figure 3. Experimental procedure

4. Data Analysis

All 40 participants completed the experiment. Data analysis was conducted using *SPSS Statistics* software. First, to verify the effectiveness of the random assignment, a one-way analysis of variance (ANOVA) was performed to examine whether there were significant differences among participants in their psychological and physiological baseline data before the TSST, as well as in the Environmental Quality (EQ) indicators recorded during the experiment. Second, to validate the effectiveness of the TSST stress induction procedure, paired-samples t-tests were conducted to compare participants' psychological and physiological indicators before and after the TSST. Furthermore, to evaluate the effect of different virtual scenarios on stress recovery, the difference between pre-test and post-test values (ΔD = pre-test − post-test) was calculated as the dependent variable and analyzed using a one-way ANOVA.

5. Results

5.1. PARTICIPANTS INFORMATION

Table 1 presents the demographic characteristics of the 40 participants. Each experimental condition included an equal number of participants, and the male female ratio was controlled across conditions to minimize potential gender effects on stress responses and recovery.

Table 1. Demographics of all participants

	Overall	A	B	C	D
Number of Participants	40	10	10	10	10
Age	28±3	28±3	27±2	28±4	29±4
Gender	--	--	--	--	--
Female	20	5	5	5	5
Male	20	5	5	5	5

5.2. EXPERIMENTAL RANDOMIZATION

HR, SBP, DBP, and STAI data from the pre-TSST and post-TSST periods were tested using ANOVA to validate the experimental randomization (Table 2). During the pre-TSST period, ANOVA results showed that there were no statistically significant differences in HR ($F = 0.863$, $p = 0.469$), SBP ($F = 0.794$, $p = 0.505$), and DBP ($F = 0.912$, $p = 0.445$) among participants. HR ($F = 0.732$, $p = 0.540$), SBP ($F = 0.052$, $p = 0.984$), DBP ($F = 0.563$, $p = 0.643$), and STAI ($F = 1.113$, $p = 0.356$) also showed no significant differences among participants during the post-TSST period. Therefore, the randomization of the experiment was considered valid.

In addition, due to the potential impact of EQs on human perception, EQ parameters were recorded once every five minutes and analyzed by a one-way ANOVA (Table 2). The results showed no significant differences in CO^2 ($F = 0.455$, $p = 0.719$), $PM_{2.5}$ ($F = 0.971$, $p = 0.438$), temperature ($F = 0.800$, $p = 0.517$), and humidity ($F = 2.111$, $p = 0.152$) across the four scenarios. Therefore, in this experiment, EQ was considered to have no significant effect on the experimental results.

Table 2. Test on whether participants' physical data and psychological data at baseline (pre-TSST), post-TSST and EQ among four environments were similar or not

Measures	Method	*F*	*df*	*P*-value
Pre-TSST				
Heart Rate (bpm)	ANOVA	0.863	3	0.469
Systolic Blood Pressure (mmHg)	ANOVA	0.794	3	0.505
Diastolic Blood Pressure(mmHg)	ANOVA	0.912	3	0.445
Post-TSST				
Heart Rate (bpm)	ANOVA	0.732	3	0.540
Systolic Blood Pressure (mmHg)	ANOVA	0.052	3	0.984
Diastolic Blood Pressure(mmHg)	ANOVA	0.563	3	0.643
State-Anxiety Inventory	ANOVA	1.113	3	0.356
EQ				
CO^2 (ppm)	ANOVA	0.455	3	0.719
$PM_{2.5}$ (μg/m3)	ANOVA	0.971	3	0.438
Temperature (℃)	ANOVA	0.800	3	0.517
Humidity (%)	ANOVA	2.111	3	0.152

Furthermore, the data collected before and during the TSST among the four scenarios were analyzed using paired-samples t-tests to measure the effectiveness of

the TSST (Table 3). The results showed that HR (T = -15.967, p < 0.001), SBP (T = -3.201, p = 0.003), and DBP (T = -3.404, p = 0.002) all showed significant differences before and during the TSST. Thus, participants' stress levels were effectively increased after the TSST experience.

Table 3. Test on whether 10-minutes of TSST produced significant stress on participants

Measures	Method	*T*	*df*	*P*-value
Heart Rate (bpm)	Paired t-test	-15.967	39	0.000
Systolic Blood Pressure (mmHg)	Paired t-test	-3.201	39	0.003
Diastolic Blood Pressure(mmHg)	Paired t-test	-3.404	39	0.002

5.3. STRESS RECOVERY EFFECTS

Figure 4 and 5 show the median, quartiles, and maximum and minimum values of pre-test and post-test data for HR, DBP, SBP and STAI.

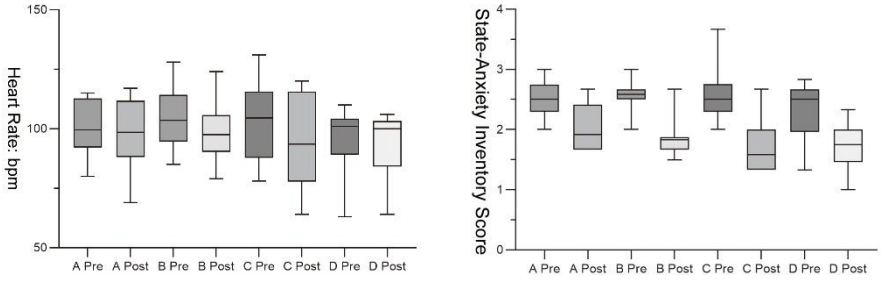


Figure 4. Pre-test and post-test of heart rate and blood pressure data among four scenarios

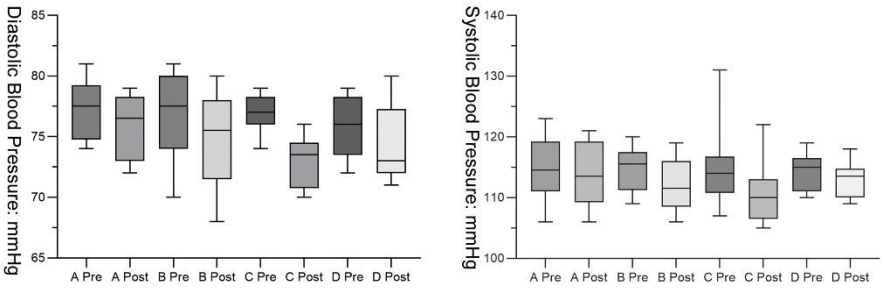


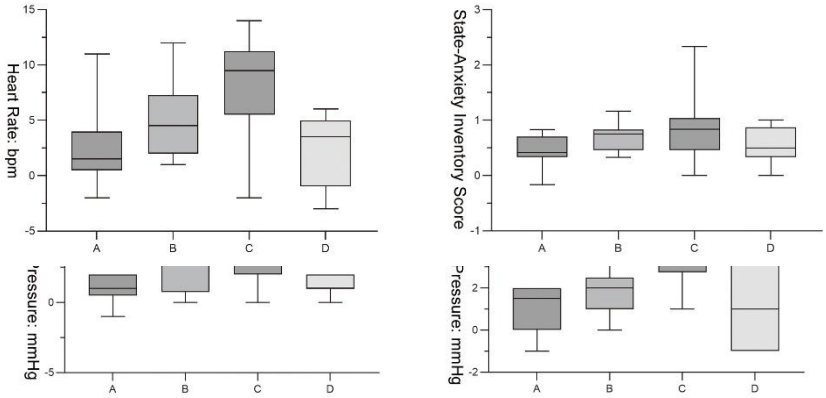
Figure 5. Pre-test and post-test of diastolic and systolic blood pressure data among four scenarios

The ΔD (ΔD = pre-test - post-test) of the psychological and physiological data was analyzed using a one-way ANOVA to investigate the differences in the stress recovery effects of the four scenarios on the participants (Figure 6 & Figure 7). Physiological data, i.e., HR, DBP and SBP showed statistically significant differences among the four scenarios (p < 0.05), while psychological data, i.e., STAI did not show statistically significant differences (p = 0.201). The ΔD of HR (F = 5.558, p = 0.003) showed that $C_{\text{medium-density}}$ ($M \pm SD = 8.300 \pm 4.644$) demonstrated the greatest stress recovery effect, followed by $B_{\text{low-density}}$ ($M \pm SD = 5.200 \pm 3.676$). $A_{\text{non-biophilic}}$ ($M \pm SD = 2.400 \pm 3.596$)

had poorer stress recovery than B $_{\text{low-density}}$, and the worst effect was observed in D $_{\text{high-density}}$ ($M \pm SD = 2.200 \pm 3.327$). The results of one-way ANOVA for STAI ΔD ($F = 1.623$, $p = 0.201$) showed that C $_{\text{medium-density}}$ ($M \pm SD = 0.850 \pm 0.631$) had the best stress recovery and the worst was A $_{\text{non-biophilic}}$ ($M \pm SD = 0.467 \pm 0.302$). B $_{\text{low-density}}$ ($M \pm SD = 0.700 \pm 0.258$) and D $_{\text{high-density}}$ ($M \pm SD = 0.583 \pm 0.326$) showed moderate stress recovery effects.

Figure 6. Δ D one-way ANOVA results for HR and STAI among the four scenarios

The results of the one-way ANOVA for SBP ΔD ($F = 5.562$, $p = 0.003$) indicated that C $_{\text{medium-density}}$ ($M \pm SD = 4.400 \pm 3.239$) achieved the greatest stress recovery among the four scenarios, followed by B $_{\text{low-density}}$ ($M \pm SD = 2.800 \pm 2.300$). D $_{\text{high-density}}$ ($M \pm$



$SD = 1.300 \pm 0.675$) showed poorer stress recovery than scenarios B and C, while A $_{\text{non-biophilic}}$ ($M \pm SD = 1.000 \pm 1.155$) exhibited the weakest recovery effect. ΔD of DBP ($F = 4.553$, $p = 0.008$) results showed that C $_{\text{medium-density}}$ ($M \pm SD = 4.100 \pm 1.969$) had the best stress recovery, followed by B $_{\text{low-density}}$ ($M \pm SD = 2.200 \pm 1.989$), D $_{\text{high-density}}$ ($M \pm SD = 1.500 \pm 2.550$) and A $_{\text{non-biophilic}}$ ($M \pm SD = 1.100 \pm 1.101$), respectively.

Figure 7. Δ D one-way ANOVA results for SBP and DBP among the four scenarios

6. Discussion

This study investigated the effects of different biophilic densities in transitional spaces on human stress recovery. Forty participants were randomly assigned to four VR scenarios: a non-biophilic space (control), low-density, medium-density, and high-density biophilic spaces. During the experiment, both psychological data (STAI) and physiological data (HR, DBP, SBP) were collected to evaluate the recovery effect across scenarios. The results revealed significant differences in HR, DBP, and SBP among the four scenarios ($p < 0.05$), indicating that biophilic density significantly affected physiological recovery, while the STAI differences were not statistically significant. Overall, the medium-density biophilic space demonstrated the most effective stress recovery, followed by the low-density, high-density, and non-biophilic conditions. These findings align with most previous studies conducted in indoor and outdoor environments, suggesting an inverted U-shaped relationship between biophilic

density and stress recovery (Tekin et al., 2022; Suppakittpaisarn et al., 2023). Moderate biophilic elements most effectively promote physical and mental recovery, whereas excessive biophilic elements may induce visual clutter, a sense of spatial closure, and perceptual overload, thereby reducing their restorative effects.

This pattern can be interpreted through several environmental psychology frameworks. Specifically, the *Prospect–Refuge Theory* (Appleton, 1975) posits that humans are biologically inclined to seek a balance between prospect (an unobstructed view allowing for the detection of opportunities or threats) and refuge (a protected location offering concealment and safety). Medium-density biophilic spaces achieve an ideal equilibrium between openness and enclosure, promoting comfort and safety. In contrast, low-density or non-biophilic spaces may lack sufficient refuge, while high-density spaces may obstruct prospect. Additionally, *Environmental Load Theory* (Berlyne, 1971) provides a cognitive-based explanation. This framework posits that the level of environmental complexity is critical: environments that are too simple fail to provide the necessary stimulation for positive affective responses, whereas environments that are excessively complex increase the individual's cognitive and sensory load, ultimately hindering restorative processes. Thus, the medium-density biophilic space, with its optimal level of stimulus complexity, is best suited for facilitating both psychological and physiological stress recovery.

7. Conclusion

This study investigated the effects of varying biophilic design densities within transitional spaces on human stress recovery, in response to the growing environmental stress and health challenges associated with urbanization. The results revealed that the medium-density biophilic environment produced the most effective stress recovery, compared with low-density and high-density settings. This finding presents an inverted U-shaped relationship and can be understood through environmental psychology perspectives. Medium-density environments offer the optimal degree of sensory stimulation, which is greater than that provided by low-density environments, while also reducing the perceptual overload caused by excessive complexity in high-density environments. These findings offer important empirical insights for the biophilic design of transitional spaces. They demonstrate that optimizing the density of natural elements can enhance the restorative potential of built environments and contribute positively to public mental well-being in rapidly urbanizing contexts.

References

Al Horr, Y., Arif, M., Kaushik, A., Mazroei, A., Katafygiotou, M., & Elsarrag, E. (2016). Occupant productivity and office indoor environment quality: A review of the literature. Building and Environment, 105, 369–389. https://doi.org/10.1016/j.buildenv.2016.05.024

Appleton, J. (1975). *The experience of landscape*. London: Wiley.

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.

Gao, L., & Mandryk, R. L. (2011). The acute effects of immersive virtual reality on cognitive performance. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2657–2666. https://doi.org/10.1145/1978942.1979336

Gehl, J. (2011). *Life between buildings: Using public space* (6th ed.). Washington, DC: Island Press.

Gillis, K., & Gatersleben, B. (2015). A review of the psychological literature on the health and wellbeing benefits of biophilic design. *Buildings*, 5(3), 948–963. https://doi.org/10.3390/buildings5030948

Kellert, S. R. (2008). Dimensions, elements, and attributes of biophilic design. In S. Kellert, J. Heerwagen, & M. Mador (Eds.), *Biophilic design: The theory, science, and practice of bringing buildings to life* (pp. 3–19). New York, NY: Wiley.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'—A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1–2), 76–81. https://doi.org/10.1159/000119004

Li, Z., Huang, X., & White, M. (2022). Effects of the visual character of transitional spaces on human stress recovery in a virtual reality environment. *International Journal of Environmental Research and Public Health*, 19(20), 13143. https://doi.org/10.3390/ijerph192013143

Li, Z., Lee, J. H., Yao, L., & Ostwald, M. J. (2025). Impact of built environments on human perception: A systematic review of physiological measures and machine learning. *Journal of Building Engineering*, 112319.

Li, Z., Zu, X., Lee, J. H., & Ostwald, M. J. (2024). The impact of visual character on perceived stress levels: An intelligent approach applied to university campus design. In *Multimodality in Architecture: Collaboration, Technology and Education* (pp. 229–244). Cham: Springer Nature Switzerland.

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the STAI (STAI-S). *British Journal of Clinical Psychology*, 31(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

Peters, K., Elands, B., & Buijs, A. (2010). Social interactions in urban parks: Stimulating social cohesion? *Urban Forestry & Urban Greening*, 9(2), 93–100. https://doi.org/10.1016/j.ufug.2009.11.003

Spielberger, C. D., & Gonzalez-Reigosa, F. (1971). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Suppakittpaisarn, P., Browning, M. H. E. M., & Sullivan, W. C. (2023). Vegetation density and restorative benefits: Evidence from a cross-cultural virtual reality experiment. *Landscape and Urban Planning*, 232, 104685. https://doi.org/10.1016/j.landurbplan.2023.104685

Tekin, B. H., Manavoglu, E., & Porras, E. (2022). A systematic review and conceptual framework of biophilic design in the built environment. *International Journal of Environmental Research and Public Health*, 19(9), 5338. https://doi.org/10.3390/ijerph19095338

Winnicott, D. W. (1971). *Playing and reality*. London, UK: Tavistock.

Yan, J., Diakité, A. A., Zlatanova, S., & Aleksandrov, M. (2019). Top-bounded spaces formed by the built environment for navigation systems. *ISPRS International Journal of Geo-Information*, 8(5), 224.

Rhizomatic Learning In Architectural Education

Integrating Humanistic Computation and Cultural Contexts through Case-Based Inquiry

Anetta Kepczynska-Walczak[1] and Bartosz M. Walczak[2]
[1,2]*Lodz University of Technology*
[1]*anetta.kepczynska-walczak@p.lodz.pl, 0000-0003-4125-2012*
[2]*bartosz.walczak@p.lodz.pl, 0000-0002-9429-9626*

Abstract. Architectural education is increasingly influenced by digital technologies and AI, yet its pedagogical models often remain hierarchical. This paper explores how rhizomatic learning – a non-linear, networked model – can enhance architectural pedagogy in computationally enriched environments. It examines how humanistic computation and culture encoding support adaptive, inclusive learning ecosystems. Through a qualitative, multiple-case study approach, two courses were analyzed: a heritage-focused graduate studio and a computational design course integrating BIM, algorithmic thinking, and AI. The comparison revealed complementary strengths – cultural depth in the first, technical proficiency in the second – alongside notable gaps. The study concludes that a balanced rhizomatic pedagogy, where computation serves humanistic inquiry rather than technical ends, offers a robust path toward reflective and ethically engaged design education in the age of AI.

Keywords. Rhizomatic Learning, Computational Design Education, Humanistic Computation, Culture Encoding, Architectural Pedagogy

1. Introduction

1.1. BACKGROUND

G. W. Leibniz, often called “the last universal genius,” symbolized an era when one person could grasp the entirety of human knowledge (Look, 2008). Since then, scientific knowledge has grown exponentially. Already in the 1960s, D. J. de Solla Price estimated that it doubled every 10–15 years (de Solla Price, 1963). Today, this acceleration – combined with the volatility and complexity of the VUCA world – has profound implications for education.

In such a context, the focus shifts from memorizing facts to cultivating meta-skills like creativity, critical thinking, and resilience (Minciu et al., 2025). Traditional, hierarchical pedagogies are increasingly inadequate. Rhizomatic learning, based on Deleuze and Guattari’s (1987) concept of the rhizome, offers a non-linear, networked

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 225-234.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

model of knowledge formation. Learners co-create knowledge through interaction, navigating ambiguity and complexity creatively. Studies confirm its value in fostering adaptability and inclusive learning (Gutiérrez-Ujaque, 2024; de Vries et al., 2022; Jahani et al., 2023).

In architectural education, however, a paradox remains. Despite the growing use of digital tools and AI, teaching methods often follow linear, siloed structures. Design studios and technical courses are separated, limiting opportunities for integrated, reflective learning. This gap calls for pedagogical models that embrace the networked nature of digital environments.

1.2. PROBLEM STATEMENT

Architectural education must reconceptualize the relationship between computational tools and learning structures. Current models isolate computation from cultural and conceptual dimensions, resulting in fragmented learning experiences that do not reflect the complexity of contemporary design practice.

1.3. AIM OF THE PAPER

This paper explores how rhizomatic learning can reshape architectural pedagogy through two case studies: a heritage-focused graduate design studio and a computational design course integrating BIM, algorithmic thinking, and AI. By analyzing their strengths and limitations, the paper proposes a more integrated model that merges humanistic, cultural, and technical dimensions of architectural education.

2. Theoretical Framework

2.1. RHIZOMATIC LEARNING

Rhizomatic learning, based on Deleuze and Guattari's (1987) concept of the rhizome, challenges hierarchical models of education by promoting non-linear, networked knowledge formation. In this approach, learners co-create dynamic curricula shaped by interest, context, and collaboration (Cormier, 2008). It emphasizes adaptability, diversity, and emergence – qualities essential in complex, unpredictable environments.

2.2. COMPUTATIONAL DESIGN AND RHIZOMATIC THINKING

Computational design tools such as BIM, parametric modeling, and AI inherently support rhizomatic structures through iterative, feedback-driven processes. However, architectural education often treats these tools as procedural rather than conceptual, limiting their potential to foster critical and creative engagement (Lin & Hou, 2022; Tabrizi et al., 2023). Embedding rhizomatic principles into digital pedagogy can bridge this gap.

2.3. HUMANISTIC COMPUTATION AND CULTURAL ENCODING

Humanistic computation foregrounds ethical, interpretive, and cultural dimensions of digital practice (Asino et al., 2017). Cultural encoding involves integrating local and

historical knowledge into computational workflows, enriching design with diverse epistemologies. Together, these frameworks support inclusive, reflective learning environments where computation becomes a medium for cultural and humanistic inquiry.

2.4. TOWARDS AN INTEGRATIVE FRAMEWORK

The intersection of rhizomatic learning, humanistic computation, and cultural encoding suggests a new paradigm for architectural pedagogy: one where digital tools and networks act as connective tissues between diverse knowledge domains, rather than isolated technical modules. In this integrative framework, computation supports the creation of distributed, adaptive, and culturally situated learning ecosystems. The role of the educator shifts from instructor to facilitator, curating conditions for emergent connections among students, technologies, and cultural contexts.

This theoretical foundation informs the comparative analysis of two pedagogical experiments presented in the following sections: (1) a graduate design studio engaging with heritage and context, and (2) a set of computational design classes introducing BIM, algorithmic thinking, and AI. The analysis applies core principles of rhizomatic learning – multiplicity, connectivity, heterogeneity, and decentralization – to assess how each environment enables or constrains the emergence of networked, humanistic, and culturally embedded learning.

3. Methodology

3.1. RESEARCH APPROACH

This research adopts a comparative, interpretive case study approach to examine how rhizomatic learning principles can inform computationally supported architectural education. The study does not aim to measure learning outcomes quantitatively, but rather to critically interpret the pedagogical dynamics and structural conditions that enable or hinder rhizomatic learning. It combines conceptual framework mapping with empirical observation and qualitative reflection drawn from two pedagogical settings led by the authors. Such an approach follows Cresswell's (2013) notion of theoretical sampling, in which cases are selected for their conceptual richness and their ability to illuminate broader theoretical questions rather than represent a statistical population. In this study, the cases are chosen to embody contrasting relationships between computation and rhizomatic learning – one emphasizing cultural and conceptual exploration, the other emphasizing technical and algorithmic development.

3.2. CASE STUDY SELECTION

The two selected case studies represent complementary pedagogical environments within architectural education: (1) design studio; (2) computational design classes. Together, these cases enable an examination of the tension between conceptual rhizomatic learning and computational implementation – revealing the challenges of integrating both dimensions within a coherent pedagogical ecosystem.

3.3. ANALYTICAL FRAMEWORK

The analysis employs a conceptual matrix derived from the four key principles of rhizomatic learning proposed by Deleuze and Guattari (1987) and adapted for educational contexts by Cormier (2008), as shown in the Table 1:

Table 1. Analytical Framework

Rhizomatic Principle	Educational Dimension	Indicators in Architectural Context
Multiplicity	Diversity of learning pathways and outputs	Variety of representational media, disciplinary references, and computational methods used
Connectivity	Degree of networked collaboration	Peer learning, cross-project exchanges, integration of shared digital platforms
Heterogeneity	Inclusion of different epistemologies	Blending of technical, cultural, and interpretive approaches
Decentralization	Learner agency and non-hierarchical knowledge formation	Distributed authorship, co-creation, and instructor-as-facilitator dynamics

Each case study is examined through the aforementioned criteria, supported by qualitative evidence including studio documentation, student work, reflective reports, and teaching observations. The framework enables a structured comparison of how rhizomatic principles manifest – or fail to manifest – within distinct pedagogical and computational configurations. The operationalization of indicators in the project under discussion involves the examination of their occurrence in individual case studies (analyzed courses). This will enable the determination of their fulfillment of the characteristics of a rhizomatic teaching process.

3.4. ANALYTICAL PROCESS

The analysis proceeds in three stages:

- Descriptive Mapping – Documenting the organizational structure, learning tasks, and digital tools used in each case.
- Interpretive Evaluation – Assessing each case through the rhizomatic framework to identify patterns of connectivity, multiplicity, and decentralization.
- Critical Synthesis – Comparing findings to reveal systemic gaps and potential complementarities, leading to design principles for a more integrated model of computationally supported rhizomatic pedagogy.

This interpretive methodology enables the study to remain conceptually grounded while offering actionable insights for rethinking how computational learning environments can more effectively support rhizomatic, humanistic, and culturally embedded approaches to architectural education.

4. Results and discussion

4.1. CASE STUDY 1: HERITAGE DESIGN STUDIO

This course was part of a graduate architecture module and served as a pilot for a living lab within the Horizon Europe WoodStock project. The module comprised lectures and seminars, as well as a design studio, with a focus on the conservation of built heritage. The 15-week design studio was valued at 6 ECTS credits, equivalent to approximately 180 hours of student effort. The students were presented with a design problem that had to be solved in a real-life setting. The task entailed the conceptualization and design of an infill structure within a densely populated urban area that was of considerable heritage significance. Additionally, the necessity to utilize bio-based building materials, particularly wood, posed a significant challenge. Ultimately, the students were required to present their ideas to multiple tutors, thereby simulating the intricacy of the authentic design process. The utilization of digital tools and computation design did not constitute a primary focus of the course.



Figure 1. Heritage design studio. Visualizations of the concepts focusing on the use of timber (left) and the sensitive new architectural forms (right).

The rhizomatic nature of the course was evident in the manner in which the students approached the task. A subset of the population was attracted to exploring the opportunities offered by building materials. In contrast, another subset focused more on the impact of the heritage context on the dialogue between old and new architectural forms (Figure 1). Students found the variety of feedback from instructors with differing backgrounds and areas of expertise to be a significant challenge. These instructors specialized in various aspects of the built heritage, including urban regeneration, adaptive reuse, wood construction, and preservation. It was imperative that the students responded to these comments with critical analysis and provided a rationale for their design decisions.

4.2. CASE STUDY 2: COMPUTATIONAL DESIGN CLASSES

As part of the undergraduate program, the course focused specifically on BIM for sustainable design, aligning with the New European Bauhaus Academy Alliance's requirements. The 15-week course was worth 3 ECTS credits. It consisted of two hours of lab work and one hour of e-learning per week, as well as self-study. The course aimed to introduce students to BIM, energy analysis, and AI-assisted visualization. Accordingly, the course was structured into three interconnected modules. The first module was delivered in an e-learning format and introduced students to the fundamental concepts and workflows of BIM; the basics of algorithmic thinking in relation to BIM processes; and the application of simulation tools for designing low-energy buildings within a BIM environment (Figure 2). The second module, conducted in a laboratory setting, provided hands-on experience designing architectural components with BIM software. This module included creating interactive schedules for building elements and using analytical tools for sun path analysis and building performance evaluation. The final module required students to develop an information model of a low-energy architectural object of intermediate complexity. This phase emphasized presenting design solutions using BIM and AI tools and culminated in critically evaluating the outcomes in the context of sustainable design principles (Figure 3).

As is evident, students were presented with the opportunity to acquire a variety of professional competencies and insights. Participants were granted autonomy in determining the extent to which they would delve into specific issues. Consequently, the mapping of these interconnections exhibited a truly rhizomatic character. This phenomenon was particularly evident during the students' dialogue with the AI system to obtain the desired visualization of the designed building.

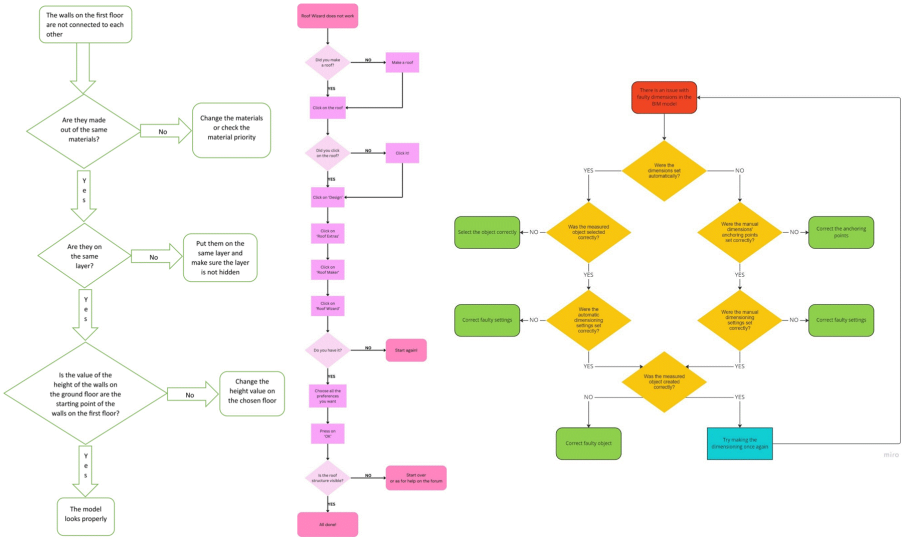


Figure 2. Computational Design Classes – student-created algorithms illustrating BIM workflows. Task: define a BIM-related problem and visualize its solution using ANSI/ISO-compliant flowcharts based on personal project experience.

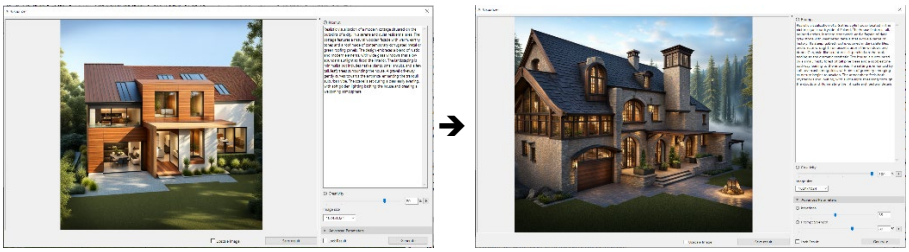


Figure 3. Computational Design Classes – AI-generated visualizations showing iterative human–machine interaction. Students tested ArchiCAD's AI Visualizer using prompts and scale adjustments based on their BIM models.

4.3. COMPARATIVE DISCUSSION

The two case studies represent distinct yet complementary applications of rhizomatic learning principles within architectural education. While both courses encouraged non-linear, self-directed, and networked learning, their pedagogical foci and integration of computational tools differed substantially, revealing both the potential and the limitations of rhizomatic learning in computationally enhanced environments.

4.3.1. Learning Structures and Rhizomatic Patterns

In the Heritage Design Studio, the rhizomatic character emerged organically from the diversity of student approaches and feedback loops. Students navigated multiple knowledge nodes – heritage theory, material experimentation, contextual analysis, and stakeholder dialogue – without a singular prescribed trajectory. Knowledge was co-constructed through iterative critique sessions involving tutors from diverse disciplinary backgrounds. However, computation here served primarily as a representational medium rather than an epistemic or generative tool. As such, the rhizomatic network remained rooted in conceptual and cultural exploration, while computational engagement was peripheral.

Conversely, in the Computational Design Classes, the rhizome was more deliberately structured. The modular organization of the course – spanning BIM, algorithmic thinking, and AI-assisted visualization – provided semi-autonomous "learning nodes" that students could explore to varying depths. The open-endedness of tool engagement, particularly in the dialogue between students and AI systems for visualization or data interpretation, reflected a more explicit enactment of rhizomatic learning in a digital environment. Students were encouraged to establish their own connections among the modules, thereby constructing personalized learning pathways that blended computation and design thinking.

4.3.2. Humanistic Computation and Culture Encoding

A key divergence between the two case studies lies in the integration of humanistic computation. In the Heritage Design Studio, computation did not mediate the cultural or ethical dimensions of design but functioned as a technical instrument. Humanistic dimensions – context, heritage, and materiality – were foregrounded, yet they were processed through human judgment and discourse rather than computational

reasoning. By contrast, in the Computational Design Classes, human–machine interaction itself became a locus of cultural negotiation. Students' dialogue with AI tools – especially in generating visualizations or simulating environmental performance – revealed how computation can embody humanistic inquiry when used reflexively. The act of questioning, modifying, and critically interpreting AI outputs transformed computation from a neutral instrument into a collaborator in meaning-making. This dynamic aligns with Deleuze and Guattari's (1987) view of the rhizome as an assemblage where human and non-human agents co-produce knowledge.

4.3.3. Limitations and Lessons Learned

The results of this study should be interpreted with caution due to the presence of confounding factors between cases, such as credits (6 ECTS vs. 3 ECTS), instructor teams, student level (graduate vs. undergraduate), and modalities (studio vs. modular labs). Consequently, this study employs a theoretical sampling approach rather than a causal comparison method.

Despite their success in fostering autonomy and exploration, both courses exhibited limitations when examined through a rhizomatic lens. In the Heritage Design Studio, the lack of structured computational integration limited the students' ability to map and visualize their learning trajectories. The result was a fragmentation of knowledge nodes without sufficient computational mediation to expose or connect them. Introducing parametric or generative tools – capable of mapping relationships between heritage elements, materials, and design intentions – could reinforce the rhizomatic network and make cultural encoding computationally tangible.

In the Computational Design Classes, the opposite challenge occurred: computation dominated at the expense of cultural or critical reflection. Students navigated digital tools with autonomy but often without sufficient grounding in ethical or contextual dimensions of design. The risk, therefore, was the reduction of rhizomatic exploration to technical iteration rather than meaningful inquiry. Integrating reflective practices – such as design diaries, peer discussions, or culturally situated case examples – could strengthen the humanistic and interpretive aspects of computation.

The comparison of the two case studies, as shown in the Table 2 (below), reveals that rhizomatic learning in architectural education can emerge both organically and through deliberate computational structuring. The Heritage Design Studio demonstrated how cultural, ethical, and contextual dimensions can generate a spontaneous rhizome of inquiry, yet lacked the computational mechanisms to visualize or extend these connections. Conversely, the Computational Design Classes established a more explicit digital rhizome through modular learning nodes, but often at the expense of cultural reflection and critical interpretation. Together, these cases illustrate that neither cultural immersion nor technical mastery alone fulfils the potential of rhizomatic pedagogy. An integrated model – where computation serves as a medium for humanistic exploration and reflection – could transform architectural education into a truly adaptive, inclusive, and networked ecosystem. Such a synthesis directly aligns with the vision of advancing computational design through culturally and pedagogically informed frameworks. These observations pertain to the total duration of both courses and encompass all operationalized indicators of multiplicity, connectivity, heterogeneity, and decentralization, as delineated in the methodology.

Table 2. Comparison of the Two Case Studies through a Rhizomatic Learning Lens

Dimension	Heritage Design Studio	Computational Design Classes	Key Insights
Educational Focus	Graduate studio on heritage and infill architecture	Undergraduate course on BIM, algorithms, and AI	Rhizomatic learning applies across levels; focus shapes its expression
Structure	Open-ended, multi-tutor studio	Modular course with digital tools (BIM, algorithms, AI)	Organic vs. structured rhizomatic networks
Computation Role	CAD, Visualization tool	Central to modeling and analysis	Balance needed between representation and reflection
Learning Patterns	Emergent, feedback-driven	Modular, student-directed	Both require mechanisms for reflection and connection
Humanistic & Cultural Integration	Strong cultural focus, low digital mediation	Some human–machine dialogue, limited cultural depth	Computation should support cultural inquiry
Student Agency	High autonomy via critique	Autonomy via tools and labs	Digital tools enhance collaboration and traceability
Limitations	Weak computational integration	Limited cultural reflection	Integration of both dimensions is essential
Potential Enhancements	Add parametric/AI tools for heritage mapping	Embed reflective practices in digital workflows	Mutual enrichment strengthens pedagogy
Pedagogical Contribution	Spontaneous, culturally embedded rhizome	Structured, digitally mediated rhizome	Hybrid model offers adaptive, inclusive education

5. Conclusions

The comparative study demonstrates that rhizomatic learning offers a productive framework for rethinking architectural education in the digital age. Both case studies revealed that when learning unfolds through multiple, interconnected pathways, students gain agency and develop adaptive strategies for navigating complexity. Yet they also exposed a persistent gap between cultural contextualization and computational integration: heritage-oriented studios tend to underuse computation as a generative medium, while computational courses risk losing touch with the humanistic and reflective dimensions of design.

Bridging this gap requires an integrated rhizomatic model in which computation becomes a partner in inquiry – an active participant in shaping, visualizing, and reflecting on cultural and ethical dimensions of architecture. Such an approach can transform digital tools from instruments of efficiency into catalysts of critical and collaborative thinking. Looking ahead, further research should explore how AI-driven systems, data visualization, and generative models can support plural, culturally embedded learning trajectories. Rhizomatic pedagogy thus points toward a more inclusive, reflective, and resilient architecture curriculum capable of evolving with the

technological and societal transformations of the twenty-first century.

Attribution Statement

ChatGPT (OpenAI, 2024) and Copilot (GitHub, 2024) were used to (1) suggest the initial structure of the manuscript and (2) to improve flow. DeepL (DeepL SE., 2024) was used to correct errors in spelling and grammar.

References

Asino, T.I., Giacumo, L.A., & Chen, V. (2017). Culture as a design "next": Theoretical frameworks to guide new design, development, and research of learning environments, *The Design Journal*, 20(1), 875–885.

Cormier, D. (2008). Rhizomatic education: Community as curriculum. *Innovate: Journal of Online Education*, 4(5).

Creswell, J.W. (2013). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

DeepL SE. (2024). DeepL Write [AI writing assistant]. https://www.deepl.com/write.

Deleuze, G., & Guattari, F. (1987). *A Thousand Plateaus: Capitalism and Schizophrenia*. University of Minnesota Press.

de Solla Price, D.J. (1963). *Little Science, Big Science*, New York: Columbia University Press.

de Vries, E. W., Delnooz, P., Velthuijsen, H., & Pinxten, R. (2022). Addressing creativity in higher education from a rhizomatic perspective. *Innovations in Education and Teaching International*, 60(6), 906–917. https://doi.org/10.1080/14703297.2022.2103583.

GitHub (Microsoft). (2024). GitHub Copilot [AI code generation tool]. https://github.com/features/copilot.

Globa, A., Yildirim, M., Gough, P., & Weir, S. (2025). Reducing Barriers to Learning Digital Design Tools: Improving practices of teaching and learning software in architecture and design. In *ARCHITECTURAL INFORMATICS - Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA)*, Tokyo.

Jahani, J., Darabi Emarati, A., Marzooqi, R., & Shafiee Sarvestani, M. (2023). Rhizomatic-based educational model and its effect on creative thinking. *The Journal of New Thoughts on Education*, 19(2), 7-33. https://doi.org/10.22051/jontoe.2021.25369.2593.

Le Grange, L.L.L. (2011). Sustainability and Higher Education: From arborescent to rhizomatic thinking. *Educational Philosophy and Theory*, 43(7), 742–754. https://doi.org/10.1111/j.1469-5812.2008.00503.x.

Lin, H.T., & Hou, J.H. (2022). Exploring the Topological System of Dougong. In *POST-CARBON - Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA)*, Hong Kong

Look, B.C. (2008). *Gottfried Wilhelm Leibniz*. *Stanford Encyclopedia of Philosophy.* Retrieved November 10, 2025, from https://plato.stanford.edu/entries/leibniz/.

Minciu, M., Veith, C., Dobrea, R. C., & Ciocoiu, C. N. (2025). The Challenges of the VUCA World and the Education System: The Need for Change to Ensure Sustainable Learning Process. *Sustainability*, 17(14), 6600. https://doi.org/10.3390/su17146600.

OpenAI. (2024). ChatGPT Model: GPT-4 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/.

Tabrizi Toktam B., Gocer, O., Sadrieh, A., & Globa, A. (2023). Leveraging AI to instruct architecture students on circular design techniques and life cycle assessment. In *9th International Conference on Higher Education Advances (HEAd'23)*. Universitat Politècnica de València.

Evaluating a Simulation Game for Energy-Efficient Design

Computational Thinking and Usability Outcomes in a University Course

Jayedi Aman[1], Delaney O'Brien[2], Katey Lam[3], Laura B. Cole[4], Jong Bum Kim[5], Laura Zangori[6], Anindita Laz Banti[7], Pegah Sepehri[8]
[1,5,6,7,8]*University of Missouri Columbia,* [2,3,4]*Colorado State University.*
1jayediaman@missouri.edu, 0000-0002-6128-8293
2d.obrien@colostate.edu, 0009-0005-6710-7130
3Katey.Lam@colostate.edu, 0009-0003-1648-1291
4laura.cole@colostate.edu, 0000-0001-5730-1881
5kimjongb@missouri.edu, 0000-0001-8762-9554
6zangoril@missouri.edu, 0000-0002-7512-5559
7alkxz@missouri.edu, 0009-0006-5931-7661
8psxh2@missouri.edu, 0009-0008-1814-9977

Abstract. This study explores a simulation-based learning game to support computational-thinking engagement in a university-level architectural technology course. Originally developed for K-12 education, the game was adapted to facilitate iterative design reasoning, data interpretation, and performance-cost trade-off exploration within constrained instructional time. A mixed-methods approach combined pre/post surveys of self-reported computational-thinking confidence and system usability with student interviews. Results indicate increased engagement with computational-thinking processes alongside positive usability perceptions supporting iterative exploration. Qualitative findings illustrate how students articulated abstraction, problem decomposition, and iterative reasoning during gameplay. As a single-course pilot relying on self-reported measures, the findings provide insight into students' engagement with computational reasoning in a game-based learning context.

Keywords. Game-based learning, computational thinking, usability, learning experience, green building education

1. Introduction

1.1. BACKGROUND

Game-based learning has been an effective approach for engaging students in complex design and scientific concepts through direct interaction and feedback (Ding and Yu, 2024). In architectural education, such methods link design choices with measurable environmental outcomes, helping learners understand relationships among form,

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 235-244.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

material, energy performance, and cost. These interactive environments can promote analytical reasoning and systems thinking that are central to design education. The Build it Green (BIG) simulation game was originally developed for middle-school classrooms to introduce energy-efficient design principles through a simplified framework combining physics-based simulation, machine-learning prediction, and cost estimation.

Although the BIG simulation game (Kim et al., 2025) was developed for middle-school contexts, its application beyond that setting remains underexplored. Extending its use to higher education is important because few accessible tools in undergraduate curricula explicitly link design decisions with energy performance. University-level design courses require more advanced reasoning, quantitative interpretation, and evidence-based decision making. Embedded within a higher-education assignment, the game may support these skills by providing an iterative environment that connects design reasoning with data-driven outcomes. While prior studies on game-based learning emphasize engagement and motivation, fewer examine how such tools shape students' reasoning during design tasks. This study addresses that gap by evaluating the adaptability of the BIG simulation game in a university sustainable technologies course.

This study adopts two complementary lenses, computational thinking (CT) and usability, to evaluate interaction quality and students' self-assessed computational problem-solving. Drawing on a CT modelling framework (Yadav et al., 2019), CT is used to examine how learners structure and refine reasoning through abstraction, decomposition, pattern recognition, and debugging during gameplay. Usability (Lund, 2001) focuses on students' perceptions of clarity, ease of use, and responsiveness of the digital environment. Together, these dimensions provide complementary perspectives on student experience in game-based design tasks.

This paper describes the CT and usability framework, course context, data collection, and analysis methods, followed by key findings and implications for adapting simulation-based games in architectural education. Attention is given to how students articulated and applied computational reasoning within the game environment.

1.2. PURPOSE, RESEARCH QUESTIONS, AND HYPOTHESES

The purpose of this study is to evaluate the adaptation of the BIG simulation game for higher education by examining its capacity to support students' reasoning processes and provide a usable interface. The study investigates whether the simulation functions effectively within a university design context by focusing on reasoning engagement and user experience as indicators of adaptation success. The primary research questions are: (1) Do students report higher levels of CT-related confidence after engaging with the game-based simulation? and (2) Do students rate the game interface as usable? The hypotheses are: (H1) students will report higher levels of CT-related confidence following gameplay, and (H2) students will report high usability ratings, reflecting a clear and usable interface.

2. BIG Simulation Game and Adaptation Framework

2.1. THE BIG SIMULATION GAME

The BIG simulation game connects principles of energy-efficient design with computational modelling and interactive visualization. It integrates three technical layers, energy simulation, machine-learning-based prediction, and cost estimation, within a single interface to support exploratory design tasks. Figure 1 presents the integrated framework.

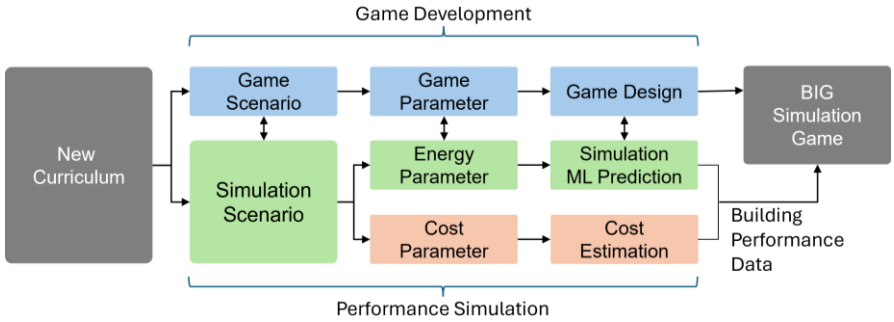


Figure 1. Integration framework linking BIG curriculum, simulation modules, and game interface

The framework links classroom activities with two primary components: simulation processing and interface interaction. Scenario parameters guide design exploration, while the simulation layer processes passive and active energy use, renewable energy contribution, and cost through validated engines. Results are returned to the interface as near real-time feedback, enabling students to interpret performance metrics while iteratively adjusting design parameters. The game interface (Figure 2) allows users to define spatial, material, and system parameters and compare resulting performance. The left panel manages design inputs such as envelope, glazing, and HVAC within a fixed budget constraint, while the right panel displays feedback on energy use, cost, and emissions. This structure supports observation of cause-effect relationships during iterative design exploration.

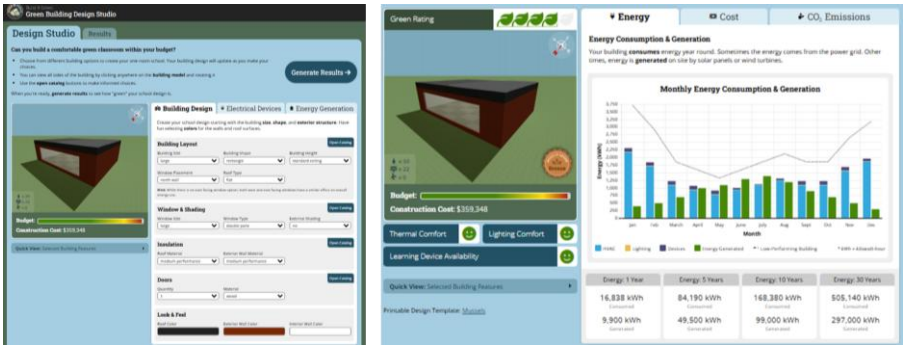


Figure 2. BIG simulation interface: design configuration (left), and real-time feedback (right)

2.2. OBJECTIVES

The simulation game test in the higher-education classroom addressed three objectives:

- **Evaluation of Computational Thinking:** Align course activities and gameplay with CT constructs to capture students' reasoning processes.
- **Evaluation of Usability:** Examine interface clarity, responsiveness, and feedback to understand user experience during gameplay.
- **Assessment of Adaptation Needs:** Use student feedback on content, parameter options, and reporting features to identify criteria for refining the BIG simulation game for higher-education use.

3. Game in the Classroom

3.1. COURSE CONTEXT AND PARTICIPANTS

The test of the BIG simulation game was implemented in a sustainability course at a large public university in the Midwestern United States. The course introduces principles of building systems, environmental design, and energy performance modelling through simulation-based activities. The study was conducted under an approved Institutional Review Board protocol (IRB #2127312). Forty-six undergraduate students participated, representing majors in architecture, interior design, and civil engineering. A priori power analysis ($\alpha = .05$, power = .90, medium effect size $d = 0.5$) indicated that a minimum of 44 participants were required across tests, and the final sample of 46 students met this threshold. The simulation activity was integrated into in-class activities and assignments over five weeks. Prior to these sessions, students completed introductory lessons on envelope performance, HVAC systems, and renewable-energy strategies. The BIG game served as a bridge between conceptual lectures and applied performance analysis, allowing students to iterate design parameters and observe quantitative feedback related to energy use, cost, and comfort.

3.2. ACTIVITY DESIGN AND DATA COLLECTION ALIGNMENT

The five-week instructional sequence aligned course content, simulation use, and reflection activities in a progressively structured format (Figure 3). The design progressed from guided exploration to more independent analysis while collecting quantitative and qualitative evidence of student engagement.

- **Week 1 Introduction:** Students received a short tutorial and explored key interface parameters affecting energy and cost. They discussed initial observations on Canvas and completed a pre-survey on CT measures.
- **Week 2 Iterations:** With instructor guidance, students completed approx. fifteen simulation trials to examine relationships among envelope, glazing, and HVAC decisions. Trial logs and reflections were submitted on Canvas, and gameplay data

were captured automatically. A usability survey concluded this phase (Figure 4).

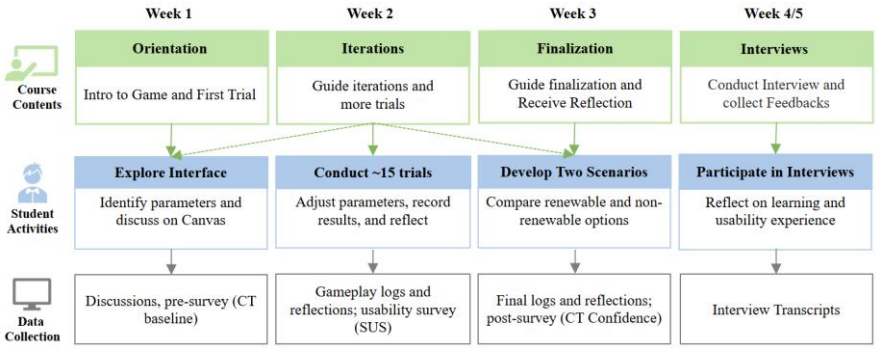


Figure 3. Five-week instructional sequence

- **Week 3 Finalization:** Students created two design scenarios, one with renewable energy and one without, compared performance and cost outcomes, and submitted written reflections. A post-survey on CT measures followed.
- **Weeks 4-5 Feedback:** Volunteers joined short Zoom interviews discussing their reasoning strategies and interface experience. Transcripts provided qualitative context for the quantitative results.



Figure 4. Classroom session during Week 2

3.3. DATA COLLECTION

Data collection focused on CT practices and usability perceptions as the two primary dimensions for evaluating how effectively the simulation functioned in the higher-education context. Questionnaire items for CT constructs were adapted from Yadav et al. (2019), which identifies abstraction, decomposition, pattern recognition, and debugging as central components of computational problem solving. Usability was assessed using the System Usability Scale (Brooke, 1996) and selected items from the USE Questionnaire (Lund, 2001), emphasizing usefulness, ease of use, and satisfaction. Table 1 summarizes how each construct was contextualized within gameplay and linked to observable actions.

Table 1. Questionnaire Mapping of Computational Thinking and Usability Constructs

Category	Construct	Game Context	Code / Item (abbrev.)
CT	Abstraction	Focus on key parameters (e.g., insulation, HVAC).	**AB1** Identify design problems solvable by computation. **AB2** Generate and interpret data. **AB3** Visualize data/processes.
	Decomposition	Divide into subsystems (envelope, lighting, HVAC).	**DE1** Break complex problems into smaller parts.
	Pattern Recognition	Detect recurring input-output relations.	**PR1** Select appropriate tools for energy analysis.
	Debugging	Refine parameters when results are unsatisfactory.	**DB1** Test models. **DB2** Iterate for improvement. **DB3** Troubleshoot errors.
Usability	Usefulness	Game supports learning and course goals.	**UF1** Class materials aided understanding. **UF2** Game aided course learning. **UF3** Game-course link unclear (Reversed: R).
	Ease of Use	Simplicity and intuitiveness of interface.	**EU1** Game easy to use. **EU2** Too complex (R). **EU3** Needed much learning (R). **EU4** Needed technical help (R).
	Satisfaction	Enjoyment and willingness to reuse.	**SA1** Plan to use in future work. **SA2** Would recommend to others.

Data were collected through four sources: (i) pre- and post-surveys measured students' CT practices on five-point Likert scales; (ii) usability surveys were administered on five-point Likert scales; (iii) written reflections submitted during the simulation activities captured students' reasoning strategies and interpretation of simulation feedback; (iv) follow-up interviews provided additional qualitative insight into students' decision processes and interface experience.

3.4. DATA ANALYSIS AND PARTICIPANT FEEDBACK PROCESSING

This paper reports findings from pre- and post-surveys of CT constructs, usability data, written reflections, and interview responses, with analysis combining quantitative comparison and qualitative coding.

Pre- and post-surveys of CT constructs were analysed using descriptive statistics, paired-samples t-tests, and effect-size estimates across abstraction, decomposition, pattern recognition, and debugging. Items used five-point Likert scales ranging from 1 (not at all) to 5 (a great amount). Only matched responses were included. Normality was examined with the Shapiro-Wilk test, and Wilcoxon signed-rank tests were applied when assumptions were not met. Higher post-scores reflected increased self-reported confidence in CT-related practices.

Reliability was assessed with Cronbach's alpha, using 0.7 as a commonly accepted

threshold. For usability, System Usability Scale (SUS) responses were summarised at item and subscale levels. Items used five-point scales ranging from 1 (strongly disagree) to 5 (strongly agree) and were converted to the standard 0 to 100 SUS metric after adjusting reverse-coded items. Following Bangor et al. (2008), values above 70 indicate good usability. Reliability was evaluated with Cronbach's alpha consistent with CT.

Qualitative reflections were analysed in MAXQDA using a hybrid coding strategy that combined deductive codes aligned with CT and usability constructs with inductive codes for emergent themes. Two researchers collaboratively refined the codebook to support consistency, and themes were synthesized in relation to the quantitative findings.

4. Results

4.1. CT OUTCOMES

Quantitative results indicated an increase in students' CT scores following use of the BIG simulation game. The overall CT mean rose from 2.85 to 3.35 on a five-point scale, reflecting higher self-reported confidence in reasoning through design tasks. This improvement was statistically significant with a medium effect size. Reliability for the CT scale was high in both phases ($\alpha = .90$ pre, $\alpha = .95$ post), confirming strong internal consistency. Table 2 presents the construct-level summary. Increases were observed across four CT dimensions, with more pronounced changes in abstraction, pattern recognition, and debugging.

Table 2. CT pre-post (M = Mean; SD = Standard deviation; Δ = Change from pre to post; CI = Confidence interval)

Construct	M (SD) Pre	M (SD) Post	Δ (95 % CI)	p-value
Overall CT	2.85 (0.68)	3.35 (0.72)	+0.50 [2.90, 3.80]	$p < 0.05$
Abstraction	3.12 (0.75)	3.54 (0.93)	+0.42 [2.88, 3.84]	$p < 0.05$
Decomposition	3.24 (0.76)	3.62 (0.95)	+0.38 [3.00, 3.93]	$p < 0.05$
Pattern Recognition	2.49 (0.80)	3.35 (0.89)	+0.86 [2.23, 3.64]	$p < 0.05$
Debugging	2.57 (0.94)	3.07 (1.04)	+0.50 [2.27, 3.41]	$p < 0.05$

Figure 5 shows the item-level trends, where increases are visible across six items. These patterns indicate that changes across several items were statistically reliable, suggesting systematic differences between pre- and post-survey responses.

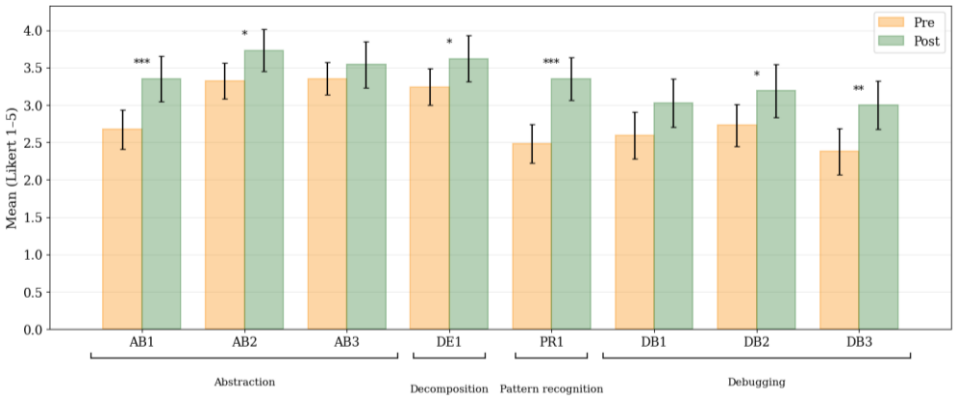


*Figure 5. pre- and post-test scores of CT constructs (*p < .05, ** p < .01, and *** p < .001)*

4.2. USABILITY OUTCOMES

Students evaluated the BIG simulation game as clear and easy to use. The overall System Usability Scale (SUS) score averaged 75, placing the tool in the "Good" usability range. Table 3 summarizes the subscale results, which show strong ratings for usefulness and ease of use, with more variation in satisfaction items related to enjoyment and long-term preference. Internal consistency for the usability items was acceptable ($\alpha = .79$ overall), supporting reliability for short scales.

Table 3. SUS summary (M = Mean score; SD = Standard deviation; CI = Confidence interval)

Construct	**M (SD)**	**95 % CI**	**Interpretation**
SUS Overall	75 (15)	[69, 81]	Good / Above average
Usefulness	77 (19)	[70, 84]	Good / Above average
Ease of Use	79 (19)	[72, 86]	Good / Above average
Satisfaction	63 (20)	[56, 70]	OK / Marginal

Figure 6 illustrates how most responses clustered in the upper interpretation bands for usefulness (UF1-UF3) and ease of use (EU1-EU4). Most item distributions fell within the 'Good' interpretation band, indicating consistently positive usability perceptions across participants. Satisfaction items (SA1-SA2) exhibited wider spread, indicating that while the interface was viewed as clear and functional, students differed somewhat in their enjoyment of the experience.

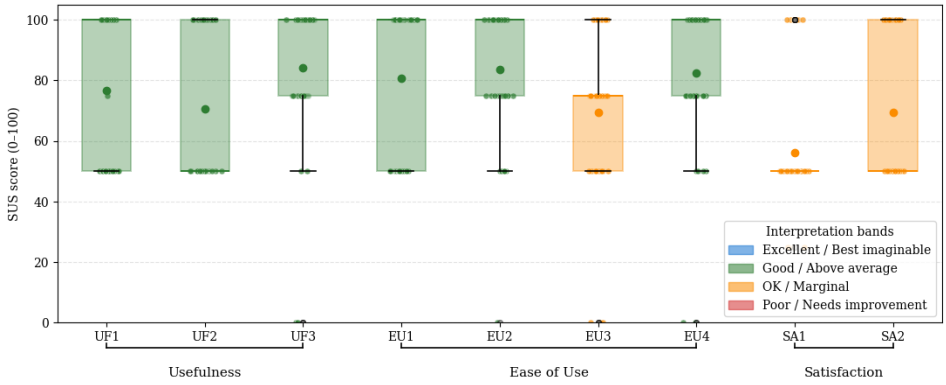


Figure 6. Distribution of SUS scores across Usability dimensions

4.3. QUALITATIVE REFLECTIONS

Post-activity interviews indicated shifts in how students described their computational reasoning while working with the simulation. Participants often described a movement from guessing to isolating individual variables and observing their effects. One student explained, "At first I changed things randomly, but then I began adjusting one parameter at a time to see what mattered," while another noted, "Changing one thing and seeing the result immediately helped me understand where the inefficiencies were." These comments illustrate increased articulation of abstraction and debugging as students refined decisions through iterative exploration. Students also highlighted how the interface supported this reasoning. The immediate feedback and visual clarity were frequently mentioned as central to their process. A participant stated, "It was easy to understand the results because the numbers and graphs updated right away," and another reflected, "Everything was consistent. When I did something, it made sense why I got the result I did." These remarks suggest that the layout and real-time updates helped students focus on interpreting outcomes rather than navigating the tool. Interviewees also identified areas for refinement. One student commented, "Some categories were intuitive, but others needed more explanation," and another suggested that more detailed result breakdowns would improve understanding. These insights point to opportunities for improving parameter clarity and supporting more detailed interpretation within the simulation environment.

5. Discussion

The findings suggest that the simulation supported the two dimensions central to this study. Students reported higher confidence in CT-related practices after gameplay and evaluated the interface as clear and easy to use. These outcomes indicate that the simulation provided a setting in which students could test ideas, monitor the consequences of design choices, and concentrate on reasoning processes without being constrained by interface complexity.

The qualitative reflections offer insight into how a tool originally designed for middle-school learners functions within an undergraduate design context. Students

emphasized the value of immediate feedback and consistent visual presentation, which supported interpretation of performance outcomes during iterative trials. They also identified situations where additional support was needed, particularly when interpreting compound effects or navigating parameters that lacked detailed explanation. These observations highlight specific adaptation considerations for the settings, where learners bring broader prior knowledge and expect clearer articulation of how design parameters interact. These findings have implications for the design of educational simulation tools. For computing and educational-technology researchers, the results underscore the importance of combining real-time feedback with interpretable visual structures that support analytical reasoning. Features such as adjustable complexity, more detailed result breakdowns, and expanded explanation layers may better align simulation environments with the analytical expectations of university-level design courses. Such refinements also contribute to the development of clearer feedback loops in interactive modelling environments.

This study focuses on a single undergraduate cohort and draws on self-reported measures of reasoning and usability. Ongoing phases of the broader research project will extend this work through additional qualitative analysis and implementation across multiple course contexts, with the goal of further examining how students interpret simulation feedback and informing continued development of the platform.

Acknowledgements

This material is based upon work supported by the National Science Foundation (Award #2201204). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the NSF.

References

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction, 24(6), 574–594.

Brooke, J. (1995). SUS: A quick and dirty usability scale. Usability Eval. Ind., 189.

Ding, A.-C. E., & Yu, C.-H. (2024). Serious game-based learning and learning by making games: Types of game-based pedagogies and student gaming hours impact students' science learning outcomes. Computers & Education, 218, 105075.

Kim, J. B., Aman, J., Zangori, L., Cole, L., Young, H., & Catherine, S. (2025). Green Building Design Studio Game Development with Parametric Simulation and ML Prediction for Green Building Education in Rural Middle Schools. Proceedings of the 30th CAADRIA Conference, Tokyo, 22-29 March 2025, Volume 1, Pp. 479–488

Lund, A. (2001). Measuring Usability with the USE Questionnaire. Usability and User Experience Newsletter of the STC Usability SIG, 8.

Yadav, A., Larimor, R., Rick, K. M., & Schwarz, C. (2019). Integrating computational thinking in elementary classrooms: Introducing a toolkit to support teachers. Society for Information Technology & Teacher Education International Conference.

Transfer Learning for Cross-Context Climate Action Insights

From Digital Citizen Participation to Climate-Aware Decision-Making

Cem Ataman[1]
[1]*Eindhoven University of Technology*
[1]*c.ataman@tue.nl , 0000-0001-8355-7119*

Abstract. The accelerating climate crisis demands timely and collective action; yet, achieving broad and meaningful citizen engagement remains a persistent challenge. Although digital participation platforms generate extensive civic discourse, their capacity to reveal climate-relevant insights is still underexploited. This study proposes a systematic method for extracting environmental knowledge from large-scale participation datasets, even when climate issues are not explicitly foregrounded. It addresses two central questions: how climate-relevant information can be reliably detected within general urban debates, and how such information can be transferred across cities without additional data collection. To investigate this, a pre-trained BERT model is fine-tuned on a climate-specific corpus and applied to participation datasets from Hamburg and Madrid of differing scale to identify and categorize environmental discourse. The approach produces coherent climate-related themes and demonstrates robust cross-context performance. The results show that climate concerns are consistently embedded within broader civic discussions and can be systematically uncovered through Natural Language Processing (NLP) and Transfer Learning (TL). By transforming existing participation data into actionable environmental knowledge, the study advances scalable and evidence-informed approaches that strengthen climate-aware urban design and planning.

Keywords. Climate Discourse Analysis, Digital Participation, Transfer Learning, Domain Adaptation, Cross-Context Data Analysis

1. Introduction

The accelerating climate crisis forces cities worldwide to advance more sustainable and resilient futures. Achieving these transformations requires not only technical innovation but also socially legitimate and participatory decision-making processes (Mendizabal et al., 2018). Urban authorities increasingly recognize that citizens' perspectives are essential for developing equitable and broadly supported climate strategies. However, facilitating large-scale, inclusive, and meaningful citizen engagement on environmental issues remains a persistent challenge (Kiss et al., 2022).

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 245-254.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Participation practices are highly context dependent, resource intensive, and often constrained by limited institutional capacity (Ataman et al., 2025). As a result, collecting new climate-specific datasets is rarely feasible, especially since citizens often lack the expertise needed to engage with specialized environmental topics.

At the same time, many cities already operate digital participation platforms that gather extensive civic discourse on a wide range of policy issues, including mobility, housing, public space, and local services (Ataman et al., 2022). Although these platforms are not explicitly designed for environmental policymaking, their datasets frequently contain valuable, yet implicit, insights related to climate action. Climate-relevant concerns are often embedded within broader debates, but their thematic structure and linguistic framing do not clearly identify them as environmental discussions. Consequently, there is a pressing need for analytical approaches that can utilize existing participation datasets and extract climate-related insights without initiating new rounds of data collection. Despite this potential, methods for systematically identifying and transferring such implicit environmental knowledge across urban contexts remain underdeveloped. Existing approaches to climate data analytics primarily rely on explicitly environmental datasets, thereby overlooking the informational richness of everyday civic discourse (Espinoza & Aronczyk, 2021).

This paper addresses these gaps by developing a transfer learning framework that applies NLP to uncover climate-relevant patterns within large-scale participation data. A pre-trained language model is domain-adapted using a climate-specific scientific corpus and applied to datasets from Madrid and Hamburg to detect environmental insights that remain implicit within general civic debates.

Consequently, the contribution of the paper is threefold: i) introducing a climate-adapted embedding model that strengthens the identification of climate semantics in heterogeneous participation datasets, ii) developing a scalable analytical pipeline that integrates semantic filtering, climate-aware embeddings, and clustering to extract interpretable environmental themes, and iii) proposing a cross-city evaluation approach based on centroid similarity, enabling assessment of how climate discourse transfers across different contexts. Consequently, it offers an interpretable and transferable method that links digital participation data to climate-aware urban decision-making, supporting the use of existing public input without additional rounds of engagement.

2. Urban Climate Resilience: Integrating Policy, Citizen Engagement, and Digital Tools

Urban climate resilience refers to the ability of cities to anticipate, absorb, and adapt to climate risks through coordinated planning, governance, and civic participation. Recent studies highlight that resilience depends not only on technical measures but also on social legitimacy and collective knowledge (Istrate et al., 2025). Digital infrastructures such as digital twins and participation platforms increasingly support this transition by enabling public engagement and data-informed decisions (Berigüete et al., 2024). Yet, major gaps persist in integrating policy frameworks, citizen perspectives, and technological tools, and utilizing these embedded insights is essential for advancing more inclusive, adaptive, and context-sensitive climate governance.

2.1. THE CLIMATE CRISIS IN URBAN DESIGN AND PLANNING

Urban design and planning policies have become central arenas for confronting the accelerating climate crisis, as cities face mounting pressures from extreme weather events, rising temperatures, and persistent environmental inequities. International and regional frameworks such as the IPCC assessments, the EU Green Deal (European Commission, 2025), and the New Urban Agenda (The New Urban Agenda, 2025) increasingly guide urban policy, positioning cities as key actors in climate mitigation and adaptation. These frameworks promote systemic approaches that move beyond short-term measures and support planning practices that embed sustainability, resilience, and long-term ecological responsibility in the built environment.

In parallel, climate-sensitive planning tools have expanded considerably, enabling cities to incorporate nature-based solutions, sustainable mobility systems, energy-efficient retrofits, and justice-oriented design strategies into urban development (Istrate et al., 2025). Such tools respond directly to the need for adaptive infrastructure and equitable climate action. However, the successful implementation of these measures is often complicated by governance challenges, including political contestation, misinformation, and persistent gaps in public trust (Ataman et al., 2024). As climate policies become more ambitious, ensuring their social legitimacy requires a deep understanding of how communities perceive risks, trade-offs, and policy measures.

Traditional environmental datasets provide essential diagnostic information on emissions and climate risks, yet they offer limited insight into how citizens understand and respond to climate measures (Ali & Thakkar, 2023). They overlook the social, cultural, and political factors that shape public acceptance, which makes it difficult to anticipate reactions to mobility transitions, land-use reforms, or energy retrofits. Addressing these challenges requires integrating civic perspectives that reveal how climate policies intersect with concerns such as housing, mobility, and affordability. Because such insights rarely appear in environmental datasets, new analytical approaches are needed to link established policy frameworks with the rich, unstructured discourse emerging from contemporary participation systems.

2.2. DIGITAL CITIZEN PARTICIPATION FOR CLIMATE MITIGATION

As climate concerns become central to urban planning, digital participation platforms have become important for broadening engagement and supporting more inclusive climate mitigation strategies. The evolution from early e-consultation systems to contemporary civic-tech and crowdsourcing platforms has reshaped how public input is collected and integrated into policy processes (Ataman et al., 2022). These tools enable scalable, accessible, and continuous communication, allowing citizens to contribute asynchronously and across diverse social groups and geographic areas.

Despite these advantages, digital participation remains underused in climate policy contexts. Engagement is shaped by demographic and spatial inequalities (Hastie, 2021), and climate-specific consultations often require technical knowledge that many citizens do not possess, leading to lower participation (Ainscough & Willis, 2024). Resource constraints and participation fatigue further limit the feasibility of launching new climate-focused engagement rounds. Moreover, existing platforms already generate extensive civic discourse with substantial but largely untapped relevance for

climate mitigation. Environmental concerns often appear indirectly in broader debates, yet remain difficult to detect because they are rarely expressed in environmental terms. Conventional analytical methods struggle with this heterogeneous, unstructured text, limiting planners' ability to link everyday concerns to climate objectives.

This gap reflects a broader limitation in the literature: most research focuses on explicitly environmental participation, overlooking climate-relevant insights embedded in general urban debates. Making these perspectives analytically accessible is essential for advancing climate mitigation strategies that are socially grounded, contextually informed, and aligned with lived experience.

2.3. CIVIC DISCOURSE FOR CLIMATE-AWARE URBAN KNOWLEDGE

Digital participation platforms generate large volumes of text-rich, locally grounded civic data, yet transforming these unstructured contributions into actionable knowledge remains a substantial analytical challenge. A growing body of research has examined computational approaches such as topic modeling, sentiment analysis, clustering, and argument mining to systematize public input and identify patterns in large-scale datasets (Ataman et al., 2023; Krasnov & Sen, 2019). These techniques extend the analytical scope of traditional qualitative assessments by enabling scalable interpretation of collective preferences and public reasoning. Yet, their application to climate-related insights is still limited, particularly when environmental concerns appear implicitly in multi-issue discussions rather than as explicit thematic content.

Recent advances in NLP have expanded the methodological possibilities for analyzing civic discourse. Unsupervised topic modeling, zero-shot classification, and transformer-based architectures such as BERT offer powerful means of detecting latent themes, identifying subtle semantic cues, and mapping the structure of public debates (Devlin et al., 2019). Large Language Models, in particular, introduce new capacities for capturing contextual meaning, disambiguating noisy text, and interpreting implicit references embedded within heterogeneous contributions (Ataman et al., 2023). Yet, participation datasets continue to pose persistent challenges: they are often fragmented, ambiguous, and shaped by overlapping issue framings, which complicates efforts to isolate climate-relevant information using standard NLP techniques.

A key limitation in current research is the focus on single-city analyses, with little attention to whether models trained in one urban context can generalize to another. This gap is critical for climate research, where shared environmental pressures intersect with distinct socio-political conditions. Emerging work on TL demonstrates that adapted language models can detect environmental semantics across contexts (Mishra et al., 2025). Yet systematic methods that connect broad civic discourse to climate themes at scale remain underdeveloped, limiting cross-context urban policy analysis.

3. Methodology

This study introduces a pipeline for extracting climate-related insights from citizen participation datasets. Its methodology consists of five main components: i) construction of a climate knowledge corpus, ii) domain-adaptive training of a climate-aware embedding model, iii) dataset preparation for two cities, iv) climate-specific text filtering, and v) embedding-based topic analysis and cross-city comparison.

3.1. CONSTRUCTION OF THE CLIMATE KNOWLEDGE CORPUS

To train a domain-adaptive text encoder, a specialized corpus representing contemporary climate action knowledge was required. Scientific publications relating to climate change, adaptation, mitigation, and environmental policy were retrieved using the Semantic Scholar API. The query set covered a broad range of climate-relevant themes, including "climate adaptation", "urban heat", "flood management", "renewable energy transition", "air quality", and "urban green infrastructure".

A total of 787 PDFs were extracted and processed through automated parsing, cleaning, and segmentation. Tables, references, and residual OCR artifacts were removed, yielding 17,034 coherent textual segments. These segments form the Climate Corpus and constitute the domain knowledge required for TL. The final corpus was stored in CSV format to ensure reproducibility. Comparable corpora can be assembled using other types of sources, such as open-access reports, municipal climate strategies, or policy documents, enabling replication by cities with limited NLP infrastructure.

3.2. DOMAIN-ADAPTIVE CLIMATE-AWARE ENCODER TRAINING

A domain-adaptive pretraining procedure was implemented to produce a climate-sensitive embedding model. The general-purpose all-MiniLM-L6-v2 Sentence-Transformers encoder served as the starting point. The model was then optimized using contrastive learning on the Climate Corpus. Positive pairs were constructed by linking adjacent or semantically similar segments from the same publication. Training was conducted for two epochs with a batch size of 64 using cosine similarity loss. The resulting model, referred to as the Climate Embedder, is designed to capture climate-related semantics more effectively than the baseline encoder. Both models were subsequently used to embed participation datasets for comparative evaluation.

3.3. CITIZEN PARTICIPATION DATASETS

The study draws on two large-scale citizen participation datasets from Madrid and Hamburg, which illustrate different modes of digital engagement while providing rich textual material for analysis. In Madrid, the data originate from Decide Madrid, an open-source participation platform launched by the City Council in 2015 (Decide Madrid, 2023). The dataset includes 3,838 debate topics, over 31,000 comments and replies, and participation from more than 470,000 registered users. For this study, debate descriptions and user comments were extracted, translated from Spanish into English using the 'googletrans' library in Python, and subsequently cleaned, deduplicated, and consolidated into a final dataset of 119,811 entries.

In Hamburg, the dataset was sourced from the Digital Participation System, introduced in 2016 to support project-based engagement (DIPAS, 2023). In this study, 28 DIPAS project were included, translated from German into English using the same library, and processed through filtering, cleaning, and deduplication, resulting in 4,020 entries. In both datasets, each textual entry was treated as a contribution representing a resident's input. The datasets differ substantially in size, reflecting differences in platform design, participation intensity, and institutional context, and this imbalance is treated as a condition for evaluating cross-context transferability of climate discourse representations rather than direct quantitative comparison.

3.4. CLIMATE-SPECIFIC TEXT FILTERING

Citizen participation platforms contain diverse content, much of which is not climate-related. To isolate climate-relevant texts, a two-stage filtering pipeline combining lexical and semantic methods was implemented.

- Strict Keyword Pre-filtering: A list of climate-related keywords (e.g., heat, flooding, air quality, renewable energy, trees, pollution, drought) was used to screen the datasets. This step removed clearly irrelevant content while retaining a broad candidate set. In Madrid, 18,697 out of 119,811 texts (approximately 15.6 percent) passed the keyword filter. In Hamburg, 839 out of 4,020 texts (approximately 20.9 percent) were retained as climate candidates.
- Zero-Shot Semantic Classification: Remaining texts were classified using a DeBERTa-v3 zero-shot classifier fine-tuned on MNLI, FEVER, and ANLI (MoritzLaurer/deberta-v3-base-mnli-fever-anli). Texts were evaluated against eight climate-related labels: climate impacts, urban heat, flooding and stormwater, air quality, energy and emissions, water scarcity, green infrastructure, and sustainable mobility. A ninth label, "not climate-related", served as a negative class. A confidence threshold of 0.35 was applied, and only texts assigned to at least one climate-relevant label were retained. Tokenization used hard truncation at 256 tokens to ensure computational efficiency. This yielded 13,932 climate-relevant texts for Madrid and 839 for Hamburg.

3.5. EMBEDDING AND TOPIC MODELING

Following climate filtering, the remaining texts were embedded using both models: the baseline MiniLM encoder and the domain-adapted Climate Embedder. Embeddings were generated in batches and stored in NumPy format for downstream analysis.

- Clustering Climate Topics: To identify thematic structures, k-means clustering was applied to the climate embeddings. The number of clusters was determined through silhouette analysis and qualitative inspection of cluster content. For each city, K values between 5 and 10 were tested, and K=5 offered a suitable balance between separation and interpretability. For each model and city, clustering produced size distributions, representative text samples, and silhouette scores that measured cluster separation. These outputs supported the assessment of topic distinctiveness and the contribution of domain adaptation.
- Cross-City Topic Generalization: To evaluate TL and cross-context generalization, cluster structures in Madrid and Hamburg were compared using embeddings from the Climate Embedder. Climate texts from each city were clustered independently with K=5, and centroids were computed in the embedding space. Cross-city alignment was evaluated in two ways. First, a global cross-city similarity score was computed as the mean cosine similarity across all Madrid-Hamburg embedding pairs. Second, a centroid similarity matrix was constructed by comparing the five Madrid centroids with the five Hamburg centroids. This matrix enabled identification of best-matching topic pairs and provided an assessment of how strongly specific themes generalize across cities. Comparison between baseline and

adapted embeddings, combined with qualitative interpretation of representative texts, supported an assessment of whether domain adaptation improved climate insight extraction and strengthened topic clarity across contexts.

4. Results and Discussion

This section presents results from the climate-filtering pipeline, embedding-based topic modeling, and cross-city generalization. It examines climate-relevant subsets from Madrid and Hamburg, the quality of topic structures, and the alignment of climate themes across contexts. Table 1 summarises key outcomes and performance scores.

Table 1. Summary of Key Dataset, Filtering, Clustering, and Cross-City Results.

Metric	Madrid	Hamburg	Metric definition
Total texts	119,811	4,020	Cleaned, translated, deduplicated
Climate-filtered texts	13,932	839	After keyword + zero-shot filtering
Climate-share (%)	11.6%	20.9%	Portion of all texts
Optimal clusters (K)	5	5	Based on silhouette sweep
Silhouette (Baseline)	0.0405	0.0568	Topic separation
Silhouette (Climate Embedder)	0.0417	0.0265	Domain-adaptation impact
Global cross-city similarity	0.307		Mean cosine similarity
Centroid similarity range	0.586-0.920		Cross-city topic alignment
Strongest cluster match	0.920 (Madrid C3, Hamburg C1)		Mobility/traffic-environment theme

The filtering pipeline produced substantially smaller yet climate-focused datasets. In Madrid, 13,932 texts (11.6%) were identified as climate-related, while Hamburg yielded 839 texts (20.9%). Despite differences in dataset size and platform design, the retained materials covered a broad range of climate-relevant themes, including green infrastructure, sustainable mobility, air quality, and climate concerns. This provided a robust basis for the later embedding and clustering analysis, with cross-city results interpreted in terms of semantic alignment rather than direct proportional comparison.

4.1. EMBEDDING-BASED CLIMATE TOPIC MODELING

K-means clustering was applied to both embedding models across K=5-10, with silhouette analysis indicating that K=5 produced the most stable separation in both cities. In Madrid, silhouette values reached 0.0405 for the baseline model and 0.0417 for the climate-adapted model, suggesting a modest improvement when domain knowledge is incorporated. In Hamburg, where the dataset is much smaller and thematically focused, the baseline model achieved the highest score (0.0568), while the climate-adapted model reached 0.0265. This asymmetry is expected: domain adaptation primarily benefits large and diverse datasets, whereas smaller project-specific corpora already display strong internal coherence.

Despite low silhouette scores, which are typical for high-dimensional short-text clustering, qualitative inspection indicated clear and interpretable climate-related themes (Table 2). This interpretability assessment is based on systematic inspection of representative texts per cluster and is intended as an internal validation of semantic coherence rather than an external policy validation. In Madrid, clusters centered on water scarcity, air quality and emissions, sustainable mobility, green infrastructure, and environmental cleanliness. The climate-adapted model produced tighter groupings,

particularly for mobility and pollution topics. In Hamburg, the clusters reflected discussions on street redesign, noise and air pollution, urban greenery and biodiversity, playground and park design, and riverfront and flood-related concerns.

Across both cities, climate themes emerged where of environmental performance and everyday urban conditions, indicating that citizen-generated climate discourse is embedded in broader concerns related to mobility, comfort, and neighborhood quality.

4.2. CROSS-CITY GENERALIZATION OF CLIMATE TOPICS

Cross-context generalization was assessed by comparing cluster centroids from Madrid and Hamburg (K=5 each) using cosine similarity, yielding a 5×5 similarity matrix and an aggregated global similarity score that function as computational indicators of semantic alignment rather than expert-led or policy-based validation.

Table 2. Cross-City Climate Topic Structures and Similarity Scores (K = 5)

Topics	Madrid Theme	Hamburg Theme	Cosine Similarity	Interpretation
T1. Green Infrastructure, Shade, Urban Microclimate	Need for more trees, shade deficits, drought stress, microclimate discomfort.	Tree canopy protection, street greening, shade for public spaces, biodiversity concerns.	0.705	Strong thematic alignment: both cities discuss the role of vegetation in heat mitigation.
T2. Sustainable Mobility, Cycling and Street Conflicts	Unsafe bike lanes, pedestrian-cyclist conflict, traffic calming.	Narrow cycling paths, dangerous crossings, traffic calming zones.	0.920	Highest cross-city similarity: mobility and safety concerns are shared and similar.
T3. Air Quality, Noise, Traffic Emissions	Noise pollution, car emissions, complaints about traffic discomfort.	Noise exposure on arterial roads, local air pollution, emissions affecting sleep.	0.892	Mobility–emissions links appear in both datasets as similar stressors.
T4. Public Cleanliness, Waste, Environmental Stewardship	Waste separation, littering, maintenance issues, community responsibility.	Cleanliness of streets, litter, local stewardship, green and public areas.	0.723	Shared framing of environmental cleanliness as climate-resilient cities.
T5. Energy, Lighting, Efficiency, Urban Technology	Energy savings, lighting efficiency, pollution reduction.	Efficient lighting, renewable energy potentials, climate-adaptive infrastructure.	0.763	Both cities link energy transitions to environmental benefits.

The global similarity score of 0.307 indicates a moderate degree of shared semantic structure across the two cities despite substantial differences in language, platform design, and participation patterns. Centroid-level similarities were considerably higher, ranging from 0.586 to 0.920, highlighting strong thematic alignment in several areas. Madrid Cluster 3 and Hamburg Cluster 1 revealed the highest correspondence (T2=0.920), reflecting a shared discourse on traffic safety, cycling conditions, and the environmental impacts of motorized mobility. Madrid Cluster 2 and Hamburg Cluster 2 also aligned closely (T3=0.892), both concerned noise, air pollution, and emissions.

These results demonstrate that core climate-related concerns, such as sustainable mobility, green infrastructure, and neighborhood-level environmental quality, exhibit consistent semantic structures across contexts. The domain-adapted model supported this alignment by producing embeddings that more clearly separated climate themes, thereby improving the comparability of topics between cities.

4.3. TRANSFER LEARNING IN URBAN CLIMATE DISCOURSE

The results highlight how climate discourse in digital participation data contains stable semantic structures that can be captured through domain-adapted embeddings. The Climate Embedder improved topic clarity in the more heterogeneous Madrid dataset, suggesting that domain adaptation is beneficial when climate themes are embedded within broad civic conversations. In Hamburg, where climate discussions were already

highly focused, the baseline model performed comparably well, indicating that model gains are sensitive to dataset characteristics rather than universally additive.

The strong cross-city alignment of clusters demonstrates that key climate concerns are expressed in similar semantic forms across different participation settings. This points to a transferable structure in how citizens articulate climate issues such as mobility emissions, heat, air quality, and green infrastructure. The ability to detect these patterns indicates that insights learned in one city can support interpretation in another, which is central to the value of TL in planning and policy contexts.

Meanwhile, the differences in silhouette scores and thematic distributions underline the importance of considering local discourse conditions, platform design, and project focus when applying domain-adaptive models. The discussion therefore suggests that while TL can enhance the extraction and comparison of climate-relevant topics, its effectiveness depends on the diversity, scale, and framing of the participation data.

5. Conclusion

This study presents that large-scale digital participation platforms contain significant climate-relevant knowledge that can be systematically uncovered through TL. By combining domain-adaptive pretraining, semantic climate filtering, and embedding-based topic analysis, the proposed pipeline identifies environmental concerns even when climate issues are not explicitly discussed by participants. Across Madrid and Hamburg, the Climate Embedder enhanced the clarity and coherence of climate themes, proving that domain adaptation strengthens the detection of nuanced discussions on heat, flooding, pollution, green infrastructure, and sustainable mobility.

A key contribution is the demonstration that climate discourse exhibits consistent semantic structures across the examined cities, indicating the potential for broader applicability across regions and languages. The strong alignment between Madrid and Hamburg clusters indicates that insights extracted in one context can meaningfully support interpretation in another, offering a scalable route for cross-city learning in climate planning. This ability to operationalize implicit civic knowledge helps bridge the gap between public engagement data and climate-informed decision-making.

Several limitations remain, including translation noise, participation biases, the imbalance in dataset size between cities, and the dependence of topic structure on clustering choices. While this imbalance constrains quantitative cross-city comparison, it does not affect the primary objective of the study, which is to assess semantic transferability of climate discourse across contexts. While the empirical evaluation focuses on two European cities, the study is positioned as a methodological feasibility demonstration rather than a claim of global generalizability. Expanding to multilingual settings, integrating richer contextual data, and involving planners in validation would further strengthen applicability. Overall, this work establishes a practical pathway for transforming unstructured civic dialogue into actionable knowledge, supporting more anticipatory, transparent, and participatory urban climate strategies.

Acknowledgments

This research was supported by the Department of the Built Environment at Eindhoven University of Technology. The author would like to thank the Municipality of

Hamburg, in particular the Behörde für Stadtentwicklung und Wohnen, for sharing their datasets, and the City of Madrid for making its dataset publicly available.

References

Ainscough, J., & Willis, R. (2024). Embedding deliberation: Guiding the use of deliberative mini-publics in climate policy-making. Climate Policy, 24(6), 828–842.

Ali, A. H., & Thakkar, R. (2023). Climate Changes through Data Science: Understanding and Mitigating Environmental Crisis. Mesopotamian Journal of Big Data, 2023, 125–137.

Ataman, C., Herthogs, P., Tunçer, B., & Perrault, S. (2025). From insight to action: An integrated assessment framework for digital citizen participation in data-centric urban practices. Cities, 156, 105545. https://doi.org/10.1016/j.cities.2024.105545

Ataman, C., Perrault, S., & Tunçer, B. (2024). Fostering Inclusive Urban Design and Planning: Enhanced Trust and Reliability through Descriptive Instructions for Digital Citizen Participation. Journal of Urban Technology, 31(4–5), 25–48.

Ataman, C., Tunçer, B., & Perrault, S. (2023). Transforming Large-Scale Participation Data Through Topic Modelling in Urban Design Processes. In M. Turrin, C. Andriotis, & A. Rafiee (Eds.), CAAD Futures. INTERCONNECTIONS: Co-computing Beyond Boundaries (pp. 271–286). Springer Nature Switzerland.

Ataman, C., Tunçer, B., & Perrault, S. T. (2022). Asynchronous Digital Participation in Urban Design Processes: Qualitative Data Exploration and Analysis With Natural Language Processing. POST-CARBON - Proceedings of the 27th CAADRIA Conference, 383–392.

Berigüete, F. E., Santos, J. S., & Rodriguez Cantalapiedra, I. (2024). Digital Revolution: Emerging Technologies for Enhancing Citizen Engagement in Urban and Environmental Management. Land, 13(11), 1921. https://doi.org/10.3390/land13111921

Decide Madrid. (2023). https://decide.madrid.es

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171-4186.

DIPAS. (2023). https://www.hamburg.de/dipas/

Espinoza, M. I., & Aronczyk, M. (2021). Big data for climate action or climate action for big data? Big Data & Society, 8(1), 2053951720982032.

European Commission. (2025). The European Green Deal. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

Hastie, C. (2021). The Perpetuation of Inequality: The Role of Community Engagement. Sociological Research Online, 26(3), 759–767.

Istrate, A.-L., Popartan, L.-A., Auerbach, J., Gaspari, J., & Tavangar, M. R. (2025). Collaborative research for transitioning to Climate-Neutral Cities – contouring a prospective framework for integrated planning. Planning Theory, 24(2), 163–182.

Kiss, B., Sekulova, F., Hörschelmann, K., Salk, C. F., Takahashi, W., & Wamsler, C. (2022). Citizen participation in the governance of nature-based solutions. Environmental Policy and Governance, 32(3), 247–272. https://doi.org/10.1002/eet.1987

Krasnov, F., & Sen, A. (2019). The Number of Topics Optimization: Clustering Approach. Machine Learning and Knowledge Extraction, 1(1), 416–426.

Mendizabal, M., Heidrich, O., Feliu, E., García-Blanco, G., & Mendizabal, A. (2018). Stimulating urban transition and transformation to achieve sustainable and resilient cities. Renewable and Sustainable Energy Reviews, 94, 410–418.

Mishra, P., Narayanasamy, S. K., & Srinivasan, K. (2025). Context-Aware Embedded Language Transformers for Evaluating Climate Change-Based Sustainable Development Goals. IEEE Access, 13, 65757–65775. https://doi.org/10.1109/ACCESS.2025.3559548

The New Urban Agenda. (2025). Habitat III. https://habitat3.org/the-new-urban-agenda/

Computational Sensory Environments

Comparing Photo-Textured LOD2 Models and LOD3-Equivalent Detailed Models in Urban VR

A Preliminary Study on Effects on Gaze Behaviour and Spatial Impressions

Kosuke Kobayashi[1] and Masatoyo Ogasawara[2]
[1,2] *Tokyo Denki University.*
[1] *25fma32@dendai.ac.jp, 0009-0002-0733-409X*
[2] *masatoyo@mail.dendai.ac.jp, 0000-0002-6796-319X*

Abstract. Highly realistic urban VR models are increasingly used for spatial evaluation and public participation, yet it remains unclear how geometric detail and photographic appearance contribute to street-level experience. This study compares a photo-textured LOD2 model and an LOD3-equivalent model with added façade relief in a VR representation of a 100 m commercial street derived from PLATEAU data. Fourteen architecture students experienced both conditions in PCVR using Unity and an HTC Vive Pro Eye along an identical automated route, while eye tracking recorded fixation data. After each walkthrough, participants rated realism, immersion, information clarity, spatial clarity, and visual load on seven-point scales. The LOD3-equivalent model was consistently rated higher in realism, immersion, information clarity, and spatial clarity, with no significant difference in visual load. Global gaze dispersion, σ_total, was significantly lower in LOD3, indicating more concentrated viewing, while total fixation count did not differ. AOI analysis showed shorter fixation duration on ground-floor glass in LOD3 and negative associations between increased canopy fixations and gains in key ratings. These findings suggest that added façade relief affects not only perceived quality but also attention allocation, and that which façade elements attract gaze is critical for streetscape impressions in urban VR.

Keywords. Urban virtual reality, 3D city models, Level of detail, Eye tracking

1. Background

In recent years, VR models with heightened visual realism have gained attention as tools for supporting spatial evaluation and consensus building in architecture and urban planning. Across Europe and Asia, three-dimensional city models compliant with the OGC CityGML standard have been widely developed and deployed, with growing use in design support and participatory processes. In Japan, the PLATEAU project has advanced the development and public release of such models, accelerating their

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 257-266.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

adoption. The level-of-detail, or LOD, framework—which defines stages of building representation—is a key factor shaping VR experience quality: LOD2 represents the three-dimensional building form, while LOD3 captures façade relief and fine exterior details. However, the effects of these differences on spatial cognition and gaze behaviour remain underexplored.

2. Objectives

This study compares a photo-textured LOD2 model with an LOD3-equivalent model incorporating detailed façade relief, in order to clarify differences in gaze behaviour and subjective impression ratings in an urban VR environment. It examines how visual realism affects attention to architectural façades and spatial comprehension, and aims to provide practical guidance for selecting an appropriate level of representation fidelity under real-world constraints.

3. Related Work

The level of detail (LOD) of 3D city models has been discussed as a framework for balancing visual accuracy, simulation fidelity, and modelling cost. Biljecki et al. (2016) reviewed the CityGML 2.0 definitions of LOD0–4, identifying ambiguities in geometric definitions and showing that LOD2 is widely adopted in practice due to its favourable trade-off between acquisition cost and applicability. More recently, objective approaches to characterising or classifying model detail in production environments have been proposed (Rodriguez-Garcia et al., 2025). In this paper, "LOD2" and "LOD3-equivalent" are used as practical labels for our representation conditions rather than strict CityGML-compliant categories.

Regarding urban-scale 3D visualisation and user experience, Spanjar et al. (2024) compared multiple urban visualisations using eye tracking and questionnaires with non-experts, employing detail variations that do not align with CityGML LOD definitions, and reported that fixation metrics and impression ratings do not necessarily differ substantially across levels of detail. In a related eye-tracking study using streetscape photographs, Spanjar and Suurenbroek (2020) showed that human-scale design elements play an important role in guiding visual attention and streetscape evaluation. Studies comparing real environments and immersive virtual environments have also examined differences in experience between VR and real-world settings (Hermund et al., 2023).

Methodological research has discussed the application of eye tracking in VR and its limitations, including the need for careful fixation detection under head-mounted display conditions (Adhanom et al., 2023), and has examined parameter settings for I-VT approaches in VR such as velocity thresholds for reliable fixation identification (Zhao et al., 2025). At the system level, VR studies have analysed relationships among performance metrics such as frame rate, latency, and rendering complexity (Geris et al., 2024), while XR surveys have summarised rendering strategies aimed at balancing visual fidelity and computational constraints (Hu et al., 2025).

Few studies, however, have used real-world city models such as PLATEAU to directly compare a photo-textured LOD2 model and an LOD3-equivalent model with added façade relief under identical VR street-walking conditions. This

paper compares these two representation conditions to examine gaze behavior and subjective experience in urban VR.

4. Methods

4.1. RESEARCH DESIGN (FIGURE 1)

We adopted a within-subjects design comparing two VR conditions of the same 100 m shopping-street segment near a university campus in the Tokyo metropolitan area. Figure 1 provides an overview of the study area and the two model conditions. Participants experienced both conditions in a PCVR setup, and condition order was counterbalanced.

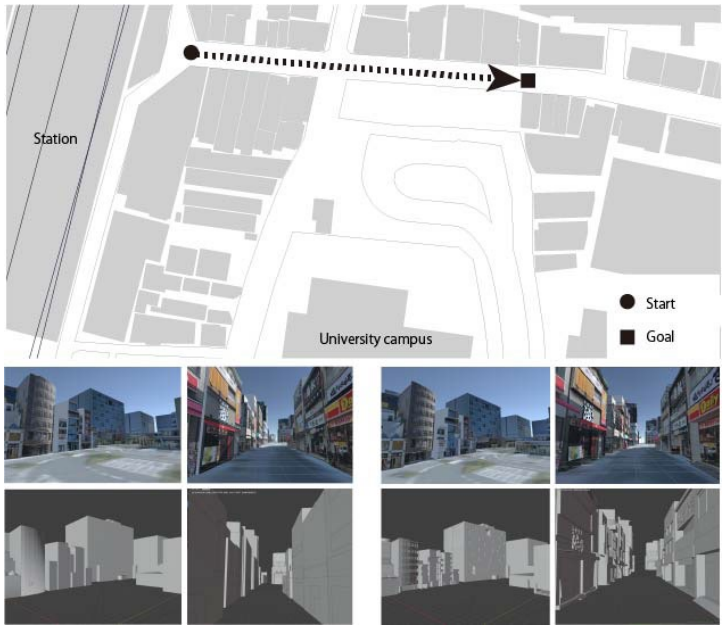


Figure 1. Overview of the study area and the two LOD conditions

4.2. STIMULUS MODELS AND LOD CONDITIONS (FIGURE 2)

The stimulus models for the street segment in Figure 1 were built from PLATEAU LOD2 building and road data. For the experiment, building meshes were exported as FBX, edited in Blender, and re-UV-mapped before being re-imported into Unity. An LOD3-equivalent model was also prepared and imported into Unity.

For LOD2-Photo, façade photographs were captured on site and rectified to a fronto-parallel view using planar homography in Python and OpenCV, based on four manually selected corner points per façade plane. The rectified images were then applied as high-resolution textures and UV-mapped onto the corresponding planar façade surfaces. The LOD2-Photo workflow for façade photo rectification and texture mapping is shown in Figure 2.

For the LOD3-equivalent condition, façade relief elements such as wall-signage thickness, canopies, and recesses around openings were created by simple extrusion on top of the LOD2 geometry, followed by minor UV adjustments. Freestanding signs visible only in photographs were removed to avoid double representation. Across conditions, building volumes, opening positions, and the layout of major wall signage were kept consistent. Both models were presented in Unity with an HTC Vive Pro Eye.



Figure 2. Photo-textured LOD2 workflow (façade photo rectification and texture mapping).

4.3. AOI DEFINITION AND GAZE LOGGING

We defined three AOIs focusing on façade elements with the largest representational differences between LOD2-Photo and LOD3: wall signage, canopy, and ground-floor glass. In Unity, colliders were assigned to these elements and the AOI ID was logged when the gaze ray hit a collider, while all other hits were recorded as non-AOI. Because façade textures can include mixed content such as reflections and interior views, we did not treat the entire façade as one AOI and instead quantified façade-level viewing using global gaze dispersion metrics.

Gaze data were collected with an HTC Vive Pro Eye via the SRanipal SDK. Binocular gaze vectors were sampled at 120 Hz, and a custom script logged the gaze hit point (world coordinates), hit object, AOI ID, and timestamp to CSV. Fixations were extracted using the I-VT method with an angular-velocity threshold of 100 deg/s and a minimum fixation duration of 120 ms; fixation onset, offset, duration, representative position, and AOI ID were then computed.

4.4. PARTICIPANTS AND PROCEDURE

Participants were 14 architecture students, including undergraduates and master's students. All were familiar with the campus area and had everyday experience of the target street. The experiment was conducted individually in a laboratory PCVR setup. Each participant completed two VR walkthroughs, LOD2-Photo and LOD3, with the order counterbalanced. In each walkthrough, the viewpoint moved along the same predefined 100 m route at a constant speed of 1.2 m/s for about 1.5 minutes. Participants were asked to observe the street and buildings while avoiding excessive backward head turns. After each walkthrough, participants completed the questionnaire.

4.5. MEASURES AND STATISTICAL ANALYSIS

4.5.1. Gaze metrics

From fixation data, we computed AOI-wise fixation count and total fixation duration. Global gaze dispersion was assessed by projecting gaze intersection points onto a vertical plane and calculating sigma total, defined as an integrated standard deviation around the centroid, and range area, defined as the convex-hull area of gaze points. Total scene-wide fixation count was also analysed.

4.5.2. Subjective ratings and statistical tests

Item scores were averaged to yield scale scores for Realism, Immersion, Information clarity, Spatial clarity, and Visual load. Differences between LOD2-Photo and LOD3 in subjective scales and gaze metrics were tested using paired-samples t-tests with alpha set to .05, and Cohen's d for dependent samples, dz, was reported. Pearson correlations were computed between within-subject change scores, defined as delta equals LOD3 minus LOD2, for gaze metrics and subjective scales.

4.5.3. Runtime performance metrics

To estimate the runtime cost of geometric detail, we replayed the same auto-walk route under free viewing and logged PCVR performance (Windows 10 Pro 10.0.19045; Intel Core i9-9900K; 48 GB RAM; NVIDIA GeForce RTX 3070; Unity 2022.3.12f1). After excluding a 10 s warm-up, we analysed the next 60 s and recorded frame time (Time.unscaledDeltaTime) and triangle count (median) using Unity profiling counters. We computed mean and 99th-percentile frame time per run and averaged these across three standing runs per condition.

5. Results

5.1. SUBJECTIVE RATINGS (TABLE 1)

Subjective ratings for the 14 participants are summarised in Table 1. The LOD3-equivalent model scored higher on realism, immersion, information clarity, and spatial clarity, while visual load showed no significant difference.

Table 1. Subjective Ratings for LOD2 and LOD3

Item	LOD2 (Mean ± SD)	LOD3 (Mean ± SD)	Diff (L3–L2)	t(13)	p	dz
Realism	3.88 ± 1.35	5.71 ± 0.61	1.83	5.8	< .001	1.55
Immersion	3.76 ± 1.38	5.67 ± 0.80	1.90	5.12	< .001	1.37
Information clarity	4.40 ± 1.20	6.17 ± 0.58	1.76	6.1	< .001	1.63
Spatial clarity	4.33 ± 1.21	6.07 ± 0.49	1.74	5.18	< .001	1.38
Visual load	2.54 ± 1.54	2.86 ± 1.92	0.32	0.74	0.472	0.20

5.2. GAZE METRICS (TABLE 2, FIGURE 3)

AOI-wise fixation durations are summarised in Table 2. Figure 3 shows global gaze dispersion for the two conditions. Gaze dispersion decreased under LOD3, indicating more concentrated viewing patterns than in LOD2-Photo. Fixation duration on ground-floor glass was significantly shorter in LOD3, whereas fixation durations on wall signage and canopies showed no significant differences. This suggests that added façade relief partly redistributed attention away from glass toward other articulated façade elements.

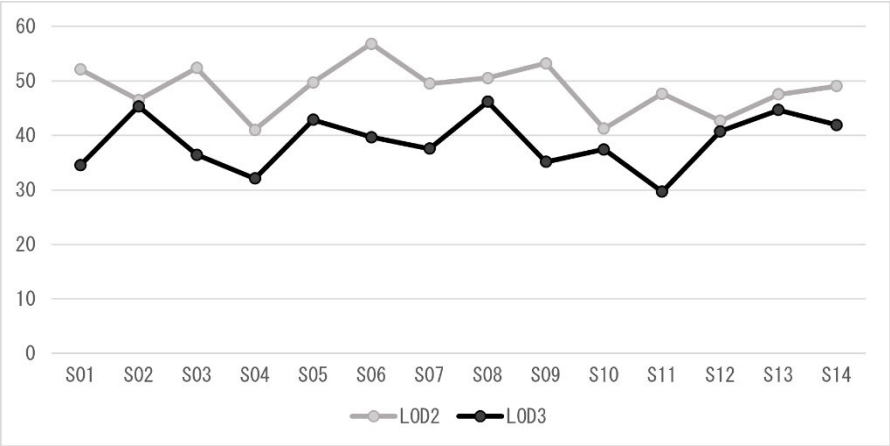


Figure 3. Global gaze dispersion (σ_total) for LOD2 and LOD3 (N = 14)

Table 2. AOI-wise fixation duration (s) for LOD2 and LOD3

AOI	LOD2 (Mean ± SD, s)	LOD3 (Mean ± SD, s)	Diff (L3–L2)	t(13)	p	dz
Wall signage	5.56 ± 2.67	5.85 ± 2.94	0.29	0.41	.691	0.11
Canopy	0.88 ± 0.51	1.21 ± 0.83	0.34	1.62	.128	0.43
Ground-floor glass	3.36 ± 1.77	2.76 ± 1.14	–0.59	–2.24	.043	–0.60

5.3. RELATIONSHIP BETWEEN CHANGES IN GAZE METRICS AND SUBJECTIVE RATINGS (TABLE 3, FIGURE 4)

Correlations between within-subject changes in gaze metrics and subjective ratings are summarised in Table 3 for 14 participants, where delta denotes LOD3 minus LOD2. Figure 4 shows a representative relationship between changes in canopy fixation count and changes in immersion. Larger increases in canopy fixations were associated with smaller gains in realism, immersion, and spatial clarity, while other correlations were generally weak.

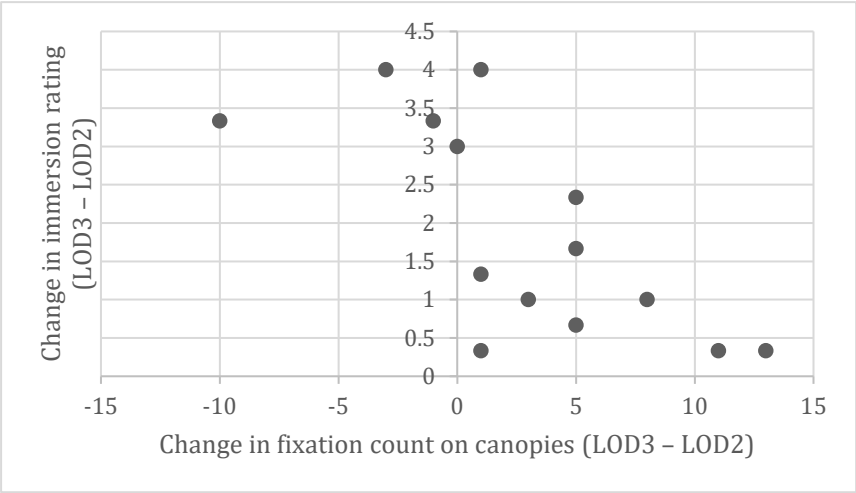


Figure 4. Δ Canopy fixations vs. Δimmersion

Table 3. Pearson correlations between change scores in subjective ratings and gaze metrics, N = 14.

Metric	ΔRealism	ΔImmersion	ΔInfo	ΔSpatial	ΔLoad
Δσ_total	.20	.50	.13	.45	–.54
Δrange_area	.06	.19	.09	.36	–.06
ΔFixation count	.31	.20	.36	.47	.15
ΔWall-signage duration	–.27	0.00	.02	.01	–.13
ΔCanopy duration	–.11	.03	.12	.19	.40
ΔGround-floor-glass duration	.03	.37	.30	.19	.06
ΔWall-signage count	–.38	–.28	.05	–.16	.20
ΔCanopy count	–.73	–.71	–.42	–.74	.42
ΔGround-floor-glass count	–.03	–.10	.08	.14	.52

5.4. RUNTIME PERFORMANCE (TABLE 4)

To examine the runtime cost of geometric detail, we logged performance metrics in the PCVR setup while replaying the same auto-walk route under free viewing. After excluding an initial 10 s warm-up, metrics were computed over the subsequent 60 s for each run and the run-level statistics were averaged across three standing runs per condition. As shown in Table 4, mean frame time was almost identical between conditions (≈11.1 ms), and the 99th-percentile frame time differed only marginally, suggesting no noticeable loss of stability in our setup under the tested route and settings. The LOD3-equivalent condition, however, increased geometric complexity, reflected in a higher median triangle count.

Table 4. Runtime performance (PCVR; standing, free viewing; 60 s after 10 s warm-up; n=3 runs/condition).

Condition	Avg frametime (ms)	99th frametime (ms)	Triangles (median)
LOD2-Photo	11.106	11.663	399,561
LOD3-equivalent	11.106	11.672	600,271

6. Discussion

6.1. RELATIONSHIP BETWEEN LOD AND VISUAL REALISM

Rather than assuming that higher LOD is always better, our results clarify how much experiential gain façade relief provides and how gaze is redistributed across façade elements, which supports practical LOD selection under modelling and runtime constraints. Subjective results indicate that the visual realism provided by the LOD3-equivalent model substantially improved the quality of spatial experience. Added 3D elements such as façade relief, canopies, and wall signage appear to reinforce a sense of actual urban façades and facilitate depth perception.

By contrast, photo-textured LOD2 models lack explicit shape cues and tend to be perceived as relatively flat surfaces, making façade structure harder to read and limiting gains in immersion and spatial understanding.

6.2. GAZE DISPERSION AND THE ROLE OF FAÇADE ELEMENTS

The decrease in σ_total suggests stronger gaze concentration in the LOD3-equivalent model, reflecting a shift toward salient façade elements rather than a simple narrowing of view. Wall signage attracted substantial gaze in both conditions, indicating robust visual salience independent of LOD. In LOD3, canopies gained clearer 3D shape and tended to attract more gaze, reflecting the salience of protruding elements. At the same time, fixation duration on ground-floor glass decreased, suggesting that in LOD2 participants relied more on window views to infer internal depth, whereas in LOD3 they could understand spatial structure from the 3D façade as a whole. Overall, the LOD3 façade composition appears to reorganise the semantic distribution of gaze.

6.3. RELATIONSHIP BETWEEN SUBJECTIVE EVALUATION AND GAZE BEHAVIOUR

ΔCanopy count showed negative correlations with ΔRealism and ΔImmersion, implying that participants whose gaze shifted disproportionately toward canopies in LOD3 tended to exhibit smaller improvements in subjective ratings. As canopies are local elements, excessive attention to them may hinder grasping the overall street configuration and diminish perceived depth and continuity. This suggests that gaze quantity alone does not predict subjective evaluation, and that where attention is directed matters for design support and visual assessment in urban VR.

Although Spanjar et al. reported limited experiential differences across detail variants, our comparison isolates façade relief cues by contrasting photographic

appearance with added geometric articulation under the same VR street-walkthrough task. Differences in LOD operationalisation, task context, and participant profile can plausibly explain why we observed clearer perceptual and gaze effects.

6.4. LIMITATIONS AND FUTURE WORK

This study has several limitations. Participants were familiar with the target street and the sample size was small, N equals 14, which may limit generalisability. Participant expertise and familiarity may bias outcomes in multiple directions. As architecture students, participants may be more sensitive to façade articulation and depth cues, which could amplify the perceived benefits of geometric relief. On the other hand, high familiarity with the street may reduce exploratory viewing and compress rating differences, producing a ceiling effect. The route included segments with different spatial openness and street widths, so the measured gaze and ratings likely reflect a mixture of spatial types and the effect of LOD may depend on street patterns such as narrow street canyons and open areas. The VR scenes lacked dynamic elements such as pedestrians or vehicles, and runtime performance was evaluated only in PCVR on a single auto-walk route under free viewing. Future work will include stricter performance tests and replication on standalone platforms, examine unfamiliar environments with dynamic elements, and segment routes by spatial configuration to compare street patterns. Because authoring cost may shift as AI-assisted modelling and photogrammetry mature, future work should quantify how hybrid workflows that combine automated reconstruction with minimal manual refinement change the cost–benefit balance.

7. Conclusion

This study compared a photo-textured LOD2 model and an LOD3-equivalent model with added geometric detail in VR, analysing differences in gaze behaviour and subjective evaluation. In relative terms, LOD3 increased key ratings by approximately 40 to 51 percent compared to LOD2-Photo as shown in Table 1, while visual load did not increase. These results indicate that adding geometric relief can enhance experiential quality without extra perceived strain.

Gaze analyses showed that σ_total decreased significantly in LOD3 and that fixation duration on ground-floor glass was reduced, while fixation volumes for wall signage and canopies changed only modestly. Correlation results further suggested that increases in canopy fixations were negatively related to gains in immersion, realism, and spatial clarity, highlighting that which façade elements attract gaze is critical for spatial impressions.

Overall, the findings show that geometric detail can meaningfully influence urban VR experience and should be weighed against modelling effort and computational load. A selective combination of photo-textured LOD2 and targeted LOD3-equivalent relief on salient façade elements can provide meaningful experiential gains while keeping production effort and computational load within practical limits.

Attribution Statement

The authors used ChatGPT (OpenAI) to support language editing. All ideas, analyses, and conclusions are the authors' own, and the authors take full responsibility for the final text.

References

Adhanom, I. B., MacNeilage, P., & Folmer, E. (2023). Eye tracking in virtual reality: A broad review of applications and challenges. Virtual Reality, 27(2), 1481–1505. https://doi.org/10.1007/s10055-022-00738-z

Biljecki, F., Ledoux, H., & Stoter, J. (2016). An improved LOD specification for 3D building models. Computers, Environment and Urban Systems, 59, 25–37. https://doi.org/10.1016/j.compenvurbsys.2016.04.005

Geris, A., Cukurbasi, B., Kilinc, M., & Teke, O. (2024). Balancing performance and comfort in virtual reality: A study of FPS, latency, and batch values. Software: Practice and Experience, 54(12), 2336–2348. https://doi.org/10.1002/spe.3356

Hermund, A., Lombardi, D., & Ramsøy, T. Z. (2023). The cognitive experience of architecture in real-life and virtual reality: A pilot study between architecture and neuroscience. In Digital Design Reconsidered: Proceedings of the 41st Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 41) (Vol. 2, pp. 863–872). eCAADe. https://papers.cumincad.org/data/works/att/ecaade2023_106.pdf

Hu, X., Wu, X., Ma, M., Xu, X., Gu, Y., Wang, G., Xu, Y., Meng, X., & Wang, L. (2025). Efficient VR rendering: Survey on foveated, stereo, cloud, and low-power rendering techniques. Virtual Reality & Intelligent Hardware, 7(5), 421–452. https://doi.org/10.1016/j.vrih.2025.08.003

Rodriguez-Garcia, B., Miguel-Alonso, I., Guillen-Sanz, H., & Bustillo, A. (2025). LoDCalculator: A level of detail classification software for 3D models in the Blender environment. SoftwareX, 30, 102107. https://doi.org/10.1016/j.softx.2025.102107

Spanjar, G., & Suurenbroek, F. (2020). Eye-tracking the city: Matching the design of streetscapes in high-rise environments with users' visual experiences. Journal of Digital Landscape Architecture, 5, 374–385. https://doi.org/10.14627/537690038

Spanjar, G., Suurenbroek, F., & Reynolds, R. (2024). The non-experts' experience of 3D city visualisations: Lessons for urban design practice. Journal of Digital Landscape Architecture, 2024(9), 471–481. https://doi.org/10.14627/537752043

Zhao, J., Jepma, A., Wallgrün, J. O., & Klippel, A. (2025). Optimizing velocity thresholds for fixation detection in virtual reality using the I-VT algorithm. In Proceedings of the 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2025) (pp. 90–96). IEEE. https://doi.org/10.1109/VRW66409.2025.00027

Ministry of Land, Infrastructure, Transport and Tourism (MLIT). (n.d.). Project PLATEAU (3D city models) [Data set]. Retrieved 2026-01-24, from Project PLATEAU website.

Linking Spatial Preference to Egocentric Information Availability

An Information Scape Approach to User-Centred Design

Ye Yint Aung[1], Shen-Guan Shih[2], Lucky Shin-Jyun Tsaih[3] and Kevin Harsono[4]
[1,2,3] *National Taiwan University of Science and Technology*
[4]*National Yang Ming Chiao Tung University*
[1]*11013803@mail.ntust.edu.tw, 0009-0008-3716-1891*
[2]*sgshih@ mail.ntust.edu.tw, 0000-0001-7108-9960*
[3]*akustx@ mail.ntust.edu.tw, 0000-0002-2468-0202*
[4]*kevinharsono.nycu@nycu.edu.tw, 0000-0003-1336-3380*

Abstract. Grasping the nature of spatial preference enables designers to anticipate the way people experience, engage with and adapt to spaces, leading to support in user-centred design process. This study introduces a new perspective of human spatial preference behaviour framed from the lens of individual information acquisition perspective, and its association with spatial preference behaviour in dining setting. We posit that human spatial preferences are not accidental but emerge from the limited information individuals can obtain and interpret to support their survival, opportunity seeking, and sense of security within the environment. Based on this notion, three egocentric information scape features were proposed to conceptualize different aspects of information acquisition scenarios. To evaluate these features, seat preference pattern was collected from the case study. Statistical analyses were conducted to access the influence of the proposed features on spatial preference patterns. The ordinary least squares (OLS) regression accounted for 67.9% of the variance in seating choice, exhibiting lower MSE and RMSE values compared to the support vector regression (SVR) model.

Keywords. Information scape, Isovist, Outdegree, Seat preference, User-centric design.

1. Introduction

Space in architecture extends beyond its geometric configuration; it embodies socio-spatial attributes that shape and are shaped by human experience. Understanding how individuals perceive, interpret, and adapt to their surroundings allows designers to

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 267-276.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

anticipate patterns of spatial behaviour and create environments responsive to user needs, an essential foundation for user-centric design. While a plethora of literature has addressed on design elements such as indoor design qualities, comfort, and environmental performance (Pan et al., 2025), the underlying human instinct for spatial comprehension, intertwined with preference formation in the built environment, has received comparatively limited scholarly attention.

This study introduces a new perspective of human spatial preference behaviour framed through the lens of information acquisition within the information scape. The information scape is conceptualized as a fabric of information which individuals utilize for their decision-making, orientation and survival in built environment (Aung et al., 2024). It is hypothesized that spatial positioning decisions are intrinsically connected to the significance and relevance of the available information in the particular environment. Individual instinctively seeks advantageous position and orientation from which meaningful information can be obtained, whether to ensure safety, identify opportunities or establish visual advantage.

While previous studies have employed discrete measures such as uniform, forward, and rearward information availability derived from Isovist analysis (Aung and Shih, 2023; Aung et al., 2025), holistic perspectives on egocentric information acquisition remain underexplored. Addressing this gap, the present study conceptualizes egocentric information availability features and investigates their influence on spatial preference within a dining context using statistical modelling. The ordinary least squares (OLS) regression accounted for 67.9% of the variance in seating choice, with lower MSE and RMSE values compared to support vector regression (SVR) model.

2. Background

Individuals possess natural instinct to prioritise relevant desired visual information than other from their surroundings (Carrasco, 2011). This selective process, known as, attention, enable the visual system to focus on certain information while filtering out others (Tünçok et al., 2025). While such attention improves cognitive performance at the focus of attention, diminishing processing accuracy in other regions of the visual field (Anton-Erxleben and Carrasco, 2013; Dosher and Lu, 2000). Directional-weighted information availability (IDR) portrays this scenario. It quantifies the amount of information perceived in a given direction. The highest weight is given to the forward-facing view with a gradual decrease toward the sides and rear, reflecting the nature of human spatio-visual attention.

Information availability is not uniform across space; its likelihood decreases as the distance between the individual and the information source increases. Distance-weighted information availability (IDS) informed by proxemics theory (Hall, 1966), which posits that spatial behaviour and perception are influenced by the varying degree of psychological and sensory impact associated with distance. Elements located within closer proximity convey stronger informational and social significance, while those farther away contribute less to immediate spatial understanding. This weighting coefficient follows a distance decay function commonly applied in spatial interaction models (Halás et al., 2014; Verma and Ukkusuri, 2025).

Spatial experience is also influenced by surrounding social dimension. The

outdegree metric, adapted from social network analysis (Sailer et al., 2021), quantifies the number of potential social connections or the likelihood of human presence that can be observed from a given location within the forward viewshed. In this study, this measure, conceptualised as social information, captures how visual access to others can shape spatial preference and influence individuals' positioning within the environment.

The Isovist, introduced by Benedikt (1979) as a geometrical representation of the visual field from a given location, has received extensive scholarly attention in spatial configuration research to describe how people visually perceive and interact with their environments (Dawes and Ostwald, 2021). However, conventional Isovist depicts panoramic visibility, which does not accurately capture the egocentric nature of human perception. To address this limitation, the present study extends Isovist by incorporating directional and distance-weighted coefficients to conceptualise information acquisition from an individual's perspective.

3. Methodology

The methodology of this study involves three phrases: (1) the conceptualization of egocentric information scape features derived from Isovist analysis (2) the collection of spatial preference behaviour in a dining environment using static snapshot method, and (3) the application of statistical techniques, including correlation analysis, ordinary least squares (OLS) regression and support vector regression (SVR) models, to examine the influence of these information features on spatial preference behaviours. Isovist analyses were performed in Grasshopper, a visual programming tools integrated within Rhinoceros 3D modelling software, while all statistical analyses were conducted using Python.

3.1. EGOCENTRIC INFORMATION AVAILABILITY

3.1.1. Directional-weighted information availability

Direction-weighted information availability quantifies the weighted amount of information perceived and prioritize in a particular facing direction, reflecting the directional bias of human perception within a given spatial context. In this measure, the highest coefficient was given to the forward-facing view with a gradual decrease toward the sides and rear, reflecting the nature of human directional attention.

$$IDR = \frac{1}{n} \sum w.L$$

IDR represents direction-weighted information availability of facing position, w denotes the coefficient of directional intensity of n-th radial segment, and L represents the corresponding radial length of an Isovist.

$$w = 1 - \mathrm{d}.(1 - b)/\pi\,, \qquad w \in [b, 1]$$

The coefficient w was computed using a linear transformation via the remap function in Grasshopper, where maximum $w = 1$, when angular distance $\mathrm{d} = 0$ (i.e., the ray is aligned with observer facing direction, and minimum $w = \mathrm{b}$ when $\mathrm{d} = \pi$ (i.e., the ray is directly behind). It offers the radial lengths of an Isovist into a specific range, ensuring that directional weighting reflects perceptual biases in spatial information

acquisition, in Figure 1 (b). The directional coefficient is assumed to be highest in the facing direction and minimized in the rearward direction, where the likelihood of accessing information is lowest. The coefficient range varies depending on the research context; however, for this study, we employed the upper bound value of 1.0 for forward facing and the lower bound value, 0.3 for the rearward direction.

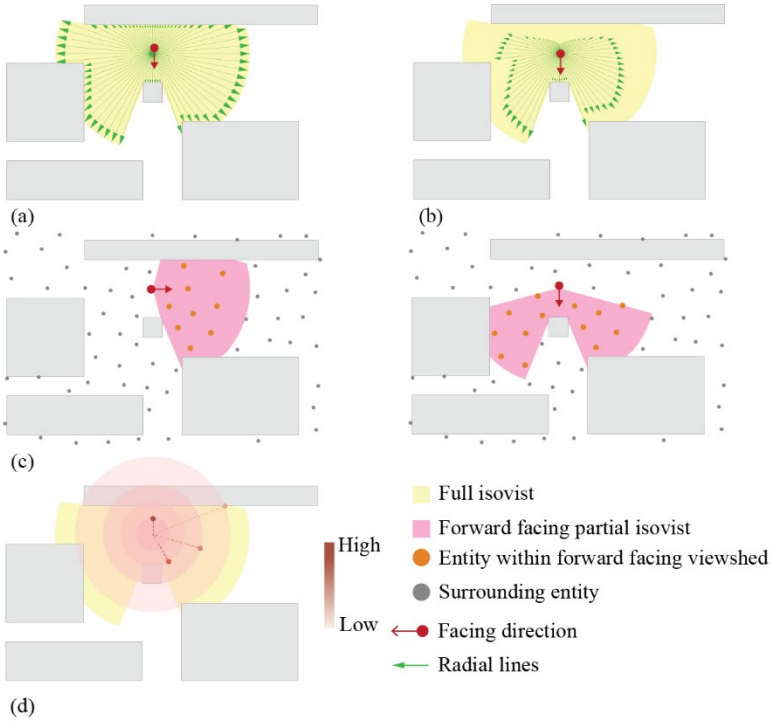


Figure 1 (a) Conventional Isovist viewshed (b) directional-weighted information availability (c) outdegree in two different directions in similar environment (d) distance-weighted information availability

3.1.2. Distance-weighted information availability

Distance-weighted information availability quantifies the strength and significance of information availability based on their spatial proximity to the individual from a specific location in the environment. Information availability diminishes as the observer's distance from the source increases, while it augments as proximity improves as illustrated in Figure 1 (d). This weight is derived from the distance decay function (Halás et al., 2014).

$$\text{IDS} = \sum RL_i \,.\, e^{-\propto RL_i}$$

The term IDS represents the distance-weighted information availability. This metric is computed by the average radial length (*RL*) of a given location (*i*) and the coefficient of weighted information accessibility (α). The coefficient α serves a parameter in an exponential decay function, reflecting the rate at which information

accessibility changes with distance. Its value is context-dependent, may vary according to information intensity or environmental conditions. In this study, α is set to 0.005, representing a controlled exponential decline in information availability as the average radial length increases.

3.1.3. Social information

Out-degree, originating from social network analysis (Sailer et al., 2021), provide a quantitative measure of directional connectivity, indicating the extent to which an individual maintains outgoing links with neighbouring people or groups within a social context. In this study, it is conceptualized as the numbers of immediate social information cues accessible within individual's 170 degree forward-facing viewshed. As illustrated in Figure 1(c), this measure captures how the amount of perceivable social information varies with facing orientation, even when individual occupy the same spatial position with a uniformly distributed social environment.

3.2. SPATIAL PREFERENCE PATTERN

A static snapshot method was employed as the primary observational approach to document seating preference patterns within the dining environment. Data were collected over 11-day period between 8 and 18 November 2023, covering both weekdays and weekends to capture a reprehensive range of individuals' behaviours. During each observation session, two photographs were taken to record the overall seating configuration. In total, 464 individual seating preferences were documented. Participants were free to select any unoccupied seat, allowing their spontaneous choices to reflect natural spatial preferences and social dynamics. This unstructured decision-making process provided valuable insights into how individuals assess positional advantage, interpersonal distance, and visual orientation when selecting their preferred seating locations.

3.3. STATISTICAL MODELS

Pearson's correlation analysis was first performed to examine the associations between seating preference patterns and the proposed information scape features, as well as to assess inter-feature correlations for model selection. Variance inflation factors (VIFs) were calculated to detect potential multicollinearity among the independent variables. Subsequently, two statistical models, Ordinary Least Squares (OLS) regression and Support Vector (SVR) Regression, were employed to compare model performance and evaluate the influence of the proposed features on seating preferences. Because SVR is sensitive to variables scaling, all input features were standardized prior to model training using the *StandardScaler* function from scikit-learn (Pedregosa et al., 2011). To ensure comparability between models, both analyses focus on linear relationships, with SVR hyperparameter optimized using a linear kernel (c = 0.1, gamma = 0.1, kernel = linear). Model performance was evaluated using the adjusted R-squared, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) metrics. The model with superior overall performance was further examined.

4. Result

4.1. EGOCENTRIC INFORMATION AVAILABILITY AND SEAT PREFERENCE PATTERN

Orientation plays a significant role in shaping information acquisition scenario, particularly in relation to visual information. The Figure 2 illustrated the social and directional-weighted information availability in two different seats configuration in the case study environment. The level of social information varies according to both orientation and spatial arrangement of seats. For example, as shown in Figure 2 (a), the window seat from the left yields the social information value of 16, whereas the central seat from the right records a higher value of 27. The comparison between conventional Isovist and directional-weighted information features was depicted in Figure 2 (b) and (c).

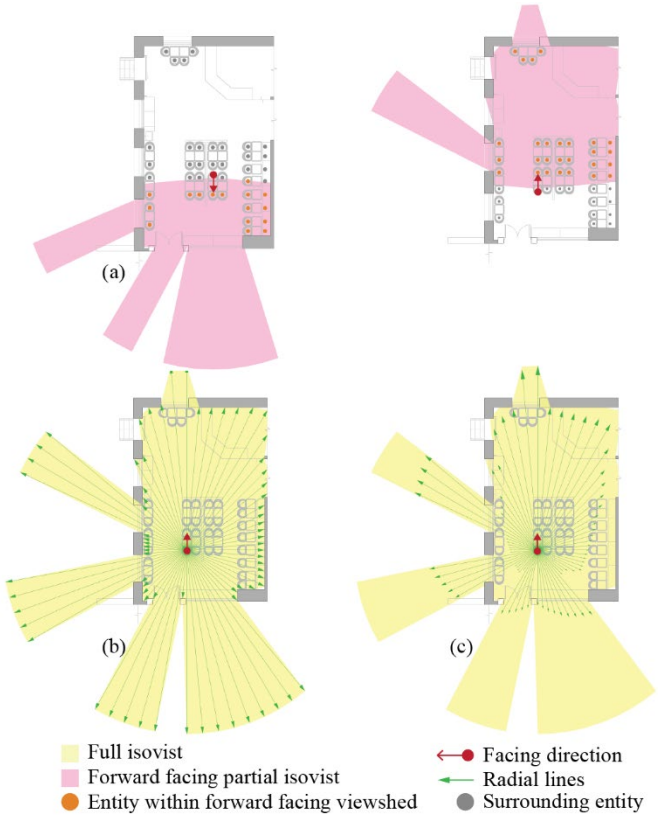


Figure 2 (a) Social information in two different seats, and comparison between (b) full isovist and its radials (c) directional-weighted information availability

Figure 3 presented the normalized gradient map of each information availability features and seat occupancy pattern. The highest values of directional-weighted found in the central seating area, while the lowest occur at the corner seats, as shown in Figure

3(a). However, an opposite trend emerges for distance-weighted information availability, with lower values found in the central and window-adjacent seats and higher values at seats against the wall, as illustrated in Figure 3(b). Similarly, social information index reaches at its peaks at seat positioned against the wall, while the lower values are recorded for seats facing the wall, as shown in Figure 3(c).

The Figure 3(d) illustrated the seat choice patterns across 11-days observation period, revealing that central seats facing the cashier or reception and seats against the wall, exhibit higher occupancy compared to the seat facing the wall. The more frequently occupied seats tend to offer advantages informational conditions such as clear visibility of the cashier and the comprehensive overview of the environment, while still maintaining a sense of awareness of the rear space. In contrast, seats facing the wall provide limited forward-facing information and naturally restrict access to visual information from behind, which may contribute to their lower occupancy.

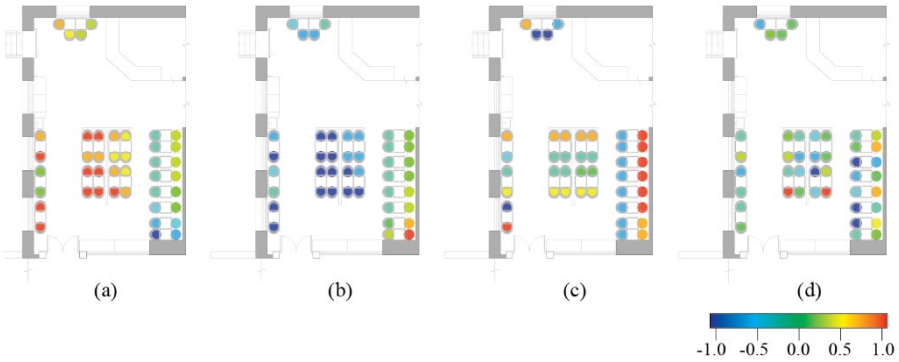


Figure 3 Normalized gradient maps of (a) directional-weighted (b) distance-weighted (c) social information, and (d) seat preference pattern

4.2. CORRELATION ANALYSIS

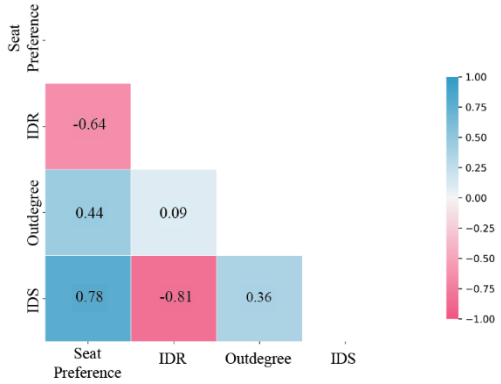


Figure 4 Correlation matrix of seat preference and information scape features

Figure 4 presents the result of Pearson's correlation examining the association between

spatial preference patterns and proposed information scape features. The results indicated that all correlations exhibited statistically significance. Specifically, distance-weighted information availability, IDS and outdegree (representing social information) both showed positive associations with seating preference. For instance, IDS demonstrated a strong positive correlation with seat preference ($r = .78, p < .001$), while outdegree exhibited a moderate positive correlation ($r = .44$, $p < .001$). In contrast, directional-weighted information availability, IDS displayed a strong negative correlation with seating preference ($r = -.64$, $p < .001$), suggesting that seats with lower directional exposure is associated with higher seat preference.

Beyond their correlations with seating preference, the analysis also revealed interrelationships among the proposed features. A strong negative correlation was found between IDR and IDS ($r = -.81, p < .001$), whereas outdegree and IDS exhibited a modest positive correlation ($r = .36, p < .001$). Given these results, variance inflation factors (VIF) were subsequently calculated to assess potential multicollinearity among the variables prior to selecting appropriate regression models for further analysis.

4.3. MODELS: OLS REGRESSION AND SVR REGRESSION

According to the results of VIF study, no collinearity was detected among the features, as all VIF values were below 10. The resultant regression models were summarized in Table 1, indicating that OLS model outperformed SVR model (using a Linear kernel). The OLS model explained 67.7% of the variance in seating choice and achieved lower MSE and RMSE values, compared to 60% explained by the SVR model. Therefore, the OLS model was selected for subsequent analyses and further interpretation.

The result of OLS regression, incorporating the three proposed egocentric information scape features, as shown in Table 2. Among these variables, Outdegree and IDS exhibited positive coefficients, suggesting that seats with higher levels of social visibility and distance-weighted informational access were more likely to be selected. Although the coefficients for both features exhibited statistically significant ($p < .001$). In contrast, IDR showed a negative coefficient with strong statistical significance ($p < .001$), indicating that seats characterized by lower directional-weighted information availability were more frequently preferred.

Table 1 Comparison of performance of OLS and SVR regression

Metric	OLS Model	SVR Model
Mean Square Error	52.12	59.78
Root Mean Square Error	7.22	7.73
R-squared	0.679	0.60

Table 2 OLS regression result

	Coefficient		Std err	t	P-value (P>|t|)	VIF
Constant	32.421	***	7.532	4.304	0.000	

IDR	-0.087	***	0.015	-5.734	0.000	5.88
Outdegree	0.426	***	0.051	8.330	0.000	6.07
IDS	0.314	***	0.080	3.947	0.000	6.63
R-squared	0.679					
Adjusted R-squared	0.677					

Note: Significance $*p < .01$, $**p < .05$, $***p < .001$

5. Conclusion

Architecture does not merely accommodate human behaviour, it subtly script and reshapes it. Understanding spatial preference allows designers to anticipate how people perceive, occupy and adapt to space. Although prior studies have focus on design quality and environmental comfort, the fundamental human instinct governing spatial understanding and preference in architectural settings has been insufficiently addressed. This study aims to introduce a new perspective of human spatial preference behaviour framed from the standpoint of an individual's information acquisition perspective, and its associations with spatial preference pattern in dining setting as a case study. We posit that human spatial preferences are not accidental but emerge from the limited information individuals can access and interpret to support their survival, opportunity seeking, and sense of security within the environment. Building upon this premise, three egocentric information scape features were proposed to conceptualize the dynamics of information acquisition. To validate these features, seat preference pattern was collected from a restaurant case study. Subsequently, a series of statistical models were employed to examine the influence of the proposed features on spatial preference patterns. The ordinary least squares (OLS) regression accounted for 67.9% of the variance in seating choice, exhibiting lower MSE and RMSE values compared to the support vector regression (SVR) model.

This study also has several limitations. First, it examined only linear relationships between egocentric information scape features and spatial preference patterns. Future research should investigate potential nonlinear relationships using advanced machine learning models to capture more complex behavioural pattern in different spatial configurations such as open-space offices, retail spaces, transit hubs and public spaces. Second, due to the fact that this study employed a dining environment as a case study, the findings may be context-specific. Further investigations across diverse spatial and cultural settings are necessary to validate and generalize the proposed framework. Third, this study primarily focused on the association between information acquisition scenarios and seating preferences, without explicitly incorporating environmental factors. Future studies integrating information scape features with these environmental variables could provide a more comprehensive understanding of human spatial behaviour. Despite these limitations, this study demonstrates the potential influence of information scape features on spatial preference, and offers designers valuable insights for developing more user-centred design and information-responsive configurations.

Acknowledgements

This work was partially supported by National Science and Technology Council,

R.O.C, (NSTC 114-2635-E-011-001).

Attribution Statement

ChatGPT (OpenAI, 2025) was used to (1) improve academic writing flow (2) to correct errors in spelling and grammar.

References

Anton-Erxleben, K., and Carrasco, M. (2013). Attentional enhancement of spatial resolution: Linking behavioural and neurophysiological evidence. Nature Reviews Neuroscience, 14(3), 188–200. https://doi.org/10.1038/nrn3443

Aung, Y. Y., and Shih, S.-G. (2023). On The Visual Aspect of The Information Scape in The Built Environment: Shilin Night Market as an Example. Proceedings of the 5th International Conference on Civil and Building Engineering Informatics, 486–493. https://civil.eng.chula.ac.th/iccbei2023/proceedings/

Aung, Y. Y., Shih, S.-G., Harsono, K., and Cheng, T. W. (2024). On the Information Scape of Spaces. Proceedings of the 29th International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2024), 3, 109–118.

Aung, Y. Y., Shih, S.-G., Tsaih, L. S.-J., Harsono, K., Wagiri, F., and Stavric, M. (2025). Examining Spatial Preference with Information Scape Using Partial Isovists. Nexus Network Journal, 27(3), 537–559. https://doi.org/10.1007/s00004-025-00823-1

Benedikt, M. L. (1979). To Take Hold of Space: Isovists and Isovist Fields. Environment and Planning B: Planning and Design, 6(1), 47–65. https://doi.org/10.1068/b060047

Carrasco, M. (2011). Visual attention: The past 25 years. Vision Research, 51(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Dawes, M. J., and Ostwald, M. J. (2021). Isovists: Spatio-visual Mathematics in Architecture. In B. Sriraman (Ed.), Handbook of the Mathematics of the Arts and Sciences (pp. 1419–1431). Springer International Publishing. https://doi.org/10.1007/978-3-319-57072-3_5

Dosher, B. A., and Lu, Z.-L. (2000). Noise Exclusion in Spatial Attention. Psychological Science, 11(2), 139–146. https://doi.org/10.1111/1467-9280.00229

Halás, M., Klapka, P., and Kladivo, P. (2014). Distance-decay functions for daily travel-to-work flows. Journal of Transport Geography, 35, 107–119. https://doi.org/10.1016/j.jtrangeo.2014.02.001

Hall, E. T. (1966). The hidden dimension (Vol. 609). Anchor.

Pan, J., Cho, T. Y., Sun, M., Steemers, K., and Bardhan, R. (2025). Environmental and spatial dynamics in a flexible workspace for hybrid work: A data-driven design framework. Building and Environment, 270, 112544. https://doi.org/10.1016/j.buildenv.2025.112544

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(85), 2825–2830.

Sailer, K., Koutsolampros, P., and Pachilova, R. (2021). Differential perceptions of teamwork, focused work and perceived productivity as an effect of desk characteristics within a workplace layout. PLOS ONE, 16(4), e0250058. https://doi.org/10.1371/journal.pone.0250058

Tünçok, E., Carrasco, M., and Winawer, J. (2025). Spatial attention selectively alters visual cortical representation during target anticipation. Nature Communications, 16(1), 8746. https://doi.org/10.1038/s41467-025-63795-3

Sketching Beyond The Page

Hybrid Design Workflows Using Augmented Reality

Alexandros Konstantinou[1], Georgios-Marios Bokos[2] and Dimitrios Papanikolaou[3]
[1,2,3]*National Technical University of Athens*
[1]*alexkwnnou@gmail.com, 0009-0005-1300-5331*
[2]*giorgos.mpokos@gmail.com, 0009-0002-5211-7158*
[3]*dimitri.c.pap@gmail.com, 0000-0002-6832-463X*

Abstract. This paper investigates the contribution of augmented reality (AR) to the architectural design process focusing on the sketch as a medium for creative exploration during early design stages. The aim is to examine how AR can bridge the gap between the two-dimensional nature of initial sketches and three-dimensional spatial perception, thereby enhancing conceptual thinking in the early phases of design development. The experimental framework was structured in two phases. In the first, AR sketching was applied in a controlled environment to explore the bidirectional translation between 2D sketches and 3D form, without external constraints. In the second phase, the same techniques were applied to a physical 1:50 scale model, with the goal of designing an organic canopy structure. The user study, for the second phase, involved seven architecture students, who followed the same iterative workflow combining sketching, AR spatial drawing, and physical model-making. Findings from an exploratory user study that AR supports enhanced spatial awareness, iterative feedback, and creative ideation within the conceptual design phase. Rather than proposing a replacement of the design process, the study investigates how it can be meaningfully enriched through the integration of AR technologies.

Keywords. Augmented Reality, Design Workflow, AR sketching, Spatial Perception, Real-time Spatial feedback, Conceptual design

1. Introduction

In contemporary architectural practice, technology is reshaping both representational tools and the ways design ideas are conceived and developed. AR has matured over the past decade, moving from novelty toward integration in education, visualization, and construction. Yet, its potential as a generative medium for early-stage sketching and spatial ideation remains underexplored. This research addresses this gap by examining how AR can function as an intermediary between analog sketches and

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 277-286.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

spatial design thinking in the conceptual phase.

A common early design practice is visiting a site and capturing spatial impressions through quick sketches that document relationships, movement, scale, and key reference points. However, two challenges persist. First, on-site sketches are typically imprecise, lacking accurate three-dimensional information, which makes spatial interpretation in the studio difficult. Topography, view corridors, or other spatial conditions are often flattened or misread when translated into 2D drawings. Second, in the studio, a disconnect remains between freehand sketching and early physical modeling. Although both support ideation, they tend to operate in parallel rather than interactively.

AR has the capacity to address these limitations. By enabling designers to sketch directly in space -on site or in the studio- it creates a real-time bridge between observation, representation, and fabrication. In doing so, AR sketching can strengthen the feedback loop between drawing, modeling, and conceptual design thinking, potentially amplifying the early stages of the design process.

2. Background

2.1. RELATED WORK

Over the past decade, several studies have explored AR as an educational or conceptual design tool. Schnabel (2009) and Oxman (2008) highlight how mixed reality environments foster iterative thinking and hybrid workflows that merge digital and physical exploration. For example, Kovacs et al. (2023) introduce Mr. Sketch, an experimental AR interface enabling freehand 3D line drawing to support conceptual exploration. In contrast, present study positions AR as a mediating layer between analog sketching and physical model-making. Similarly, Rasoulzadeh et al. (2024) present Strokes2Surface, prioritizes the computational reconstruction of geometry from sketches, the present research avoids formal reconstruction or optimization. Instead, it investigates how imprecision and manual translation between media contribute to conceptual exploration during early design stages.

Most studies focus on advanced headset-based systems or fabrication-oriented applications (Kyaw et al. 2025), leaving the domain of lightweight, handheld AR interfaces underexplored. This study contributes to filling this gap by investigating accessible, low-cost AR sketching tools as part of a hybrid analog- digital workflow. By framing AR as an interactive sketching environment rather than a representational medium, this research situates itself within a growing discourse on hybrid workflows that blend physical and digital authorship. The focus lies not on technological novelty but on methodological insight how AR transforms the designer's spatial perception and thinking processes during the earliest phases of design.

2.2. CONTRIBUTIONS

This paper presents the following contributions to the field of architectural design and representation:

- Proposes a hybrid analog- digital design workflow that integrates augmented reality

sketching with traditional drawing and physical model-making during early-stage architectural design.

- Demonstrates a method for constructing physical models directly from AR sketches, without relying on digital fabrication tools.
- Introduces a proposed hybrid design workflow and evaluates its use through a preliminary exploratory study.
- Provides preliminary qualitative insights into the pedagogical and creative affordances of AR sketching in conceptual design, identifying opportunities and limitations that inform future, more systematic investigations.

3. Methods

3.1. DISPLAY AND SOFTWARE SELECTION

To implement AR technology, we selected a portable handheld device (smartphone) as the primary display interface, in order to avoid the high cost and physical intrusiveness of headsets, while maintaining accessibility for architecture students. Head-mounted displays were considered unsuitable due to their obstructive nature during the design process. In contrast, handheld devices provide a transparent "window" through which users can interact with augmented content while retaining full awareness of their environment. Among several commercially available AR applications tested, those selected for use were evaluated based on model stability, ease of spatial sketching, and support for iterative interaction.

3.2. SYSTEM AND TOOL DESIGN

We propose a structured workflow that integrates AR sketching, analog modelling, and spatial collaboration into the early stages of architectural design. The system is articulated through a series of interrelated steps that enable the designer to alternate fluidly between physical and digital environments, using accessible tools and iterative strategies. The proposed workflow comprises the following stages:

- **Tool construction (step 1):** To ensure a minimal level of alignment between user gestures and the AR projection, a custom "digital pencil" was developed. This lightweight, needle-like extension was attached to the back of the smartphone, positioned directly beneath the rear camera. Its 12 cm length was calibrated through repeated testing to approximate the depth offset between the virtual drawing plane and the physical surface visible on the screen. However, precision remained limited due to the lack of real-time depth tracking or sensor calibration. While sufficient for exploratory use, future iterations could incorporate sensor-based or computer-vision methods to improve accuracy.
- **On-site sketching and data capture (Optional, step 2):** A user may begin by sketching directly in the physical site, using AR to overlay spatial lines, gestures, or references.
- **Transfer to studio environment (Optional, step 3):** The spatial data or

impressions captured on-site may be saved and brought into the design studio, serving as the conceptual seed for further exploration.

- **Physical construction (step 4):** Based on AR sketches projected in space, the user constructs physical representations of the form using simple materials (e.g., wire, tape, foam). The AR model functions as a live reference or scaffold during this step.
- **Design refinement through collaboration (step 5):** In the studio environment, additional design ideas are introduced and sketched directly in 3D space through the AR tool. These interventions are typically done in collaboration with a partner, enabling feedback, adjustments, and spatial negotiation.
- **Iteration and feedback:** Steps (4) and (5) are repeated as needed, allowing the project to evolve through cycles of spatial testing, physical modelling, and sketch-based augmentation.
- **Evaluation and reflection:** The resulting form is evaluated both spatially and conceptually, with attention to its coherence, adaptability, and responsiveness to the original constraints and intentions. Observations are recorded for analysis.

This system emphasizes low-cost, embodied experimentation, using accessible AR tools to enrich the designer's spatial intuition and bridge the gap between drawing, modelling, and making.

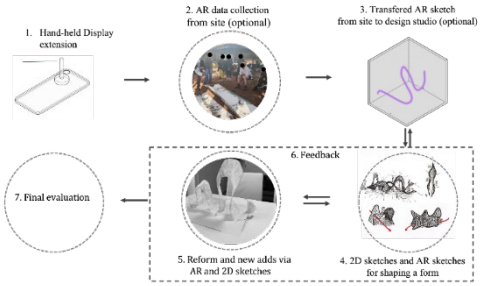


Figure.1 Proposed structured workflow integrating AR into the architectural design process (step by step)

3.3. EXPERIMENT

The experiment was designed to test the ability of the proposed AR-based workflow to address two critical challenges in early-stage architectural design.

The transfer of spatial data and impressions from the physical site to the design studio and the integration of freehand sketching with physical modeling through iterative feedback loops.

To explore these challenges, we applied the proposed methodology that combines traditional 2D sketching, spatial drawing in augmented reality, and analog model-making. The experiment was structured in two phases, each targeting a specific aspect of the research.

The First Phase, conducted in a controlled, neutral setting (a blank white surface), focused on isolating the feedback loop between 2D sketches and 3D spatial

interpretations. This allowed us to assess how effectively AR sketching supports transitions between dimensions without contextual interference, serving as a foundational calibration of the workflow.

The Second Phase, which was used for user study, introduced a synthetic design environment with spatial constraints and a scaled site model. Here, we engaged in a complete design scenario, using the AR workflow to capture, generate, and refine architectural ideas through real-time interaction between sketches, projections, and physical models.

3.3.1. ***First Phase***

We began by creating 2D sketches inspired by diverse architectural references. These sketches were then projected into AR environment, where we tested their spatial potential as three-dimensional forms. Following this, we reversed the process: we generated new 2D sketches based on the AR forms we had created, with the goal of reinforcing our geometric understanding.

Through this iterative feedback loop between 2D sketches and AR-based spatial interpretations, preliminary design ideas began to take shape. Then, we translated these ideas into physical models made of wire and masking tape, using AR projections as visual guides during construction. The first phase served as an experimental foundation for addressing the research's primary objectives: AR sketching, translating the initial sketch into three-dimensional space, and interactive feedback during construction supported by augmented reality.



Figure 2. Example of sketching in AR environment in the first phase (designing an organic form, app: DRAW)

3.3.2. ***Second Phase***

In the Second Phase, the Design Challenge chosen to be addressed through the use of AR is an architectural brief focusing on small scale constructions. According to the design brief, the selected site must accommodate the design of a terminal station for the suburban railway. The platforms should be covered with a canopy structure.

The process began with a site analysis that provided specific goals and needs, enabling the integration of actual architectural constraints into the experimental framework. A topographic physical model was then constructed, at 1:50 scale, featuring elevation differences and specific design requirements (such as height limits, column locations, orientation, etc.). This physical model supported the design development by anchoring it in material presence and scale.

Concurrently, reference points and axes were established both physically and digitally within the model. The design process unfolded through three successive design iterations, during which the same cyclical interaction and feedback loop between 2D sketches, AR sketches, and physical construction was applied, aiming to develop the initial compositional concept of a three-part canopy structure.

Iterative feedback across media

Starting with two-dimensional sketches, a series of conceptual ideas were developed and subsequently translated into AR sketches, creating a feedback loop between AR and 2D design. Through multiple iterations and adjustments, an initial form was generated, leading to the beginning of an assembly process guided by augmented reality projection.

After creating both 2D and AR representations of our form, the assembly process began, guided by the AR projection. Following the completion of the first form, the design process restarted based on the existing structure, aiming to integrate new shapes into it. The explorations extended beyond predefined constraints and design requirements, incorporating the limitation of embedding the first stage geometry into the new configuration. Various forms were tested until one was selected and constructed using the same method. A third form was then added, following the same steps used in the creation of the previous two.

4. User Study

To evaluate how users perceived the proposed AR based workflow, a user study was conducted with architecture students. Unlike the initial experiment, which tested the internal structure of the process, this study examined its clarity, usability, and cognitive impact when applied by others. Participants followed the same steps as in the second phase, creating spatial forms through iterative feedback between sketches, AR projections, and physical models under identical design constraints for consistent comparison. The focus was on how participants navigated across media and how AR affected their spatial thinking, rather than on the design outcomes. Each task required collaboration between at least two individuals to manage construction and real-time AR feedback.

Qualitative data from observations, interviews, and questionnaires revealed evolving spatial reasoning and diverse interaction strategies. The study functioned as a reflective lens, highlighting pedagogical and design implications beyond individual experience.



Figure 3. Photos and screenshots from the process of User study. The first two photos show how AR sketching interacts with the physical topography model and anchors. The third photo is a screenshot of an AR sketch from a student's try. In the fourth photo a pair of students try to assembly their design with the aid of AR projection. (app: Draw in Air)

4.1. RESULTS AND DISCUSSION

4.1.1. Observational Analysis

Through direct observation, semi-structured interviews, and questionnaires, the study explored how spatial thinking developed and what interaction strategies emerged in the AR environment. Distinct behavioural phases were observed. Initially, most participants struggled to conceptualize in augmented space. AR sketching was unfamiliar, causing hesitation in early drawings and 3D attempts. Over time, participants gained fluency, enabling freer creative thinking. Overall, while initial engagement was marked by uncertainty and limited spatial clarity, repeated experimentation led to normalization of the medium. The study relied on a structured questionnaire with both quantitative and qualitative elements, capturing multiple dimensions of participant experience throughout the process.

Table 2: Quantitative Questions samples.

Questions	Strongly disagree %	Disagree %	Neither agree nor disagree %	Agree %	Strongly agree %
The transition from sketch to AR was easy	-	-	28,57	57,14	14,28
Using AR can help with spatial perception	-	-	14,28	57,14	28,57
AR positively influences inspiration and production of creative ideas	-	14,28	-	57,14	28,57
AR helped you better communicate your design idea to other people	-	-	28,57	71,42	-
The use of AR helped to better understand the design scale	-	14,28	14,28	57,14	14,28

Regarding the closed ended questions, most participants reported no prior experience with AR in architectural contexts. The transition from paper based sketching to the augmented environment was described as easy or very easy. Participants generally agreed that AR enhanced their understanding of the spatial dimension of their initial concept. When asked about its usefulness for testing ideas during early design stages, four participants found it very useful, while three rated it as useful. There was strong agreement on AR's ability to support design communication, with most indicating that it prompted them to reconsider or further develop their proposals. Several participants also noted that AR improved their understanding of design scale.

4.1.2. Qualitative feedback and analysis

The analysis of the qualitative data revealed complex dimensions of participants'

experience with the experimental use of AR in the design process. Despite their lack of prior exposure to architectural applications of AR, participants gradually became familiar with the medium. As one participant noted: “At first, I didn’t know how to approach it, but soon it started to feel familiar.”

All participants identified AR as having a positive impact on their understanding of volume and spatial integration within their proposals. One student commented: “AR helped me understand how much space the form actually takes up on the model,” while another remarked: “No matter how good a sketch is, you can’t fully grasp the space until you see it three-dimensionally.” For some, transitioning from paper to spatial representation even led to a reconsideration of their initial intentions: “AR sketching added new parameters—it wasn’t like I had imagined it when I first drew it on paper.”

Beyond its supportive function, AR also appeared to stimulate imagination and creative thinking. One participant reflected: “As soon as I started sketching in space, I unlocked a part of my thinking. Overall, AR contributed to the activation of creative processes by enabling a fluid exchange between physical and digital modes of design, enhancing conceptual flexibility and spatial awareness.

Interestingly, participants developed a range of interaction strategies with the tool. One student relied almost entirely on AR to compose the form, while another used it selectively—as a means of verifying memory or ensuring geometric accuracy.

Regarding future use, participants expressed a strong interest in employing AR in other design stages, including client presentations, final visualizations, and even conceptual development on-site. One student suggested: “I’d like to use it to present ideas to others—it’s much easier for them to understand that way.”

Table 3: Qualitative Questions samples.

	Question: How did you feel when you were "drawing in augmented space"? Was this experience unfamiliar to you or did it seem natural?
Participant E	" Initially it felt unfamiliar because I haven't used AR before but it was easy. In the end it felt no different moving a pencil through space."
	Question: At what point in the process did you feel you had the clearest picture of your form? On paper or in AR? Why?
Participant A	"I felt I had the clearest picture in AR since it helped me to add more information on my design."
	Question: Do you think that translating the 2D sketch into an AR environment helped you better understand the form of your original idea spatially? If so, how?
Participant D	"Working on a sheet of paper forces you to approach the design as a projection, viewing it from a single axis, whereas in augmented reality, you engage with the problem simultaneously from multiple axes."
	Question: Have you ever felt that AR “limited” or “liberated” you in your thinking? How?
Participant B	"I felt liberated in my thinking because I could visually represent my design in 3D space."

4.1.3. Limitations and Future Work

While this study offers insights into the relationship between AR and spatial representation in early architectural design, it presents several limitations in scope and methodology.

First, the small participant sample limits the generalizability of the findings. Although a reduced group enabled in-depth qualitative observation, the results remain indicative rather than conclusive. Likewise, the short duration of the experiment prevented the study of long-term engagement, learning curves, or the evolution of design thinking.

Second, the experimental workflow, physical model, AR sketch, and 2D drawing restricted the ways participants could interact with the system. This approach revealed key behaviours but did not capture the full range of AR affordances, especially in more complex or open ended scenarios.

Third, the study did not systematically document or evaluate final design outcomes. Since the focus was on process rather than product, the influence of AR on formal quality or design resolution was not assessed.

Based on these limitations, the following directions are proposed for future research:

- Diversification of the participant sample, including individuals of varying experience levels, age groups, and design approaches.
- Extension of the experiment over time to monitor shifts in familiarity, ideation patterns, and decision-making.
- Integration of advanced AR technologies, such as headsets or spatial tracking, to examine how different interfaces influence spatial understanding.
- Comparative analysis with other tools, both digital and analog, to better articulate AR's distinct role in conceptual design.
- Evaluation of design outcomes, through structured assessment tools and external expert review, to measure the qualitative influence of AR on design synthesis.

This paper represents an initial step toward understanding hybrid AR workflows in conceptual design. Future research could extend its use across additional design phases including early ideation, site analysis, and presentation supporting a more fluid dialogue between physical and digital spatial thinking.

4.1.4. Conclusion

This research has highlighted the potential of integrating AR into early architectural design, particularly in enhancing spatial understanding, conceptual development, and the transfer of three-dimensional data from field to studio. It also demonstrated AR’s value as a tool for developing ideas directly in space, positioning it as an intermediate medium within the evolving logic of the design process.

The experiment with architecture students showed that even without prior experience, participants integrated AR into their design thinking. Observation, questionnaires, and qualitative analysis revealed how AR enhanced spatial perception,

informed form through feedback, and stimulated creative imagination. Rather than functioning as a static visualization tool, AR became a generative and expressive medium for architectural exploration.

Despite technical limitations and its early integration into education, the findings suggest AR can be a valuable instrument during conceptual design bridging the gap between site-based spatial experience and studio abstraction. Its ability to mediate between drawing, space, and design intent makes it a promising tool for a more intuitive, embodied, and integrated architectural education.

Attribution statement

The author acknowledges the use of OpenAI ChatGPT (GPT-5) for limited language assistance (sentence-level clarity, minor edits, and copy-editing). All conceptual framing, arguments, analysis, and conclusions in this paper are authored by the humans listed as authors, who retained full oversight and responsibility for the content. No AI was used to conduct analysis, produce results, or formulate arguments.

References

Azuma, R. (1997). A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments, 6, 355-385. https://doi.org/10.1162/pres.1997.6.4.355

Kidik, Aysegül & Asiliskender, Burak. (2024). Augmented Reality Technologies in Architectural Design Education: A Systematic Literature Review. 10.12955/pss.v4.394.

B. I. Kovacs, I. Erb, H. Kaufmann and P. Ferschin, "MR.Sketch. Immediate 3D Sketching via Mixed Reality Drawing Canvases," 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Sydney, Australia, 2023, pp. 10-19, doi: 10.1109/ISMAR59233.2023.00015.

Kyaw, Alexander Htet, Arvin Xu, Sasa Zivkovic, Gwyllim Jahn, Cameron Newnham, and Nick Van Den Berg. 2025. “AR Glulam: Accurate Augmented Reality Using Multiple Fiducial Markers for Glulam Fabrication.” arXiv. https://doi.org/10.48550/arXiv.2502.08566.

Milgram, Paul, and Fumio Kishino. 1994. ‘A Taxonomy of Mixed Reality Visual Displays’. IEICE Transactions on Information E77-D (12): 1321–29.

Oxman, Rivka. (2006). Theory and design in the first digital age. Design Studies. 27. 10.1016/j.destud.2005.11.002.

Oxman, Rivka. 2008. “Digital Architecture as a Challenge for Design Pedagogy: Theory, Knowledge, Models and Medium.” Design Studies 29 (March):99–120.

Rasoulzadeh, S., Wimmer, M., Stauss, P. and Kovacic, I. (2024), Strokes2Surface: Recovering Curve Networks From 4D Architectural Design Sketches. Computer Graphics Forum, 43: e15054. https://doi.org/10.1111/cgf.15054

Schnabel, Marc Aurel. 2009. “Virtuality as a Design Input.” International Journal of Architectural Computing 7 (4): 535–554.

Tetiranont, Chutima, Kanokwan Rattanaporn, and Phongsathon Kraisornsuthasinee. 2024. Augmented Reality Integration in the Early Architectural Design Process. Journal of Architectural Computing and Innovation 12 (1): 45–62. https://doi.org/xxxx.

An Embodied Interaction Framework for Analytical Engagement in Virtual Architectural Exhibitions

Yueying Zhang[1] and Tiantian Lo[2]
[1,2]*The Hong Kong Polytechnic University.*
[1]*yueying2025.zhang@connect.polyu.hk, 0000-0002-6892-8214*
[2]*tt2lo@polyu.edu.hk, 0000-0002-1992-0777*

Abstract. Online exhibitions, particularly graduation shows, have become a vital medium for architectural communication by extending engagement beyond physical constraints. Although embodied spatial cognition suggests that interaction can support situated interpretation, many implementations still function as static galleries that rely on audiences already knowing how to read plans, infer scale, and grasp project context. This study addresses that gap by proposing an Embodied Interaction Framework, developed and examined in an online graduation show set within a virtual model of a campus building, to shift visitors from passive viewers to active agents of analysis and critique. The framework comprises three levels, with spatial controls enabling perspective shifts across scales, object-level tools supporting inspection and comparison, and a contextual layer coupling exhibits with spatiotemporal architectural settings to guide interpretation. Based on qualitative feedback following a session of free exploration, this framework enables analytical engagement by first anchoring users in a familiar spatial presence and then providing interaction that transcends physical constraints to deepen comprehension beyond static display. This study offers a transferable model applying to virtual exhibition design and the broader practice of embodied space design.

Keywords. Virtual Exhibition, Online Graduation Show, Analytical Engagement, Embodied Spatial Cognition, Embodied Interaction

1. Background

1.1. EMBODIED COGNITION IN VIRTUAL ENVIRONMENTS

Embodied Cognition holds that spatial understanding arises from the body's actions and senses within the environment rather than from detached mental processing (Varela et al., 1992). In architecture, this shifts interpretation from passive or purely visual reception to bodily engagement with space, so buildings are understood through moving, sensing, and acting within them (Pallasmaa, 2012).

Recent advances in high-bandwidth networking and real-time rendering have turned virtual spaces, which resemble built environments with readable 3D form and

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 287-296.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

depth, into a new medium for social activities, and this shift was further accelerated during the COVID-19 pandemic (Ververidis et al., 2022). Although virtual environments mostly offer visual and auditory input, research indicates that the body is implicitly involved in perceiving motion, scale, and spatial depth, which supports embodied cognition through a sense of presence (Wiepke & Heinemann, 2024).

However, most studies still test whether virtual spaces can reproduce aspects of embodied cognition, while giving limited attention to the potential of interaction to enhance these effects, along with mechanisms. This study argues that virtual space is valuable not only as a digital substitute for physical settings but as a medium that enables embodied interaction with spatial layouts and objects (Figure 1), which opens new ways to interpret architectural experience.

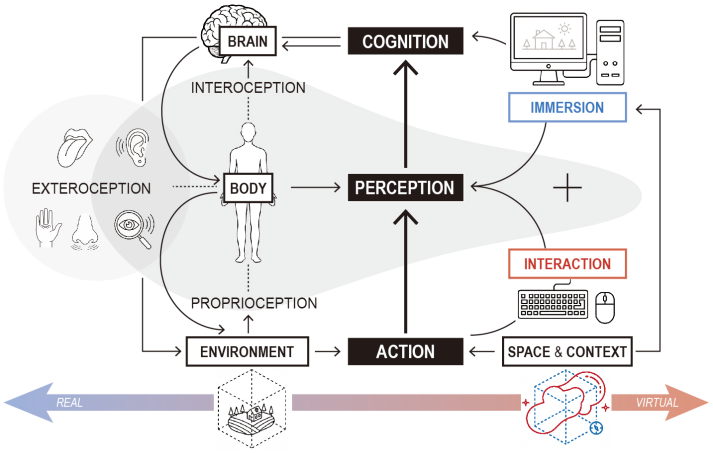


Figure 1. Conceptual model of embodied cognition in virtual space

1.2. ARCHITECTURAL EXHIBITIONS AND ONLINE GRAD SHOW

Architectural exhibitions are curated to guide how design works are read as arguments about form, use, and social value (Iacovou & Artopoulos, 2024), which are inherently processes of embodied spatial cognition where understanding arises through bodily engagement rather than purely visual observation. Visitors move through the gallery, act on displayed objects, and perceive panels, layouts, and models that together constitute the field of experience (Falk & Dierking, 2016). Through this continuous loop of action and perception, visitors can develop an understanding of design outcomes and interpret the curatorial narrative.

Within architectural education, digital reviews and exhibitions have become common practice and continue to play a central role in post-pandemic learning environments (Parrinello & Borucka, 2023). In this context, online graduation exhibitions have emerged as an important platform for showcasing student work (Wang & Han, 2022). According to a previous survey conducted by the author examining 100 online grad shows in 2020-2021 from architecture schools worldwide (Figure 2), despite the immersive and interactive potential of current virtual environments, most (80%) used web catalogues that remained limited in the capacity to represent the full complexity of architectural projects.

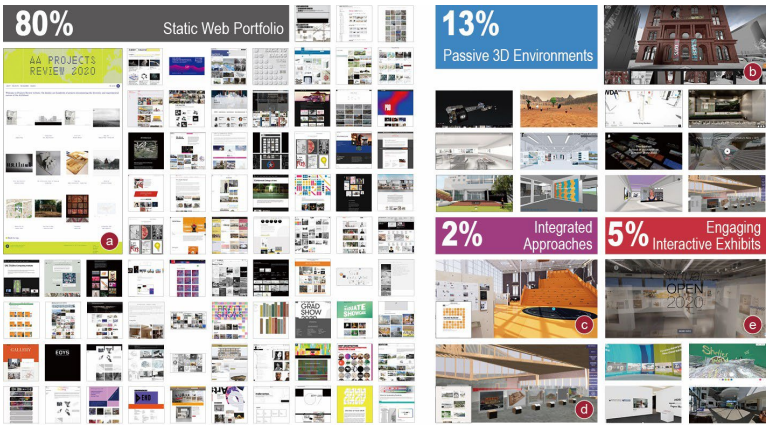


Figure 2. Typological survey of 100 online architectural grad shows from the onset of the pandemic

On the one hand, the annual grad shows serve not merely as a gallery but as a critical pedagogical rite of passage, a social forum, and an informal space for academic exchange. Therefore, during periods of pandemic isolation, some schools (totally 20%) like Cooper Union (Figure 2b), TU Delft (Figure 2c) and the University of Liverpool (Figure 2d) adopted immersive 3D environments through reconstructing familiar academic spaces, with the aim to help regroup the school community and communicate ideas to wider audiences (Cooper Union Architecture, 2020). However, most implementations focused on visual simulation and preset navigation within a 3D scene rather than on free roaming, while lacking interaction with exhibits.

On the other hand, many web catalogues and passive 3D environments simply move physical display boards or models into the virtual realm. Although architecture is inherently three-dimensional, presentations that rely on 2D drawings expect audiences to interpret plan and section drawings, judge scale, and picture material effects from fragmented representations instead of encountering space directly. Since architectural studio projects cover diverse themes and priorities, online grad shows face added complexity when communicating multiscale information, such as site context and spatial layout, structure, and materials responding to diverse social issues (Iranmanesh & Onur, 2021). However, only 5% of the shows enabled dynamic roaming and direct interaction with exhibits (Figure 2e), which limited the available actions to clicking information panels or performing model rotation and zoom, leading to shallow engagement and inconsistent understanding.

1.3. SCOPE AND OBJECTIVES

The online architectural graduation show serves as an ideal testbed for examining the role and mechanisms of interaction in supporting embodied cognition within virtual environments. It combines complex spatial content with varied forms of engagement that are suitable for studying embodied experience. Meanwhile, as established, its current applications rarely make full use of this potential and often provide only limited opportunities for deep cognitive engagement. This study, therefore, moves beyond basic assessments of presence or usability and focuses on the design, implementation,

and evaluation of an Embodied Interaction Framework intended to support richer modes of engagement within an online architectural graduation show. This paper addresses the following core research questions (RQs):

- RQ1: How can the abstract concept of embodied cognition be translated into a concrete and multi-level Embodied Interaction Framework specifically tailored for architectural analysis and critique?
- RQ2: To what extent does this new framework shift the user's role from a passive viewer and deepen their comprehension of the architectural works?
- RQ3: What design principles for virtual architectural exhibitions can be derived from the application and evaluation of this framework?

2. Methodology

2.1. CASE STUDY

The case study was built on a broader design concept for a virtual campus at a leading architecture school in Asia, intended to host projects across all academic years (Figure 3). This paper focused on one building modelled as a virtual environment, which was selected because it serves as the daily activity space for the architecture undergraduates, and therefore offered a highly recognizable and relatable context for the exhibition. The content hosted within this building prototype was the annual bachelor's graduation show, including all 22 graduating design studios with different themes.

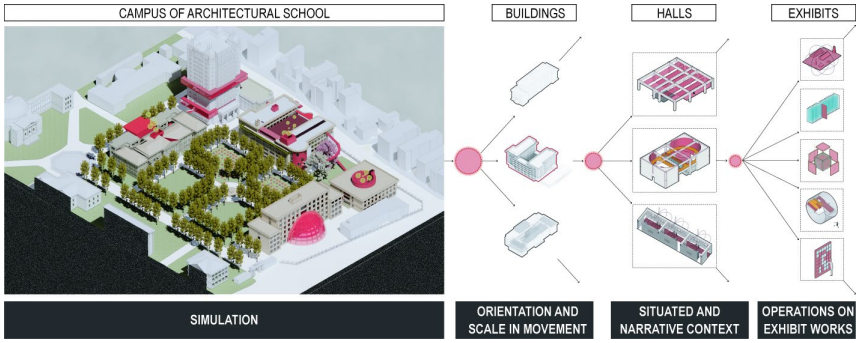


Figure 3. Overview of the virtual graduation show prototype

This paper analysed the characteristics of each project and its site and placed them inside a building whose spatial form closely mirrors the real setting, so that visitors retained a conventional embodied sense of being there. Following the hierarchy of architectural exhibition design, this prototype structured the visitor's journey in virtual space through three layers of interaction structure, including movement across the campus and into rooms, roaming and reading within galleries, and the direct interpretation of specific architectural works. The detailed interaction design is presented in Section 3.

2.2. TECHNICAL WORKFLOW

This study used Unreal Engine 5 (UE5) for its advanced real-time rendering and Blueprint visual scripting system that supported complex interaction design. The virtual campus and exhibition halls were first modelled in Rhinoceros and then imported into UE through the Datasmith plugin. Then, scenes were organized into levels, and a first-person camera was configured to provide immersion on a desktop screen. Interactive functions with mouse and keyboard input were implemented through Blueprint scripting. Finally, a stand-alone application was packaged for user evaluation.

2.3. EVALUATION METHOD

A stratified sample of 12 participants was recruited to capture a diverse range of perspectives, including architecture students familiar with the projects, faculty members involved in assessment, and non-professionals. After a short introduction, participants freely explored the online exhibition and interacted with the prototype at their own pace. A semi-structured interview then examined usability, sense of spatial presence, and perceived understanding of the architectural projects.

Based on a comprehensive review of the design together with user feedback, the study examined how interaction can enhance embodied cognition and thereby support visitors' understanding of works in virtual exhibitions, and finally synthesized the Embodied Interaction Framework.

3. Embodied Interaction Design Framework

To translate the abstract principles of embodied cognition into a concrete design strategy that addresses the passive nature of most online exhibitions, this prototype proposes the Embodied Interaction Design Framework, including three connected layers that guide visitors from initial orientation to hands-on examination, and finally to situated interpretation.

3.1. ORIENTATION AND SCALE IN MOVEMENT

First of all, this prototype treats movement and scale as one continuous experience so that visitors arrive at the right place and distance when needed (Figure 4). Cross-scale navigation complements first-person walking by using spatial hyperlinks through clear portals and carefully placed teleports that take visitors directly to the next room (Figure 4a-1,2). Furthermore, vertical change is delivered through elevating platforms or a locally gravity-free cylinder that permits inverted walking or roof-down viewing in selected zones (a-3). The prototype extends spatial perception and perspective shift, turning movement itself into an exploratory act and revealing relations among roof, structure, and floor that are difficult to see under normal gravity, while zone limits are kept to maintain orientation.

In addition, the prototype lets the visitor adjust bodily scale and viewpoint in ways that stay connected to movement. For example, a rabbit-hole path (a-4) shrinks the avatar to model scale, allows entry into a project interior, and then returns to room scale in one fluid sequence. Role switching (a-5) complements this by offering avatars with different body sizes, such as a child, which changes eye height, reach, and field of view, and therefore supports empathetic reading of space.

The prototype also makes the displays responsive to the visitor's movement, as architectural exhibitions integrate exhibits as constitutive elements of the gallery space (Figure 4b). Specifically, a panoramic panel system follows the moving viewer and keeps captions easy to read without blocking exhibits, which helps maintain immersion during motion and reduces the need to stop and refocus.

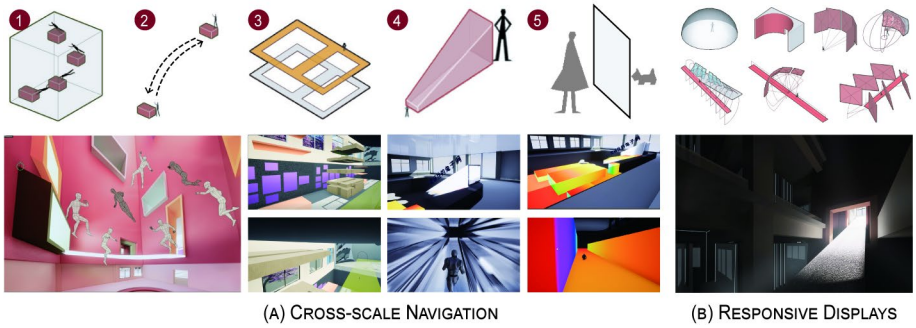


Figure 4. Spatial-level interaction mechanics and UE screenshots

3.2. OPERATIONS ON EXHIBIT WORKS

After visitors are settled in the space, the prototype shifts from seeing to examining by turning traditional drafting techniques (like sectioning or exploding views) into direct actions on the model. Slicing tools transform sectioning from a fixed representation into a real-time operation, allowing planar cuts at any angle or height and helping users examine interior relationships more directly in situ (Figure 5a). Rotation tools then give full control of model attitude (Figure 5b), with free rotation for quick scanning, rotation around a fixed axis which stabilises orientation for steady comparisons of silhouette, roof profile, and floor relationships, and view toggles for switching among standard reference views.

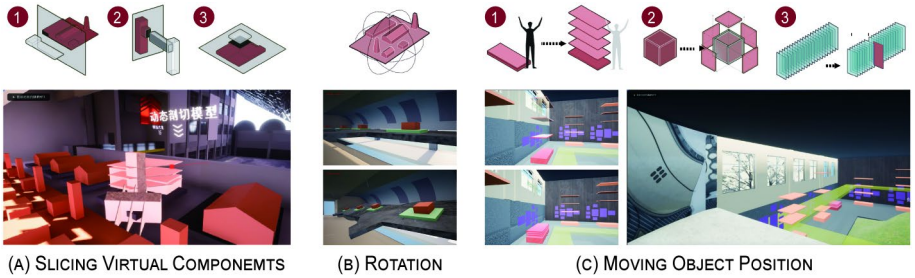


Figure 5. Object-level interaction tools and UE screenshots

Meanwhile, to understand the architectural composition of the whole, the prototype facilitates controlled movement of its constituent parts, thereby exposing underlying design principles and structural logic (Figure 5c). Progressive unfolding shows how a project grows or opens from inside to outside, skin and core can be separated to inspect envelope and interior systems side by side, modular arrays can be filtered and rearranged to reveal patterns, and stacked drawings can be fanned into 3D space so that 2D evidence stays with the model. A simple enlarge tool further adjusts the viewing

scale from territory to full-scale detail to make joints legible without leaving the scene.

Apart from letting visitors set up the reading, this prototype also allows interactions for judgment. Modular arrays can be filtered and rearranged to reveal patterns and to test alternative organizations (Figure 6b). Likewise, material switching at a basic level allows quick trials of primary finishes that support personal judgments about effect and fit (Figure 6b-2). Stacked drawings can be pulled into three-dimensional space so that plans, sections, and elevations remain next to the model, and visitors can then reorganize these drawings to form a personalized reading path (Figure 6a).

3.3. SITUATED AND NARRATIVE CONTEXT

As the interpretation of an architectural work depends on its setting as much as on the object itself, this prototype provides interactions to modify contextual elements so that visitors can read the work within its environment. While the virtual gallery follows a familiar layout where multiple design works share a common site model, visitors can place different proposals onto this shared model to observe and compare environmental relationships. For renovation or expansion projects, states can also be toggled in place to compare existing and proposed conditions directly (Figure 6b-1).

Besides, dynamic thematic and temporal environments are used to create situated narratives. Multiple temporal states of the campus are embedded along a single path and linked to exhibit themes. For example, a room themed around traditional timber construction opens to a spring scene with blossom views, while a hall on ruin renovation is set in a sparse winter landscape (Figure 5c). These curated settings also function as transitions. Along a corridor, a visitor may encounter the campus in autumn foliage (Figure 4c) or the first light of dawn that evokes an all-night drawing session, turning shared memories into scenes that bridge one theme to the next.

In addition, user-driven environmental modifiers allow direct control over surrounding conditions (Figure 6c). Visitors can select weather (c-1), lighting (c-2), and time of season (c-3) to simulate a design under specific atmospheres. As architecture is experienced over time rather than as a single static object, the prototype supports evaluation beyond a fixed rendering or a single curated moment and extends interpretation across multiple scenarios by varying conditions in place.

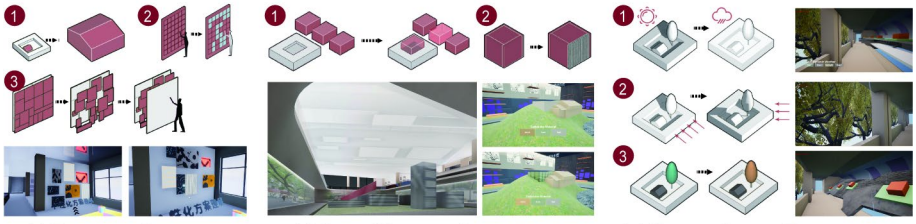


Figure 6. Contextual interaction concepts and UE screenshots

4. Results

4.1. FEEDBACK ON SPATIAL INTERACTION

Participants agreed that the virtual campus provided a clear sense of spatial grounding, which functioned as an essential cognitive anchor to establish a stable sense of presence. Familiar forms, materials, lighting, and atmosphere supported orientation and encouraged most users to navigate through free roaming rather than hyperlink navigation, treating the exhibition as a continuous spatial journey. Alumni and staff reported strong emotional responses linked to shared routines and everyday experiences on campus, while participants unfamiliar with the real campus occasionally felt disoriented and suggested that clearer visual cues would be helpful for wider audiences.

Additionally, participants were pleasantly surprised by the spatial interventions. Added circulation and viewing spaces were considered well integrated yet distinct enough to offer new spatial experiences. Tools that altered body scale and viewing height were seen as helpful for understanding spatial hierarchy and the relationship between overall form and interior spaces. Role switching between different body sizes and identities was described as useful for imagining how diverse users might inhabit the same environment and for developing a more empathetic reading of the projects.

4.2. FEEDBACK ON OBJECT INTERACTION

Participants responded positively to the interactive tools for examining architectural models. Many felt that real-time slicing made sectional and structural logic clearer and appreciated being able to place cuts at chosen heights and angles. They reported that this level of control allowed them to uncover spatial sequences, assemblies, and voids with greater precision than in static drawings.

The controlled movement of building components received strong feedback as well. Participants found that unfolding, separating, or regrouping parts helped them grasp organisational principles that are often difficult to perceive in conventional exhibitions. They especially valued the ability to view envelopes and interior systems side by side and to reveal modular patterns through filtering and rearrangement. Users also appreciated being able to enlarge specific areas without leaving the scene, which helped them inspect details at a clearer scale.

Furthermore, users felt that evidence and objects remained aligned and easier to interpret when plans, sections, and elevations could be pulled into the environment and kept next to the model. Several noted that organising drawings into a personalised reading path reduced the need for constant switching between media. They suggested that this alignment between documents and models should be extended across more exhibits, as it supported smoother transitions from inspection to interpretation and encouraged more deliberate comparison and judgment.

4.3. FEEDBACK ON CONTEXTUAL INTERACTION

Participants agreed that contextual settings such as weather, lighting, and seasonal atmosphere supported a clearer understanding of each project. They reported that this specific context helped them connect formal decisions with environmental and programmatic intentions rather than reading the works as isolated objects.

Users also valued situations where multiple design schemes or temporal states were related to one another. They indicated that comparative models, for example, placing

different proposals on a shared site, made it easier to judge strengths and weaknesses across alternatives. A faculty member suggested further development of comparative settings, including side-by-side arrangements of different options and overlays of historical and proposed states, as they felt these configurations encouraged more nuanced and critical interpretation of architectural transformation.

5. Discussion

5.1. BALANCING REALISM AND ANALYTICAL VALUE

In response to RQ1, results indicated that the designed interactions aligned with how participants oriented themselves in space, examined design logic, and made sense of project narratives. This framework successfully shifts users from passive viewers to active agents (RQ2), which lies in a balance in which the realistic spatial simulation provides an essential cognitive anchor for presence, while the object and contextual tools further provide analytical affordances. Finally, this study proposes a design guideline for virtual architectural exhibitions, linking realistic embodied grounding with interactive tools beyond physical limitations and forming a continuous loop between interaction and cognition for analytical engagement (Figure 7).

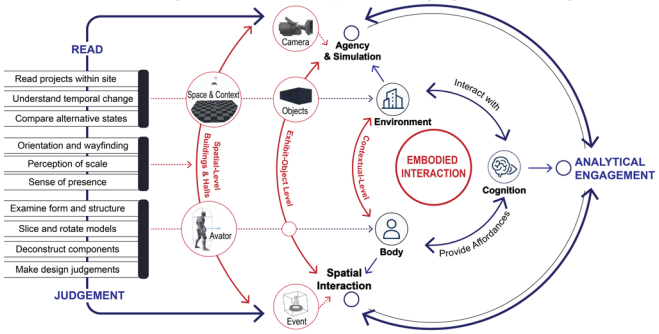


Figure 7. Embodied Interaction Framework for analytical engagement in virtual exhibition design

5.2. LIMITATIONS AND FUTURE WORK

Nevertheless, the study relied predominantly on qualitative analysis, and the quantitative evidence was not analysed in a systematic way. Future work should include larger and more diverse samples and incorporate behavioural, experiential, and questionnaire-based metrics to triangulate the findings. The prototype also functioned as a single-user system, which limited real-time social interaction and reduced opportunities to capture collaborative interpretation processes and interoperable data traces. Future development should introduce multi-user capability and integrate external data sources such as BIM or GIS, and adapt this framework to the specific pedagogical goals of different architectural topics.

6. Conclusion

This study proposed and evaluated an Embodied Interaction Framework for virtual architectural exhibitions, in order to provide users with the agency to move beyond

passive viewing and perform deliberate acts of analysis, interpretation, and critique. This paper demonstrates that by structuring interaction at spatial, object, and contextual levels, a virtual exhibition can be transformed from achieving simple immersive participation toward enabling deep analytical engagement. Besides, it argues that the next generation of digital architectural spaces must be conceived not as simulations to be viewed, but as interactive and participatory environments designed for actively embodied cognitive construction.

Acknowledgement

The author also acknowledges Chuanfei Yu at Southeast University for his guidance on the design aspects of this work.

Attribution Statement

GPT-5 (OpenAI, 2025) was used for minor language refinement, and Gemini 2.5 Pro (Google, 2025) was used to generate icon drafts.

References

Cooper Union Architecture. (2020, June 17). End of Year Show. https://archeoys2020.cooper.edu/

Falk, J. H., & Dierking, L. D. (2016). The Museum Experience Revisited. Routledge. https://doi.org/10.4324/9781315417851

Google. (2025). Gemini 2.5 Pro [Large language model]. https://gemini.google.com/

Iacovou, P., & Artopoulos, G. (2024). The architecture exhibition as a stage of mediated narratives. *Digital Creativity*. https://www.tandfonline.com/doi/abs/10.1080/14626268.2024.2304811

Iranmanesh, A., & Onur, Z. (2021). Mandatory Virtual Design Studio for All: Exploring the Transformations of Architectural Education amidst the Global Pandemic. *International Journal of Art & Design Education*, 40(1), 251–267. https://doi.org/10.1111/jade.12350

Lee, J. M. (2025). Embodied Learning in Architecture: A Design Studio Model Utilizing Extended Reality. *Buildings*, 15(13), 2158. https://doi.org/10.3390/buildings15132158

OpenAI. (2025). GPT-5 [Large language model]. https://openai.com

Pallasmaa, J. (2012). The Eyes of the Skin: Architecture and the Senses (Third Edition). John Wiley & Sons Inc.

Parrinello, S., & Borucka, J. (2023). VREA Project—A digital curator for architecture and digital perspectives for heritage management and enhancement. https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-289-2023

Varela, F. J., Rosch, E., & Thompson, E. (1992). Embodied Mind: Cognitive Science and Human Experience. The MIT Press.

Ververidis, D., Nikolopoulos, S., & Kompatsiaris, I. (2022). A Review of Collaborative Virtual Reality Systems for the Architecture, Engineering, and Construction Industry. *Architecture*, 2(3), 476–496. https://doi.org/10.3390/architecture2030027

Wang, Z., & Han, H. (2022). Design Strategy for the Curation of Digital Exhibition Experience. In C. Stephanidis, M. Antona, & S. Ntoa (Eds.), *HCI International 2022 Posters* (pp. 320–327). Springer International Publishing. https://doi.org/10.1007/978-3-031-06391-6_42

Wiepke, A., & Heinemann, B. (2024). A systematic literature review on user factors to support the sense of presence in virtual reality learning environments. *Computers & Education: X Reality*, 4, 100064. https://doi.org/10.1016/j.cexr.2024.100064

Phygital Gamification for Co-creation Environmental Education

An XR-Based System for Urban Waste Cognition and Collaboration in Smart Cities

Yuhao Li[1], Tian Tian Lo[2], Zilu Wang[3], Usman Khalid Rafique[4] and Xi Chen[5]
[1,2,3,4,5]*The Hong Kong Polytechnic University.*
[1]*yuhao33.li@connect.polyu.hk, 0009-0000-0619-7455*
[2]*tt2lo@polyu.edu.hk, 0000-0002-1992-0777*
[3]*ziluu.wang@connect.polyu.hk, 0009-0001-3239-1283*
[4]*tameem.ghani@gmail.com, 0009-0004-1667-877X*
[5]*xi-ceci9.chen@connect.polyu.hk, 0009-0009-3798-6225*

Abstract. The motivation-behaviour gap in urban waste education in megacities such as Hong Kong remains a pressing issue. Learning is repositioned from passive knowledge transfer toward immersive and collaborative practice through a physical-digital (phygital), extended reality (XR) based dual player system. The proposed framework integrates historical reconstruction with gamified mechanics across past–present–future timelines to elicit empathy, sustain motivation, and translate cognition into action. The research investigates how a system combining phygital interaction and multi-ending narratives can catalyze collective empathy and promote sustainable behaviour in complex urban contexts. A prototype has been implemented that merges role-based tasks, mirrored devices, real-time feedback, and cross-temporal collaboration. Player decisions generate dynamic environmental visualizations and outcome summaries encompassing environmental, social, and emotional dimensions, providing a multifaceted evaluation of impact and engagement. Preliminary findings suggest improved contextual understanding, stronger affective engagement through co-creation, and better transfer of sustainable decision patterns to real-world practice. The resulting framework establishes a coherent interaction loop for empirical evaluation and offers a scalable pathway for XR-enabled and socially engaged environmental education applicable to diverse community settings.

Keywords. XR gamification, phygital interaction, collaborative gamification, environmental cognition, collective collaboration

1. Introduction

With the rapid pace of urbanization, traditional environmental education faces significant challenges in fostering deep public understanding of urban waste issues and encouraging sustained behavioural change. In megacities such as Hong Kong, a

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 297-306.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

cross-contextual gap in the perception of historical waste crises has weakened intrinsic motivation for environmental action. As a high-density metropolis, Hong Kong's waste generation has accelerated almost in parallel with its urban development. According to the Hong Kong Environmental Protection Department's 2023 Solid Waste Monitoring Report, the city disposed of an average of 10,884 tons of municipal solid waste per day in 2023, about 67% of which still went to landfills despite a 33% recycling rate (EPD, 2024).

Conventional environmental education models, which emphasize knowledge transmission and policy advocacy, excel in information delivery but struggle to attract voluntary public participation and evoke emotional engagement. As immersive media technologies such as XR and Mixed Reality (MR) become more accessible, new forms of interactive and experiential learning have emerged. Enhanced participation and emotional resonance have been shown to significantly strengthen cognitive activation and cultural empathy (Song et al., 2025), while team-based gamified mechanisms can improve engagement, creativity, and cross-temporal collaboration in sustainable design education (Basak et al., 2025). Traditional approaches often stress theoretical knowledge but neglect lived experience, making it difficult to induce deep behavioural transformation; immersive education, by contrast, demonstrates greater potential in bridging that gap.

In this study, gamification serves as the conceptual foundation for reinterpreting environmental education. Systematic game mechanics can transform cognition and behaviour when integrated into an interactive, XR-based platform that supports phygital collaboration and multi-user participation. Historical contextual reconstruction fosters collective empathy and collective responsibility, offering a fresh approach to urban waste education.

2. Background

2.1. THE SOCIAL AND COGNITIVE DILEMMA OF URBAN WASTE

Public participation in waste reduction remains limited despite long-term promotional efforts. Hong Kong's fast-paced consumer culture and shifting communication patterns have widened the generational gap in environmental awareness. Older generations, shaped by past material scarcity, tend toward frugality and reuse, whereas younger citizens often lack historical empathy and contextual understanding. Traditional education fails to bridge this cognitive divide, leaving a disconnection between environmental knowledge and action. Urban waste thus represents not only an ecological but also a cultural issue, in which inadequate incentives, information gaps, and regulatory complexity continue to impede consistent sustainable behaviour (Judmaier et al., 2024).

2.2. THEORETICAL FOUNDATIONS OF GAMIFICATION AND IMMERSIVE LEARNING

Immersive and gamified learning have been shown to enhance motivation, self-efficacy, and behavioural internalization (Makransky et al., 2021). Evidence from digital-heritage studies further indicates that combining gamified design with

collaborative human–computer interaction enhances immersion and learning outcomes, encouraging social engagement (Tromp et al., 2025). Collectively, these findings form a solid theoretical foundation for XR-based phygital education, where immersive media drive emotional engagement while gamified processes sustain motivation and persistence.

2.3. MECHANISMS OF CO-CREATION AND HISTORICAL RECONSTRUCTION

The concept of co-creation has been widely applied in healthcare, engineering, architecture, and technology. Under immersive conditions, a dual player collaborative environment allows participants to reconstruct historical events and social issues in blended physical–virtual spaces, transforming unidirectional experiential learning into a bidirectional co-constructive process that fosters knowledge exchange, empathy, and creative synergy.

XR-based interdisciplinary co-creation benefits the design industry while offering an analogous model for social learning in education, where participants jointly experience, express, and reflect through historical reconstruction and problem resolution to achieve cognitive transfer and affective collaborative development (Hatana, El-Jarn & Southern, 2020). Co-creation thus functions as a process of social learning and cultural transmission that emphasizes shared meaning and collective responsibility. In mixed-age contexts, the dual player mechanism mediates generational exchange through spatial and narrative collaboration, deepening mutual understanding between historical and contemporary perspectives. Shared values and common goals unite participants from different disciplines, forming socially engaged networks that drive innovation for the common good (Orr et al., 2021).

From a historical reconstruction perspective, XR technologies enable virtual re-enactment of cultural memory and civic engagement. Roaro (2021) demonstrated that combining 3D modelling, archival documentation, and oral histories can reconstruct local collective memory in immersive VR form. Similarly, integrating 3D surveying, historical archives, and HBIM models transforms tangible and intangible heritage into interactive learning resources, allowing users to perceive social and spatial changes across eras (Banfi et al., 2021).

In summary, research on educational interaction and digital heritage underscores the potential of immersive technologies in co-creation and reconstruction. Base on this, the XR-enabled mechanism enhances these processes through interactive historical reconstructions, enabling collaboration across generations and the transmission of knowledge and cultural responsibility. This represents cognitive innovation through technological advancement and the reintegration of affective and social learning.

3. Methodology

This study centres on a dual player co-creation and historical reconstruction mechanism and accordingly designs and develops a phygital co-creation system that integrates virtual narrative with physical interaction. Through deeply embedded gamification, the system seamlessly integrates virtual historical scenarios with tangible interaction devices, leveraging the motivational, immersive, and collaborative

affordances of games. By means of a gamified collaboration framework, it combines virtual historical scenes with real-world interactive apparatus, thereby constructing an interaction platform that simultaneously supports immersive experience, social collaboration, and environmental education. In the experimental setting, two players wear VR headsets and collaborate within separate yet logically connected spaces. Standing back-to-back, they operate task-specific physical devices placed on their respective tables (see Figure 1).

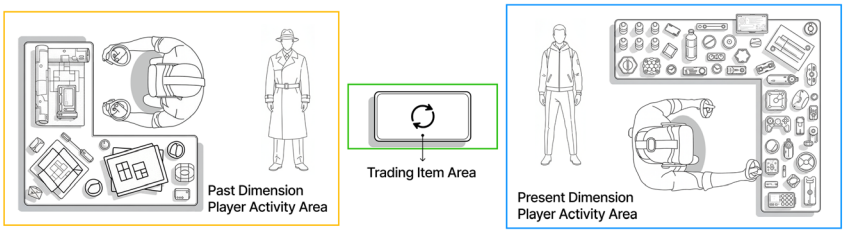


Figure 1. Dual player activity layout

The overall structure of the system adopts a tripartite temporal logic comprising the past, the present, and the future. Players assume the roles of urban citizens from different time periods and undertake tasks such as waste sorting and resource recovery. During gameplay, each player encounters obstacles that cannot be resolved within their own temporal context; at such moments, the system prompts the player situated in the other temporal context to perform corresponding actions. Through temporal coordination, these actions assist the counterpart and enable cooperative co-creation between the two players. Player behaviours and decisions not only affect the state of the contemporary virtual environment but also generate cascading feedback within the "future" dimension. In this way, immersive experience facilitates a dynamic transition from individual cognition to collective action, fostering sustainable ecological thinking and a sense of civic responsibility through interaction.

3.1. DUAL PLAYER CO-CREATIVE FRAMEWORK

To construct and validate the co-creation mechanism, the system establishes a dual player experimental environment that combines tangible interaction devices with a virtual narrative interface. Cantered on gamified design logic, this environment is organized around four key elements: task-driven progression, situational simulation, real-time feedback, and dual player collaboration. Through role differentiation and task interdependence, players are guided to engage in sustained exploration, communication, and cooperation within a multidimensional space where the virtual and the physical are interwoven.

Physical interactive devices are placed on each desk and serve as control interfaces corresponding to actions in the virtual world. The system situates the two players in distinct temporal and spatial versions of Hong Kong:

- Past — the player experiences the developmental stage of Hong Kong's 20th-century waste treatment and recycling systems, characterized by material scarcity and limited environmental awareness;

- Present — the player faces challenges of waste increase, recycling pressure, and public participation within modern urban contexts;
- Future — serves as the system's feedback dimension, dynamically visualizing the long-term environmental and social consequences of both players' decisions.

The experimental area is divided into a "past dimension activity zone" and a "present dimension activity zone." The layout incorporates multiple operable physical objects; through sensors and embedded hardware components, the system enables real-time mapping between the physical and virtual environments. Each object has a corresponding virtual model and functionality within the VR environment.

Within the system design, an original Co-evolution Chain mechanism is introduced to govern dual player task dependencies and cross-temporal collaboration. The mechanism operates through a cyclic logic of request, response, feedback, and correction. When the present-dimension player encounters a contextual obstacle that cannot be resolved independently, the system automatically generates a help signal and forwards it to the past-dimension player. The past-dimension player then performs task operations in the historical scene related to resource repair or recycling. Once the task is correctly completed, the system immediately applies a temporal correction and regenerates the environmental state, thereby removing the obstacle faced by the present-dimension player (see Figure 2). Completion of this cycle advances task progression and establishes cross-temporal interdependence and behavioural symbiosis at the level of causal structure. Through this mechanism, the interactions of the two players at the semantic, task, and data layers are integrated into a dynamic co-evolutionary entity.

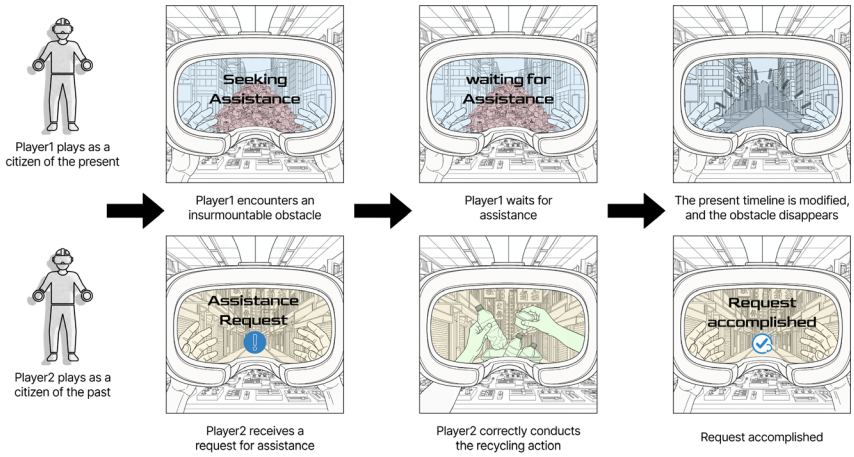


Figure 2. Co-evolution Chain workflow

During task execution, players must use their respective virtual and physical tools to achieve the task objectives. Because the two players differ in temporal context, equipment, and task pathway, they must rely on verbal communication and information description to jointly resolve obstacles that span temporal dimensions. Successful repair of basic resource-cycling facilities by the player situated in the past

increases the waste-sorting efficiency of the player situated in the present. Conversely, behavioural choices in the modern context, such as policy adjustments or waste-sorting habits, alter the availability of materials in the historical timeline.

This mechanism of cross-temporal communication not only reinforces the cooperative logic of co-creation but also enables players, through interaction, to understand the deep causal relationships between individual decisions and environmental systems.

3.2. PHYGITAL GAMEPLAY AND CONTENT DESIGN

In this system, the phygital layer provides the technical foundation for the integration of virtual and physical environments and for co-creation. Relying on embedded environmental sensors, motion capture, and spatial positioning modules, the system constructs a real-time updatable digital-twin interaction framework. This framework dynamically maps actions, object states, and data parameters from the physical world into the virtual city space, thereby forming a closed loop of action and feedback. In this way, the system achieves bidirectional perception and real-time coordination between physical and virtual scenes, enabling participants to experience the corresponding virtual changes immediately when operating physical devices, and thus to complete a continuous progression from perception to cognition.

Using button triggers and magnetic sensing, the system transmits signals from the physical end in real time to the digital-twin model of the virtual city. When the past-context player repairs recycling equipment or adjusts energy allocation parameters through physical operations, this physical event triggers a visual evolution in the virtual city. Resource pathways are reactivated, waste treatment systems are restored, and the present-context player can immediately observe ecological improvements within the virtual environment.

In this process, tactile feedback from physical devices, visual rendering in the virtual space, and streams of sensor data jointly constitute the core expression of the phygital layer. The virtual narrative is driven by physical kinetic input, mediated by data circuits, and oriented toward social reflection. Ultimately, through the intertwining of bodily operation and visual experience, players develop multisensory understanding of urban environmental systems and cross-temporal collaboration, thereby achieving digital-twin-driven social education and co-creative learning.

3.3. MULTI-ENDING NARRATIVE SYSTEM

The system constructs a multi-ending narrative through a dynamic logical chain of action, feedback generation, and feedback guidance, thereby supporting the continuous impact of player behavior on the urban ecosystem. It is important to emphasize that player choices are not constrained to a single direction. Users may engage in positive recycling and sorting behaviors, or they may make decisions that negatively affect the environment based on short-term interests. This open behavioral model is designed to examine value orientations and psychological motivations within a virtual recycling context. By simulating ecological degradation resulting from different behavioral outcomes, the system presents the environmental consequences of individual decisions in an intuitive manner, thereby strengthening reflective awareness and a sense of

environmental responsibility at the experiential level.

With respect to task structure, the system comprises two categories: main tasks and side tasks. A main task constitutes mandatory content, and its outcome directly determines the trajectory of environmental evolution in the virtual city and the game ending. A side task is optional. Completing side tasks can reduce the difficulty of the main task to some extent or unlock hidden options and special narrative triggers. Conversely, if side tasks are ignored or left unfinished, the system may present unexpected negative effects in subsequent stages, which reinforces understanding of the long-term consequences of environmental behaviour. This hierarchical task system makes the game structure more strategic and exploratory, and it enhances player agency and engagement in practices of waste recycling and resource circulation.

Players' final outcomes are evaluated through three interrelated dimensions — Environmental, Social, and Emotional — which together reflect the overall quality of intergenerational co-creation (The multi-ending operating logic is shown in Figure 3). Based on quantitative algorithms and dynamic visualization, the system generates a comprehensive feedback report encompassing the following aspects:

- Environmental Dimension: measures players' effectiveness in waste collection, classification accuracy, and resource recycling efficiency, representing the ecological sustainability of the virtual city;
- Social Dimension: evaluates the depth of collaboration, communication frequency, and the coherence of joint policy decisions across timeframes, reflecting the strength of intergenerational cooperation;
- Emotional Dimension: analyzes affective engagement, empathy, and motivational consistency throughout gameplay, illustrating players' attitudinal orientation toward collective environmental responsibility.

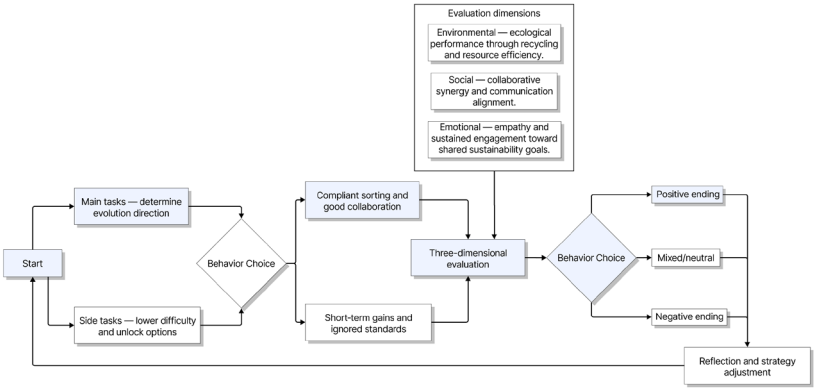


Figure 3. Multi-ending decision and evaluation workflow

The future timeline functions as a public feedback system that is open to both players. When they demonstrate positive collaboration and positive behavior, the virtual city exhibits positive developments such as ecological recovery, improved energy circulation, and increased citizen well-being, which strengthens the sense of achievement and the willingness for sustained participation. Conversely, if the players

lack coordination, overconsume resources, or disregard sorting standards, the virtual city presents negative outcomes including pollution spread, accumulation of waste, and a decline in social trust. These outcomes prompt reflection and adjustment of behaviour.

Through this mechanism, the system enhances playability and situational tension at the narrative level, and it also strengthens causal experience and moral feedback at the level of educational psychology. Through repeated experience, players gradually complete a learning loop from knowledge understanding to emotional resonance and then to behavioural transformation, thereby achieving a three-stage progression from cognitive learning to behavioural practice and finally to social responsibility.

4. Discussion

This study presents a prototype system designed for dual player co-creation and historical reconstruction, grounded in the Co-evolution Chain mechanism and a phygital gameplay framework. The system constructs an experimental narrative field based on three interconnected timelines—past, present, and future—within which players collaboratively explore, negotiate, and make decisions. Modules for narrative and task management, collaboration and state persistence, and dynamic environmental visualization jointly support decision tracking, feedback generation, and path retrospection, enabling the causal relationships during the collaborative process to be logged, tracked and explained. The open-ending pipeline enables divergent and convergent narrative outcomes, providing traceable causal chains of negotiation, concession, and conflict that can be analysed for pedagogical reflection and methodological validation.

Although formal user experiments and behavioural data collection have not yet been conducted, the design framework demonstrates clear research potential across cognitive, affective, and behavioural dimensions. Cognitively, the tripartite timeline and immersive structure help transform fragmented knowledge into a systemic understanding of urban waste governance. Affectively, dual player co-creation fosters empathy, shared experience, and emotional resonance, enhancing social memory and environmental identity. Behaviourally, the combination of task-driven interaction and immediate visual feedback may strengthen intrinsic motivation for sustainable action and lay the foundation for transferring environmental decisions from virtual spaces to real-world practices.

To validate these assumptions, future research will adopt a phased assessment strategy. The research will first carry out small-scale trial sessions under controlled conditions. Heuristic evaluation, observational studies and semi-structured interviews will evaluate operational fluency, immersion, and task comprehension, with outcomes informing iterative optimization of task logic and interface design. On this basis, the subsequent phases will use mixed-methods approaches—integrating questionnaires, interviews, and behavioural log analysis—will systematically trace user experience, emotional engagement, and motivational trajectories. Statistical and pattern-recognition techniques will help identify common and atypical decision patterns, shedding light on the cognitive mechanisms and social dynamics that influence collaborative behaviour.

When considered against current practices in environmental education and

XR-based interaction, the proposed system addresses three primary methodological gaps. First, traditional educational games rely heavily on linear narratives or fixed endings, which obscure negotiation and conflict processes and make learning outcomes difficult to trace. Second, most existing systems operate only within virtual domains, providing limited evidence for behavioural transfer to real contexts and lacking mechanisms to connect physical acts with virtual states. Third, research methods often overlook social-cognitive analysis, focusing on individual immersion rather than collective strategy or inter-agent causality. By introducing the dual player phygital mechanism and open-ended narrative structure, this study provides a collaborative instrumental framework for environmental education that can be embedded in school lessons, community workshops, or municipal public engagement programs. The system can be deployed in various forms in different contexts. In addition to the immersive XR experience, it can also be deployed for guided assistance, community shared interactive screen, community co-creation workshops, and other forms, so as to support participants of different age and educational backgrounds. As a result, this study has the research potential to transform from a conceptual prototype to a scalable practice tool, creating a replicable framework for sustainable game-based learning.

Nevertheless, the prototype is still at an early stage, and no large-scale user data or longitudinal evidence have been obtained. Current assessments are primarily theoretical, awaiting empirical validation. The focus on urban waste governance confines contextual generalizability, and system scalability may be constrained by dependencies on specific hardware and spatial configurations. Furthermore, the long-term effects of emotional resonance and behavioural transfer require longitudinal tracking to verify. Future studies will therefore expand testing scope, diversify participant groups, and adopt cross-scenario evaluations to improve methodological rigor and ensure the broader applicability of the phygital educational framework.

5. Conclusion

This study proposes and implements an XR co-creation framework cantered on open endings and dual player phygital gameplay, aiming to transform environmental education from passive knowledge transmission into a collaborative, traceable, and interpretable participatory experience. The system is supported by a tripartite timeline narrative, task and narrative management, dual player collaboration with linkage between physical and virtual components, and visualized ending retrospection. These pillars connect the collaborative process with multiple endings and construct an integrated intervention pathway across cognition, emotion, and behaviour. The framework provides an operational experimental basis for examining how social interaction influences environmental decision making and holds potential for cross domain application. Future work will validate learning outcomes and behavioural persistence through empirical data, improve the accessibility and scalability of interfaces and devices, and advance deployment in schools, cultural institutions, and community programs.

Acknowledgement

The work described in this paper was substantially supported by a grant from the Hong Kong Polytechnic University (Project No. P0046262). The project was supported by

New Life Plastics Ltd, Hong Kong, for recycled plastic raw materials. ChatGPT

Attribution Statement

(OpenAI, 2025) was used to (1) assist in translation and (2) polish the language for clarity and fluency (All research design, analytical decisions, and interpretations were conducted by the author.)

References

Banfi, F., Bolognesi, C. M., Bonini, J. A., & Mandelli, A. (2021). The virtual historical reconstruction of the Cerchia dei Navigli of Milan: From historical archives, 3D survey and HBIM to the virtual visual storytelling. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021, 39–46. https://doi.org/10.5194/isprs-archives-XLVI-M-1-2021-39-2021

Basak, A., Kumar, S., Upadhyay, P., & Banerjee, S. (2025). Balance: A gamified toolkit for sustainability-oriented design thinking. The International Journal of Art & Design Education. https://doi.org/10.1111/jade.12601

El-Jarn, H., & Southern, G. (2020). Can co-creation in extended reality technologies facilitate the design process? Journal of Work-Applied Management, 12(2), 191–205. https://doi.org/10.1108/JWAM-04-2020-0022

Environmental Protection Department, HKSAR. (2024, October). Monitoring of solid waste in Hong Kong: Waste statistics for 2023 [Statistical report]. https://www.wastereduction.gov.hk/sites/default/files/resources_centre/waste_statistics/msw2023_sc.pdf

Judmaier, P., Ladner, S., Taurer, F., Mynha, C., Mauthner, B., Salhofer, S., Beigl, P., & Huber-Humer, M. (2025). Motivational apps and gamification in waste management. Österreichische Wasser- und Abfallwirtschaft, 77(1–2), 32–40. https://doi.org/10.1007/s00506-024-01094-9

Makransky, G., & Petersen, G. B. (2021). The Cognitive-Affective Model of Immersive Learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. Educational Psychology Review, 33(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

OpenAI. (2025). ChatGPT Model: GPT-4 [Large Language Model]. https://chatgpt.com/

Orr, N., Matthews, B., See, Z. S., Burrell, A., Day, J., & Seengal, D. (2021). Transdisciplinarity in extended reality (XR) research design: Technological transformation and social good (co-creation session at XR + Creativity Symposium, University of Newcastle, 2020). Virtual Creativity, 11(1), 163–179. https://doi.org/10.1386/vcr_00048_1

Roaro, E. (2021). Cinema memories in 3D modelling and virtual reality storytelling: The Odeon Cinema in Udine. Alphaville, 21, 113–130. https://doi.org/10.33178/alpha.21.07

Song, Z., Wyke, T. W., Sun, X., & Evans, L. (2025). A phenomenology study of transmedia museum storytelling and gamification through XR technologies. Digital Applications in Archaeology and Cultural Heritage, 38, Article e00444. https://doi.org/10.1016/j.daach.2025.e00444

Tromp, J., Schofield, D., Raesian Paryani, P., Povade, M., Eaglesham, C., Torres, J. C., Johnson, T., Jürivete, T., Lauer, N., Reyes-Lecuna, A., Gonzalez-Rodriguez, M., & De-Molina-Tanco, L. (2025). Designing and evaluating XR cultural heritage applications with human-computer interaction methods: Insights from ten international case studies. Applied Sciences, 15(14), 7973. https://doi.org/10.3390/app15147973

Neuro-Adaptive Lighting Architecture for Emotional State Modulation

Real-Time Facial Expression-Driven Closed-Loop System with EEG Validation

Tiansu Wang[1], Tianyi Bo[2], Zheng Hou[3], Chen Hong[4], Jing Huang[5] and Donghui Dai[6]
[1,3,5,6]*Harbin Institute of Technology*
[2]*Rice University School of Architecture and Building Technology*
[4]*Xuancheng Vocational and Technical College*
[1]*25b356005@stu.hit.edu.cn, 0009-0006-8234-289X*
[2]*harperbo1219@gmail.com, 0009-0007-4969-7872*
[3]*2370771908@qq.com, 0009-0007-3944-3289*
[4]*hc2011301539@outlook.com, 0009-0006-2052-783X*
[5]*2966288@qq.com, 0009-0003-7932-4859*
[6]*daidonghui@hit.edu.cn, 0000-0002-9605-1130*

Abstract. Addressing the critical translational gap between physiological light effects and emotional intervention, we developed a novel neuro-interactive lighting system employing a "recognition–response–verification" framework. Our system utilizes real-time facial expression recognition to dynamically adjust LED lighting parameters, with efficacy quantified through EEG α/β power ratio in a controlled within-subjects experiment (N=30). Results demonstrate that specific lighting parameters exert targeted, statistically significant effects on emotional neurophysiology: relaxation-oriented lighting significantly increased the EEG α/β power ratio ($p<0.05$) with corresponding facial features indicating reduced arousal, while activating lighting produced inverse neural patterns. The closed-loop system successfully translated real-time facial cues into adaptive light feedback, demonstrating stable emotion regulation that effectively guided participants from high-stress to calmer states, establishing a new paradigm for neuro-adaptive environmental intervention.

Keywords. Neuro-Interactive Lighting, Affective Computing, Electroencephalography (EEG), Facial Expression Recognition, Emotional Regulation, Restorative Environmental Design

1. Introduction

The escalating prevalence of socio-psychological stress-related issues underscores a pressing need for restorative environmental design (Ulrich, 1984; Kaplan, 1995).

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 307-316.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Recognized for its potential to facilitate psychological recovery through spatial intervention, restorative design is increasingly becoming a convergent research focus spanning architecture, environmental psychology, and neuroscience (Sanjuan et al., 2017; Lyu et al., 2022). As one of the most ubiquitous and controllable elements of environmental stimuli, light profoundly influences not only the visual quality of a space but also physiological homeostasis and emotional balance (Korshunov et al., 2017; Terajima et al., 2018). The emerging concept of Human-Centric Lighting emphasizes that lighting parameters—such as correlated colour temperature (CCT), illuminance, and dynamic variation—can significantly modulate autonomic nervous system activity and central nervous system function, thereby regulating affective perception (Zhu et al., 2023; Ru et al., 2019).

However, current lighting-emotion research predominantly relies on subjective self-report measures (VAS, SAM) susceptible to individual differences and bias (Bradley & Lang, 1994; Russell, 1989; Wang et al., 2018). While objective physiological measures like Electroencephalography (EEG) have been introduced, most findings remain observational, establishing unidirectional correlations rather than causal interventions (Chellappa et al., 2011; Borghini et al., 2014). A critical translational gap persists between measuring physiological changes and actively intervening in physiological states, confining lighting applications in emotional remediation largely to experimental stages and hindering the development of personalized, adaptive lighting therapy devices (de Kort & IJsselsteijn, 2008; Boyce, 2014).

Addressing this research gap, this study proposes an interactive lighting system integrating facial expression recognition and EEG validation to establish a neuro-interactive framework based on "recognition–response–verification" logic. The system utilizes facial capture for real-time emotion recognition and LED matrix for dynamic lighting parameter adjustment, concurrently validating emotional intervention efficacy through EEG α/β power ratio analysis. This research aims to: (1) elucidate lighting-EEG mapping relationships and neural regulatory mechanisms; (2) develop a facial recognition-based interactive lighting system with real-time emotion feedback; (3) validate system effectiveness and reproducibility through pre/post-exposure EEG experiments. By integrating multimodal signal fusion and cross-scale validation, this study establishes a novel technical pathway for intelligent luminous environments with quantifiable emotional states, controllable feedback, and interpretable mechanisms, while laying experimental foundations for the neuroscientific advancement of restorative architectural design

2. Literature Review

The neurobiological mechanisms through which light regulates emotion via non-visual pathways have been substantiated by molecular neuroscience and cognitive neuroimaging (Huang et al., 2019; Sabbah et al., 2022). Melanopsin, expressed by intrinsically photosensitive retinal ganglion cells (ipRGCs), responds to short-wavelength light (460–480 nm) (Lucas et al., 2020). These ipRGCs project via the retinohypothalamic tract to the suprachiasmatic nucleus (SCN), which, through the hypothalamic-pituitary-adrenal (HPA) axis, regulates cortisol secretion and targets key emotion-related brain regions such as the prefrontal cortex (PFC) and amygdala

(AMY), forming a definitive "light-neural-emotion" regulatory pathway (St Hilaire et al., 2022; Brown et al., 2021). The synergistic action of ipRGCs with classical cone/rod photoreceptors enables lighting parameters like CCT and illuminance to bidirectionally regulate emotional arousal and valence by modulating GABA and serotonin release, providing a clear neurochemical basis for light-based emotional intervention (Sonoda et al., 2020).

Current research exhibits three core limitations: Firstly, subjective assessment tools are vulnerable to social desirability bias and contextual cognitive load, impeding the capture of subconscious emotional fluctuations (Alarco & Fonseca, 2019). Secondly, while EEG technology has confirmed the α/β power ratio (8–13 Hz / 14–30 Hz) as a potential emotional biomarker, studies predominantly focus on linear associations between single lighting parameters and EEG metrics, lacking systematic analysis of neural responses under multi-parameter coupling. Furthermore, a consensus on the precise mapping between the α/β ratio and emotional valence remains elusive (Gable et al., 2021). Thirdly, contemporary Human-Centric Lighting practices often rely on static parameter pre-sets, lacking the closed-loop logic integrating real-time emotion recognition, dynamic light adjustment, and intervention efficacy verification (Zhang et al., 2023).Multimodal fusion in emotion recognition technology offers a pathway to overcome these limitations. Techniques based on Facial Action Unit (AU) recognition can accurately capture micro-expression features, achieving emotion classification accuracy of 89%–94%, thus meeting real-time interaction requirements . However, existing research has not deeply integrated this technology with neurophysiological validation, resulting in adaptive lighting systems whose intervention logic lacks direct neuroscientific underpinning.

Addressing these research gaps, this study constructs a tripartite interactive framework of "Facial Expression Recognition – EEG Validation – Dynamic Light Feedback": utilizing facial micro-expressions as real-time emotional input, enabling precise control of lighting parameters via an LED matrix, and employing the α/β power ratio and PFC-AMY functional connectivity strength as neuro-validation indicators to decipher the emotional intervention mechanisms of lighting patterns (Zhang et al., 2020). This framework transforms the luminous environment into an active emotional regulation medium. Through dynamic "human-light-brain" coupling, it achieves quantifiable, controllable, and interpretable emotion regulation, advancing Human-Centric Lighting from "parameter optimization" towards "neural adaptation" (Liu et al., 2023), thereby laying a theoretical and experimental foundation for the precision design of restorative spaces.

3. Methodology

3.1 EXPERIMENTAL DESIGN AND DATA ACQUISITION

This study established an integrated "Lighting Design – Multimodal Data Acquisition – Closed-Loop Verification" framework to systematically investigate luminous environment effects on emotion. Seven lighting modes spanning rhythm, CCT, luminance, and spatial dimensions were designed and preliminarily screened to identify four core modes eliciting significant neural responses. Synchronized data acquisition employed a 32-channel EEG system (ErgoLAB) and MediaPipe-OpenCV

facial capture module, analyzing α/β power ratios (8–12 Hz/13–30 Hz) and facial parameters including size and centroid coordinates. Thirty healthy participants underwent a three-stage protocol (baseline induction, resting-state recording, interactive lighting experience, post-intervention reassessment) in controlled environmental conditions. The system achieved real-time LED mode adaptation via serial communication based on facial emotion recognition, while paired-sample t-tests and facial-EEG coupling validation constructed a quantitative assessment framework for emotional regulation efficacy.

3.2 LIGHTING DESIGN

Based on comprehensive literature review and preliminary experimental validation, we systematically developed and evaluated seven distinct lighting modes (Flowing Light, Digital Rain, Concentric Diffusion, Random Sparkle, Block Flicker, Heartbeat Pulse, and Flame Simulation) engineered through an Arduino-controlled LED matrix. These modes encompass diverse parameters including rhythmic frequencies, luminance variations, CCT distributions, and spatial diffusion patterns. Initial EEG mechanism analysis comparing α/β ratio modulation and subjective emotional responses across all conditions identified four modes exhibiting the most significant neural responses—Flowing Light, Random Sparkle, Concentric Diffusion, and Block Flicker—which were subsequently selected for formal experimental validation. The remaining three modes demonstrate system extensibility, highlighting the potential for diverse visual interactions in lighting-based neuro-adaptive interventions..

The experimental protocol incorporated four distinct lighting paradigms: ① Flowing Light employing low-frequency rhythmic stimulation with cool-tone blue-purple gradient and continuous sliding trajectory at 24fps; ② Random Sparkle generating high-uncertainty stimuli through multi-point random activation and full-spectral colour variations; ③ Concentric Diffusion utilizing blue-green gradient expansion with medium-low frequency intensity and alternating focal points; and ④ Block Flicker creating high-contrast rhythms via 1.6Hz alternating flickers across red, green, blue, and yellow quadrants. The four lighting modes are shown in Figure 1.

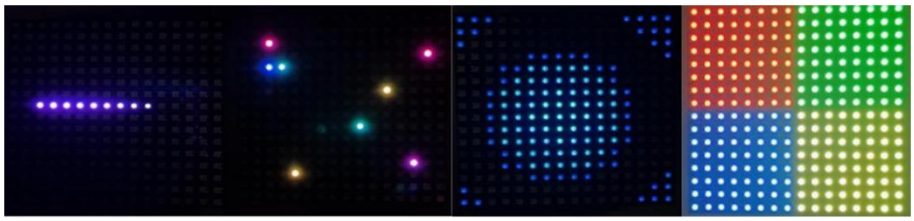


Figure 1. Four lighting modes in experiment

3.3 LIGHTING AND EEG MECHANISM STUDY

EEG data were acquired using the ErgoLAB platform (v3.17.20) with a 32-channel electrode cap positioned across frontal (Fp1, FpZ, F3, Fz, FC5), parietal (CP5, P7, Pz, POz), temporal (T7), and occipital (O1) regions. Electrodes were prepared with saline solution and interfaced with the amplifier. Impedance was adjusted to $\leqslant 0.2\ \mathrm{M}\Omega$. System integrity was validated using standard motion tests (blinks, mouth openings,

and chewing), which elicited expected amplitude changes.

Participants were first exposed to urban traffic and crowd stimuli to induce standardized psychological stress, establishing a baseline of high stress and arousal. Subsequently, in an acoustically and visually controlled environment, participants viewed 30-second lighting design stimuli while maintaining natural gaze and seated posture. To minimize order effects, stimuli were presented in randomized sequence with 1-minute inter-stimulus intervals for EEG baseline recovery. EEG acquisition focused on alpha (8–12 Hz) and beta (13–30 Hz) wave dynamics, with the α/β power ratio serving as the primary metric for quantifying emotional regulatory effects and elucidating the mechanistic basis of stress alleviation

3.4 FACIAL CAPTURE AND INTERACTIVE SETUP

This study utilized the MediaPipe Face Detection module (mp.solutions.face_detection.FaceDetection,model_selection=0,min_detection_confidence=0.7) in conjunction with OpenCV (cv2.VideoCapture(0)) to capture the facial region in real-time and extract bounding box parameters. Face size and centroid coordinates (face_center_x, face_center_y) were calculated as indirect proxies for emotional state.

A closed-loop communication framework was established via serial communication between PC and Arduino (9600 baud), with facial feature data encoded in "state,size,centerX,centerY\n" format to trigger real-time LED matrix responses. While facial features directly controlled lighting parameters, they were subsequently correlated with EEG measures and subjective emotion reports to validate emotional state monitoring. Existing literature confirms facial features as reliable emotion indicators (>90% static recognition accuracy). During emotional transitions, significant coupling emerged between facial parameters and EEG biomarkers: under relaxation-oriented lighting, facial area expansion and centroid coordinate normalization indicated muscle relaxation, while high-arousal conditions produced facial contraction and upward centroid shift, demonstrating quantitative correspondence with EEG-based arousal states. System stability was verified through synchronized timestamp recording of facial features and lighting mode switches, with emotion regulation efficacy validated via pre-/post-intervention α/β ratio analysis from independent EEG exposure experiments.

3.5 PRE-/POST-EXPOSURE VALIDATION EXPERIMENT DESIGN

This experiment validated the emotional regulation efficacy of the interactive lighting system by comparing pre- and post-exposure changes in the EEG α/β power ratio. Thirty university students with normal or corrected-to-normal vision and no history of neurological disorders were recruited. The experiment was conducted in a closed, sound-attenuated room to exclude external interference.

The experiment consisted of three sequential phases: (1) Initial State Acquisition: Resting-state EEG data were recorded while participants were in their natural mood, establishing a veridical emotional baseline. (2) Interactive Lighting Exposure: as Figure 2 shows that participants were positioned facing the LED matrix. The system, via real-time facial region detection and bounding box parameter extraction, identified

the participant's emotional state and matched it to one of four predefined regulation modes. The state-to-mode mapping was as follows: Frustrated → warm yellow, low-luminance, slowly diffusing light; Slightly Frustrated → neutral cool-white light with rhythmic breathing-like variation; High Arousal → cool blue-green, softly dynamic light; and a Subtle Mode of neutral, slightly warm, low-luminance static light. Each assigned lighting mode was presented for a duration of 30 seconds, with successive presentations interspersed by 1-minute rest intervals to mitigate potential order effects. (3) Post-Exposure Measurement: Resting-state EEG data were acquired immediately following the interactive session to assess the residual or altered neural state

EEG data were acquired and exported via the ErgoLAB platform, focusing on alpha (8 – 12 Hz) and beta (13 – 30 Hz) power, with the α/β ratio as the core neural indicator. The experimental hypothesis posited that successful system intervention would be indicated by enhanced alpha power, attenuated beta power, or an increased α/β ratio post-exposure, signifying a positive emotional regulation effect. Data underwent bandpass filtering and artifact rejection before analysis using paired-sample t-tests to determine the significance of pre-/post-exposure differences, providing quantitative support for system efficacy



Figure 2. Experiment process

4. Result and Discussion

4.1 RESULT

Through multimodal analysis integrating EEG and facial expression recognition, this study systematically confirms the targeted regulatory effects of lighting design on emotional states, clarifying the associated physiological mechanisms and quantitative characterization patterns. The core findings are as follows:

Firstly, multimodal experiments substantiate that specific lighting designs can exert targeted emotional regulation. A key finding is that lighting intended to promote relaxation—characterized by low CCT and gentle dynamics—significantly increased the EEG α/β power ratio. This neural change was coupled with an increase in recognized face size and specific shifts in centroid coordinates, effectively guiding the transition from an anxious towards a calmer state. Conversely, activating lighting exhibited the opposite trend, shifting participants from calmness towards excitement. The controlled experimental system evaluated these regulatory effects, demonstrating a specific mapping relationship between lighting modes and emotional states. Specifically, among the four regulation modes, Group A lighting induced a marked relaxation effect on EEG indicators: mean alpha power increased significantly from a baseline of 14.38 μV^2 to 23.12 μV^2 post-intervention, with a concurrently elevated α/β ratio. Simultaneously, facial expression analysis showed shifts in face region

parameters associated with emotional arousal towards decreased activity intensity, further supporting the participant's transition from a high-arousal to a more subdued state. In contrast, Group D lighting elicited a physiological response pattern consistent with emotional activation. Under this condition, mean alpha power decreased from a baseline of 14.72 μV^2 to 12.20 μV^2, and the α/β ratio remained at a relatively low value of 0.52. Facial analysis concurrently indicated that participants maintained a higher level of emotional arousal, with some individuals even showing enhanced stress-related expression features,all details are shown in Table 1.

Table 1. The variations of α and β in power as well as their ratios among the four lighting modes.

Group	α1mean	α2mean	β1mean	β2mean	α1m/β1m	α2m/β2m
A	14.38	23.12	23.6	15.41	0.61	1.5
B	13.26	16.07	23.54	21.57	0.56	0.75
C	15.01	19.79	22.83	18.26	0.66	1.08
D	14.72	12.2	22.08	23.52	0.67	0.52

Regarding the neural mechanisms of light regulation, pre-/post-exposure data demonstrate that the experimental system can effectively modulate participant emotions by matching lighting modes: The system targeted four emotional states with corresponding lighting interventions: frustrated states received warm yellow, low-luminance, slowly diffusing light; slightly frustrated states utilized neutral cool-white light with rhythmic breathing-like variation; high-arousal states employed cool blue-green, softly dynamic light; and a subtle mode applied neutral, slightly warm, low-luminance static light—all contributed to emotional balancing, evidenced by favourable shifts in the α/β ratio. Paired-sample t-tests confirmed statistically significant differences in the α/β ratio between post-exposure and initial states, validating the system's effectiveness and stability in emotion regulation. A representative case study illustrates this dynamic process. Participant S21 exhibited typical stress markers at baseline: alpha power was low, beta power was high, resulting in a low α/β ratio, consistent with high arousal. Following a 5-minute exposure to Group A lighting, their EEG pattern shifted markedly: alpha power rose, beta power decreased, and the α/β ratio increased substantially, as Figure 3 shows , clearly tracing the neural transition from a high-pressure to a calmer state and providing physiological evidence for light-mediated emotion regulation.

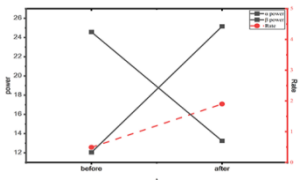


Figure 3.Changes in EEG parameters of S21 before and after the experiment

4.2. DISCUSSION

This study contributes to environmental emotion regulation through multimodal integration and closed-loop interaction frameworks. By transcending traditional

subjective reporting limitations, the research establishes objective multi-evidence pathways for emotional state assessment through quantitative coupling of lighting parameters, EEG α/β power ratios, and facial features. Four validated lighting modes capable of eliciting specific neural responses clarify emotional advantages of low-CCT gentle rhythmic illumination and activating effects of high-frequency high-brightness lighting, partially bridging knowledge gaps in neural emotion regulation mechanisms. The developed interactive lighting system achieves critical technological advancement from static parameter presets to real-time emotional adaptation, informing restorative architectural design. While affirming these findings, we also acknowledge several limitations of the present study. The participant pool was currently restricted to 30 healthy young volunteers, precluding stratified analysis based on variables like age, gender, and emotional baseline, which limits the generalizability of the conclusions. Furthermore, the experiments were conducted in a controlled laboratory setting, failing to account for potential synergistic regulatory effects arising from the interaction of light with other environmental factors present in real-world scenarios. From a technical perspective, the current EEG analysis focused predominantly on the α/β power ratio, lacking in-depth analysis of functional connectivity within key brain networks. Simultaneously, facial expression analysis relied on basic parameters, not yet incorporating the fine-grained decoding of Facial Action Units, potentially missing subtle nuances in emotional expression.

Despite these limitations, this study offers significant theoretical and practical contributions. Theoretically, it advances understanding of the intrinsic "light-neural-emotion" pathway and establishes a novel paradigm for environmental psychology-neuroscience integration. Practically, the validated lighting-emotion response mapping provides quantitative design guidelines for architectural, interior, and healthcare luminous environments, enabling scalable Human-Centric Lighting applications in emotional remediation. Future research should prioritize multi-center longitudinal studies with expanded samples, brain imaging integration for neural mechanism exploration, algorithm refinement for higher-precision emotion modeling, and synergistic environmental element analysis to enhance intervention efficacy and advance restorative design solutions.

5. Conclusion

This study proposes a closed-loop neuro-interactive lighting paradigm that achieves precise emotional regulation through dynamic "human-light-brain" coupling. Our threefold contribution encompasses: (1) elucidating specific mappings between lighting parameters and EEG α/β power ratio neural correlates; (2) engineering a real-time facial expression-driven adaptive lighting system; and (3) demonstrating causal efficacy through rigorous pre-/post-intervention EEG validation. The integration of facial recognition and neurophysiological monitoring creates a robust cross-verified framework that bridges real-time intervention with objective efficacy assessment, fundamentally advancing beyond passive stimulus-response paradigms toward active neuro-environmental regulation. This work provides the experimental and theoretical foundation for precision light therapy and adaptive architecture, opening new avenues for non-pharmacological mental health interventions and establishing light as an active therapeutic modality in next-generation restorative environments.

References

Alarco, S. M., & Fonseca, M. J. (2019). Emotions Recognition Using EEG Signals: A Survey. *IEEE Access*, 7, 116779-116805. https://doi.org/10.1109/ACCESS.2019.2936142

Borghini, G., Piccardi, L., Schiaratura, L., D'Innocenzo, V., Matarazzo, L., & Veltrini, E. (2014). Assessing engagement through psychophysiological and self-report measures. *Proceedings of the 2014 International Conference on Multimodal Interaction*, 2014, 51-58. https://doi.org/10.1145/2663204.2663248

Boyce, P. R. (2014). Human factors in lighting. *CRC Press*. https://doi.org/10.1201/b17130

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59. https://doi.org/10.1016/0005-7916(94)90063-9

Brown, T. M., Thapan, K., Arendt, J., Revell, V. L., & Skene, D. J. (2021). S‐cone Contribution to the Acute Melatonin Suppression Response in Humans. *Journal of Physiology*, 599(6), 1427-1440. https://doi.org/10.1113/JP279647

Chellappa, S. L., Steiner, R., Oelhafen, P., Schulz, E., & Cajochen, C. (2011). Non-visual effects of light on melatonin, alertness and cognitive performance: Can blue-enriched light keep us alert? *PLoS ONE*, 6(1), e18178. https://doi.org/10.1371/journal.pone.0018178

de Kort, Y. A., & IJsselsteijn, W. A. (2008). Illuminating the experience: Lighting in a mediated world. Philosophical Transactions of the Royal Society B: *Biological Sciences*, 363(1499), 2739-2753. https://doi.org/10.1098/rstb.2008.0005

Gable, P. A., Paul, K., & Pourtois, G. (2021). Utilizing Electroencephalography (EEG) to Investigate Positive Affect. *Neuroscience & Biobehavioral Reviews*, 126, 153-165. https://doi.org/10.1016/j.neubiorev.2021.01.009

Huang, L., Xi, Y., Peng, Y., Yang, Y., Huang, X., Fu, Y., ... & So, K. F. (2019). A Visual Circuit Related to Habenula Underlies the Antidepressive Effects of Light Therapy. *Science*, 363(6434), 1417-1421. https://doi.org/10.1126/science.aax0229

Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology*, 15(3), 169-182. https://doi.org/10.1016/0272-4944(95)90001-2

Korshunov, K. S., Blakemore, L. J., & Trombley, P. Q. (2017). Dopamine: A Modulator of Circadian Rhythms in the Central Nervous System. *Frontiers in Neuroscience*, 11, 36. https://doi.org/10.3389/fnins.2017.00036

Liu, D., Dai, W., Zhang, H., Jin, X., Cao, J., & Kong, W. (2023). Brain-Machine Coupled Learning Method for Facial Emotion Recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 34(5), 2789-2802. https://doi.org/10.1109/TNNLS.2023.3239657

Lucas, R. J., Allen, A. E., Milosavljevic, N., Storchi, R., & Woelders, T. (2020). Can We See with Melanopsin? *Progress in Neurobiology*, 193, 101814. https://doi.org/10.1016/j.pneurobio.2020.101814

Lyu, K., de Dear, R., Brambilla, A., & Globa, A. (2022). Restorative Benefits of Semi-Outdoor Environments at the Workplace: Does the Thermal Realm Matter? *Building and Environment*, 213, 108542. https://doi.org/10.1016/j.buildenv.2022.108542

Piho, L., & Tjahjadi, T. (2020). A Mutual Information Based Adaptive Windowing of Informative EEG for Emotion Recognition. *IEEE Access*, 8, 185230-185252.

Roy, V., Shukla, P. K., Gupta, A. K., Goel, V., Shukla, P. K., & Shukla, S. (2020). Taxonomy on EEG Artifacts Removal Methods, Issues, and Healthcare Applications. *Biomedical Signal Processing and Control*, 63, 102149.

Ru, T., de Kort, Y. A. W., Smolders, K. C. H. J., & Chen, Q. (2019). Non-image forming effects of illuminance and correlated color temperature of office light on alertness, mood, and performance across cognitive domains. *Lighting Research & Technology*, 51(8), 1235-1254. https://doi.org/10.1177/1477153518809372

Russell, J. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, 57(3), 493-502. https://doi.org/10.1037/0022-3514.57.3.493

Sabbah, S., Worden, M. S., Laniado, D. D., Berson, D. M., & Sanes, J. N. (2022). Luxotonic Signals in Human Prefrontal Cortex As a Possible Substrate for Effects of Light on Mood and Cognition. *Cerebral Cortex*, 32(5), 2721-2732. https://doi.org/10.1093/cercor/bhab416

Sanjuan, C., Subiza-Perez, M., & Vozmediano, L. (2017). Restoration and the City: the Role of Public Urban Squares. *Frontiers in Psychology*, 8, 1469. https://doi.org/10.3389/fpsyg.2017.01469

Sonoda, T., Li, J. Y., Hayes, N. W., Chan, J. C., Okabe, Y., Belin, S., ... & Schmidt, T. M. (2020). A Noncanonical Inhibitory Circuit Dampens Behavioral Sensitivity to Light. *Science*, 367(6484), 527-531. https://doi.org/10.1126/science.aax0229

St Hilaire, M. A., Amundadottir, M. L., Rahman, S. A., Rajaratnam, S. M. W., Ruger, M., Brainard, G. C., ... & Lockley, S. W. (2022). The Spectral Sensitivity of Human Circadian Phase Resetting and Melatonin Suppression to Light Changes Dynamically with Light Duration. *Journal of Biological Rhythms*, 37(2), 149-162. https://doi.org/10.1177/07487304211050708

Terajima, H., Yoshitane, H., Yoshikawa, T., Shigeyoshi, Y., & Fukada, Y. (2018). A-to-I RNA Editing Enzyme ADAR2 Regulates Light-Induced Circadian Phase-Shift. *PLOS ONE*, 13(3), e0194278. https://doi.org/10.1371/journal.pone.0194278

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science*, 224(4647), 420-421. https://doi.org/10.1126/science.224.4647.420

Wang, Q., He, X., Li, N., Chen, J., & Li, H. (2018). The effects of lighting on human emotions: A review. *Building and Environment*, 141, 343-352. https://doi.org/10.1016/j.buildenv.2018.06.015

Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion Recognition Using Multi-Modal Data and Machine Learning Techniques: A Tutorial and Review. *IEEE Access*, 8, 185230-185252. https://doi.org/10.1109/ACCESS.2020.3025453

Zhang, X., Qiao, Y., Wang, J., & Chen, D. (2023). Lighting Environmental Assessment in Enclosed Spaces Based on Emotional Model. *Building and Environment*, 226, 109432. https://doi.org/10.1016/j.buildenv.2023.109432

Zhu, F., Lin, Y., Huang, W., Lu, T., Liu, Z., Ji, X., Kang, A., Wang, H., & Chen, Z. (2023). Multi-primary Human-Centric Lighting Based on the Optical Power Ratio and the CCT Super-Smooth Switching Algorithms. *Applied Sciences*, 13(5), 2973. https://doi.org/10.3390/app13052973

Examining Critical Factors of Nighttime Lighting Design for Cultural Heritage Sites using Digital Twins

Ting Lu Chiang[1], Hao-Yun Chi[2] and Tsung-Hsien Wang[3]
[1]*Brunel University of London*
[2]*China University of Technology*
[3]*University of Sheffield*
[1]*Ting.LuChiang@brunel.ac.uk, 0009-0006-2358-512X*
[2]*chihyn@gm.cute.edu.tw, 0000-0003-3179-7887*
[3]*tsung-hsien.wang@sheffield.ac.uk, 0000-0002-4500-8415*

Abstract. Nighttime lighting plays a crucial role in visibility, safety, and spatial orientation, especially at historic sites where lighting must balance functional needs with the preservation of historical atmosphere. Although previous research has examined how illuminance and colour temperature influence perception and visual comfort, findings remain inconsistent, and studies focusing on nighttime historic environments are limited. This study investigates the impacts of different colour temperatures and illuminance levels on pleasantness, pedestrian recognition, detail recognition, and obstacle detection in a historic-site context. A high-resolution digital twin of Hobe Fort in Taiwan was constructed to simulate controlled nighttime scenarios. Through animated guided walkthroughs and task-based questionnaires, results show that among critical lighting factors influencing perceptual clarity and visual-attention patterns, illuminance emerges as the dominant factor across all tasks. This research presents empirical evidence for enhancing nighttime lighting design in cultural heritage environments. It demonstrates the effectiveness of digital-twin-based assessment for evaluating perceptual and safety-related outcomes without requiring physical intervention on protected sites.

Keywords. Nighttime lighting, cultural heritage, digital twins

1. Introduction

Nighttime lighting in cultural heritage sites must balance visual safety, perceptual clarity, and the preservation of historical ambience. Existing guidelines and standards in Taiwan and the United Kingdom, such as CNS 15015(2024), BS 5489(2024), and EN 13201(2015), provide technical recommendations for illuminance levels and lighting characteristics. However, the influence of historical readability on perception remains insufficiently studied, particularly in heritage contexts where lighting must ensure safety and interpretive clarity. Previous studies examined the relationships between factors such as illuminance, colour temperature, and luminance with

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 317-326.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

pedestrian attention and nighttime visibility (Caminada & van Bommel, 1984). Nevertheless, very limited studies explored how these variables influence visual attention and perceptual experience in a cultural heritage setting.

Using terrestrial LiDAR scanning, the architectural geometry and material of the Hobe Fort in Tamsui, Taiwan, were reconstructed to create a high-resolution digital twin. Using these simulated experiments enabled controlled testing of nighttime lighting design scenarios without physically intervening in the protected site. With eye-tracking experiments and participants' subjective responses, this study examined lighting performance from both physiological and perceptual perspectives. As a result, the empirical evidence illuminates how lighting parameters influence visual behaviour and perceptual clarity in cultural heritage environments. It further demonstrates the value of digital-twin–based assessment as a user-centred evaluation approach for nighttime lighting design in historically sensitive settings.

2. Related Works

2.1. LIGHTING STANDARDS AND DESIGN PRINCIPLES

In Taiwan, CNS 15015 stipulates luminaires for outdoor landscape lighting, aiming to improve overall outdoor illumination quality and ensure pedestrian safety and comfort. BS 5489 in the UK emphasises minimising energy consumption and reducing environmental impacts. Across the European Union, EN 13201 specifies performance requirements for roadway lighting based on the visual needs of road users and incorporates environmental conditions into lighting design considerations.

Illuminance is a key indicator of adequate lighting. BS 5489 builds upon the EN 13201 series as a national guideline to provide road users with a safe, comfortable, and visually attractive nighttime environment. According to EN 13201-2, nighttime lighting is categorised into three series: M series: motorised traffic routes, C series: conflict areas where pedestrian–vehicle interactions occur, and P series: pedestrian and low-speed traffic areas. Within the P series, six classes (P1–P6) are defined based on usage frequency and safety needs. In Taiwan, guidelines for outdoor nighttime lighting refer to the CNS 15015 standard, Luminaires for Outdoor Landscape Lighting. This standard recommends appropriate illuminance levels for various lighting objectives or zones and specifies average illuminance requirements.

Regarding colour temperature, BS 5489 states that lighting should, where possible, create a pleasant and attractive nocturnal ambience while complying with local regulations and guidance. EN 13201 does not explicitly regulate Correlated Color Temperature (CCT). CNS 15015 recommends using shields or low-blue-light luminaires in contexts where lighting may affect residents, astronomers, or biological organisms.

2.2. VISUAL ATTENTION IN LIGHTING STUDIES

Many early studies analysed factors that influence visual attractiveness or attention levels, typically examining the effects of colour, shape, brightness, and other variables. However, the degree to which these factors attract attention has not reached a consistent conclusion (Turatto & Galfano, 2000). Subsequent research frequently recorded

participants' eye-tracking (ET) data, and these tests have been widely applied across various fields, including pedestrian safety (Jiang et al., 2021), driving safety (Li et al., 2025), and indoor lighting quality (Amundadottir et al., 2017).

Providing appropriate illuminance in a space not only enhances people's ability to recognise objects at night but also influences attention and reading performance. Research has shown that brightness positively affects outdoor pedestrians' attention during both daytime and nighttime, with a stronger correlation observed at night (Jiang et al., 2021). However, indoor studies reported that increasing brightness does not necessarily improve decision-making efficiency during guided indoor reading tasks. Colour temperature can also be a critical indicator in artificial lighting design. Both illuminance and colour temperature can significantly influence nighttime visual attention induced by environmental elements, and the visual effects produced by different CCT levels vary substantially across environmental contexts (Yang & Lyu, 2025).

Beyond illuminating historic buildings, meeting the safety requirements of outdoor nighttime lighting is also crucial. Previous research has identified tasks such as obstacle detection, person identification, and subjective ratings of spatial pleasantness as criteria for evaluating adequate lighting (Caminada & van Bommel, 1984). Viewing time is considered an important factor in such evaluations. Observations for pedestrian to recognise facial features at varying distances and fixation durations of other pedestrians. Compared with unlimited viewing durations, such measurements more accurately reflect participants' real-world behaviour (Fotios et al., 2015).

2.3. PERCEPTUAL AND MEMORY RECALL

ET captures actual visual responses across different environmental conditions through recording participants' gaze behaviour (Holmqvist et al., 2011). However, this method does not allow for the direct determination of whether participants' fixations and saccades influence their subjective memory. After ET experiments, collecting participants' recollections and identifying observed details through questionnaires helps verify what they truly perceived (Rusnak et al., 2024).

Previous research shown that participants may repeatedly encounter similar colours, materials, shapes, or patterns in the environment, leading to confusion in their actual recollections (Koutstaal & Schacter, 1997). As such, this study employs ET to verify the content of fixations and saccades, thereby excluding false memories.

3. Methodology

This study examines how people's subjective perceptions and actual gaze behaviours vary across different lighting conditions to uncover the critical factors in nighttime lighting design. First, we constructed a digital twin of the heritage site using point-cloud scanning and employed it for nighttime lighting analysis. Luminaires were modelled in the digital twin using two lighting modes, warm (2700K) and natural (3900K), to meet required illuminance levels.

With these nighttime lighting models, we then arranged a series of test tasks using

prepared guided-path animations and representative static images. Participants were subsequently asked to watch the rendered videos to experience the nighttime lighting effects and, after the video ended, to review the task-related scenes through static images. During this process, eye-tracking data were recorded, and participants completed questionnaires to recall and document what they had seen. Thirty participants (fifteen females, fifteen males; age range 18–35 years) were recruited. For exploratory subgroup analyses, participants were grouped by gender and by age to examine potential differences in subjective ratings and task performance across lighting conditions.

The final step examined participants' questionnaire responses with fixation data and gaze heatmaps to verify their physiological reactions within the simulated environment. The objective is to evaluate participants' performance on obstacle recognition, person identification, perceived pleasantness, text legibility, and the perception of details and materials under different lighting conditions.

3.1. CASE STUDY: HOBE FORT, TAMSUI, TAIWAN

The site used for this study is Hobe Fort, a late Qing Dynasty defensive structure that occupies a modest area overlooking the Tamsui River estuary in northern Taiwan. The fort comprises key military elements, including artillery platforms, vaulted tunnels, and enclosing stone walls, reflecting the strategic defensive planning of its time.

3.2. 3D SCANNING AND MODEL RECONSTRUCTION

The study employed a Matterport Pro3 LiDAR scanner to conduct high-precision 3D scanning and capture information on the heritage building's geometry, colour, surrounding landscape, and natural environment. A total of 400 scanning stations were set up throughout the interior corridors, gun emplacements, exterior ramps, and all-terrain nodes of the fort, enabling the construction of a complete indoor–outdoor digital twin of the entire site. At each position, LiDAR point-cloud data and 360° RGB images were captured simultaneously to enable automatic registration and spatial alignment.

The Time-of-Flight (ToF) LiDAR in Matterport Pro3 provides ±20 mm accuracy at distances up to 100 metres. Each scanning station required approximately 20 seconds to acquire LiDAR depth data and capture images using the 20-megapixel RGB sensor, resulting in a 134.2-megapixel 360° panorama. As shown in Figure 1., the study employed an Apple iPad Pro to monitor on-site scanning progress, verify station-to-station alignment accuracy, and manage cloud uploads simultaneously.



Figure 1. (Left) iPad Pro M4; Matterport Pro 3 Scan Process; (Right) Exported models.

After scanning, the exported dataset includes geometric meshes (OBJ), colour

textures (JPG), material references (MTL), coordinate data (XYZ), and scanning logs (CSV). These data were imported into Rhino for visualization and lighting layout and subsequently rendered and presented using Enscape.

3.3. CASE STUDY: HOBE FORT, TAMSUI, TAIWAN

For the outdoor heritage nighttime lighting design, this study developed two lighting versions based on the colour-temperature guidelines specified in CNS 15015. The first version uses 2700K warm light, creating a calm and approachable spatial atmosphere that enhances the natural colours of the cultural heritage structure at night. The second version employs 3900K neutral light, offering a brighter and clearer presentation that emphasizes detail visibility.

Both versions were designed to ensure that the illuminance distribution meets the CNS 15015 recommendations, 5–10 lx for pedestrian paths and 10–15 lx for main walkways. Lighting levels were adjusted according to the visual tasks required in different scene areas, and the overall layout was designed to maintain reasonable illuminance uniformity. The luminaire arrangement was tested to ensure a visually comfortable lighting environment while supporting both leisurely viewing of historical details and safe walking tasks.

3.4. RENDERING AND ANIMATION PREPARATION

The study used Enscape Render to produce animated walking sequences. The path begins on the external access road leading toward the gate, proceeds through the gate into the central courtyard, and then circulates along the inner side of the fortress wall. Elements encountered along the route include various pavement materials, the outer gate inscription "北門鎖鑰", the main fortress wall, and heritage architectural components. For testing purposes, the study added pedestrian models, a tall lamp model, and garden light models to support the task-based evaluations.

To investigate how the colour temperature of light sources influences visual attention and perceptual experience at night, the study selected three nighttime scenes of the Hobe Fort (Figure 2.) for eye-tracking analysis. These scenes represent the fort's entrance pathway, the arched entry passage, and the inner corridor adjacent to the side arcade and gun emplacement chamber.



Figure 2. Selected scenes in 2700 & 3900K (yellow & neutral white light) Environmental

3.5. EYE-TRACKING EXPERIMENT SETUP

To measure participants' visual behavior while viewing simulated nighttime scenes of the heritage environment, this study utilized the Tobii Pro Spark 60 for eye-tracking. The Tobii Pro Spark 60 is a screen-based, non-intrusive eye tracker operating at a

sampling frequency of 60 Hz, offering sufficient temporal resolution for studying environmental perception, architectural detail detection, and visual attention. All participants reported normal or corrected-to-normal vision, and the screen-based setup supported natural viewing while minimizing behavioral interference.

3.6. TASKS AND MEASURES

This experiment designed a series of visual tasks to explore how different lighting conditions affect participants' perceptual abilities and visual attention in historical heritage environments. These tasks reflect both safety-related visual needs and the need to access historical information. First, participants were asked to identify transitional areas between sidewalks and gravel surfaces to assess their ability to identify ground textures and avoid potential obstacles. Embedded human models in the virtual environment served as targets for examining person identification.

To measure the accessibility of historical heritage, participants were asked to identify information on signs near the fortress entrance, recognize the text and colour of the inscription "North Gate Guardian" on the outer lintel, and observe the material and textural details of the walls and gate structures. These tasks reflect typical needs for interpreting cultural monuments. Finally, participants rated their subjective comfort and sensitivity, reflecting their overall emotional responses to lighting conditions.

3.7. DATA ANALYSIS

To analyze eye-movement events and visual attention, areas of Interest (AOIs) were defined for specific architectural elements—such as stone textures, inscriptions, structural openings, and lighting fixtures—allowing for the extraction of fixation-based metrics within these zones. Resulting heatmaps, gaze plots, and scan path visualizations facilitated both qualitative interpretation and quantitative comparison across the 2700K and 3900K lighting conditions. These datasets provide a robust empirical basis for evaluating the effects of lighting colour temperature on visual attention, environmental legibility, and perceptual experience within digitally reconstructed cultural heritage contexts.

For the questionnaire component of the eye-tracking experiment, participants were asked to rate each reviewed scene after watching the walking animation. A five-point Likert scale was used (1 indicating "very low" and 5 indicating "very high") to evaluate task-related dimensions, including the perceived clarity of historical architectural features, recognition of approaching pedestrians, identification of ground conditions, avoidance of obstacles, and overall environmental pleasantness. The scores reflected participants' level of acceptance regarding the lighting conditions within the heritage site. Through the questionnaire survey, responses from 30 participants were collected across the various scenes at Hobe Fort. A reliability analysis confirmed the internal consistency of the scale. Statistical analyses were then conducted to elucidate participants' fundamental needs for nighttime lighting in Hobe Fort, and their perceptual experience of historical readability.

4. Results

To assess the questionnaire's reliability, Cronbach's α was used to evaluate internal

consistency. According to the classification criteria proposed by Cuieford (1965), a Cronbach's α value of < 0.35 indicates low reliability; a value of $0.35 \leq \alpha < 0.70$ falls within the acceptable range, and a value of $\alpha \geq 0.70$ indicates high reliability. This study employed $\alpha \geq 0.70$ as the criterion for determining that the scale possesses adequate internal consistency.

4.1. QUESTIONAIRE RESULTS

Analysis From the pleasantness questionnaire, the 2700K scene yielded a Cronbach's α of .863, indicating high reliability. The 3900K scene produced an overall Cronbach's α of .782, demonstrating good internal consistency of the scale.

Table 1. presents the spatial pleasantness results. A paired-samples t-test showed higher pleasantness under 3900K than 2700K, with significant differences in Scenes A, C, and D ($p < .05$). Exploratory subgrouping by gender and age suggested systematic variation: male participants tended to report higher pleasantness and perceptual-clarity ratings across both CCTs, while female participants showed a larger pleasantness increase when shifting from 2700K to 3900K. Younger participants (age<25) also demonstrated a stronger preference shift toward 3900K and larger gains in perceived clarity/heritage focus than the 25–35 group. Scenes B, E, and F did not show statistical significance on the CCT difference, yet the 3900K is still superior to 2700K.

Table 1. Spatial pleasantness results comparison by t-test.

Scene	CT*	Mean	Std. Dev*	MDiff*	t	df	p Value
A	2700K 3900K	2.933 3.633	1.368	-0.700	-2.802	29	.009
B	2700K 3900K	3.667 3.867	1.448	-0.200	-0.757	29	.455
C	2700K 3900K	3.167 3.767	1.545	-0.600	-2.127	29	.042
D	2700K 3900K	2.367 3.067	1.745	-0.700	-2.197	29	.036
E	2700K 3900K	2.200 2.533	2.057	-0.333	-0.888	29	.382
F	2700K 3900K	3.167 3.533	1.974	-0.367	-1.018	29	.317

*CT-Color Temperature; Std.Dev.- Standard Deviation; MDiff- Mean Difference

In Table 2., illuminance shows a significant effect on person identification and detail recognition. Under both the 2700K and 3900K lighting conditions, participants demonstrated markedly higher clarity and responsiveness in identifying human figures in the high-illuminance condition, and the differences were all statistically significant ($p < .05$). Similarly, participants' detail-recognition performance in the high-

illuminance setting was also significantly better than in the low-illuminance setting ($p < .05$). These findings indicates that illuminance is a key factor influencing person identification and detail-recognition ability. Increasing illuminance effectively enhances participants' efficiency in identifying object contours, textures, and fine features, thereby improving overall visual quality.

Table 2. Person identification and detail recognition results comparison by t-test.

	CT*	Illum*	Mean	Std. Dev*	MDiff*	t	df	p Value
Person Identification	2700K	High Low	3.167 1.867	1.343	1.300	5.302	29	<.001
	3900K	High Low	3.400 2.033	1.450	1.445	5.163	29	<.001
Detail Recognition	2700K	High Low	3.600 2.567	1.033	1.033	5.477	29	<.001
	3900K	High Low	3.800 3.200	1.133	0.600	2.902	29	.007

*CT-Color Temperature; Illuminance- Standard Deviation; MDiff- Mean Difference

This study compared participants' performance on the three obstacle-recognition answers in scene D, E, F across the 2700K and 3900K lighting conditions. Obstacle-recognition scores under the 3900K condition were consistently higher than those under 2700K, indicating that the higher colour temperature enhanced participants' ability to identify obstacles.

The differences across the three scenes were mostly "trend-level differences" and did not reach statistical significance ($p < .05$). Compared with warm lighting (2700K), the neutral lighting (3900K) still provides better contrast and visual clarity, thereby facilitating improved obstacle detection. These findings support the application of higher colour-temperature lighting in scenarios that require obstacle recognition.

4.2. VISUAL ATTENTION

In both 2700K and 3900K scenarios (figure 3.), participants exhibited a consistent pattern of visual attention across all three scenes. Historical architectural textures and materials constituted the dominant focus of visual attention, yielding the highest fixation duration (≈67–85% of AOI time) and the highest fixation occurrence rates (near 100%). Signage and indicators received minimal fixation, whereas obstacles and ground irregularities captured moderate but secondary attention. Across all six fixation-count datasets, historical texture & materials consistently produced the highest fixation counts and a 100% fixation rate under both 2700K and 3900K lighting conditions. This indicates architectural heritage details as a dominant and stable visual attraction across all scenes.

For obstacles and ground-level changes, the 2700K condition showed slightly higher fixation averages (Scene 2: 2.40; Scene 3: 2.00; Scene 6: 3.90) compared with 3900K (Scene 2: 2.70; Scene 3: 1.77; Scene 6: 3.00). However, the differences were

relatively small, suggesting that hazard detection behavior remains generally consistent across color temperatures. Human features and signage received the fewest fixations under both lighting conditions, with signage attracting fixations from only 13.33% (2700K) and 26.67% (3900K) of participants, indicating low visual saliency within the studied environments.

Overall, 3900K scenes produced the stronger visual engagement with heritage-related elements, demonstrated by the highest fixation count (17.50) and largest variance (57.57), implying both strong attraction and large individual differences.

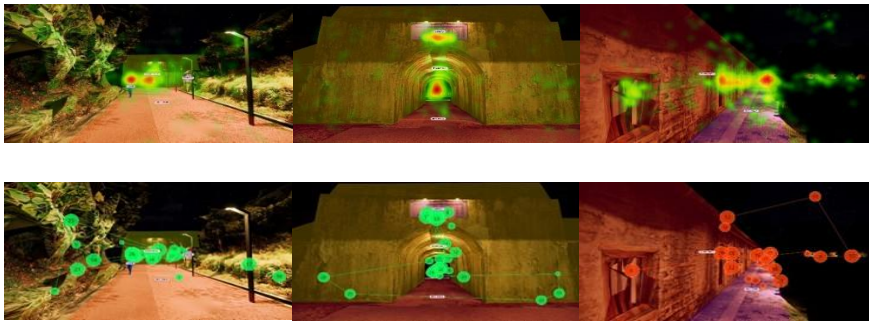


Figure 3. (Top) Color temperature 2700K in 3 scenes' Gaze plot, path; (Bottom) Color temperature 2700K in 3 scenes' eye-tracking heat map; (left) scene1; (middle) scene 2; (right) scene3

5. Conclusion

The results of this study indicate that illuminance has the greatest impact on perceptual performance in nighttime historical environments. Higher illuminance consistently improved the accuracy of person detection, detail recognition, and obstacle detection, confirming its central role in visual clarity and safety. The influence of colour temperature was comparatively smaller, with 3900K showing slightly better performance in visual comfort and a modest advantage in obstacle detection.

Eye-tracking data further demonstrated that participants exhibited highly consistent visual behaviour under both lighting conditions. Their attention was primarily directed toward the textures and materials of the historical architecture, followed by ground-level changes, while people and signage attracted minimal attention. This suggests that scene content and task relevance dominated visual-attention patterns beyond CCT.

Overall, the findings show that illuminance has a stronger influence than colour temperature, and that a neutral lighting condition provides slightly better perceptual outcomes. The study further demonstrates how digital-twin simulation can support lighting design in heritage contexts by enabling controlled, repeatable testing of CCT and illuminance, rapid iteration of luminaire layouts against standards, and risk-free evaluation of safety and interpretive readability without physical intervention on protected sites. The findings are derived from screen-based simulations and were not validated through an in-situ lighting installation at the heritage site. Future work should calibrate digital-twin photometry with field measurements and verify the outcomes under real-world nighttime conditions.

References

BSMI (Bureau of Standards, Metrology and Inspection). (2018). CNS 15015: Lighting design—Roadway and outdoor lighting. Ministry of Economic Affairs.

British Standards Institution. (2020). BS 5489-1:2020: Code of practice for the design of road lighting. BSI Standards Limited.

European Committee for Standardization. (2015). EN 13201-2:2015: Road lighting—Part 2: Performance requirements. CEN.

Amundadottir, M. L., Rockcastle, S., Sarey Khanie, M., & Andersen, M. (2017). A human-centric approach to assess daylight in buildings for non-visual health potential, visual interest and gaze behavior. Building and Environment, 113, 5–21. https://doi.org/10.1016/j.buildenv.2016.09.033

Caminada, J. F., & van Bommel, W. J. M. (1984). New Lighting Criteria for Residential Areas. Journal of the Illuminating Engineering Society, 13(4), 350–358. https://doi.org/10.1080/00994480.1984.10748787

Fotios, S., Yang, B., & Uttley, J. (2015). Observing other pedestrians: Investigating the typical distance and duration of fixation. Lighting Research & Technology, 47(5), 548–564. https://doi.org/10.1177/1477153514529299

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. van de. (2011). Eye Tracking: A comprehensive guide to methods and measures. OUP Oxford.

Jiang, X., Li, J., & Yang, B. (2021). Luminance and saliency have impact on pedestrians' fixation distribution during natural walking: Evidence from mobile eye-tracker. Lighting Research & Technology, 53(4), 359–372. https://doi.org/10.1177/1477153520968158

Koutstaal, W., & Schacter, D. L. (1997). Gist-Based False Recognition of Pictures in Older and Younger Adults. Journal of Memory and Language, 37(4), 555–583. https://doi.org/10.1006/jmla.1997.2529

Li, Q., Guo, M., Ye, L., Li, P., Luan, J., & Liu, B. (2025). The Investigation of Visual Characteristics of High-Speed Railway Drivers: Perspectives of Light Environment. Journal of Advanced Transportation, 2025(1), 5116362. https://doi.org/10.1155/atr/5116362

Rusnak, M., Szmigiel, M., Geniusz, M., Koszewicz, Z., & Magdziak-Tokłowicz, M. (2024). Exploring the impact of cultural context on eye-tracking studies of architectural monuments in selected European cities: Sustainable heritage management. Journal of Cultural Heritage, 66, 326–342. https://doi.org/10.1016/j.culher.2023.12.002

Turatto, M., & Galfano, G. (2000). Color, form and luminance capture attention in visual search. Vision Research, 40(13), 1639–1643. https://doi.org/10.1016/S0042-6989(00)00061-4

Yang, C., & Lyu, C. (2025). Study the nighttime visual attention in urban waterfront public space: Insights using mobile eye-tracking. Frontiers of Architectural Research. https://doi.org/10.1016/j.foar.2025.05.002

Occupant-Driven Building Performance

Designing for Biodiversity in the Anthropocene

Multispecies Building Envelopes as Ecological Infrastructure

Narkis Joffe[1], Oren Agay-Shay[2], Lilu Golan[3], Anastasiia Slavina[4] and Yasha J. Grobman[5]
[1,2,3,4,5]*Technion – Israel Institute of Technology*
[1]*narkis.joffe@campus.technion.ac.il, 0009-0009-3380-4762*
[2]*orenagayshay@campus.technion.ac.il, 0009-0003-1522-1631*
[3]*lilu.golan@campus.technion.ac.il, 0009-0003-1713-5446*
[4]*anastasiias@campus.technion.ac.il, 0009-0007-4751-1296*
[5]*yasha@technion.ac.il, 0000-0003-4683-4601*

Abstract. The ongoing rapid urbanization has significantly disrupted natural ecosystems, resulting in habitat loss, diminished biodiversity, and an increasing detachment between humans and non-human organisms. It also leads to a decrease in human well-being as exposure to nature is acknowledged as one of the factors in promoting well-being and resilience. This paper presents a design framework and a case study for a multispecies building envelope tile system that simultaneously promotes biodiversity and enhances environmental building performance. Within this framework, it explores how ecological data can be employed in architectural design via parametric design and digital fabrication to generate eco-architectural building envelopes. The study's methodology progresses through three main stages: Ecological, environmental and spatial data collection, the development of a multispecies design methodology for building envelope tiles, and the fabrication methodology for the multispecies tile system. The results demonstrate the feasibility of the suggested framework to integrate ecological data into digital design process to increase the ecological and environmental performance of the building envelope.

Keywords. Multispecies Design, Architectural Design Process, Building Envelope, Computational Design, Additive Manufacturing

1. Introduction

Urban growth has led to the disruption and reduction of natural habitats, intensifying the climate crisis and ecological challenges. Cities have become ecologically disturbed areas, excluding non-human species from their habitats. Non-human organisms are included mainly when they serve human aesthetic or productive purposes. This fosters alienation, a phenomenon referred to as "extinction of experience" (Soga & Gaston, 2016). This also leads to decreased in human well-being as exposure to nature is

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 329-338.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

acknowledged as a factor promoting well-being and resilience (Bratman et al., 2012).

Human-made structures create new topography on facades and rooftops. Green facades and walls are gaining increasing importance as sustainable elements. However, current applications focus mainly on plants and often prioritize aesthetics, demand high maintenance. Recognizing the ecological potential of building envelopes promotes a nature-centred design approach that integrates multi-species organisms (Flora and Fauna) into the urban fabric (Grobman et al., 2023). The paper presents the results of research that develops a framework for the design and fabrication methodologies of a modular multispecies envelope cladding system that simultaneously promotes biodiversity and enhances environmental building performance.

2. Background

2.1. STATE-OF-THE-ART RESEARCH OF MULTISPECIES DESIGN

Integrating habitats onto building facades is a complex task that requires expertise across multiple fields (de Wilde and Bleil de Souza, 2022). While extensive research has been conducted on integrating multispecies at urban scale, few studies focus on multispecies integration on façades at the building scale.

Grobman et al. (2023) reviewed research in this field and introduced the concepts and challenges of multispecies integration in architecture. Multispecies design advocates treating plants and animals as co-inhabitants and demonstrates the potential for creating shared habitats. According to Grobman, significant challenges remain for true multispecies integration in architectural design, among them are the anthropocentric focus on human needs at the expense of non-human habitats, knowledge gaps, limited data, and a lack of methodologies for designing environments for non-human species.

The EU-funded ECOLOPES (ECOlogical building enveLOPES) project developed a conceptual and computational framework for multispecies integration in architectural design, providing new tools for embedding ecological data and digital design processes into building envelope development (Weisser et al., 2023). While it laid the theoretical and methodological foundation for linking ecological and architectural parameters, the present research builds upon this work and extends it to the building envelope tile scale, as ECOLOPES project stayed in the schematic scale of tile distribution over the building envelope.

Recent research on multispecies design highlights the important influence of building façades on other species (Oloś, 2023). However, none of this research focuses on the additional species themselves, and most regard their integration as a side effect of greenery.

In practice, multispecies design exists in varying levels of development, from singular habitat element, which is an independence habitat attached independently to an existing structure, through case study parts of an envelope system, and varied multispecies façade that refers to a single facade, or Entire building envelope.

In existing projects that focus on independence habitat elements, most projects address only one species. For example, Tevaplanter's "Teva-tile" (see Figure 1), which presents a geometrical design for plant growth, a bird habitat called the "Birdhouse

rooftile" (see figure 1). Most of these type of element are commercial products with no published research, monitoring, or data found on their success.



Figure 1, left to right: TevaTile (Source: Tevaplenter, 2024, Kickstarter URL), Bird Apartment (Source: Nendo, 2012, URL), Bee-Brick (Source: Greenandblue, 2024, URL), Birdhouse (Source: klaas kuiken, 2012, Designboom URL) Beecosystem (Source: beecosystem 2014, URL)

As opposed to a single multispecies element, a multispecies single façade refers to an entire section of a building envelope, or a free-standing large element, or a part of an envelope system. For insect-related designs, "La Bourdonnerie," a project by the Réserves Naturelles de France (RNF) and Atelier Clac in Dijon, France, serves as a representative example of an add-on modular element for an existing façade that features modular units designed for insects and pollinators (see figure 2). Another approach involves utilizing the facade itself to incorporate integrated habitats within the insulation gap while maintaining insulation benefits for the building as appears in a project by Saint-Gobain company and NABU union in Germany, employed a double-skin insulation façade, to create nesting areas for birds (see figure 2). Similarly, Larikova's research (see figure 2), developed a 3D-printed facade that enables internal nesting zones while enhancing facade insulation (Larikova et al., 2022). Most examples focus on a specific species group and do not produce a supportive ecological design that accounts for local interspecies relationships.



Figure 2, left to right: La Bourdonnerie (Source: Atelier Clac, 2015, URL) Nest Boxes (Source: Saint-Gobain& NABU, 2018, URL), Additively Manufactured Urban Multispecies Façades (Adapted from Luliia Larikova et al. 2022, JFDE URL), Mellor Primary School (Source: Sarah Wigglesworth Architects, 2015, URL), Hosting life:a research (Source: ChartierDalix, 2014, URL)

Examples of entire building envelopes are rare, and those that do exist are typically designed to incorporate additional species primarily for human benefit. One example is the Mellor Primary School habitat wall by Sarah Wigglesworth Architects (see figure 2), which integrates nest boxes, insect hotels, and planters into a unified façade system. Another is the work of the French firm Chartier-Dalix, which incorporates planting spaces within the building envelope and has developed prototypes integrating vegetation into brick or concrete walls (see figure 2). However, no scientific research evaluating the outcomes of these projects was found.

3. 3. Gaps and aim of the research

While existing research laid the theoretical and methodological foundation for linking ecological and architectural parameters, it did not advance to the design, fabrication,

and performance evaluation of a complete multispecies envelope system. Building upon this groundwork, the present study extends the current multispecies approach by developing and testing an integrative workflow that connects ecological data acquisition, parametric design, and digital fabrication toward the realization of a modular multispecies tile system for architectural application.

4. Methodology

The research consists of three main stages: Ecological, environmental and spatial data collection, development of multispecies design methodology for building envelope tiles, developing fabrication methodology for multispecies tiles and testing these processes in a case study design. The research methodology is presented in Figure 3.

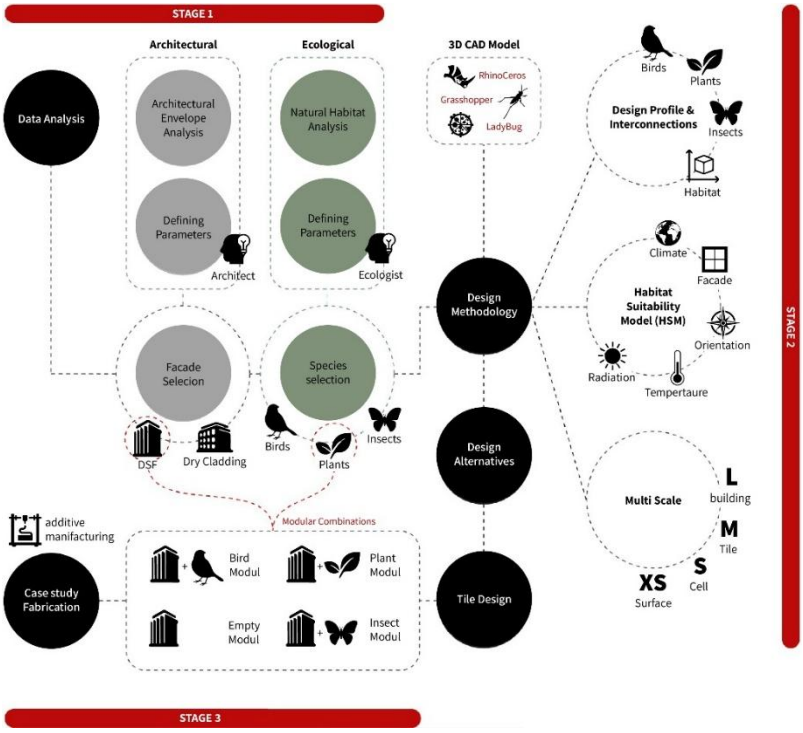


Figure 3. Research Methodology Diagram

The aim of the first stage is to define the ecological, environmental and spatial parameters and their interconnection for developing a multispecies design methodology. The needs were examined from two perspectives: the first ecological-biological, and the second spatial-architectural.

The ecological guidelines were developed based on ecological and biological data retrieved from a review of academic data, including local and governmental sources. This was followed by interviews with local ecologists that refined the list according to defined parameters, such as adaptation to the specific habitat area, ability to integrate

into the urban environment, human acceptance (not defined as pests), living environment including food sources, physical structure, and interactions with other species. From the list of potential species, the physical dimensions of their habitats were examined, such as root cell size and form, shelter or nesting area, and possible adaptation and integration onto a building façade. Among the selected species, those that coexist within an interconnected ecosystem were chosen, supporting one another in terms of food, shelter, nesting, and attractiveness to other desired species.

We evaluated multispecies building envelope feasibility by defining envelope parameters and surveying dense-urban façade typologies. Each façade was characterized (geometry, structure, openings, proximity, available depth/volume), its role and aesthetics assessed, and species/human risks hazard-mapped from literature and construction guidelines. The result of this stage was the selection of nine local species, including three plants, two birds, and four insects, that best met the defined criteria, along with the identification of a potential façade type suitable for integration.

The second stage formalized a multi-scale, multispecies design process progressing from the building scale to the tile, cell, and surface/texture scales. At the building scale, we classified envelope types according to building characteristics and defined a distribution strategy informed by a parallel project (removed for review). Multiple envelope alternatives were generated for distinct typologies. From these, we selected a tile system and specified façade configurations, connection methods, and modular options to accommodate variable dimensions. The process then moved to the tile scale, where layout and attachment were optimized for constructability and aesthetic coherence. At the cell scale, the dimensions of planters and bird nests governed minimum and optimal tile sizes. At the surface scale, we introduced three levels of geometric complexity enabled by additive manufacturing. These morphologies were designed to enhance water management and thermal insulation. Together, these steps yield a scalable workflow for multispecies integration across architectural scales.

The third stage focused on developing a fabrication methodology for multispecies tiles and examination of the design result in a case study design. The fabrication process was carried out using additive manufacturing technology. The design was developed with parametric control of the toolpath, using Rhinoceros and the Grasshopper plugin for G-code generation. To achieve varying levels of morphological complexity, the process replaced conventional slicing methods with a design approach based on custom print paths. This method, based on Vargas et al., (2025) enabled the creation of new surface qualities and controlled porosity.

5. Results and discussion

5.1. STAGE 1 - ECOLOGICAL, ENVIRONMENTAL AND SPATIAL DATA COLLECTION

In the first stage, we established ecological, environmental, and spatial criteria to identify species suitable for façade integration and to clarify their habitat needs. We screened ~50 candidates across plants, birds, and insects.

Plant selection, guided by an ecologist, prioritized local adaptability and survival in minimal soil and water substrates to limit structural loads, while enabling interactions

with birds and insects (pollination, foraging, shelter). The chosen species were White Hedge-Nettle (Prasium majus L.), Podonosma orientalis (L.) Feinbrun, and Caper bush (Capparis spinosa var. aegyptia).

Bird selection, also ecologist-guided, favoured small, urban-adapted cavity nesters compatible with vertical habitats and unlikely to encourage invasive or aggressive species. We focused on non-harmful taxa experiencing declines from habitat loss, resulting in the Great Tit (Parus major) and House Sparrow (Passer domesticus).

Insects were evaluated with an entomologist and grouped as pollinators, butterflies, and beetles; criteria emphasized ecological support roles, occupancy of foliage or small cavities/crevices, and non-pest status. Aesthetic value (e.g., butterflies) and ecological pest control (e.g., beetles) were additional considerations. The selected taxa, solitary wild bees, Old World Swallowtail (Papilio machaon), Large White butterfly (Pieris brassicae), and ladybug (Coccinellidae), reflect both functional significance and documented declines, particularly among pollinators.

Following the species selection, the physical habitat conditions of each species were defined based on literature and framed according to criteria relevant to façade integration, such as nesting height above ground, the physical volume of the natural habitat, and the ability to coexist and interact with other species. In addition, preferred ecological parameters were identified and evaluated in grasshopper ladybug software using a local climate file in EPW format, including wind orientation, temperature, humidity, environmental resilience, and levels of sun and wind exposure. For each species, a design profile was developed based on its ecological requirements.

5.2. STAGE 2 - DEVELOPMENT OF MULTISPECIES DESIGN METHODOLOGY FOR BUILDING ENVELOPE TILES

The distribution model used is based on an updated version of 'Habitat Suitability Model' (HSM) by Selvan et al., (2025), which uses ecological databases map species distributions according to environmental and interspecies proximity needs. These models predict potential species distributions by statistically correlating spatial predictors, such as temperature, solar radiation, or topography, with recorded occurrences data. (see figure 4).

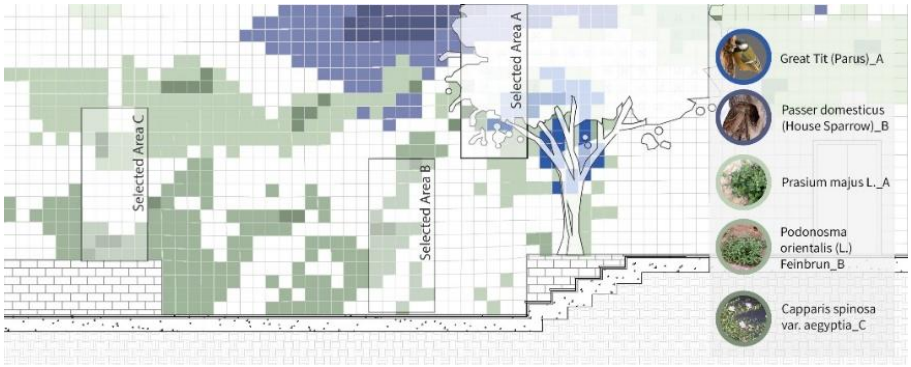


Figure 4. Case study façade species distribution - image adopted from Anastasiia Salvina.

The division across the grid considers the habitat size required for birds (25 × 25 cm with a depth of 20 cm) and for vegetation cells of a similar scale. The division is organized at the height of one floor and considers tile dimensions that allow convenient and modular installation without the need for heavy machinery or external equipment.

To maximize flexibility in the distribution of tiles across different wall configurations, a modular, multi scale design approach was developed: one that enables the creation of various relational combinations of multispecies tiles. These levels included envelope system design at the building scale, design the envelopes multispecies various tile types at the tile scale, which was informed by the cell scale (e.g., the dimensions of a root cell or a nesting cavity) and designing the texture morphology of the tile in the surface scale for increasing thermal performance, supports ecological needs and directing rain water into the planters. The modular system included four interchangeable tile types: a tile for plant cells, a tile for birds nest, a tile for insects, and an "empty" tile (See figure 5).

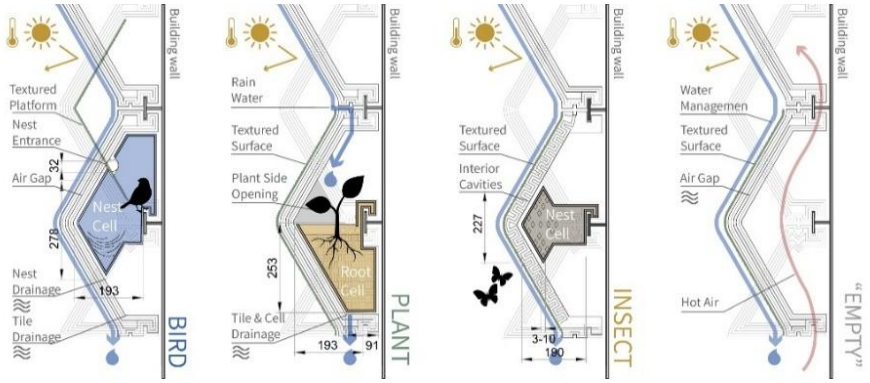


Figure 5. Tile module Detail

The planter module, featuring a side opening for plant growth and exit, was designed to imitate traditional brick walls where vegetation grows from the joints rather than from the bricks themselves. In addition, in the surface scale, the use of an angled collection surface combined with a water-lip detail enables the tile to harvest and channel rainwater for the self-irrigation of the planting cells.

The bird-nesting module features internal dimensions of 20 × 20 × 25 cm. Here too, the nest entrance and exit opening is positioned on the side, differing from conventional nesting boxes, thus providing enhanced shading and protection for the concealed entrance in terms of solar exposure and lines of sight. The nest opening was designed according to the dimensions suitable for the selected species, preventing access by invasive birds. In addition, the continuous integration with the neighbouring tile creates a perch and a convenient approach for the bird to the nest entrance.

The insect module features side-entered cavities and tunnels with diameters of 5 to 10 mm and depths of approximately 10 cm from each side.

5.3. DEVELOPING FABRICATION METHODOLOGY AND CASE STUDY FOR MULTISPECIES TILE SYSTEM

The fabrication employed Bambu Lab X1-Carbon and Prusa MK3 printers for the mock up and Wasp 3MF LDM Extruder for the 1:1 tiles. The materials used was natural terracotta clay up to 0.2 mm chamotte with an addition of water and glass water for extrudability, In a ratio of 250 ml of water and 10 ml of water glass per 10 kg of clay. The entire print is designed as a single continuous extrusion line, ensuring material strength and minimizing seam lines.



Figure 6. 1:5 wall model mock-up

The case study fabrication process, following the design stage, was carried out at multiple scales. At the building scale, a modular system was designed and 3D-printed at a 1:5 scale, including façade details and wall-connection components (See figure 6).

At the tile scale, tiles and habitat cells for plants, birds, and insects were fabricated, featuring side-entry openings to maintain a uniform façade appearance.

At the case-study fabrication stage, the three modules were printed with terracotta clay at 1:1 scale using Wasp 3MF LDM Extruder 3D printer, with overall dimensions of 62 cm in height and 20 cm in width, and 20 cm depth. The printer bed measures 1000×1000 mm, with a nozzle diameter of 8 mm and a layer height of 3.5 mm. To achieve a porous morphological range by allowing material Z-fall and controlled variations in the toolpath, the offsets vary between 8–16 mm. Printing speed 10-30 mm/s. The full-scale tiles were printed, dried using a fan for seven days, and fired in a ceramic kiln to a bisque state at 900 °C, preserving material porosity (See figure 7).

For full wall installation, complete modelling and visualization were performed according to the spatial distribution discussed in Section 4.3. The tiles were designed for dry installation on a detached supporting structure, forming a DSF system. Several prototypes were fabricated to evaluate performance for on-site installation.

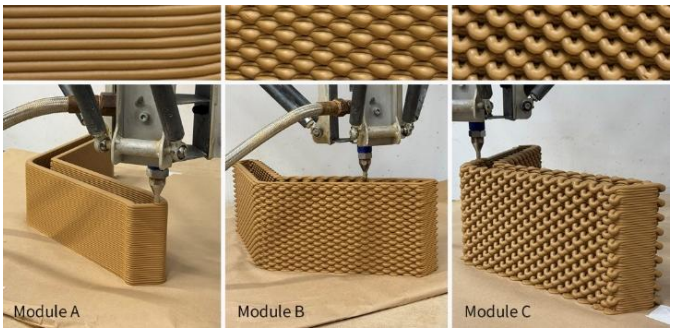


Figure 7. Clay printing process

6. Conclusions

This research proposes and demonstrates a design methodology that positions the building envelope as an active ecological interface rather than a passive boundary between humans and nature. By integrating ecological data, parametric tools, and digital fabrication processes, the study develops and tests a multispecies tile system capable of supporting birds, insects, and plants while improving the façade's environmental performance. The research shows that multispecies design can extend beyond aesthetic greening or symbolic gestures toward a scientifically informed and computationally supported framework for ecological integration at the building scale.

The study showed how ecological, environmental, and spatial data can inform architectural decision-making. The resulting workflow operates across multiple scales, from ecosystem-level species selection to façade-level distribution and tile-level fabrication, illustrating how species' biological needs can be embedded within architectural geometry, material, and performance criteria. The fabricated prototypes demonstrate the feasibility of additive manufacturing using natural clay as a bio-receptive and environmentally responsive material, opening new opportunities for adaptive, low-impact, and locally sourced construction strategies.

7. Future Research and Limitations

Future work will move beyond prototyping to longitudinal ecological monitoring and performance evaluation. Because adoption is gradual, at least two annual cycles are needed to assess species establishment and habitat stability via field observation, sensors, and digital tracking. We will also broaden the species palette, test across climatic zones, and integrate smart sensing for adaptive maintenance and design optimization. By merging computational ecology, environmental simulation, and robotic fabrication, multispecies building envelopes can contribute to the evolution of architecture as part of the urban ecosystem, a shift from constructing isolated artifacts toward cultivating living infrastructures that sustain both human and non-human life.

Acknowledgements

The author acknowledges the Design Tech Laboratory, Technion, for their support with clay 3D printing. This research was supported by the Israel Science Foundation (ISF) (Grant No. 485/25). AI Use Statement

Attribution Statement

ChatGPT (OpenAI, 2024) was used for language editing, the authors retain full responsibility for the text in the paper.

References

Bratman, G. N., Hamilton, J. P., & Daily, G. C. (2012). The impacts of nature experience on human cognitive function and mental health. Annals of the New York Academy of Sciences, 1249(1), 118–136. https://doi.org/10.1111/j.1749-6632.2011.06400.x

de Wilde, P., & Bleil de Souza, C. (2022). Interactions between buildings, building stakeholders and animals: A scoping review. Journal of Cleaner Production, 367, 133055. https://doi.org/10.1016/j.jclepro.2022.133055

Grobman, Y. J., Weisser, W., Shwartz, A., Ludwig, F., Kozlovsky, R., Ferdman, A., Perini, K., Hauck, T. E., Selvan, S. U., Saroglou, S. (Tanya), Barath, S., Schloter, M., & Windorfer, L. (2023). Architectural Multispecies Building Design: Concepts, Challenges, and Design Process. Sustainability, 15(21), 15480. https://doi.org/10.3390/su152115480

Larikova, I., Fleckenstein, J., Chokhachian, A., Auer, T., Weisser, W., & Dörfler, K. (2022). Additively Manufactured Urban Multispecies Façades for Building Renovation. Journal of Facade Design and Engineering, 10(2), 105–126. https://doi.org/10.47982/jfde.2022.powerskin.7

Oloś, G. (2023). Green facades support biodiversity in urban environment – A case study from Poland. Journal of Water and Land Development, Journal of Water and Land Development, 257–266. https://doi.org/10.24425/jwld.2023.148450

Selvan, S. U., Windorfer, L., Saroglou, S. T., Calbi, M., Culshaw, V., Roccotiello, E., Grobman, J. Y., & Barath, S. (2025). Multispecies design decision-making: A hybrid computational model for ecological building envelope optimization and evaluation. Journal of Building Engineering, 108, 112565. https://doi.org/10.1016/j.jobe.2025.112565

Soga, M., & Gaston, K. J. (2016). Extinction of experience: The loss of human–nature interactions. Frontiers in Ecology and the Environment, 14(2), 94–101. https://doi.org/10.1002/fee.1225

Vargas, J. H., Westerlind, H., Breseghello, L., & Silfwerbrand, J. (2025). Design of Porous Structures in 3D Concrete Printing Using Procedural Print Patterns: 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia. Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2025, 2, 151–160.

Weisser, W. W., Hensel, M., Barath, S., Culshaw, V., Grobman, Y. J., Hauck, T. E., Joschinski, J., Ludwig, F., Mimet, A., Perini, K., Roccotiello, E., Schloter, M., Shwartz, A., Hensel, D. S., & Vogler, V. (2023). Creating ecologically sound buildings by integrating ecology, architecture and computational design. People and Nature, 5(1), 4–20. https://doi.org/10.1002/pan3.10411

Parallel Optimization for Competing Design Scheme Exploration

A competitive island-based evolutionary algorithm for early-stage design exploration

Likai Wang[1]
[1] *Xi'an Jiaotong - Liverpool University*
[1] *wang.likai@outlook.com, 0000-0003-4054-649X*

Abstract. Evolutionary optimization has recognized potential for early-stage architectural design but is typically constrained to exploring variations of a single building typology and design scheme during an optimization process. This limitation hinders the comparative analysis of different design schemes, which is crucial in early design phases. This paper presents the Competing Steady-State Island Evolutionary Algorithm (C-SSIEA), which integrates a competitive evolutionary strategy with an island-based model. By leveraging the two strategies, C-SSIEA enables the parallel optimization of multiple distinct design schemes within a single optimization process, maintaining their diversity for comparative analysis rather than merely selecting the optimal one. A case study of a multi-functional building is presented to demonstrate its application. Results show how C-SSIEA can effectively reveal performance trade-offs, such as a high-rise tower excelling in contextual sky exposure but underperforming in solar heat gain balancing, which provides comprehensive insights to support designers in navigating competing design criteria and schemes from the outset of the design process.

Keywords. Evolutionary Optimization, Design Exploration, Early-Stage Architectural Design, Performance-Based Design

1. Introduction

During the early stages of architectural design, architects often investigate and compare multiple competing design schemes to maintain parallel thinking and pinpoint promising directions for further development (Cross, 2023; Harding & Shepherd, 2017). However, this process is constrained by the considerable time and effort required for both design creation and multi-faceted assessment, often leading to a limited exploration of the design space and potentially suboptimal outcomes. Evolutionary optimization has emerged as a powerful tool to augment this phase, automating the generation and evaluation of numerous design variants to help architects explore a large design space efficiently. By integrating performance simulations for daylighting or energy consumption, it incorporates quantitative environmental criteria directly into early design exploration and decision-making,

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 339-348.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

providing insights into performance trade-offs.

Despite its potential, a significant limitation persists in conventional applications of evolutionary optimization, in which the optimization workflow often operates on a single parametric model representing one specific design scheme, such as a twisted tower or a central courtyard layout (Bailey & Caldas, 2023; Estrado et al., 2023). This confines the optimization output to a family of variants sharing the same underlying spatial configuration and design typology. As a result, in order to support the comparison of different schemes at early design stages, it entails multiple independent optimization runs incorporating with different parametric models encoding a distinct design scheme (Venus et al., 2025; Wang et al., 2023). This approach is inefficient and also complicates data management, further hindering its integration into the fast-paced design process at early design stages.

To address this gap and better support parallel exploration and comparative analysis, this study presents an enhanced implementation that integrates a competitive evolutionary strategy (Janssen et al., 2022) with a steady-state island-based evolutionary algorithm (SSIEA) that is currently embedded in the EvoMass plugin for Rhino-Grasshopper (Wang et al., 2020). This extended version of SSIEA is referred to as the Competitive Steady-State Island Evolutionary Algorithm (C-SSIEA). The use of the competitive evolutionary strategy enables the simultaneous evolution of several distinct design schemes within a single, consolidated optimization process. At the same time, the island-based population structure maintains separate subpopulations for each scheme, preventing the premature loss of underperforming yet potentially valuable schemes. This combination facilitates a more systemic and comparative analysis alongside performance optimization, aligning the computational process more closely with the exploratory needs of early architectural design.

The remainder of this paper is structured as follows. After this introduction, a background section is presented to review related work and progress. The main body of the paper then details the implementation of C-SSIEA and demonstrates its application through a case study. The paper concludes by discussing the relevance of this development for the application of computational design in architecture.

2. Background

Evolutionary optimization offers significant potential for the early, conceptual stages of architectural design, helping architects navigate complex, ill-defined problems. This process begins with parametric modelling, where the designer defines a search space by encoding algorithmic logic. A critical challenge, however, is that most parametric models are inherently built upon specific, pre-defined spatial configurations, such as a courtyard or a streamlining massing (Bailey & Caldas, 2023; Estrado et al., 2023). This results in a low level of algorithmic abstraction of the design generative rules, which limits the typological diversity and variability of the generated designs. Consequently, a single optimization run is confined to exploring variations within one design typology, producing a family of solutions with a shared underlying configuration and ultimately restricting the scope of design discovery.

In response to this limitation and the growing role of optimization in design exploration, recent research has focused on strategies to expand the search scope for

more extensive comparative analysis. In many studies, incorporating multiple parametric models becomes a common approach to address this need. For example, studies have used different models to explore buildings with varying numbers of courtyards or to investigate massings with "fragmented," "monolithic," and "balanced" formal characteristics (Liu et al., 2020; Wang, 2022). While this approach expands the design space, it typically relies on multiple, independent optimization runs. This approach is inefficient and complicates direct comparison, as results are generated from separate evolved design populations. The root of this issue is technical: a standard evolutionary workflow is designed as a closed loop connecting a single parametric model, a simulation engine, and the evolutionary algorithm, which creates an inherent barrier to natively supporting the evolution of multiple competing models.

To overcome this obstacle, Janssen et al. proposed a method that integrates a competitive evolutionary strategy, allowing a single optimization process to evolve a diverse population from different parametric models (Janssen et al., 2022). With this competitive approach, designs from various schemes compete directly, resulting in the dominance of one scheme and the removal of others. Although this is an effective way to select design schemes, its competitive nature may not fully support the exploratory needs of early design. By focusing on selection rather than thorough comparison, the method risks losing valuable but underperforming schemes that could provide important trade-offs in design aspects not measured by the optimization metrics. Therefore, this warrants a method that enables both parallel optimization and continuous, lateral comparison of diverse design concepts.

3. Method

The development of the Competitive Steady-State Island Evolutionary Algorithm (C-SSIEA) addresses two key challenges: the difficulty of evolving multiple design schemes for comparative analysis and the limitations of existing competitive evolutionary strategies. This section details the evolutionary mechanism, workflow, and implementation of C-SSIEA within the Rhino-Grasshopper environment.

3.1. STEADY-STATE ISLAND-BASED EVOLUTIONARY ALGORITHM

To contextualize the development of C-SSIEA, it is pertinent to first outline its predecessor, the Steady-State Island-based Evolutionary Algorithm (SSIEA). SSIEA is a hybrid algorithm that integrates an island-based model with a steady-state replacement strategy (Wang et al., 2020). The island-based model divides the conventionally centralized design population into multiple, semi-isolated subpopulations. This structure counters the premature convergence to a local optimum often seen in algorithms using a single, centralized population. By evolving subpopulations in parallel, the algorithm can explore different regions of the design space simultaneously, reducing the likelihood of tracking by a local optimal solution and enhancing the diversity of the final design set (Figure 1).

However, using the island-based model alone can reduce search efficiency, as computational resources are divided among subpopulations. To mitigate this, SSIEA employs a steady-state replacement strategy instead of a generational one. When using the steady-state strategy, only a small number of low-fitness individuals are replaced

in each generation, allowing high-fitness designs to influence the population more immediately. While this strategy can potentially increase the risk of premature convergence, this effect can be counterbalanced in SSIEA by the diversity-preserving nature of the island-based model, as even if one subpopulation stagnates, others can continue to explore.

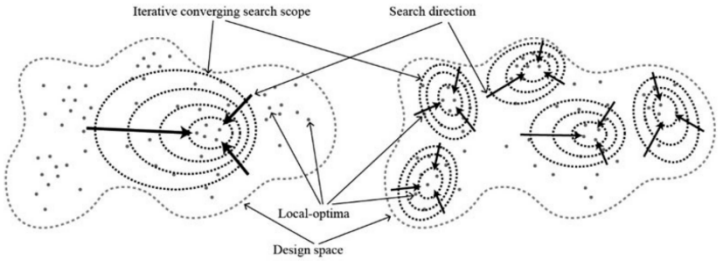


Figure 1. A comparative schematic of evolutionary optimization processes: a centralized population (left) versus an island-based model (right). Source: (Wang et al., 2020)

The current SSIEA is operated with a single parametric model for design optimization. Consequently, even though optimized designs from different subpopulations may exhibit configurational differences, they remain within the same design typology, as all are derived from the same underlying algorithmic logic. Despite this, the island-based model provides a structure that can be adapted to support the co-evolution of designs from multiple parametric models, enabling a more effective parallel optimization and comparative analysis.

3.2. COMPETITIVE EVOLUTIONARY STRATEGY

The competitive evolutionary strategy is designed to facilitate the comparative exploration of different design schemes for early-stage design. Unlike conventional evolutionary algorithms, it allows a design population to consist of individuals generated from different parametric models. A primary technical challenge this strategy has to overcome is the heterogeneous genotypical representations of these designs, which typically use different numbers and types of parameters. Consequently, conventional crossover operations, which recombine genotypes from two parent designs, are incompatible with this approach. The reproduction process, therefore, relies solely on mutation, a similar approach to those in genetic programming.

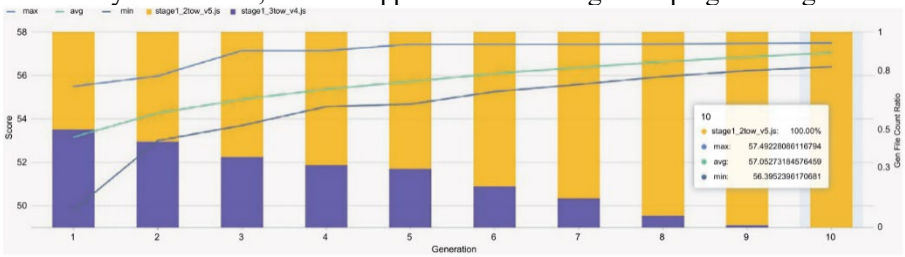


Figure 2. Evolutionary trajectory under a competitive strategy, showing the dominance of one design scheme (yellow) and the extinction of another (purple) over successive generations.

In the previous implementation by Janssen et al. (2022), the typical single, centralized population was also used, where heterogeneous designs competed directly. This leads to the rapid elimination of less-fit individuals, making the process efficient for selecting an optimal design scheme (Figure 2). However, this "winner-takes-all" approach often filters out initially inferior but potentially promising design schemes too early. As a result, significant trade-off information is lost, which is valuable for the synthesis and comparative analysis characteristic of early-stage architectural design.

3.3. IMPLEMENTATION

C-SSIEA is implemented by extending the existing SSIEA component within the EvoMass plugin in the Rhino-Grasshopper environment, sharing its user interface and connection methods. An additional checkbox in the interface allows designers to activate the C-SSIEA mode. When activated, the component automatically configures the number of subpopulations to match the number of connected generative components (Figure 3).

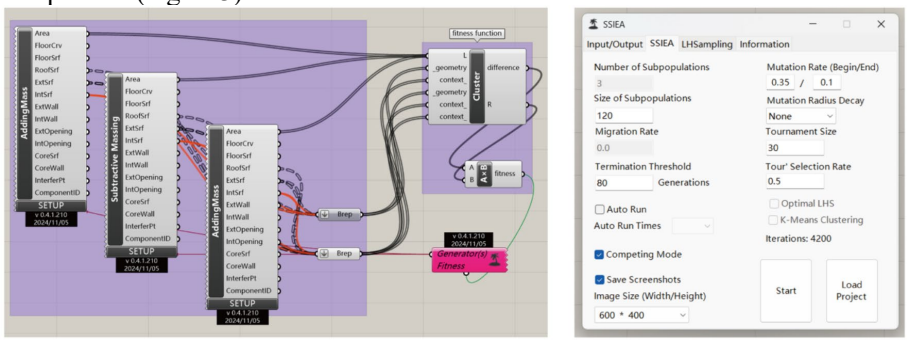


Figure 3. Grasshopper definition with C-SSIEA and multiple generative components (left) and user-interface panel of C-SSIEA (right)

While there are only two generative component types (subtractive and additive) offered in the EvoMass plugin, the C-SSIEA mode allows multiple instances of these two components to be connected. Each can be configured with a unique set of user-defined parameters to control geometrical characteristics, enabling the generation of distinct design schemes. For example, multiple additive components can be used, each parameterized to produce massing designs with "solid," "fragmented," or "balanced" formal characteristics using different sets of user-defined parameter configurations.

During an optimization driven by C-SSIEA, the process iterates through the connected parametric models to generate and evolve designs. The evolutionary process begins by generating the initial population for the first subpopulation using its associated generative component, while others are temporarily disabled. It then proceeds sequentially through the second, third, and subsequent subpopulations. After all subpopulations are processed, the cycle repeats with evolutionary operations, including selection, crossover, and mutation. This iterative process continues until a termination threshold is met, at which point the final optimized designs from all subpopulations are outputted for analysis. For analysis, two interfaces are provided to visualize the optimization progress and performance correlations, as demonstrated in

the subsequent case study (Figures 5 and 7).

4. Case Study

In order to demonstrate the efficacy of C-SSIEA for early-stage architectural design, a case study is presented. This case study describes the design of a multi-functional complex on a university campus in Suzhou, China. Two optimization runs were conducted using different environmental performance objectives to showcase how C-SSIEA reveals distinct design implications under varying criteria. For the two optimizations, all parameters for design generation and optimization remained consistent, with only the performance metrics and objective changed.

4.1. OPTIMIZATION SETUP

Three design schemes are defined using the two generative components in EvoMass, as illustrated in Figure 4. Schemes 1 and 2 utilize the additive component to create mid-rise buildings differentiated by their geometrical complexity through varying numbers and dimensions of aggregated masses. Scheme 3 employs the subtractive component to generate a high-rise tower that occupies a smaller footprint covering only half of the site due to its verticality. The target Gross Floor Area of 15,000 m² was maintained using a soft constraint, which imposed escalating penalties for significant deviations.

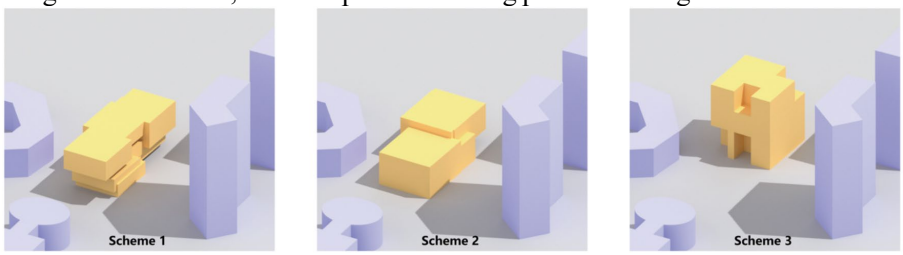


Figure 4. Example designs (randomly generated) for the three design schemes

The configuration of the evolutionary algorithm comprised three subpopulations, each corresponding to a specific design scheme. The optimization process employs a subpopulation size of 120 individuals, iterated over 80 generations. Selection is conducted via tournament selection with a size of 30 and a selection rate of 0.5, leading to a total of 4,200 iterations of design generation and evaluations per optimization run.

4.2. BUILDING-CENTRIC SOLAR PERFORMANCE OPTIMIZATION

The first optimization focused on the solar heat gain balancing of the building, which is critical for Suzhou's hot-summer-cold-winter climate. The objective is to minimize the difference between summer and winter solar irradiance values to achieve balanced annual thermal performance. With this objective, the design is aimed at reducing cooling loads in summer while enhancing passive heating in winter, within a design site complicated by surrounding high-rise and mid-rise buildings that create complex solar obstruction patterns.

Results presented in Figure 5 show that Scheme 1 significantly outperformed the

other two schemes. The performance difference between Schemes 2 and 3 is marginal, with Scheme 2 holding a slight advantage. The parallel plot reveals that both Schemes 1 and 2 achieve lower summer solar irradiance, with Scheme 1 being more effective. In contrast, the taller building designs of Scheme 3 receive higher irradiance in both seasons but achieve performance gains primarily through increased winter solar heat gain. Moreover, it can be noted that designs of Schemes 1 and 2 have higher variation of GFA, which make them easier to generate designs with lower GFA penalty.

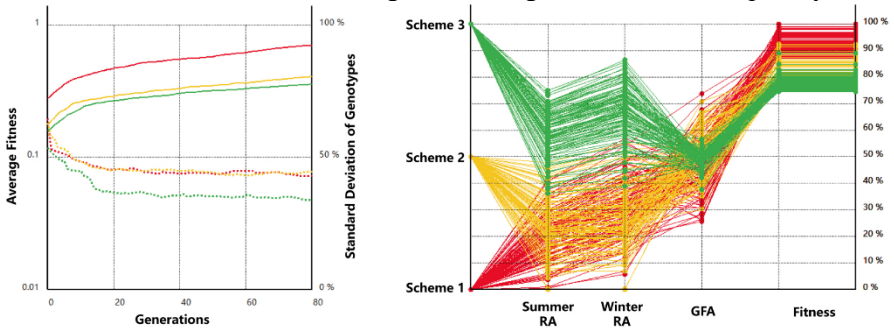


Figure 5. Progress trendlines (left) and parallel plots (right) for the three subpopulations (Red: Scheme 1, Yellow: Scheme 2, Green: Scheme 3)

The optimized designs in Figure 6 reveal distinct formal strategies. Designs of Scheme 1 consistently feature stepped volumes and cantilevered masses that enhance self-shading, particularly on the west side, which is less obstructed by adjacent buildings. Designs of Scheme 2 show stepped volumes that may aid winter solar access, but limited cantilever formation due to larger, fewer aggregated blocks. Designs of Scheme 3 incorporate recessed voids at lower levels for self-shading, though extensive vertical envelope exposure results in the lowest overall performance for this objective.

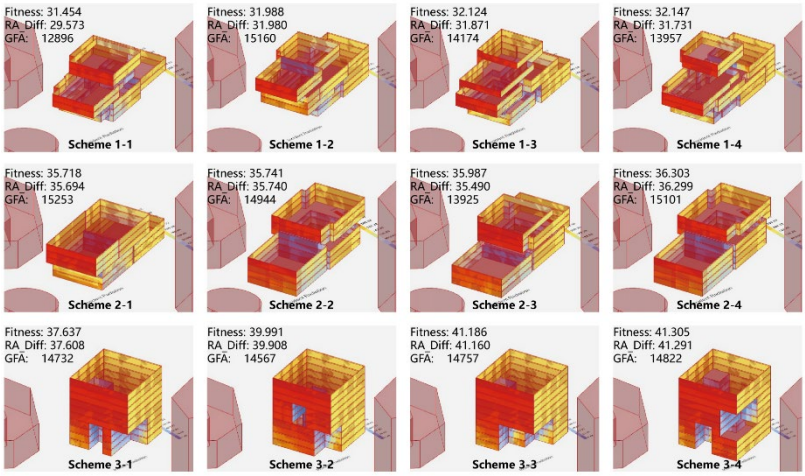


Figure 6. Optimized designs from the three design schemes

4.3. CONTEXT-CENTRIC SKY EXPOSURE OPTIMIZATION

The second optimization examines the building's impact on its immediate environment by maximizing sky exposure in adjacent outdoor spaces, a critical factor for perceptual comfort and urban heat island mitigation.

Results in Figure 7 show a reversal from the first optimization. Designs of Scheme 3, previously the worst performer, achieve superior performance by a significant margin, while designs of Scheme 1 moderately outperform those of Scheme 2. The tower typology of Scheme 3 inherently creates a smaller ground contact area, minimizing sky obstruction for the surrounding area. In addition, the parallel plot also indicates that in this optimization, the GFA variation of the designs associated with all three design schemes are similar, which suggests that the design variability during the optimization process is relatively limited.

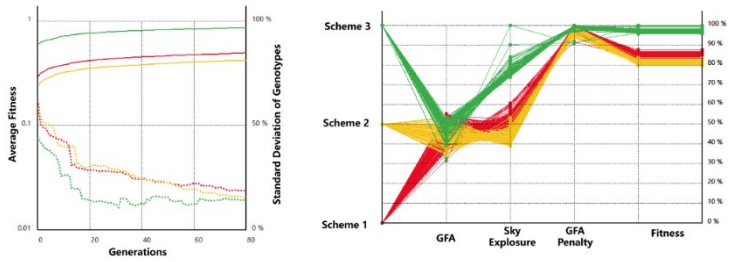


Figure 7. Progress trendlines (left) and parallel plots (right) for the three subpopulations (Red: Scheme 1, Yellow: Scheme 2, Green: Scheme 3)

The optimized designs in Figure 8 show Schemes 1 and 2 using cantilevered masses to decrease the ground contact area, with Scheme 1's greater formal flexibility providing its marginal advantage over Scheme 2. In comparison, the designs of Scheme 3 are more effective at creating a building with a smaller contact ground area. Moreover, the designs of Scheme 3 also leverage the stilt structure at the lower level of the building, which further reduces the sky obstruction for the outdoor ground.

4.4. SYNTHESIS AND DESIGN IMPLICATIONS

The two optimizations collectively demonstrate C-SSIEA's capacity to reveal critical performance trade-offs and inform early-stage design decision-making. The most salient finding is the fundamental conflict between the two environmental objectives for the given design schemes. The high-rise typology of Scheme 3 turned out to be a specialist: it was the clear winner in maximizing urban sky exposure but the clear loser for balancing solar heat gain. Conversely, the complex mid-rise massing of Scheme 1 demonstrated robust performance for solar balancing and moderate capability for sky exposure, positioning it as a more balanced, generalist solution. The consistent underperformance of Scheme 2 across both objectives strongly suggests its infeasibility for further consideration under these specific performance criteria.

Beyond selecting a single optimal scheme, C-SSIEA enables a more nuanced, multi-criteria exploration. Designers are not forced to choose a scheme blindly but can understand the inherent compromises each typology presents. For instance, if a project brief prioritizes urban environmental quality, Scheme 3 becomes a compelling option, and the optimization reveals that its key feature of verticality is the direct cause of both its strength and its weakness. This moves the decision-making process from a singular focus on a performance score to a more informed negotiation of design priorities. The methodology thus supports a reflective design practice where computational exploration discloses trade-offs, allowing architects to make strategic decisions aligned with specific project values and constraints.

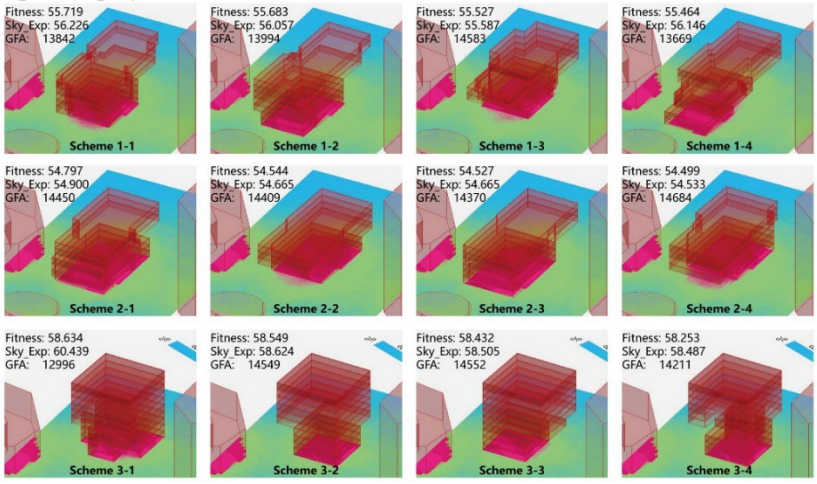


Figure 8. Optimized designs from the three design schemes

5. Discussion and Conclusion

The case study demonstrates that C-SSIEA effectively enhances early-stage architectural design exploration by enabling parallel optimization of multiple competing design schemes. The synergy between the competitive evolutionary strategy and the island-based model creates a new framework for evolving designs from different parametric models simultaneously. Integrated with EvoMass, this approach provides a streamlined exploration method for pre-design phases, allowing designers to efficiently define and compare various building massing schemes that can reflect different design concepts and prototypes. This capability is particularly valuable for addressing the data-poor and ill-defined nature of initial design processes, supporting more informed decision-making from the project outset.

The development of C-SSIEA represents an important step in the continuous evolution of EvoMass development, marking a direction apart from the previous upgrades that primarily focused on enhancing design generation capabilities. In addition, in comparison with other similar studies, C-SSIEA emphasizes a designerly application of evolutionary computation, unlike conventional approaches that typically prioritize algorithmic efficiency and optimization speed. The development of C-SSIEA reflects the priority of the quality of information provision and exploration support over

mere computational performance, addressing a crucial gap in current computational design research where exploration quality is often overshadowed by efficiency concerns.

5.1. CONCLUSION

This paper presents the development of the Competitive Steady-State Island Evolutionary Algorithm (C-SSIEA), which integrates competitive evolutionary strategies with an island-based optimization framework. This integration enables the parallel evolution of designs generated from multiple parametric models, each representing distinct architectural design schemes. The method provides an effective tool for comparative analysis of competing design concepts, significantly streamlining the application of evolutionary optimization in early-stage architectural design.

The case study illustrates C-SSIEA's potential through preliminary implementation, though it currently focuses on single-performance metric optimizations. Future research will explore the extension of this parallel comparative framework to multi-objective optimization scenarios, investigating how C-SSIEA can handle competing performance metrics while maintaining its capacity for diverse design exploration. This direction promises to further enhance the utility of the method in addressing the complex, multifaceted challenges characteristic of architectural design.

References

Bailey, E. T., & Caldas, L. (2023). Operative generative design using non-dominated sorting genetic algorithm II (NSGA-II). Automation in Construction, 155(July).

Cross, N. (2023). Design thinking: Understanding how designers think and work. Bloomsbury Publishing.

Estrado, E., Turrin, M., & Eigenraam, P. (2023). Optimization of complex-geometry high-rise buildings based on wind load analysis. SIMULATION, 99(11), 1133–1146.

Harding, J. E., & Shepherd, P. (2017). Meta-Parametric Design. Design Studies, 52, 73–95.

Janssen, P., Bui, T. D. P., & Wang, L. (2022). Möbius evolver: Competitive exploration of urban massing strategies. In I. As, P. Basu, & P. B. T.-A. I. in U. P. and D. Talwar (Eds.), Artificial Intelligence in Urban Planning and Design (pp. 293–321). Elsevier.

Liu, X., Wang, L., & Ji, G. (2020). Optimization Approaches in Performance-Based Architectural Design. 1, 599–608.

Venus, V., Wang, L., Tao, L., & Xiang, C. (2025). Performance-Based Building Design for Solar Energy Production - A rooftop-PV-integrated building design generation and optimization workflow. ARCHITECTURAL INFORMATICS, Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 3, 173–182.

Wang, L. (2022). Optimization-aided design: two approaches for reflective exploration of design search space. International Journal of Architectural Computing, 20(4), 758–776.

Wang, L., Janssen, P., Bui, T. D. P., & Chen, K. W. (2023). Comparing Design Strategies: A System for Optimization-Based Design Exploration. HUMAN-CENTRIC, Proceedings of the 28th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), 1, 221–230.

Wang, L., Janssen, P., & Ji, G. (2020). SSIEA: A hybrid evolutionary algorithm for supporting conceptual architectural design. Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM. https://doi.org/10.1017/S0890060420000281

Integrated Acoustic Geometry

Towards Multi-Objective Design and Robotic Fabrication of Acoustic Wall Systems

Jingyang Liu[1], Jiangyang Zhao[2] and Yang Song[3]
[1,2,3]*Xi'an Jiaotong - Liverpool University.*
[1]*Jingyang.Liu@xjtlu.edu.cn, 0000-0002-2671-4015*
[2]*Jiangyang.Zhao@xjtlu.edu.cn*
[3]*Yang.Song@xjtlu.edu.cn*

Abstract. This paper introduces a computational framework for designing architectural wall systems in which acoustic performance emerges directly from geometric and tectonic variation. Instead of adding acoustics after form-finding, the workflow integrates acoustic behaviour from the outset through geometric parameterization, ray-based simulation, structural assessment, and robotic fabrication. The design space couples global wall curvature with local block rotation, which enables control over diffusion, reflection, and absorption. These variables are evaluated using acoustic analysis and discrete-element structural modelling, with trade-offs explored through multi-objective optimization. Representative solutions show how distinct geometric families yield different acoustic and structural characteristics. A full-scale prototype, assembled via a six-axis robotic pick-and-place process, demonstrates the constructability of the optimized designs. Overall, the study presents a unified workflow that connects acoustic design, geometric generation, structural stability, and robotic assembly, with the aim of creating a practical approach for developing acoustically performative wall systems.

Keywords. Acoustic design, Multi-Objective optimization, Robotic fabrication, Computational geometry

1. Introduction

Acoustic performance is an important yet often under-integrated dimension of architectural design. While the geometry of a surface can shape how sound behaves, whether it is diffused, reflected, or absorbed, acoustic considerations are usually addressed only after a form has already been established. As a result, acoustic treatments are commonly applied as secondary layers rather than embedded into the tectonic and geometric logic of the wall itself.

This paper proposes a different approach. We introduce a computational

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 349-356.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

framework that integrates acoustic behaviour directly into the design, simulation, and fabrication of architectural wall systems. The framework links geometric variation, acoustic performance, structural stability, and robotic assembly into a single workflow, allowing designers to intentionally shape how sound interacts with the built environment.

At the core of the method is a multi-objective optimization process that explores variations in both global wall curvature and local block rotation. These geometric parameters collectively influence key acoustic properties such as diffusion, reflection, and absorption. By evaluating each configuration through acoustic and structural simulations, the framework identifies solutions that balance performance with material and tectonic feasibility.

To demonstrate constructability, we fabricate a full-scale prototype using a six-axis robotic pick-and-place system. This step confirms that optimized geometries can be produced with robotic precision, ensuring that the acoustic effects predicted in simulation carry through to physical production.

Together, the computational workflow and robotic fabrication pipeline reposition acoustic performance as an integral design driver rather than an afterthought. The contribution of this research lies in demonstrating how acoustic tuning, structural reasoning, geometry generation, and robotic assembly can operate simultaneously within a coherent design-to-fabrication process, enabling architects to create acoustically performative wall systems tightly coupled with their geometric and material logic.

2. Related Work

Advances in computational simulation and digital fabrication have significantly expanded the design space for acoustically performative surfaces in architecture. Early explorations demonstrated how parametric modelling could be used to shape wall geometries that influence sound behaviour, showing that surface articulation and materiality have measurable acoustic impact (Peters, 2009). More recent work has deepened this relationship by integrating simulation directly into the design workflow, allowing designers to evaluate acoustic scattering, absorption, and reflection while simultaneously exploring formal variation (Dessi-Olive & Hsu, 2022; Peters & Olesen, 2010). These studies emphasize how iterative, feedback-driven modelling can couple formal or geometric exploration with acoustic performance, demonstrating that computational tools can meaningfully support design decisions across multiple scales (Peters et al., 2019).

Parallel developments in robotic fabrication have enabled increasingly complex surface geometries to be materialized with high precision. Robotic assembly has been shown to facilitate differentiated acoustic behaviour through fine-grained control of block orientation, curvature, and surface texture (Savage et al., 2021). Such workflows highlight how automation can produce wall systems that would be infeasible to fabricate manually, while also ensuring consistency between simulated and built acoustic performance (Area & Hilmer, n.d.). Moreover, the integration of robotic pick-and-place systems, customized masonry units, and mass-customized acoustical components illustrates how robotic workflows can support the creation of adaptive and

highly responsive architectural surfaces (Williams et al., 2015).

Building on this foundation, the present study extends research in two directions. First, it couples acoustic simulation with structural assessment through a multi-objective optimization framework, enabling designers to explore trade-offs among absorption, diffusion, reflection, and stability within a single design space. Second, it demonstrates how robotic fabrication can serve as an end-to-end pipeline to construct optimized surfaces at full scale. Together, these contributions advance an integrated approach to acoustic wall design, one that links geometric variation, acoustic performance, and constructability within a unified computational workflow.

3. Method

Designing an acoustic wall requires balancing multiple performance criteria that often pull in different directions. Surfaces that maximize reflection can enrich spatial resonance, while surfaces optimized for absorption can improve clarity and reduce reverberation. Similarly, structural stability may conflict with desired acoustic effects, particularly when complex curvature or rotated modules are introduced. To manage these competing goals, we develop a workflow that integrates design space definition, performance simulation, multi-objective optimization, and robotic fabrication.

Our method consists of three phases, including (1) Defining the geometric design variables (2) Evaluating acoustic and structural performance and (3) Searching the design space through multi-objective optimization.

3.1. DESIGNING THE GEOMETRIC VARIABLES

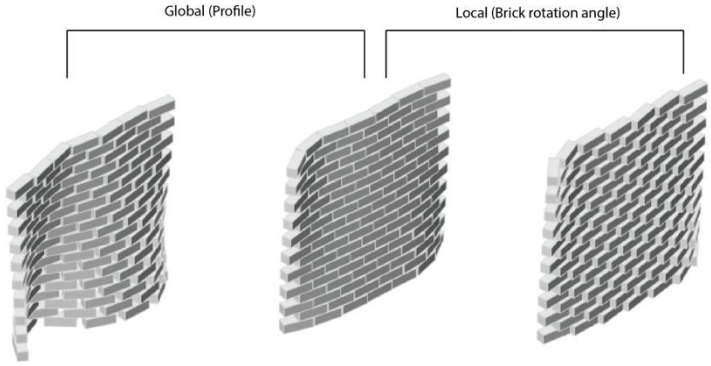


Figure 1. The wall system is governed by two scales of geometric manipulation: (left) global curvature defined by the Bézier-based wall profile, which shapes large-scale acoustic behavior through convex, concave, or hybrid forms; and (right) local brick-level rotation, which fine-tunes directional scattering and diffusion by adjusting the orientation of individual units.

The geometric definition of the wall operates at two complementary scales (Figure 1). At the global level, the overall wall profile is shaped by two quadratic Bézier curves,

one at the base and one at the top, each defined by five control points. By adjusting these curves, the designer can generate a wide range of acoustic behaviours: convex profiles that promote broad scattering, concave profiles that concentrate energy into focal regions, and hybrid profiles that combine focusing and dispersion. At the local level, each block within the wall is assigned an independent rotation angle, allowing fine-grained modulation of how sound interacts with the surface. Larger rotations intensify scattering by redirecting reflected rays, whereas smaller rotations produce smoother and more uniform reflections. Together, the Bézier-curve profiles and block-level rotations form the complete set of design variables, which provide a multi-scalar means of tuning both the formal expression and the acoustic performance of the wall system.

3.2. ACOUSTIC OBJECTIVE FUNCTIONS

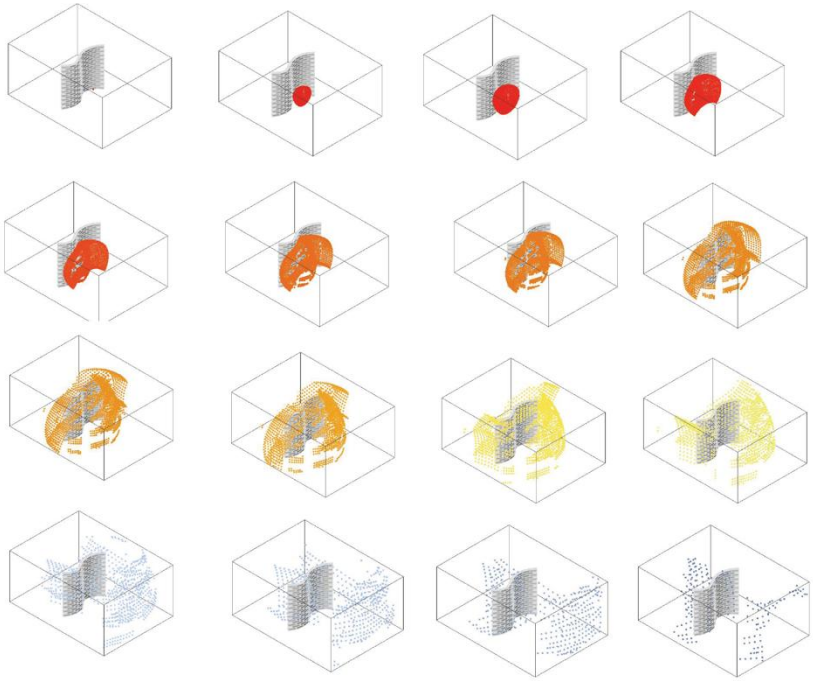


Figure 2. Acoustic ray-tracing analysis of the wall system. Sequential ray-tracing simulations show how the wall modulates sound propagation through absorption, redirection, and scattering. Varying local rotations and global curvature generate distinct dispersion fields, demonstrating the system's ability to regulate acoustic behaviour across frequency bands and spatial configurations.

To assess the performance of each design configuration, we simulate how sound rays interact with the wall and evaluate their behaviour at one or more receiver locations (Figure 2). The acoustic analysis considers three complementary criteria, including

absorption, reflection, and diffusion, each capturing a different aspect of how the wall shapes the aural environment. Absorption is quantified by comparing the energy received in a free-field condition with the energy measured when the wall is introduced between source and receiver. This allows us to determine the proportion of incident sound that is dissipated by the system, a property that directly influences reverberation and clarity in spaces such as studios or lecture rooms. Reflection, in contrast, is defined as the proportion of incident energy that is returned to the receiver after interacting with the surface. Surfaces with higher reflection coefficients are particularly relevant in performance environments where resonance and tonal richness are desired.

Diffusion measures the uniformity with which sound is scattered across different angles around the receiver. An ideal diffusive surface distributes reflected energy evenly; therefore, we evaluate diffusion by analysing the variance of sound energy across a polar sampling array. Lower variance corresponds to more even scattering, reducing acoustic hotspots and improving spatial balance. Together, these three criteria capture a multi-dimensional understanding of acoustic performance.

Since the wall system is physically constructed from stacked blocks, structural integrity is evaluated in parallel with acoustic behaviour. We adopt a simplified discrete-element approach in which each block is treated as a rigid unit and support relationships between blocks are represented as a directed graph. Blocks lacking a supporting neighbour beneath them are identified as structurally invalid. Two indicators are used to assess overall stability: the alignment and overlap between adjacent layers, which influences the continuity of load paths, and the global centre of gravity of the assembly relative to the base footprint, which affects resistance to overturning. Designs that achieve strong local alignment and maintain a compact global centre of gravity are rewarded for structural soundness. By evaluating acoustic and structural criteria together, the workflow ensures that resulting configurations are both performative and feasible for full-scale construction.

3.3. MULTI-OBJECTIVE OPTIMIZATION

We employ the NSGA-II genetic algorithm to explore the multi-dimensional trade space defined by the four acoustic and structural objectives. Using a population of 200 individuals evolved over 100 generations (Table 1), the algorithm samples a wide design space while preserving convergence toward high-performing regions. The resulting Pareto front reveals several distinct families of solutions that show the relationships between geometry and performance. Designs that prioritize absorption tend to converge toward flatter global profiles with relatively small block rotations, which reduce reflective energy and promote dissipation. In contrast, highly diffusive solutions frequently emerge from convex wall geometries combined with stronger local rotations, producing richer scattering effects. Configurations optimized for reflection often favour concave profiles that focus sound energy toward the receiver. Structural stability, meanwhile, is associated with configurations exhibiting lower rotation angles and increased overlap between stacked blocks, which support more continuous load paths. Taken together, these clusters of solutions offer designers a spectrum of viable configurations that reflect different acoustic intentions. Through navigating the design space, designers can select wall geometries tailored to the specific

needs of diverse spatial typologies (Figure 3).

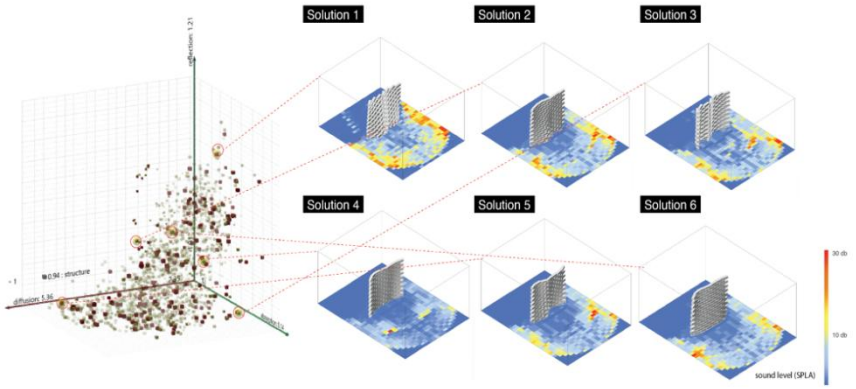


Figure 3. Multi-objective optimization results showing representative solutions from the Pareto set, mapped against key performance metrics (absorption, diffusion, reflection, structural stability)

Population Size	*Generation Size*	*Generation Count*	*Mutation Probability*	*Mutation Rate*	*Crossover Rate*
200	10	100	0.20	0.90	0.80

Table 1. NSGA-II parameter configuration for multi-objective optimization of wall geometry.

4. Result

To evaluate whether the optimized wall configurations can be constructed with the geometric fidelity required for acoustic performance, we produced a full-scale physical prototype using a six-axis UR10 robotic arm equipped with an electric parallel gripper (Figure 4). The prototype selected for fabrication corresponds to a solution on the Pareto front that prioritizes acoustic diffusion while remaining structurally stable. The robotic process was conducted through a pick-and-place workflow in which each block was retrieved from a predefined stock location, oriented according to its target rotation angle, and positioned along the wall's global profile.

A dedicated motion-planning routine was implemented to minimize toolpath length and avoid collisions, allowing the robot to assemble the full system through a continuous sequence of controlled placements. This level of precision is important for this project, as small deviations in geometry can significantly alter scattering behaviour and degrade acoustic performance.

The construction of the prototype confirms that the geometries generated through

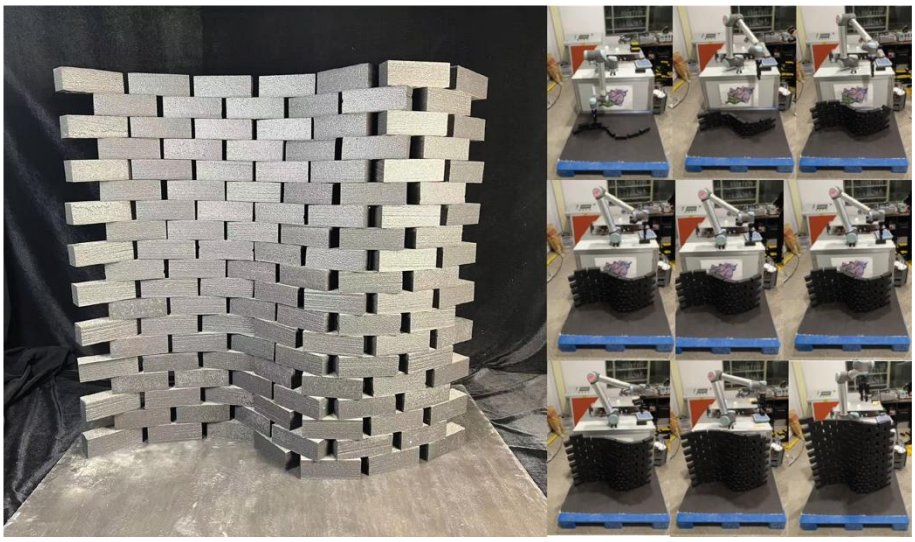


Figure 4. Robotic assembly of the optimized acoustic wall system.

multi-objective optimization are not only computationally valid but also materially and logistically feasible at 1:1 scale. The robotic fabrication results therefore provide evidence that the proposed workflow can translate optimized acoustic and structural configurations into real-world physical assemblies.

5. Conclusion

This study introduces an integrated computational framework that unifies geometric design, acoustic simulation, multi-objective optimization, and robotic fabrication for the development of acoustically performative wall systems. By embedding acoustic behaviour directly into the geometric and tectonic logic of the wall, rather than treating it as an additive layer, the framework enables designers to shape diffusion, reflection, and absorption through form itself. The inclusion of structural evaluation ensures that the resulting configurations balance performance with feasibility, which allows acoustic ambitions to be pursued without compromising stability.

The results demonstrate that the workflow can generate a diverse family of geometries attuned to different acoustic scenarios, from focused reflection to broad diffusion. The optimization process reveals the interconnected relationships between global profile, local block rotation, and multi-variable performance. The optimization gives designers actionable insight into how geometry mediates sound. The successful robotic assembly of a full-scale prototype validates the constructability of the optimized solutions. Taken together, these contributions position the framework as a robust tool for integrating acoustic reasoning into early-stage design while simultaneously bridging the gap between simulation and real-world construction.

Acknowledgements

This work was supported by the Research Development Fund (RDF-24-01-055) from Xi'an Jiaotong - Liverpool University.

References

Area, S., & Hilmer, J. (n.d.). Design-to-manufacture workflows of sound-scattering acoustic brick walls.

Dessi-Olive, J., & Hsu, T. (2022). A simulation-validated shape grammar for architectural acoustics. *Nexus Network Journal*, 24(1), 55–73. https://doi.org/10.1007/s00004-021-00569-0

Peters, B. (2009). Parametric acoustic surfaces. *In ACADIA 2009 Conference Proceedings* (pp. 174–181).

Peters, B., & Olesen, T. (2010). Integrating sound scattering measurements in the design of complex architectural surfaces. *In Proceedings of the 28th eCAADe Conference*.

Peters, B., Hoban, N., Yu, J., & Xian, Z. (2019). Improving meeting room acoustic performance through customized sound-scattering surfaces. Universitätsbibliothek der RWTH Aachen.

Savage, K., Hoban, N., & Peters, B. (2021). Gradient acoustic surfaces: The design, simulation, robotic fabrication of a prototype wall. In T. Fukuda, W. Huang, P. Janssen, K. Crolla, & S. Al-Mohannadi (Eds*.), Computer-Aided Architectural Design Futures 2021* (pp. 360–377). Springer. https://doi.org/10.1007/978-981-16-7116-0_21

Williams, N., Burry, J., Davis, D., Peters, B., De Leon, A. P., & Burry, M. (2015). FabPod: Designing with temporal flexibility and relationships to mass-customisation. Automation in Construction, 51, 124–131. https://doi.org/10.1016/j.autcon.2014.12.008

From Behavior to Office Design Insights

A Framework for Modeling Attractiveness Using Map-matched Trajectories and 3D Tag Information

Hiroyuki Hasada[1], Junya Maruyama[2], Ayumi Mukai[3], Hiroki Imai[4], Ryohei Takahashi[5], Tomoki Akiyama[6], Jingya Li[7], Hiroaki Goto[8], Shiro Takeshita[9] and Yudai Honma[10]
[1,10]*Institute of Industrial Science, The University of Tokyo.*
[2,3]*Graduate School of Engineering, The University of Tokyo.*
[4, 5, 6, 7, 8, 9]*Kozo Keikaku Engineering Inc.*
[1]*hasada@iis.u-tokyo.ac.jp, 0000-0001-7200-7322*
[2]*juntama0826@g.ecc.u-tokyo.ac.jp, 0000-0002-6703-3259*
[3]*m-ayumi@g.ecc.u-tokyo.ac.jp, 0009-0004-9561-8934*
[4]*hiroki-imai@kke.co.jp*
[5]*ryohei-takahashi@kke.co.jp*
[6]*tomoki-akiyama@kke.co.jp*
[7]*jingya-li@kke.co.jp*
[8]*hiroaki-goto@g.ecc.u-tokyo.ac.jp*
[9]*shiro-takeshita@kke.co.jp*
[10]*yudai@iis.u-tokyo.ac.jp, 0000-0002-6458-0767*

Abstract. Modern offices are expected to foster not only efficiency but also creativity, collaboration, and well-being. However, spatial layout decisions are still often guided by intuition rather than empirical behavioral evidence, leaving a gap in understanding how environmental design shapes workers' movement and space use. The aim of this study is to contribute a data-driven framework that links behavioral sensing with spatial information to quantify the attractiveness of office environments. The research addresses two questions: (1) Can we empirically estimate spatial attractiveness from observed behavior instead of relying on subjective judgments? (2) Can environmental attributes based on 3D scanning data explain attractiveness before and after layout modifications? Methodologically, workers' behavior captured via Bluetooth beacons are reconstructed into trajectories using a hidden Markov model–based map-matching process, and an inverse shortest path problem model is applied to infer link-specific attractiveness. This attractiveness is then related to 3D-tagged environmental features such as furniture, equipment, and passage widths through regression analysis. The results show that spatial openness, accessibility, and functional diversity are positively associated with attractiveness, providing interpretable evidence of how design elements influence movement. The proposed framework

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 357-366.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

contributes to human-centered, data-driven office design by translating behavioral data into actionable spatial insights.

Keywords. 3D Scanning Data, Behavioral Data, Office Design, Mathematical Optimization Model, Trajectory Map-matching

1. Introduction

1.1. BACKGROUND AND MOTIVATION

Modern offices are increasingly recognized as environments that foster not only productivity but also creativity, collaboration, and well-being. Spatial layouts strongly influence how workers move, meet, and communicate, shaping both organizational culture and daily work experiences (Yu 2023). However, design decisions are still often guided by intuition or precedent rather than objective behavioral evidence. There is an emerging need for data-driven design approaches that help practitioners understand how spatial layouts influence interaction, communication, and well-being in workplaces.

With recent advances in sensing and modeling, both behavioral and spatial data have become increasingly accessible in real office settings. Indoor positioning systems such as Bluetooth beacons, which transmit short-range signals for estimating indoor locations, allow continuous observation of workers' movements (Chen et al. 2025; Maruyama et al. 2023; Pan et al. 2023; Pollard et al. 2022), while 3D scanning technologies, which generate detailed spatial representations of interior environments, capture the geometry, furniture, and environmental features of office spaces with high precision. The research gap lies not only in collecting such data but in translating it into interpretable knowledge that supports design reasoning.

Existing approaches often focus on behavior or spatial form, such as in space syntax analysis (Hillier et al. 1976). These perspectives have rarely been integrated, leaving a gap in understanding how people respond to environments both structurally and experientially. A unified analytical framework is needed to link measured human behavior with environmental attributes to enable a more comprehensive interpretation of spatial attractiveness within office environments.

1.2. OBJECTIVE AND CONTRIBUTIONS

The objective of this study is to answer the following research questions:

(1) Can we empirically estimate the attractiveness of office spaces based on actual behaviors captured through beacon data, instead of relying on subjective judgments?

(2) Can environmental attributes based on 3D scanning data explain attractiveness before and after layout modifications?

To address them, we propose a mathematical framework that bridges indoor behavioral sensing and environmental modeling. The framework connects map-matched beacon trajectories with spatial attributes extracted from 3D-tagged environmental data, enabling the quantitative evaluation of location-specific attractiveness. It provides a means to translate workers' movement patterns into measurable indicators that reflect how design configurations influence spatial use.

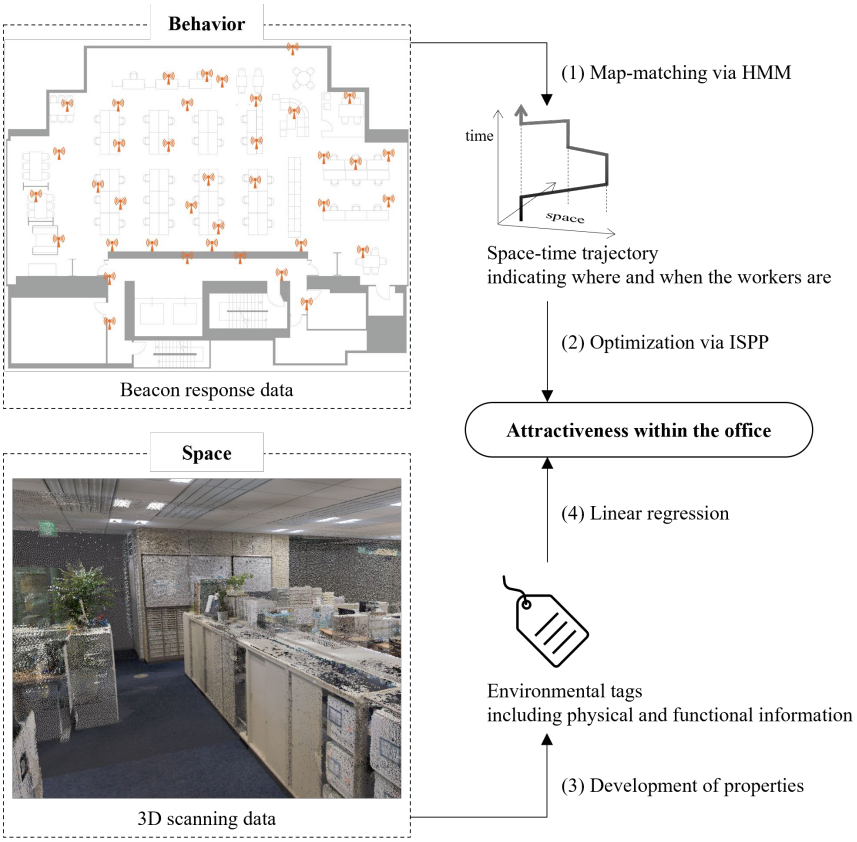


Figure 1. Proposed framework

The study establishes a foundation for human-centered, data-driven office design by transforming behavioral and spatial data into interpretable insights that support evidence-based decision-making. We demonstrate its validity through a real-world case study conducted before and after a layout modification. By comparing distributions of attractiveness across configurations, it shows how environmental changes reshape workers' behavior offering a practical analytical tool for improving office layouts.

2. Methodology

This section introduces a unified analytical framework for translating behavioral and spatial data into interpretable indicators of office attractiveness. The framework provides a quantitative structure that integrates behavioral sensing, mathematical modeling, and spatial data analysis. Specifically, it connects space–time trajectories reconstructed from beacon data with 3D-tagged environmental information, enabling the empirical estimation of how design elements correspond to office workers' spatial choices. By bridging those two data domains, the framework allows behavioral

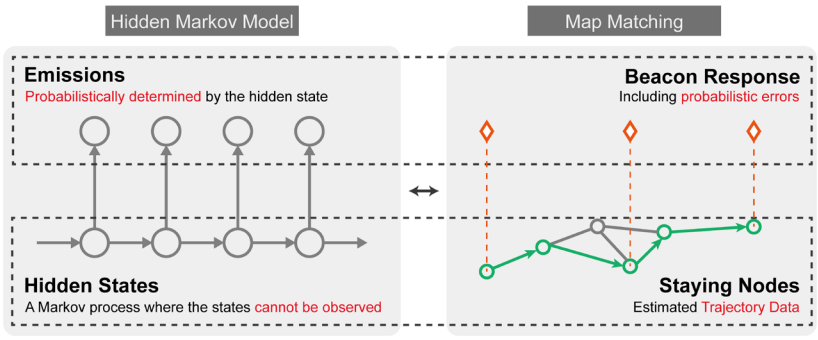


Figure 2. Concept of Map matching based on HMM

evidence to be translated into interpretable design knowledge.

As illustrated in Figure 1, the proposed framework consists of four main steps. (1) Map-matching of beacon data reconstructs continuous movement trajectories by aligning discrete signals with the most probable space-time paths through a Hidden Markov Model (HMM) that handles noise and missing readings. (2) An Inverse Shortest Path Problem (ISPP) model estimates link-specific attractiveness from the map-matched paths, by optimizing the relative tendency of workers to approach or avoid specific areas within the office. (3) Environmental attributes are developed from 3D-scanning data by tagging properties of furniture, equipment, and spatial characteristics such as width, or openness, categorized into physical and functional factors. (4) A regression model integrates the estimated attractiveness with these tagged environmental attributes to identify which spatial factors significantly influence behavioral use.

2.1. MAP-MATCHING VIA HMM

We represent continuous movement trajectories as transitions on a network, where the target office floor is represented by nodes placed at desk positions and additional nodes arranged along corridors so that the average distance between adjacent nodes is uniform (Maruyama et al. 2023). This enables detailed representation of trajectories within the indoor space. For areas used for specific purposes, such as meeting rooms, representative nodes are placed to clearly capture how these spaces are utilized.

By performing HMM–based map-matching (Zheng et al. 2018), we infer trajectories from raw beacon responses (Figure 2). The raw data is not only contaminated with noise but also include intervals with no responses despite the system being configured to communicate at 1 Hz. The role of map matching is to reconstruct a smooth trajectory from such sparse and noisy beacon observations. HMM-based map-matching fundamentally hinges on specifying transition probability $P_{\text{trans}}(x_t \mid x_{t-1})$ and emission probability $P_{\text{emis}}(y_t \mid x_t)$. In this work, we define these probability functions on the network as follows. Let $i, j \in I$ denote node IDs and $b \in B$ a beacon ID; let $d_{\text{trans}}(i,j)$ be the network distance between nodes i and j, and $d_{\text{emis}}(i,b)$ the direct distance between node i and beacon b:

$$P_{\text{trans}}(x_t \mid x_{t-1}) = \exp[-\lambda d_{\text{trans}}(x_{t-1}, x_t)^2] / \sum_{i \in I} \exp[-\lambda d_{\text{trans}}(x_{t-1}, i)^2] \quad (1)$$

$$P_{\text{emis}}(y_t \mid x_t) = \exp[-\mu d_{\text{emis}}(x_t, y_t)^2] / \sum_{b \in B} \exp[-\mu d_{\text{emis}}(x_t, b)^2] \quad (2)$$

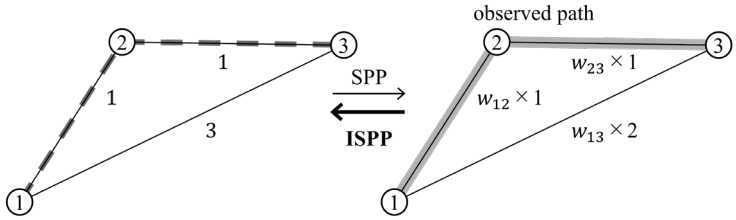


Figure 3. Concept of ISPP

2.2. OPTIMIZATION VIA ISPP

Using the workers' map-matched trajectories on the office network, ISPP is applied to estimate the attractiveness within the office. The ISPP is the inverse of the classical shortest path problem (SPP) (Burton and Toint 1992): while the SPP identifies the path with the minimum cost given link lengths, the ISPP optimizes the cost of each link, which is the product of its given length and unknown weight, that makes the observed paths the shortest (Figure 3). In this study, we assume that the workers' observed trajectories represent least-resistance paths through the office. By inferring link costs so that these trajectories become shortest under the estimated weights, ISPP quantifies the degree to which each area attracts or repels movement. The resulting link weights are interpreted as the disutility associated with traversing space, and their inverse provides a behavioral measure of spatial attractiveness. This inverse optimization enables the empirical estimation of attractiveness directly from observed behaviors without relying on prior assumptions.

The ISPP model proposed by Hasada and Honma (2025) optimizes link weights $\boldsymbol{w}$ by minimizing the following objective function that indicates the deviation between the total cost of an observed path $p_n(\boldsymbol{w})$ and that of its shortest alternative $q_n(\boldsymbol{w})$ on the space-time network for the worker n:

$$\text{min.} \quad \sum_n \big(p_n(\boldsymbol{w}) - q_n(\boldsymbol{w})\big) \tag{3}$$

The space-time network is developed from the office network so that it includes transition links connecting each pair of spatially adjacent nodes across consecutive time steps. The transition links that correspond to the same spatial location share a common cost across all time steps.

2.3. DEVELOPMENT OF PROPERTIES

We construct a Point of Interest (POI) tagging system for an office floor as part of the environmental semantics framework, using the NavVis IVION platform as the foundation. This 3D indoor mapping and visualization platform is widely used for geometric management in factories and is also well suited for acquiring 3D scanned data of office environments, while enabling operators to capture and annotate spatial information directly within immersive panoramic views. A POI refers to a discrete, spatially identifiable object, such as furniture, fixtures, or building elements. This tagging system specifies the physical and functional attributes of objects at the furniture and element level, allowing detailed spatial analysis beyond room- or zone-level abstraction.

Three operators manually create POIs for each tangible object visible in the office environment, jointly reviewing the panoramic imagery to ensure that objects are consistently and appropriately captured. Each POI is assigned the following standardized set of metadata to ensure consistency and interoperability across operators, reflecting that POI selection is jointly performed by multiple operators and subsequently reviewed by a supervisor for overall validity:

- Object ID: 4-digit code, consisting of operator ID (1 digit) and sequential number (3 digits).
- Category ID: 5-digit code, consisting of major category ID (2 digits)—four functional types, Partition, Acc. (Accessory), Fixture, MEP (Mechanical, Electrical, and Plumbing systems)—and minor category ID (3 digits)— specific object types, e.g., Desk, Partition, Lighting, Wall.
- Position: 3D coordinates extracted from NavVis IVION.
- Groundedness: 1 (on floor) or 0 (floating).
- WDH: width, depth, and height of objects, rounded to 0.5 m.
- Room Name: chosen from a fixed set (e.g., Office, Meeting Room, EV Hall).

The resulting POI dataset on NavVis IVION is exported as a CSV file, including all categorical and spatial metadata. This structured dataset is then linked directly with the office network.

2.4. LINEAR REGRESSION

We conduct a two-step regression analysis to identify both the structural determinants of spatial attractiveness and the factors explaining its variation following the layout modification. In the first step, we estimate a pooled regression model that integrates pre- and post-modification data, including a binary indicator distinguishing between the two configurations. This model reveals the baseline relationships between estimated attractiveness values and environmental attributes while controlling for overall shifts caused by the layout change. In the second step, we perform a difference in differences (DID) regression analysis focusing on the drivers of attractiveness change. For each spatial link, the change in estimated attractiveness between the two periods is regressed on the corresponding changes in explanatory variables such as tag attributes, spatial centralities such as betweenness centrality, and passage width. This procedure allows us to disentangle how modifications in environmental conditions contribute to increases or decreases in attractiveness, thereby decomposing the observed behavioral response into its underlying spatial components.

3. Results and Discussion

The analysis in this section builds on a layout modification introduced in the office of Kozo Keikaku Engineering Inc., an engineering consulting firm. Specifically, the planting area and snack box were relocated, and shelves were repositioned to widen several passages. These spatial interventions were intended to enhance local openness and flow continuity while providing experimental conditions for examining how changes in spatial affordances influence workers' behavior.

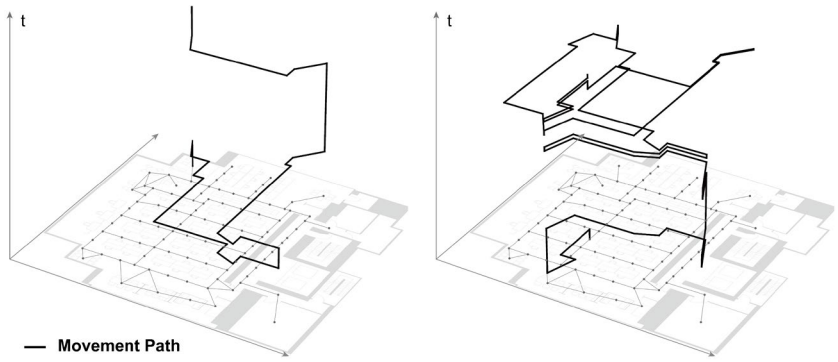


Figure 4. Examples of the estimated trajectories on the space-time network

Table 1. Summary statistics of the trajectory dataset

		Pre-modification	Post-modification
Target site	Location	Kozo Keikaku Engineering Inc., New Building, 5th Floor	
	Period	1–30 Oct. 2024 (Weekdays only)	1-15 Apr. 2025 (Weekdays only)
Beacon configuration	Transmitter count (Individual count)	40	40
	Receiver count	39	39
Map-matched trajectories	Data count	812	427
	Min. duration	7 s	1 h 20 m 59 s
	Max. duration	11 h 52 m 17 s	12h 1 m 15 s
	Avg. duration	11 h 13 m 48.1 s	11 h 23 m 31.4 s
	Med. duration	11 h 52 m 15 s	12 h 1 m 6 s
	Std. Dev. duration	1 h 49 m 4.6 s	1 h 15 m 57.1 s

3.1. MAP-MATCHING VIA HMM

Examples of the estimated results are shown in Figure 4. The x- and y-axes represent people's movements on the network, and the z-axis represents the passage of time. The details of the dataset are shown in Table 1. The parameters in Equation 1,2 of the HMM were determined based on a questionnaire survey conducted with the individuals, adopting $\lambda = 0.5$ and $\mu = 0.01$. This result priorities movement resistance over beacon proximity, suiting noisy Bluetooth environment. To reduce the burden on participants while ensuring sufficient sampling, the observation period was shortened to two weeks. The observation that most of data durations are around 12 hours simply reflects the 8:00 to 20:00 observation window.

3.2. OPTIMIZATION VIA ISPP

The ISPP model is applied using the map-matched trajectories as input to estimate link-

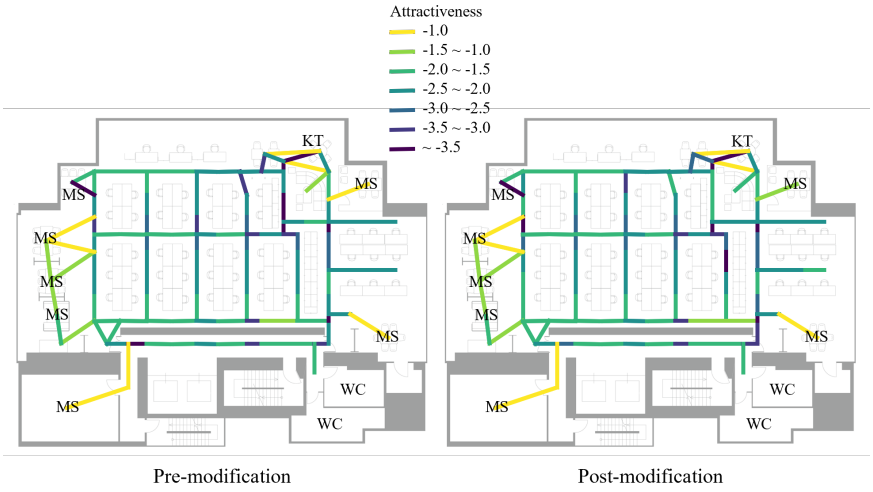
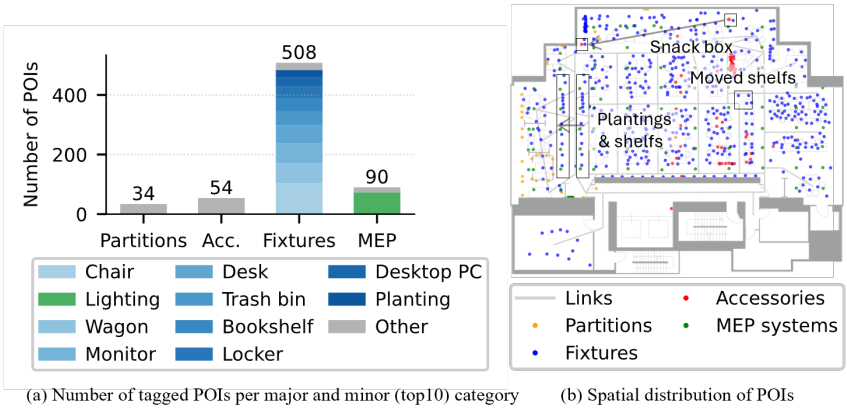


Figure 5. Estimated Attractiveness



(a) Number of tagged POIs per major and minor (top10) category (b) Spatial distribution of POIs

Figure 6. POI distribution before and after layout modification

specific attractiveness values on the network. To visualize more intuitively, the estimated link weights—which take positive values and represent spatial disutility—are sign-reversed and displayed as attractiveness values (Figure 5). In this representation, higher values indicate greater behavioral preference toward the corresponding spatial links.

The estimated attractiveness map reveals locally distinct patterns. The areas surrounding the meeting space (MS) exhibited relatively high attractiveness, suggesting that these spaces functioned as behavioral hubs. The kitchenette areas (KT) also showed higher attractiveness compared to adjacent corridors, reflecting their frequent use for informal breaks and interactions. WC is an abbreviation for water closet, meaning restroom. Overall, the spatial distribution patterns before and after the layout modification remain largely consistent, indicating that the fundamental behavioral structure of the office was stable across both configurations. A more

Table 2. Estimates of the pooled regression model

	Coef.	t-value
Const	-2.93	-6.16
Dummy for Post-modification	0.74	1.09
Passage Width	0.18	0.51
MS	0.38	1.10
Whiteboard Count within 0.5 m	0.28	0.82
Snackbox Count within 0.5 m	0.16	0.47
Plant Count within 0.5 m	0.18	0.53

Table 3. Estimates of the DID regression model

	Coef.	t-value
Const	0.74	1.12
Passage Width	-0.01	-0.02
Snackbox Count within 0.5 m	0.07	0.11
Plant Count within 0.5 m	-0.07	-0.10

1

detailed interpretation of subtle local changes in attractiveness will be discussed later through regression analysis.

3.3. DEVELOPMENT OF PROPERTIES

A total of 749 POIs is tagged across the office floor, covering both the pre- and post-modification layouts as shown in Figure 6. Most POIs are classified under Fixture and MEP Systems, reflecting the functional density of workspaces and building service infrastructure. Desk and partition elements exhibit high spatial frequency, capturing the layout structure of the open-plan office.

This dataset reflects a layout modification that was implemented between two observation periods (Figure 6). This structured tagging framework provides the semantic basis for behavioral analysis, linking spatial features with movement and utilization patterns.

3.4. LINEAR REGRESSION

To examine how the environmental attributes developed in the previous section relate to the estimated attractiveness, a linear regression analysis is conducted using variables preselected with reference to ridge regression to ensure stability under multicollinearity. The candidate explanatory variables are drawn from the tagged properties of the office environment, including physical features such as passage width and functional elements such as proximity to MS or the number of nearby objects within 0.5 m (Table 2). The pooled regression results show that all coefficients are positive, indicating that wider passages and areas containing whiteboards, plants, or snack boxes tend to have higher attractiveness. These patterns suggest that spaces combining openness and the presence of collaborative or comfort-related elements are

more likely to attract movement and short stays within the office.

The DID regression model (Table 3) yields very small t-values for all predictors, implying that the layout modification causes only marginal changes in these relationships. Given the limited explanatory power of this model, further interpretation is omitted here, although its predictive value is expected to improve as more behavioral data become available.

4. Conclusion

This study introduced a unified framework that combines behavioral and spatial data to estimate and interpret spatial attractiveness in offices. By linking map-matched trajectories from Bluetooth beacons with 3D-tagged environmental attributes, it quantitatively reveals how design elements shape workers' movement and spatial preferences, offering a data-driven basis for human-centered office design.

A limitation is that the analysis was conducted in a single office, restricting the generalizability of the findings. To address this, a follow-up experiment is underway at another building as of November 2025, applying the same framework to a similar layout modification. Multi-site results will help validate the method and advance predictive tools for evidence-based spatial planning.

References

Burton, D., & Toint, P. L. (1992). On an Instance of the Inverse Shortest Paths Problem. *Math. Program.*, 53, 45–61. https://doi.org/10.1007/BF01585693

Chen, S., Fukuda, T., & Yabuki, N. (2025). Development of an occupancy measurement system for micro-zones within open office spaces based on multi-view multi-person 3D pose estimation. *J. Build. Eng.*, 111(113037), 113037. https://doi.org/10.1016/j.jobe.2025.113037

Hasada, H., & Honma, Y. (2025). Non-parametric Inverse Estimations of Destination Preferences Based on Visitor Path Data. In *Proceedings of the Transportation Research Board 104th Annual Meeting (TRB 2025)* (TRBAM-25-01805).

Hillier, B., Leaman, A., Stansall, P., & Bedford, M. (1976). Space syntax. *Environ. Plann. B Plann. Des.*, 3(2), 147–185. https://doi.org/10.1068/b030147

Maruyama, J., Honma, Y., Nishiyama, Y., & Kawahara, Y. (2023). A Trajectory estimation method from spatially sparse and noisy beacon data based on spring dynamics. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, BuildSys '23* (pp.286-287) Association for Computating Machinery (ACM). https://doi.org/10.1145/3600100.3626263

Pan, J. et al. (2023). Future workspace needs flexibility and diversity: A machine learning-driven behavioural analysis of co-working space. *PloS One*, 18(10), e0292370. https://doi.org/10.1371/journal.pone.0292370

Pollard, B., Engelen, L., Held, F., & de Dear, R. (2022). Activity space, office space: Measuring the spatial movement of office workers. *Appl. Ergon.*, 98(103600), 103600. https://doi.org/10.1016/j.apergo.2021.103600

Yu, S. (2023). Evaluating architectural layouts for occupancy patterns and interactions using agent-based modelling as a methodology for workplace design. *Autom. Constr.*, 155(105025), 105025. https://doi.org/10.1016/j.autcon.2023.105025

Zheng, X., Pan, C., & Shang, J. (2018). Reference Points Density Invariant Map Matching for Wi-Fi Fingerprinting Positioning. In *Proceedings of 2018 26th International Conference on Geoinformation* (pp1-5). https://doi.org/10.1109/GEOINFORMATICS.2018.8557040

From Optimisation to Nudging

An Occupant-Centric Simulation Framework for Identifying Energy-Aware Shading Behaviour

Bohan Wang[1], Robert W. Lindeman[2], Peter Thompson[3], Farhang Tahmasebi[4], Yihan Huang[5], and Heide K. Lukosch[6]
[1,2,6]*HIT Lab NZ, University of Canterbury.*
[3]*Civil and Environmental Engineering, University of Canterbury.*
[4]*Bartlett School Env, Energy & Resources, University College London.*
[5]*Independent Landscape Architect.*
[1]*bohan.wang@pg.canterbury.ac.nz, 0009-0002-0550-741X*
[2]*rob.lindeman@canterbury.ac.nz, 0000-0002-0637-7701*
[3]*peter.thompson@canterbury.ac.nz, 0000-0003-4972-7916*
[4]*f.tahmasebi@ucl.ac.uk, 0000-0001-5727-2646*
[5]*yihanh@alumni.upenn.edu, 0009-0007-4250-1291*
[6]*heide.lukosch@canterbury.ac.nz, 0000-0002-9585-0723*

Abstract. Manual shading operation can substantially influence building energy performance, yet occupants adjust shades based on habits and comfort preferences rather than energy considerations. This study proposes a simulation framework that couples the parametric shading behaviour model with multi-objective optimisation to identify energy-aware shading behaviour for informing nudging strategy design. In a case study, set within the heating season of Christchurch, New Zealand, the derived energy-aware behaviour achieves approximately a 10-20% improvement in both electric-lighting savings and passive solar heat gains. Energy-aware shading patterns in this case are generally characterised by longer response delay before closing, greater hesitation when closing, and shorter post-closure persistence. An exception is observed for occupants with low daylight preference, for whom nudges aimed at shortening shade-closure persistence may have limited effects. For occupants with high daylight preference, nudges that prolong response delay appear less applicable, whereas nudges that reduce behavioural inertia after closure remain promising. Overall, the proposed framework provides an evidence-based pathway for identifying energy-aware behaviour and informing nudging strategy design for manual shading operation.

Keywords. Occupant Behaviour, Manual Shading Operation, Energy Awareness, Nudging Strategy, Multi-Objective Optimisation, Building Performance Simulation

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 367-376.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

1. Introduction

Building operations account for 30% of global energy consumption (IEA, 2023), considerably contributing to climate change. Performance-based design has advanced energy-efficient building design practice, yet a substantial energy performance gap persists between design predictions and real-world operation, which is largely due to the uncertainty of occupant behaviour. Occupants adapt to indoor environmental conditions through interactions with buildings, and such adaptive behaviours can directly influence energy use. Window shading offers a clear example: people adjust shades to satisfy their visual, thermal, and psychological comfort, and these adjustments modify daylight and solar radiation entering the room, which in turn affect energy use in artificial lighting and heating, cooling, and air conditioning (HVAC) systems. This interaction highlights the necessity of accounting for occupant behaviour at the early building design stage to bridge the performance gap and support occupant-centric design practices.

While research has shown that granting occupants direct control over building components, rather than relying entirely on automatic systems, can substantially enhance their satisfaction with the space (Lee and Brand, 2005), most optimisation studies have focused on façade designs or algorithmic shading controls. These studies overlook the fact that manually operated shading is a major driver of energy outcomes. This gap raises a critical question: can multi-objective optimisation (MOO) be extended beyond systems to guide occupants toward more energy-aware shading behaviour?

Nudging, rooted in behavioural economics, provides a complementary pathway by subtly steering decisions while preserving individual freedom of choice (Thaler and Sunstein, 2008). Applied to buildings, it holds the potential to guide manual shading actions toward more energy-aware patterns without compromising comfort and satisfaction. However, developing nudging strategies first requires defining the target behaviour, that is, identifying what constitutes energy-aware shading behaviour. This is highly context-dependent and dominated by individual preferences. To address this challenge, this paper proposes a simulation-based framework that applies MOO to identify energy-aware shading behaviours under different daylight preference conditions, establishing an evidence-based foundation for developing nudging strategies for manual window-shade operation as a case of co-designing behavioural nudges in the built environment.

2. Background: Nudging Shading Behaviour Toward Energy Awareness

Given the variations in climate, sky conditions, solar geometry, and indoor task requirements, the 'optimal' shading state is context-dependent rather than fixed. As a result, the actual performance of shading systems depends more on how they are operated than on their rated specifications. Although automatic systems can follow the 'optimal' schedule, many retrofit and small-scale projects still rely on manually controlled shading because of cost, wiring, and maintenance constraints. Even where automated controls exist, occupants often override them based on personal habits and comfort preferences (Bresa et al., 2024). This behavioural uncertainty ultimately contributes to the deviation between operational performance and design expectations.

Nudging strategies refer to measures that can be implemented easily with minimal time and financial costs (Thaler and Sunstein, 2008). Although the literature on nudging strategies in buildings appears limited, the concept has already been implicitly embedded in studies described as 'behavioural interventions' or 'behavioural changes', which have demonstrated that simple strategies such as energy feedback and social comparison can guide occupants toward more energy-aware behaviours while maintaining their perceived sense of control (Ma et al., 2018; Ayres et al., 2013). However, the literature largely tests nudges under tightly scoped conditions, with limited discussion of systematic processes for designing such strategies. This leads to two key limitations. First, the context-dependent nature of 'optimal' shading behaviour limits the transferability of nudging strategies. Second, the effects of nudging remain uncertain, while conducting field experiments often requires substantial time and financial resources. These challenges motivate the exploration of a simulation-based pre-occupancy evaluation approach, through which reliable design principles can be derived before investing extensive effort and cost in real-world testing, thereby reducing the risks associated with implementation.

3. Methodology

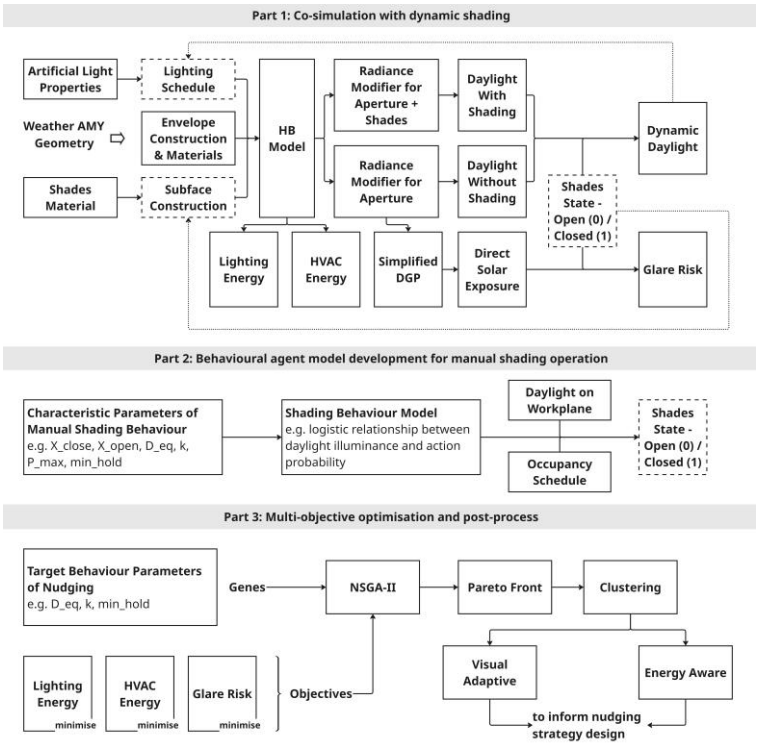


Figure 1. Proposed workflow

We propose a workflow within the Rhino-Grasshopper environment that helps

stakeholders across building design, operation, and policy domains systematically identify the key characteristics of energy-aware shading behaviour to inform the development of nudging strategies, as shown in Figure 1. The workflow comprises three parts: (1) a coupled simulation of daylighting and thermal performance with dynamic shading, (2) an agent decision model of manual shading behaviour, and (3) multi-objective optimisation with post-processing analysis. The workflow addresses the methodological gap of lacking a systematic approach for identifying energy-aware shading behaviour during nudge development.

3.1. CO-SIMULATION WITH DYNAMIC SHADES

EnergyPlus (Crawley et al., 2001) and Radiance (Ward, 1994) are two well-established simulation engines coupled through Ladybug Tools (Roudsari and Pak, 2013) in the Grasshopper environment to enable joint energy-daylight analysis within a unified parametric workflow. However, integrating a dynamic shading system into this workflow remains challenging, because Radiance is highly sensitive to boundary conditions. When geometrical or optical states (e.g., variable shading configurations) change over time, daylight evaluation often needs to be repeated at each timestep to capture these evolving conditions, making long-period simulations computationally expensive. To address this challenge, the present study adopts a state-based pre-calculation approach to integrate dynamic shading into the workflow. Two Radiance simulations, representing unshaded and fully shaded conditions, were generated to obtain sensor illuminances for both states, which are later mapped to each timestep according to the agent-defined shading schedule (see Section 3.2). The resulting dataset is used to determine the artificial lighting schedule and subsequently coupled with the energy simulations. Similarly, a simplified daylight glare probability (DGPs) is approximated using the empirical relationship $DGPs = 6.22 \cdot 10^{-5} \cdot Ev + 0.184$, where Ev is the vertical illuminance in lux (Wienold, 2009). By cross-referencing direct solar exposure and the shading schedule, we assign DGPs = 1.0 for direct sun (intolerable glare) and DGPs = 0.3 for closed shades (no glare), producing a complete DGPs time series. This addresses the known limitation of DGPs for direct-sun glare (Wienold, 2009) and enables the analysis of glare fluctuation under dynamic shading conditions. Glare Risk (GR) is defined as the proportion of occupied timesteps with DGPs > 0.40.

For the energy model, the same shading schedule is used to switch the window sub-surface construction at each timestep, and EnergyPlus dynamically switches the corresponding construction according to the shading state at that timestep. In this way, all variations in shading behaviour can be represented within a physics-based energy simulation, while the daylight analysis is fully data-driven, performed through custom scripts that process pre-computed results. This hybrid approach significantly reduces computational demand and makes the MOO of parametric shading behaviour computationally feasible.

3.2. AGENT SHADING BEHAVIOUR MODEL

To represent occupants' manual shading behaviour, this study employs a hybrid modelling approach that combines a stimulus-probability mapping with temporal

mechanisms. The stimulus-probability mapping represents the probabilistic relationship between environmental stimuli and the likelihood of shading actions (O'Brien et al., 2013), while the temporal mechanism captures the time-reflective aspects of decision-making (Harris and Hutcherson, 2022). Together, these two components allow the model to reproduce both the environmental triggers and the decision dynamics of manual shading actions.

The stimulus-probability mapping is represented by a logistic function that characterises how the likelihood of closing the shades to block sunlight varies with the average horizontal illuminance on the workplane. The illuminance level at which the agent exhibits an absolute 1% probability of closing the shades (X_{close}) corresponds to the condition under which visual discomfort is first perceived. The slope parameter k reflects the occupant's decisiveness in responding to excessive daylight. A higher k indicates a more decisive response. The maximum probability P_{max} denotes the upper limit of the likelihood to close the shades. Together, these three behavioural parameters define the logistic illuminance-probability relationship within the agent shading behaviour model:

$$P(E) = P_{max}/\{1 + exp\{-k[E - X_{close} - 1/k \cdot ln(P_{max}/0.01 - 1)]\}\}$$

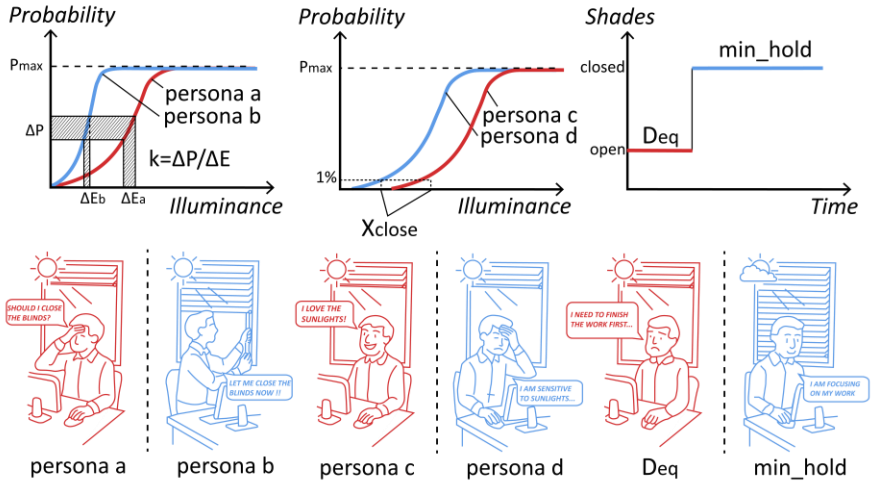


Figure 2. Conceptual illustration of behavioural parameters

The temporal mechanism captures the responsive delay (Li et al., 2022) and behavioural inertia (O'Brien et al., 2013) observed in manual shading behaviour through two parameters: D_{eq} and min_hold. D_{eq} quantifies the delay between perceiving visual discomfort and initiating a decision-making process by accumulating normalised excess illuminance over time. A larger value indicates greater tolerance for excessive daylight. min_hold defines the minimum duration for which the shades remain closed before re-evaluation, which reflects the behavioural persistence often observed in practice. Once min_hold has expired, the agent re-evaluates the environment and decides whether to open the shades based on the threshold (X_{open}). Together, these parameters allow the model to reproduce the shading decisions as a

temporal dynamic process characterised by perception-reaction delay and post-action inertia rather than an instantaneous response. Figure 2 presents a schematic of the agent shading model and the meaning of each behavioural parameter.

3.3. MULTI-OBJECTIVE OPTIMISATION

Three behavioural parameters D_{eq}, k_{raw}, and min_hold representing manual shading patterns are encoded as genes in the Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2000) based MOO process in the Wallacei plugin (Makki et al., 2018). In contrast, since X_{close} and X_{open} reflect occupants' inherent daylight preferences, and P_{max} represents the controllability (the likelihood of eventually closing the shades under extreme daylight levels), these parameters are treated as fixed scenario parameters rather than nudging-targeted decision variables. The optimisation simultaneously minimises lighting energy use, HVAC energy use and glare risk, generating Pareto-optimal sets of non-dominated solutions that represent different trade-offs between energy performance and visual comfort for each daylight preference scenario. To interpret these solutions and identify energy-aware behaviour patterns, unsupervised clustering is applied in the normalised performance space, grouping solutions with similar performance into behavioural archetypes whose characteristic parameter features and deviations inform the design of nudging strategies.

4. Case Study

For this case study, a shoebox model was created to present a 20 m² (4 m × 5 m × 3 m) private office located in Christchurch, New Zealand. To isolate the impact of façade-related lighting and energy performance, the office is assumed to be situated within the interior of a larger office building, with only one highly glazed north-facing wall (window-to-wall ratio 80%) exposed to the outdoors.

Three behavioural parameters D_{eq}, k_{raw}, and min_hold were encoded as genes in the MOO process, where k_{raw} represents the slope parameter k in logarithmic form to obtain a more uniform sampling of logistic curve steepness across the search space. In this case study, a Daylight Preference Index (DPI) was introduced to encode the shades opening and closing thresholds (X_{open}, X_{close}) as a single occupant-daylight-preference parameter. The neutral level, DPI = 0.5, represents broadly reported office occupants (Nabil and Mardaljevic, 2005), while DPI = 0.1 and DPI = 0.9 represent extreme low and high daylight-preference groups under a simple linear scaling around the neutral case. The P_{max} was fixed as 1 across all scenarios. The parameter characteristics and value ranges are summarised in Table 1. In total, 150 individuals over 20 generations (3000 evaluations) were simulated for each daylight-preference group, using a crossover probability of 0.9, crossover and mutation distribution indices of 20, and a mutation probability of 1/n (n = 3).

To simplify the simulation, the case study was conducted with a 15-minute timestep and focused solely on the heating season (April-September), during which shading mainly affects the extent to which solar heat gains (SHG) can offset space-heating demand. In this context, SHG was maximised as a surrogate objective for minimising HVAC heating energy demand in the MOO. In post-processing, LE and SHG were

combined into a weighted Energy Benefit (EB) indicator for clustering. The weighting assumes electric resistance heating (COP = 1) and that SHG can be effectively utilised during only about 65% of heating hours, which yields an effective weighting ratio of approximately $w_{LE} = 0.60$ and $w_{SHG} = 0.40$. EB then served as the clustering variable in a k-means analysis (k = 2), separating energy-aware solutions from visually adaptive solutions. Accordingly, the case study tests whether the proposed workflow

Table 1. Behavioural parameters and decision space for MOO

Parameter	Numeric domain	Interval	Type of number
D_{eq}	0.1-6.0 timesteps	0.1	float
k_{raw}	0.01-1.00	0.01	float
min_hold	1-36 timesteps	1	integer
DPI (X_{open}, X_{close})	0.1 (100, 1000) /0.5 (500, 2000) /0.9 (900, 3000)	-	fixed
P_{max}	1	-	fixed

can not only extract energy-aware shading solutions from the optimised solution set, but also identify interpretable shading-behaviour patterns based on their underlying behavioural parameters, revealing systematic parameter differences between clusters that can be translated into nudging strategies.

5. Results and Discussion

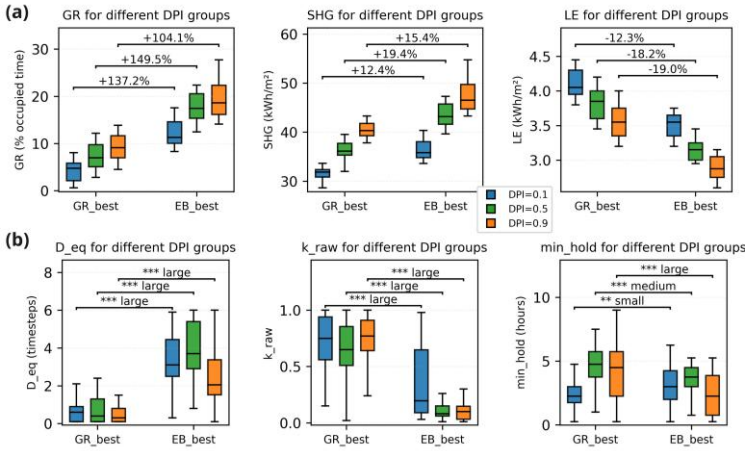


Figure 3. Clustering results for performance objectives (a) and behavioural parameters (b)

Figure 3-a shows the clustering results in the objective space, identifying two behavioural archetypes across daylight-preference groups: an energy-aware cluster

(EB_best; n = 66, 61, 50 at DPI 0.1, 0.5, 0.9) and a visual-adaptive cluster (GR_best; n = 125, 120, 155 at DPI 0.1, 0.5, 0.9), with median % changes annotated. In this case study, adopting energy-aware shading behaviour yields roughly 10-20% lower electric-lighting demand and similar-magnitude increases in passive solar energy during the heating season. The medium-daylight-preference group (DPI 0.5) shows the greatest combined improvement in LE and SHG. This is because the low-daylight-preference group (DPI 0.1) tends to keep the shades closed for long periods, limiting the behavioural parameter search space, while the high-daylight-preference group (DPI 0.9) is already near its energy-performance limit and further shade opening is constrained by glare. The pronounced relative increase in GR reflects the specific test configuration, where a single occupant is seated facing a north-oriented, highly glazed façade without any additional glare-mitigation devices.

Further analysis of the behavioural parameters is shown in Figure 3-b, which illustrates the distributions of D_{eq}, k_{raw}, and min_hold across different clusters and daylight preference groups. Overall, the energy-aware shading behaviour exhibits consistent and distinct characteristics compared with visual-adaptive behaviour. Specifically, a longer response delay before closing the shades, greater hesitation of closing the shades, and a shorter close-persistence time are all beneficial to reduce energy use during the heating season. The statistically significant differences (*** $p < 0.001$) with large effects ($|\text{Cliff's } \delta| \geq 0.474$) in the Mann-Whitney U tests support these findings. These findings are consistent with the physical expectation that, during the heating season, keeping shades open under glare-free conditions helps maximise passive solar heat and daylight use. However, an exception is observed for groups with low daylight preference, where the difference min_hold is statistically significant but exhibits only a small effect size. This may be attributed to the low X_{open}, which means that even if nudging strategies encourage them to reconsider the shading state shortly after closing, their low preference for daylight still leads them to keep the shades closed. This indicates that in this case, a nudging strategy such as notifications that remind occupants to reopen the shade after it has been closed may be less effective for occupants with low daylight preference.

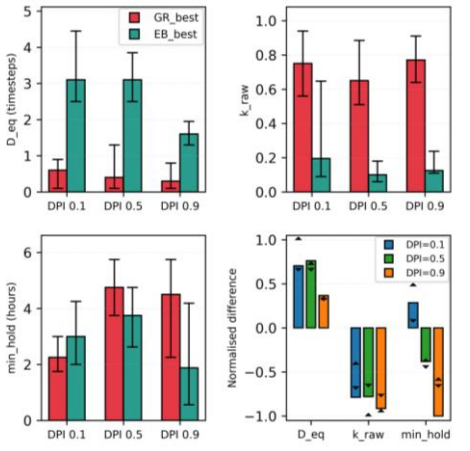


Figure 4. Behavioural gaps between shading archetypes

After filtering out cases with excessive visual discomfort (GR > 20% in this case), the remaining results reveal the behavioural gaps between the energy-aware and visual-adaptive shading patterns across the three daylight preference groups (Figure 4). The normalised differences quantify the relative 'nudging intensity' for each behavioural parameter: values approaching ±1 indicate a larger behavioural divergence and thus a stronger potential leverage point for intervention. The comparison suggests that nudging strategies aimed at increasing D_{eq} (prolonging the response delay) may be more relevant for occupants with low and medium daylight preference, whereas such interventions appear less effective for high-daylight-preference occupants, who tend to experience glare more frequently due to their higher shades-closing thresholds. In contrast, reducing k_{raw} (reflecting more hesitation) emerges as a universally important nudging direction for all groups, implying that lowering sensitivity to small daylight variations consistently supports more energy-aware shading operation. Finally, in addition to the finding that decreasing min_hold (shortening shading-close-persistence time) may not benefit low-daylight-preference occupants, the comparison results highlight its critical role for high-daylight-preference groups, who are more likely to open shades when cued by nudging strategies, thereby maximising effective solar gains and daylight utilisation.

6. Conclusion and Limitations

This study presents a simulation framework that integrates occupant behavioural modelling with multi-objective optimisation to identify energy-aware shading behaviours for informing nudging strategy development. The case study demonstrated the workflow can distinguish the characteristics of energy-aware behaviour from typical visually adaptive patterns, capture how these behaviours vary across daylight-preference groups, quantify their energy impacts in a heating context, and highlight where nudges may require stronger intervention or yield limited benefit. While these insights help prioritise high-nudging-impact conditions and guide strategy design, several limitations remain. The behavioural model relied on daylight illuminance as the sole stimulus, whereas factors such as thermal comfort, view needs, and psychological influences should be incorporated in future work. The approach was demonstrated in a single-occupant office with exclusive control, leaving shared-control dynamics in multi-occupant settings to be explored. In addition, intermediate shading states should be considered, rather than limiting the model to fully open and fully closed positions. Finally, the workflow is timestep-resolved but one-way coupled. Extending it to a closed-loop setup with bidirectional communication will be essential for representing the dynamic feedback between shading actions and indoor environmental conditions.

Attribution Statement

ChatGPT (OpenAI, 2025) was used solely for language-related assistance, including translating parts of the manuscript and refining grammar, clarity, and writing style. All AI-assisted outputs were manually reviewed, checked, and revised by the authors.

References

Ayres, I., Raseman, S., & Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. Journal of Law, Economics, & Organization, 29(5), 992–1022. https://doi.org/10.1093/jleo/ews020

Bresa, A., Zakula, T., & Ajdukovic, D. (2024). Occupant-centric control in buildings: Investigating occupant intentions and preferences for indoor environment and grid flexibility interactions. Energy and Buildings, 317, 114393. https://doi.org/10.1016/j.enbuild.2024.114393

Crawley, D. B., Lawrie, L. K., Winkelmann, F. C., Buhl, W. F., Huang, Y. J., Pedersen, C. O., Strand, R. K., Liesen, R. J., Fisher, D. E., Witte, M. J., & Glazer, J. (2001). EnergyPlus: Creating a new-generation building energy simulation program. Energy and Buildings, 33(4), 319–331. https://doi.org/10.1016/S0378-7788(00)00114-6

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In M. Schoenauer et al. (Eds.), Parallel problem solving from nature PPSN VI (pp. 849–858). Springer. https://doi.org/10.1007/3-540-45356-3_83

Harris, A., & Hutcherson, C. A. (2022). Temporal dynamics of decision making: A synthesis of computational and neurophysiological approaches. Wiley Interdisciplinary Reviews: Cognitive Science, 13(3), e1586. https://doi.org/10.1002/wcs.1586

IEA (2023), Tracking Clean Energy Progress 2023, IEA, Paris https://www.iea.org/reports/tracking-clean-energy-progress-2023, Licence: CC BY 4.0

Lee, S. Y., & Brand, J. L. (2005). Effects of control over office workspace on perceptions of the work environment and work outcomes. Journal of Environmental Psychology, 25(3), 323–333. https://doi.org/10.1016/j.jenvp.2005.08.001

Li, C., Zhu, H., Lian, X., Liu, Y., et al. (2022). Study of "time-lag" of occupant behavior occurrences for establishing an occupant-centric building control system. Building and Environment, 216, 109005. https://doi.org/10.1016/j.buildenv.2022.109005

Ma, G., Lin, J., & Li, N. (2018). Longitudinal assessment of the behavior-changing effect of app-based eco-feedback in residential buildings. Energy and Buildings, 159, 486–494. https://doi.org/10.1016/j.enbuild.2017.11.019

Makki, M., Showkatbakhsh, M., & Song, Y. (2018). Wallacei: An evolutionary and analytic engine for Grasshopper 3D [Computer software]. Wallacei. https://www.wallacei.com/

Nabil, A., & Mardaljevic, J. (2005). Useful daylight illuminance: A new paradigm for assessing daylight in buildings. Lighting Research & Technology, 37(1), 41–59. https://doi.org/10.1191/1365782805li128oa

O'Brien, W., Kapsis, K., & Athienitis, A. K. (2013). Manually-operated window shade patterns in office buildings: A critical review. Building and Environment, 60, 319–338. https://doi.org/10.1016/j.buildenv.2012.10.003

Roudsari, M. S., & Pak, M. (2013). Ladybug: A parametric environmental plugin for Grasshopper to help designers create an environmentally-conscious design. In Proceedings of BS2013: 13th Conference of International Building Performance Simulation Association (pp. 3128–3135).

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. Yale University Press.

Ward, G. J. (1994). The RADIANCE lighting simulation and rendering system. In Proceedings of the 21st annual conference on Computer graphics and interactive techniques (pp. 459–472). ACM. https://doi.org/10.1145/192161.192286

Wienold, J. (2009). Dynamic daylight glare evaluation. In Proceedings of Building Simulation 2009: 11th International IBPSA Conference (pp. 944–951). International Building Performance Simulation Association.

Regenerative Cities

A Data-Driven Generative Framework for Adaptive Interventions within Heritage Urban Contexts

Rutuja Rode[1], Sonali Rane[2], Abhijeet Manjunath[3] and Milad Showkatbakhsh[4]
[1,2,3,4]*Architectural Association School of Architecture*
[1]*roderutuja@gmail.com, 0009-0000-5861-5815*
[2]*sonalirane.str@gmail.com, 0009-0000-2275-3212*
[3]*manjunath.abhijeet@gmail.com, 0009-0002-2270-8838*
[4]*showkatbakhsh@aaschool.ac.uk, 0000-0001-5440-5175*

Abstract. This paper investigates adaptive design strategies for heritage cities experiencing rapid densification because of population influx. It proposes a computational workflow that analyses the city's network, density and environmental conditions, generates and evaluates architectural interventions, recalibrating the urban context through iterative feedback. Using generative algorithms, the system synthesises socio-cultural and environmental datasets to produce context-sensitive, climate-responsive solutions. A kit-of-parts system based on local materials enables adaptability within dense historic fabrics, allowing configurations to be assigned and optimised according to spatial and environmental parameters. The workflow extends into a predictive framework through a machine learning model trained on optimised datasets to predict future spatial configurations. These morphologies are then re-integrated into the urban network to inform the placement and impact of subsequent interventions, supporting an ongoing, adaptive evolution of the heritage fabric.

Keywords. Heritage Cities, Adaptive Design, Generative Algorithm, Multi-Objective Optimisation, Machine Learning

1. Introduction

Population influx has become an emerging challenge for heritage cities worldwide, intensifying urban density and straining existing infrastructure (Bertocchi et al., 2020). While it supports local economies, it leads to rapid conventional redevelopment practices that often compromise the city's cultural identity (Al Ameen et al., 2025). This research explores nodal halt points as resting spaces for the floating population through a cyclical computational workflow integrating generative algorithms, multi-objective optimisation, and machine learning. The system generates and reintegrates architectural interventions within the urban network to create adaptive, data-informed

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 377-386.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

design solutions. A locally derived kit-of-parts supports flexibility in dense heritage contexts, enabling configurations that respond to both spatial and environmental parameters. A machine learning model trained on optimised datasets predicts future configurations, offering a responsive alternative to conventional top-down redevelopment and large-scale demolition.

As a case study, this paper examines demographic pressures driving urban densification in Varanasi, one of the world's oldest continuously inhabited cities and the 'Religious Capital' of India. Situated along the river Ganges, Varanasi functions as a cultural and spiritual hub, drawing a continuous influx of pilgrims and tourists (Singh & Rana, 2017). This transient population places increasing pressure on the city's compact urban fabric, particularly around Dharamshalas (rest houses) that serve as transitional dwellings (Singh & Kumar, 2022).

Varanasi's inclusion on UNESCO's Tentative World Heritage List has accelerated redevelopment projects aiming to capitalise on its cultural and historical significance (Singh & Rana, 2017). Yet, conventional demolition-and-reconstruction strategies impose permanent, homogeneous solutions on an inherently transient urban condition, eroding the city's identity (Das et al., 2024). The Kashi Vishwanath Corridor project in Varanasi exemplifies this trend with its large concrete plaza designed for global appeal, disregarding human scale and climatic sensitivity in a pedestrian-centric city (Singh & Rana, 2022; Mukundan, 2020). In contrast, this study analyses the city's network, density, and environmental parameters to identify contextually responsive sites for intervention. It proposes adaptive, generative insertions using a climate-responsive kit-of-parts to enhance sociability, manage demographic flux, and preserve the spatial and cultural identity of Varanasi.

2. Literature Review

Dharamshalas are rest houses for pilgrims in Varanasi, typically located near major religious nodes along pilgrimage routes. These low-rise structures are organised around central courtyards that enable social interaction and passive ventilation. Transitional foyers mediate between street and interior, while shaded verandahs and terrace roofs enhance thermal comfort in Varanasi's hot, humid climate (Singh & Kumar, 2022; Verma & Bano, 2023). Integral to the city's cultural identity, they accommodate sleeping, cooking, and communal rituals, with courtyards seasonally transforming for storage, weddings, and gatherings (Singh & Kumar, 2022). Rising urban density has reduced access and openings, leaving courtyards as the primary light and ventilation source. As Hillier (2007) observes, courtyards act as 'convex' social nodes, and verandahs as 'linear' transitional spaces, forming a spatial gradient from public to private (Pandya, 2016) creating a hierarchy that translates into computational parameters for adaptive interventions.

3. Methodology

The methodology follows a computational workflow comprising three primary phases: site selection, morphology generation, and predictive modelling. The objective is to integrate environmental, spatial, and socio-cultural parameters derived from spatial hierarchies operationalised through physical urban typologies. The site selection phase

combines quantitative network analysis with assessments of spatial availability, evaluating building density, height variance, and environmental suitability through solar exposure, wind flow, and visibility metrics. The morphology generation stage employs a multi-objective evolutionary algorithm (MOEA) to generate voxel-based configurations optimised for spatial adjacency, environmental comfort, and compliance with local by-laws. Given the number and conflicting nature of parameters influencing morphology, the multi-objective optimisation helps identify solutions that balance competing design goals effectively.

The workflow adopts a two-stage sequential optimisation framework, where the outputs of the first stage (morphology optimisation) define and refine the design space for the second stage (fenestration optimisation). This sequential process allows the experiment to progressively evolve both form and envelope systems, integrating performance feedback from the first stage to the next. Each voxel is then categorised and assigned kit-of-parts components based on spatial role, orientation, and environmental performance, ensuring efficient outcomes in both spatial configuration and facade logic.

In the third phase, datasets derived from the optimised solutions are used to train a machine learning model, which establishes a predictive framework for generating new spatial configurations based on established performance criteria. Each reintegration of new configurations updates the network analysis to identify the next potential site, forming a feedback loop that supports adaptive urban transformation.

4. Experiment

4.1. SITE SELECTION

According to the Varanasi Development Plan 2031, five old-city zones are earmarked for redevelopment. (Singh & Rana, 2017). The first, including the Kashi Vishwanath Corridor, has already seen large-scale demolition. Zone 2 comprising major temples, religious nodes, and heritage sites, forms the next phase and serves as the case study for this research.

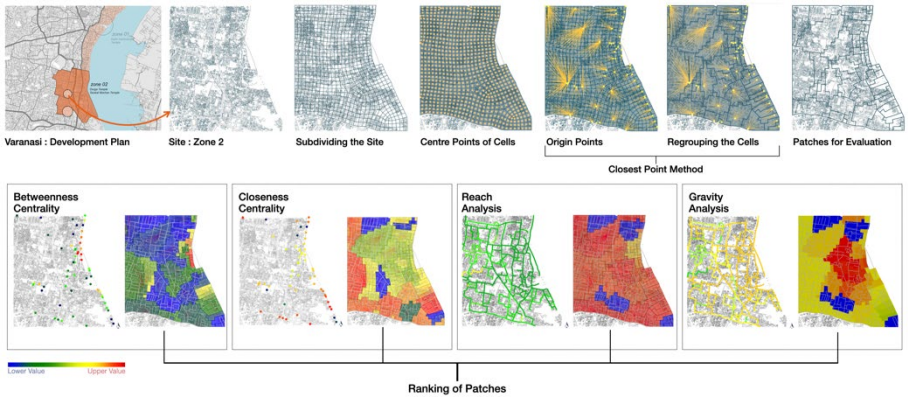


Figure 01. Site Selection process with node identification and quantitative network analysis

4.1.1. Step 1: Node Identification and Network Analysis

A 3D model of Zone 2 was developed using ArcGIS and Google Maps data, mapping key nodes such as ghats, temples, kunds, markets, parks, transport hubs, and existing rest houses. Rest houses and transport hubs were defined as 'origins', and the remaining nodes as 'destinations', forming the topological framework for network analysis. Using Space Syntax based metrics namely - Betweenness, Closeness, Reach, and Gravity, the network's connectivity and accessibility were evaluated (Numerical Network Analysis). As shown in Figure 01, the site was subdivided and regrouped into patches based on proximity to origin points, with each normalised between 0.00 and 1.00. Overlapping patches within the defined performance range were identified, and the highest-ranking patch that demonstrated strong accessibility and balanced through-movement was shortlisted for further analysis.

4.1.2. Step 2: Integrated Density, Environmental, and Visibility Analysis

As shown in Figure 02, patches were subdivided and clustered using centroid attractor points. Sub-patches under 1,000 m² were removed, retaining only major urban fragments for analysis. These were assessed for density, height variance, solar exposure, and visibility. Highly visible and therefore more solar-exposed sites were prioritised for rest house suitability. CFD analysis then identified the sub-patch with the least airflow variation. These analyses highlighted opportunities to improve site shading, self-shading and ventilation in the morphological experiment.

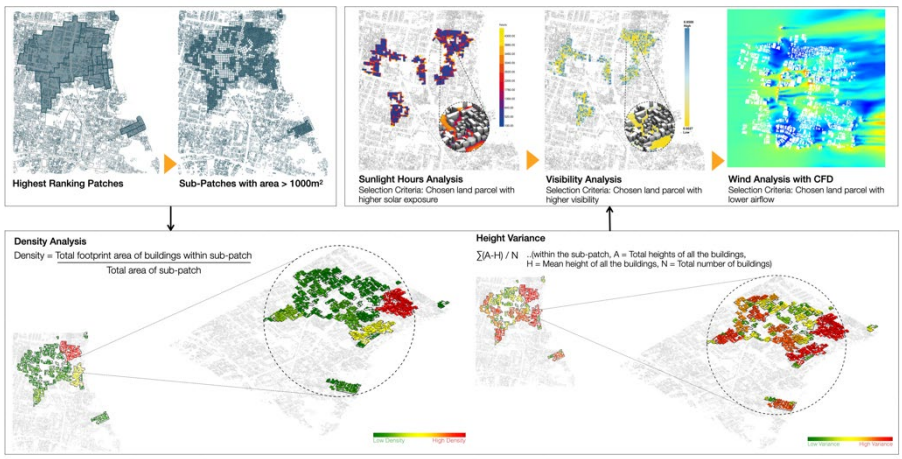


Figure 02. Density, height variance, environmental and visibility performance analysis

4.2. MULTI-OBJECTIVE EVOLUTIONARY ALGORITHM

4.2.1. Stage I : Design Objectives

Stage I focuses on the placement of the built mass within the site, balancing spatial, environmental, and contextual criteria. The fitness objectives (FOs) were formulated to control the level of demolition of the existing fabric, maximise shading suited to

Varanasi's hot and humid climate, ensure functional adjacencies, and maintain a central courtyard. Collectively, these objectives aimed to optimise the use of available FAR, control circulation compactness, and enhance the centrality of open spaces, framing wind flow and shading as key design opportunities within the dense urban fabric.

4.2.2. Stage I : Phenotype Build

As illustrated in Figure 03, Stage I establishes the built configuration through a voxel-based process. The site is subdivided to define the active search space, followed by controlled demolition (40–60%) to balance continuity and redevelopment. Setbacks, courtyards, verandahs, and fixed cores are then defined while meeting FAR requirements. Functional adjacencies and open courts are generated through multi-objective optimisation, producing iterations that balance spatial efficiency, shading, and environmental performance within Varanasi's dense urban fabric.

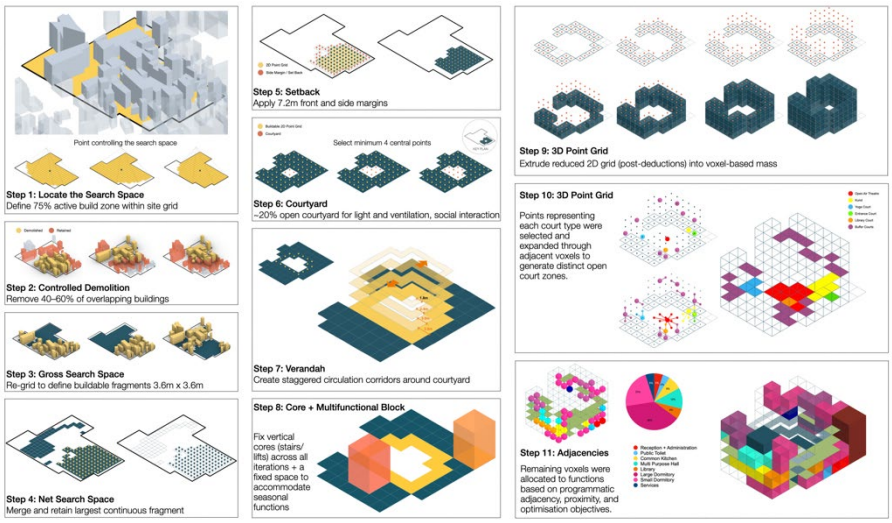


Figure 03. Phenotype Build process for Stage I of the Multi Objective Optimisation

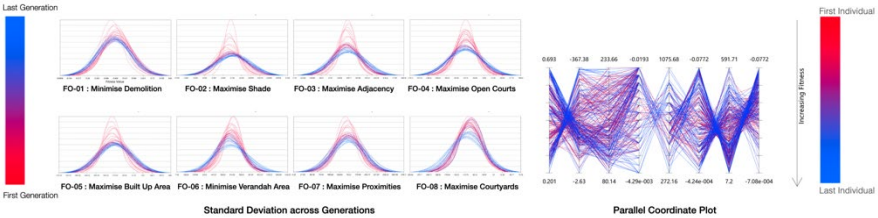


Figure 04. Standard Deviation Graph and Parallel Coordinate Plot of Stage I of the MOEA

4.2.3. Stage I : Optimisation Results

A Multi-Objective Evolutionary Optimisation was conducted in Grasshopper (Wallacei) over 20 generations with 50 individuals, yielding 230 Pareto fronts from

1,000 total iterations. Significant variation appeared across these fronts representing the best trade-offs among competing objectives, particularly in form and site placement, influenced by the broad search space. Maximising courtyard area consistently produced larger or multiple courts across phenotypes. As shown in Figure 04, the Parallel Coordinate Plot reveals clear trade-offs, with gains in some objectives reducing performance in others, while standard deviation graphs indicate rising variation over generations. Figure 05 shows most Pareto fronts clustering toward the site's south and west, increasing shading, with Gen48Ind15 and Gen49Ind19 emerging as the most balanced in built-to-open ratios.

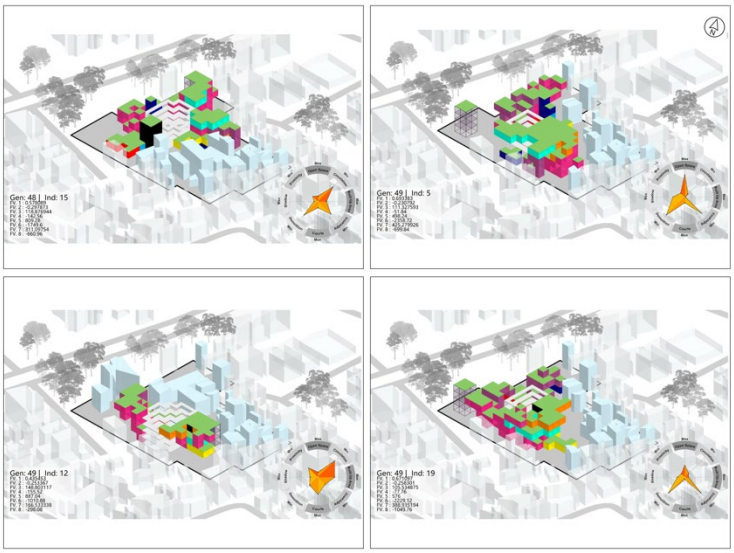


Figure 05. Four Pareto Front Solutions from Stage I of the MOEA

4.2.4. Stage II : Sequential Simulation Experiment Set Up

As discussed in the methodology section, the sequential simulation builds on the refined design space from Stage I, forming the second stage of the optimisation framework. Decoded genes and Pareto front data defined input parameters and domain limits, narrowing the search space for precision. This stage assigns kit-of-parts components like walls, fenestration, and roofs - based on environmental performance and spatial adjacency, extending the experiment from spatial morphology to material and facade logic while maintaining design continuity.

4.2.5. Stage II : Sequential Simulation Design Objectives

The fitness objectives balanced daylight, thermal comfort, and passive cooling for Varanasi's hot-humid climate. FO-01 reduced solar radiation, FO-02 maximised daylight, and FO-03 aimed to enhance passive thermal moderation through maximising shaded rammed-earth walls. Together, they informed climate-responsive kit-of-parts deployment within the dense urban fabric.

4.2.6. Stage II : Sequential Simulation Phenotype Build

As shown in Figure 06, the second-stage simulation refined the voxelised morphology through kit-of-parts assignment. Voxels were grouped by adjacency to define structural logic; shaded faces received rammed-earth panels, exposed faces were assigned timber, and panels facing courtyards incorporated openings to optimise daylight, ventilation, and circulation to open courts.



Figure 06. Phenotype Build process for Stage II of the Multi Objective Optimisation

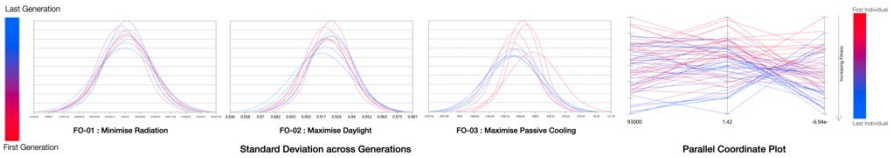


Figure 07. Standard Deviation Graph and Parallel Coordinate Plot of Stage II of the MOEA

4.2.7. Stage II: Optimisation Results

The results of the sequential simulation run over 20 generations with 10 individuals each (200 phenotypes), showed a preference for north-south orientations to reduce radiation and maximise shaded facades for rammed-earth panels. Context buildings provided additional shading to east –west sides, while open courtyards oriented north achieved optimal daylight balance. The Parallel Coordinate Plot (Figure 07) highlights clear trade-offs between objectives, demonstrating successful optimisation.

4.2.8. Selection Strategy

A weighted ranking prioritised fitness objectives by contextual relevance, with FO-03 (passive thermal moderation via shaded rammed-earth walls), followed by FO-01 (solar radiation) and FO-02 (daylight). The top eight Pareto-front solutions were selected for their shaded façades and central courtyards enhancing daylight. Three were advanced for architectural evaluation and structural testing. Finite Element Analysis (FEA using Karamba3D) assessed structural performance under floor-area-based

loads, with Gen2Ind2 showing the least displacement (see Figure 08).

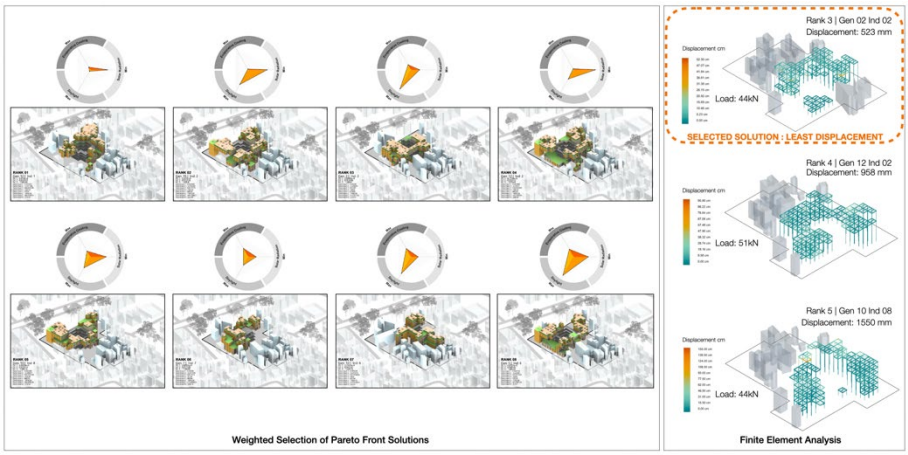


Figure 08. Selected eight Pareto Fronts from Stage II of MOEA; Finite Element Analysis

4.3. FUTURE PREDICTIONS

This section presents the predictive stage, where machine learning extends optimisation beyond the initial site. The ranked land parcels identified through the multi-layered site selection (Figure 09) form the basis for this phase. While the top-ranked site was used for the MOEA experiment, the Rank 2 and Rank 3 parcels served as testing grounds for predicting spatial configurations through a trained Artificial Neural Network (ANN), enabling performance-driven interventions across the wider urban network.

4.3.1. Neural Network Training and Feedback Integration

The Artificial Neural Network (ANN) was trained on 230 Pareto-front solutions from the morphological optimisation stage. To preserve the architectural intent and avoid unconstrained prediction, three genes defining the foundational spatial logic of the intervention - (1) search space initiation, (2) controlled demolition, and (3) courtyard centralisation, were fixed as input parameters. These genes control orientation, intervention intensity, and socio-spatial organisation, and were therefore retained as constants. Their fitness values served as performance targets, while the remaining genes acted as training outputs. Training was performed using LunchBoxML within Grasshopper, with all data normalised to a 0–1 range for consistent learning. The model learned correlations between the fixed inputs and performance outcomes, generating new configurations for the Rank 2 and Rank 3 sites, later decoded into voxelised geometries (Figure 09). When integrated into Zone 2, these interventions functioned as new network origin points, dynamically altering the Betweenness Centrality of the city's spatial system (Figure 10). Each iteration recalibrated the network, revealing emergent opportunity nodes and establishing a feedback loop between building and

urban scales for predictive, adaptive evolution.

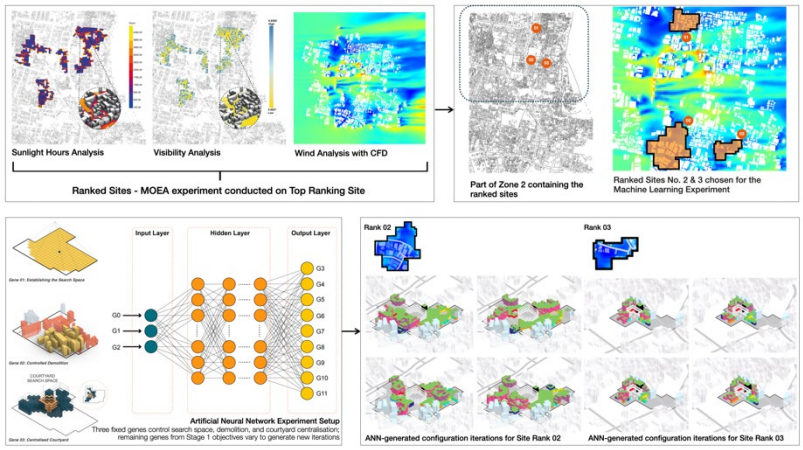


Figure 09. Predictive Stage : Ranked Sites, ANN Experiment Setup and generated iterations

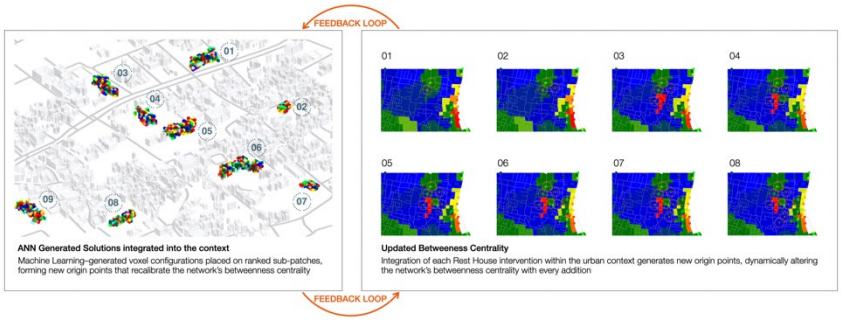


Figure 10. Recalibration of configurations into the urban context

5. Discussion

Population influx in heritage cities like Varanasi should be viewed as an opportunity for adaptive, rather than demolition-driven, redevelopment. The multi-layered methodology combining generative design, multi-objective optimisation, and machine learning integrates environmental, social, and spatial criteria to conserve cultural identity through data-driven adaptability. The morphological optimisation reduced demolition while improving shading and open–built ratios, but the overlapping objectives and linear voxel assignment limited spatial coherence. The voxel-based setup also constrained architectural expression to box-like forms, but voxel centroids can act as positional guides for exploring non-rectilinear geometries. The sequential simulation successfully linked morphology with environmental logic but relied too narrowly on solar-shading criteria; integrating spatial hierarchies and programmatic layouts could enhance usability. While this study translates socio-cultural logic through spatial hierarchies and typologies, future work could integrate pedestrian and interaction simulations to empirically test behavioural performance and social use over

time. The machine learning phase established a predictive framework for future interventions, though the selection of fixed genes remains context-dependent, requiring calibration to different climatic and cultural settings.

6. Conclusion

This research presents a replicable computational framework integrating generative design, optimisation, and prediction to enable adaptive interventions in heritage cities. By combining multi-objective evolutionary optimisation with machine learning, it establishes a feedback-driven model for iterative urban transformation. While the study demonstrates adaptability at spatial and environmental scales, future work should include physical prototyping of the kit-of-parts, on-site testing, and expanded datasets for validation. The approach contributes to computational heritage urbanism, offering a foundation for responsive, regenerative interventions that evolve with their cultural context.

Attribution Statement

ChatGPT (OpenAI, 2025) was used for language editing and grammar refinement during the preparation of this paper.

References

Al Ameen, Z., Font Basté, M. G., & Martí Audí, N. (2025). Reuse strategies of urban areas and built heritage: Case study of Karbala Old City. *Environmental Science and Sustainable Development, 10(2), 1-15.* https://doi.org/10.21625/essd.v10i2.1154

Bertocchi, D., Camatti, N., Giove, S., & van der Borg, J. (2020). Venice and overtourism: Simulating sustainable development scenarios through a tourism carrying capacity model. *Sustainability*, 12(1), 1–23. https://doi.org/10.3390/su12010100

Das, D., Choudhary, B. K., Mishra, S. V., & Aditi, A. (2024). *Varanasi: the making of a smart heritage city.* Environment and Urbanization ASIA, 15(1), 53–72. https://doi.org/10.1177/09754253241230588

Hillier, B. (2007). Space is the machine: a configurational theory of architecture. Cambridge: Cambridge University Press.

Mukundan, S. P. (2020). Changing sacredscapes: A cultural approach for a sustainable Varanasi. Doctoral dissertation, Delft University of Technology, Delft.

Pandya, Y. (2016). Courtyard houses of India. Ahmedabad: Mapin Publishing.

Singh, R. P. B., & Kumar, S. (2022). The sacred nodes of Pañchakroshī Yātrā route, Varanasi (India): Spatial perspectives and prospects for the future. Esempi di Architettura, International Journal of Architecture and Engineering, 9(2), 123–142.

Singh, R. P., & Rana, P. S. (2017). Varanasi: Heritage zones and its designation in UNESCO's World Heritage properties. Kashi Journal of Social Sciences, 7(1–2), 201–218.

Singh, R. P., & Rana, P. S. (2021). The Kashi Vishvanatha, Varanasi City, India. *ACLA Asian Cultural Landscape Association Conference Proceedings.* https://doi.org/10.53136/97912599480762

Verma, L. A., & Bano, F. (2023). Socio-environmental sustainability of traditional courtyard houses of Lucknow and Varanasi. *U. Porto Journal of Engineering*, 9(1), 72–103. https://doi.org/10.24840/2183-6493_009-001_001438.

Surrogate-Based Optimization for Climate-Responsive Design of Thermally Activated Building Envelopes

Tianxin Chang[1] and Zhi Zhuang[2]
[1,] *College of Architecture and Urban Planning, Tongji University*
[2] *Sino-German College of Applied Sciences, Tongji University*
[1]*ctianxin@tongji.edu.cn, 0009-0000-5647-1600*
[2]*zhuangzhi@tongji.edu.cn, 0000-0003-3085-6996*

Abstract. Thermally Activated Building Envelopes (TABE) offer a promising strategy for climate-responsive design, yet their complex heat-transfer processes make traditional physical models too computationally intensive for rapid optimization. This study addresses the critical challenge of enabling rapid, climate-responsive design iterations without compromising physical accuracy. A performance dataset is constructed using a thermal resistance and capacitance model across four representative Chinese climate zones, including severe cold, cold, hot-summer and cold-winter, and hot-summer and warm-winter regions. Six key design parameters are selected to train machine learning surrogate models for predicting dynamic thermal performance. Results demonstrate that the eXtreme Gradient Boosting model achieves superior accuracy and stability, effectively reducing annual simulation time from hours to milliseconds. Cross-climate validation yields R^2 values exceeding 0.85, confirming the model's robust transferability. Based on the surrogate model, a genetic algorithm performs rapid parameter optimization, translating numerical results into distinct, climate-dependent design strategies. Overall, this research establishes an integrated computational workflow linking physical modeling with intelligent optimization, providing an efficient method to support performance-oriented decision-making in the early design stages of active building envelopes.

Keywords. Thermally Activated Building Envelopes, Surrogate Modeling, Climate-Responsive Design, Performance-Driven Design, Multi-objective Optimization

1. Introduction

As climate change accelerates and the demand for carbon-neutral buildings intensifies, the ability of building envelopes to dynamically respond to diverse environmental conditions has become a central challenge in architectural design and performance engineering (Sabapathy and Gedupudi, 2024). Heating and cooling loads account for nearly half of building operational energy use. In this context, adaptive envelope systems capable of buffering temperature fluctuations and improving energy

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 387-396.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

efficiency are critical to ensuring resilience and comfort across climates (Bahrami et al., 2024). Thermally Activated Building Envelopes (TABE) represent a class of active, climate-responsive technologies that embed hydronic pipe networks within structural layers (see Figure 1). By circulating water through these embedded pipes, TABEs can store, release, and redistribute heat to smooth diurnal temperature swings, reduce peak HVAC loads, and maintain steady indoor thermal environments (Fawaier M and Bokor, 2024). Extensive studies have confirmed their potential in reducing energy consumption and improving comfort performance (Krajčík et al., 2021).

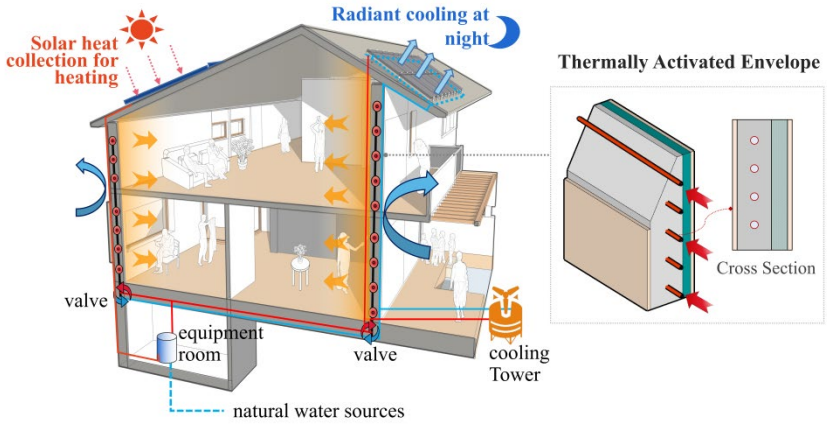


Figure 1. Thermally Activated Building Systems schematic

However, the design and optimization of TABEs remain computationally demanding because their transient behavior arises from the nonlinear coupling among geometric parameters (e.g., insulation configuration, pipe spacing, and diameter), operational settings (e.g., water supply temperatures), and external climatic boundary conditions (Zhao T et al., 2025). To model this, physics-based methods such as finite element analysis (FEA) and resistance–capacitance (RC) models are commonly used (Yang et al., 2024). While RC models are more efficient than FEA, they are still too computationally intensive to support rapid design iteration or real-time decision-making, where thousands of variants must be evaluated to find optimal solutions. This latency becomes a critical bottleneck when considering the complex coupling between the envelope and building zones, making it difficult to support rapid, adaptive design iterations.

Recent advances in machine learning (ML) have shown strong potential for accelerating building performance prediction. Prior studies have applied data-driven models to tasks such as daylight estimation and airflow or energy simulation (Shen et al., 2023) (Ahmed et al., 2024). Despite these advancements in the broader architectural field, the application of ML to solve the rapid design problem of TABE remains limited. Moreover, most existing research focuses on prediction accuracy rather than integration into the architectural design workflow. Consequently, the capacity of surrogate models to not only predict performance but also drive climate-responsive design strategies across different regions remains insufficiently

explored.

To address these challenges, this study develops a surrogate modeling framework to facilitate the rapid performance evaluation and optimization of TABEs (see Figure 2). Instead of relying on computationally expensive physical simulations, a performance dataset is constructed using a validated RC model across four representative Chinese climate zones: severe cold, cold, hot-summer–cold-winter, and hot-summer–warm-winter. Six key design parameters are selected to train ML algorithms, creating a fast-response surrogate model. This model is then coupled with a genetic algorithm to derive region-specific design strategies, transforming numerical predictions into actionable architectural guidance.

The main contributions of this study are: (1) establishing a physics-informed ML framework that achieves high-fidelity prediction with significantly reduced computational cost; (2) demonstrating robust model generalization across diverse climate zones; and (3) applying the surrogate model to derive quantitative, climate-responsive design guidelines, thereby supporting performance-oriented decision-making in the early design stages.

Overall, the study contributes a scalable, data-driven approach that bridges physical simulation and intelligent design, supporting rapid, high-accuracy performance prediction and optimization of thermally activated building envelopes across diverse climates.

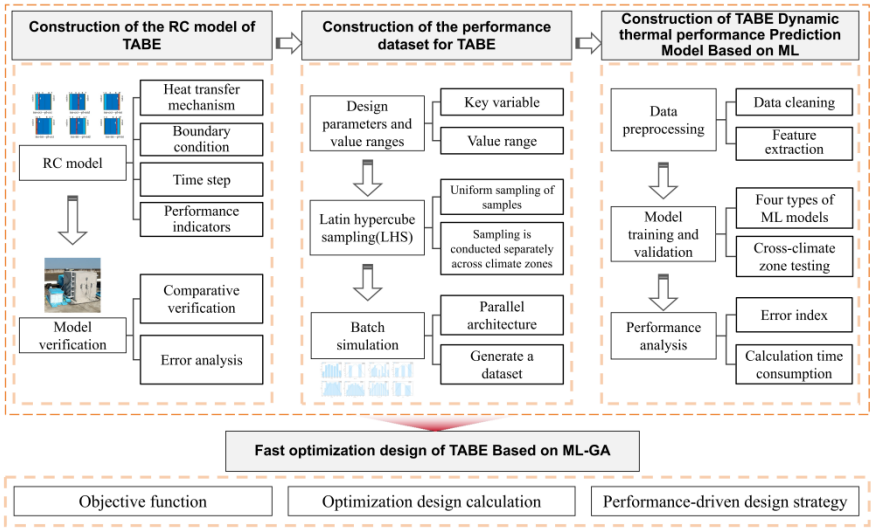


Figure 2. Workflow for constructing the TABS thermal performance surrogate model

2. Method

2.1. RC MODEL DEVELOPMENT FOR TABE HEAT TRANSFER

This study focuses on a typical Thermally Activated Building Envelope (TABE) with embedded hydronic pipes. An equivalent resistance–capacitance (RC) model is

constructed based on the thermal–electrical analogy to generate high-fidelity transient simulation data. The wall assembly consists of an exterior plaster layer, insulation layer, thermal storage layer, embedded-pipe layer, and interior plaster layer, as shown in Figure 3.

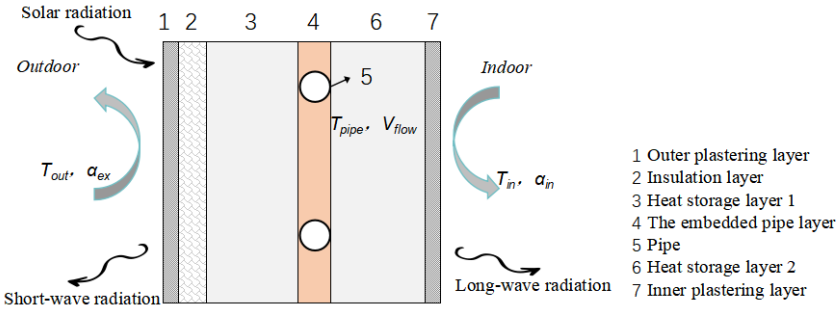


Figure 3 Layered structure of the TABE and its transient heat-transfer process

Energy balance equations are established for each layer, which is treated as a homogeneous thermal node, and interlayer coupling is represented by equivalent thermal resistance and thermal capacitance (Yang et al., 2024). Heat exchange within the embedded-pipe layer is modeled through an equivalent heat source without explicitly solving fluid flow. Outdoor solar and convective effects are combined into an equivalent exterior surface temperature, and the indoor boundary is set as a constant temperature (26 °C in summer and 18 °C in winter) (Kang et al., 2024). Hourly EPW weather data, including dry-bulb temperature and solar radiation, are linearly interpolated to 5-min intervals to capture short-term dynamics. The obtained system of ordinary differential equations is discretized by the backward Euler method and implemented in Python. Its calculation formula is as follows:

$$C_{\mathrm{i}}\frac{T_i^{n+1}-T_i^n}{\Delta t}=\sum_{j=1}^{N}\frac{T_j^{n+1}-T_i^{n+1}}{R_{ij}}+P_i^{n+1}\quad(i=1,\ldots,N) \tag{1}$$

Where C_i (J·K⁻¹) is the lumped thermal capacitance of node i; T_i (°C) is the temperature of node i at time t; R_{ij} (K·W⁻¹) is the thermal resistance between nodes i and j; P_i (W) is the equivalent heat source or sink at node i, representing pipe–wall convective exchange or convective/radiative boundary inputs; and Δt is the chosen time step—set to 300 s.

The resulting system of ordinary differential equations is discretized using the backward Euler method and implemented in Python. Model accuracy is validated against existing experimental data from the research team (Chen et al., 2024) (Chen et al., 2023), showing a maximum absolute error of less than 0.5 °C for interior surface temperature, which confirms the model's reliability for transient TABE simulations.

2.2. CONSTRUCTION OF THE DYNAMIC THERMAL PERFORMANCE DATASET

A high-dimensional dataset is generated to cover diverse climates and design

parameters. Based on sensitivity analyses (Yang et al., 2021)(Schmelas et al., 2015), six key design variables are selected: insulation thickness, insulation position, pipe-layer position, pipe diameter, cooling water supply temperature, and heating water supply temperature (Table 1). Latin Hypercube Sampling is used to generate 1000 uniform design samples. Simulations are conducted across four representative Chinese climate zones: Harbin (Severe Cold Zone, SC), Beijing (Cold Zone, C), Shanghai (Hot Summer and Cold Winter Zone, HSCW), and Guangzhou (Hot Summer and Warm Winter Zone, HSWW). For each combination, the RC model performs year-round simulation to output surface temperatures and heat fluxes. Using multicore parallel computing, approximately 4000 complete samples (105,120 time-steps each) are generated, providing a physically consistent basis for machine learning training.

Table 1 Design Parameters of TABE and their threshold ranges

Type	Input parameter	Abbreviation/symbol	Range	Size	Unit
Wall design	Thickness of insulation layer	d_ins	[20, 50]	5	mm
	Insulation layer position	pos_ins	ext/int	/	/
Embedded pipe design	The position of the pipe embedding layer	pos_pipe	ext/mid/int	/	/
	Embedded pipe diameter	d_pipe	6, 8, 10, 12, 16, 20, 22, 25, 28	/	mm
Operation control	Water temperature in cooling mode	T_w,cool	[15, 23]	1	℃
	Water temperature in heating mode	T_w,heat	[18, 30]	1	℃
	Meteorological conditions	W_city	Shanghai, Harbin, Beijing, Guangzhou	/	/

2.3. MACHINE LEARNING–BASED PREDICTION MODEL

To achieve rapid prediction of TABE dynamic performance, four types of machine learning models are constructed and compared: XGBoost, LightGBM, Random Forest, and a Deep Neural Network (DNN). The model inputs include the six design variables and climate-related features, and the outputs match those produced by the RC model. Continuous variables are standardized using Z-score normalization, and categorical variables are encoded using one-hot encoding. The dataset is split into training, validation, and test sets at a ratio of 70%, 15%, and 15%. Cross-climate extrapolation is used to assess generalization. In each test round, three climate zones form the training set while the remaining one serves as the unseen testing climate. Model accuracy is evaluated using R^2 , RMSE, and MAE (Sun et al., 2024).

2.4. TABE OPTIMIZATION METHOD BASED ON THE SURROGATE MODEL

To evaluate the surrogate model's applicability to climate-adaptive TABE design, it is integrated into a performance-oriented optimization framework (ML-GA). The optimization objective is to reduce both annual cooling/heating loads and system operating energy, forming a multi-objective problem. The six design variables serve as optimization parameters. A Genetic Algorithm (GA) is adopted as the optimization solver. The ML model functions as a fast performance evaluator, avoiding repeated year-long RC simulations. The GA procedure includes population initialization, surrogate-based fitness evaluation, Pareto-based selection, crossover and mutation, and generation of optimal solution sets and design ranges for each climate region. The ML-GA framework significantly reduces computation time and maintains reproducibility, making it suitable for integration into architectural design workflows.

3. Results and discussion

3.1. PERFORMANCE EVALUATION OF MACHINE LEARNING MODELS

This section compares the accuracy and computational efficiency of four machine learning algorithms in predicting the dynamic thermal behavior of TABE. As shown in Figure 4 and Figure 5, all models achieve high accuracy on the test set, with R^2 values above 0.95 and low RMSE and MAE values. XGBoost achieves the optimal balance between fidelity (R^2 =0.983) and speed, reducing annual simulation time from hours to milliseconds. While Deep Neural Networks (DNN) capture nonlinear relationships, gradient boosting models demonstrate superior stability under limited data scales.

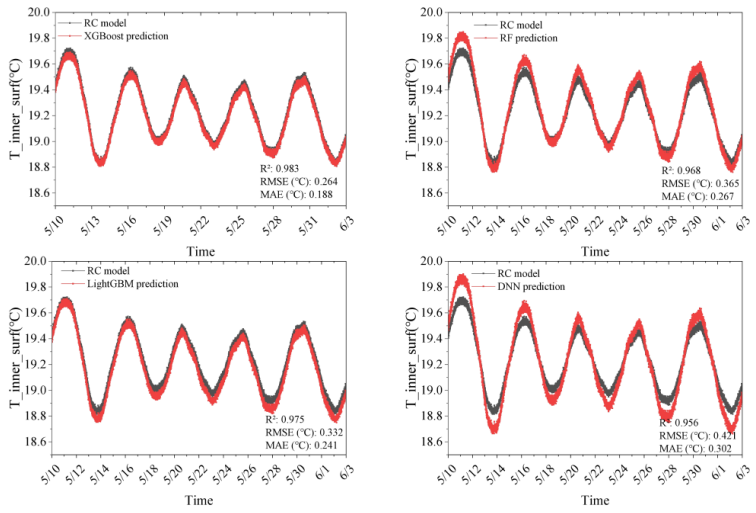


Figure 4 Prediction performance of four machine learning surrogate models

Crucially, XGBoost achieves a computational acceleration of approximately 1915 times per sample compared to the RC model. For a batch of 4,000 samples, parallel computation on a 24-core processor takes only about 80 seconds. This magnitude of acceleration transforms the workflow from a post-design verification process into an

interactive design exploration tool, allowing architects to instantly test thousands of iterations.

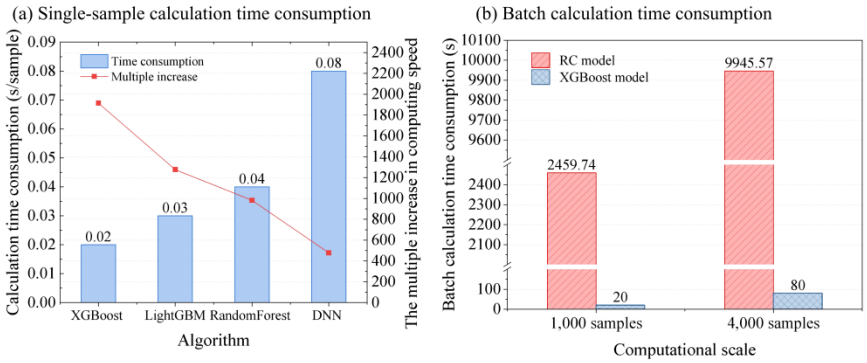


Figure 5 Comparison of computation time between machine learning models and the RC model

3.2. GENERALIZATION ABILITY ACROSS CLIMATE ZONES

Cross-climate extrapolation confirms the model's robustness for regional adaptability analysis. Three climate zones are used for training and the remaining one for testing. As shown in Figure 6, all predictions fall within the ±10% error band, and no systematic deviation or trend is observed in the residuals. This indicates that the surrogate model correctly learns the underlying physics rather than overfitting to specific weather data.

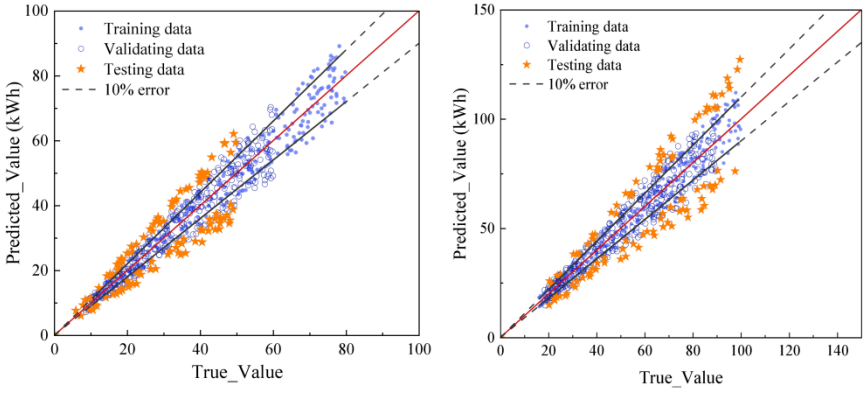


Figure 6 Prediction results of the XGBoost model

Table 2 summarizes the prediction accuracy for interior surface temperature across the four test cities. All R^2 values exceed 0.85, confirming consistent model performance under climate migration. Guangzhou and Beijing achieve the highest accuracy, suggesting that the model generalizes better in climates with relatively stable meteorological patterns. Performance in Harbin is slightly lower (R^2 = 0.87) due to

extreme boundary conditions, yet the accuracy remains sufficient for early-stage design decision-making.

Table 2 Comparison of Cross-climate Test Performance

No.	Training City	Testing city	R^2	RMSE (°C)	MAE (°C)
1	Beijing+Shanghai+Guangzhou	Harbin	0.87	0.65	0.42
2	Harbin+Shanghai+Guangzhou	Beijing	0.91	0.52	0.36
3	Harbin+Beijing+Guangzhou	Shanghai	0.89	0.60	0.39
4	Harbin+Beijing+Shanghai	Guangzhou	0.92	0.48	0.33

3.3. CLIMATE-SPECIFIC OPTIMIZATION DESIGN

Following validation, the surrogate model is integrated with a genetic algorithm to optimize TABE parameters across four climate zones, aiming to minimize annual thermal loads and system energy. The model's millisecond-level prediction speed enables the rapid evaluation of tens of thousands of candidates, facilitating efficient architectural exploration. Table 3 summarizes the recommended parameter ranges for each climate zone. The results reveal clear regional differences in optimal strategies. For example, in hot climates (Guangzhou/Shanghai), placing pipes in the exterior or middle layers effectively intercepts solar gain before it penetrates the structure, acting as an active thermal barrier. Conversely, in cold regions (Harbin/Beijing), interior-side placement is preferred to provide direct radiative heating and rapid response to occupancy demands. Regarding facade and insulation strategies, severe cold regions require thicker insulation layers with flexible positioning options (interior or exterior), whereas hot-summer and warm-winter regions utilize thinner insulation profiles. This indicates that in warmer climates, TABE systems can be integrated with lighter facade assemblies, thereby reducing material costs. The optimized water temperatures demonstrate that moderate supply temperatures are sufficient to maintain comfort when combined with optimized layer positioning, supporting the potential use of low-grade renewable energy sources. Overall, the surrogate-based optimization approach greatly enhances computational efficiency and provides quantitative guidance for TABE design in diverse climates, supporting reproducible workflows for complex active envelope systems.

Table 3 Recommended TABE design ranges for different climate zones

Parameters	Unit	Harbin (SC)	Beijing (C)	Shanghai (HSCW)	Guangzhou (HSWW)
d_ins	mm	45–50	40–50	35–45	20–30
pos_ins	—	int/ext	int / ext	ext	ext
pos_pipe	—	int / mid	mid / int	mid / ext	ext / mid
d_pipe	mm	12–20	12–20	10–16	10–16
T_w,cool	°C	16–20	16–21	18–22	20–23
T_w,heat	°C	25–30	22–28	20–24	18–22

4. Conclusions

This study develops a performance-driven design framework for Thermally Activated Building Envelopes (TABE), addressing the challenges posed by their nonlinear thermal response and the computational latency associated with traditional physical models. By utilizing an RC-based physical model to generate a high-dimensional dataset and replacing hours-long simulations with millisecond-level XGBoost inference, the framework enables rapid, climate-adaptive optimization. Cross-climate evaluations demonstrate that the surrogate model maintains robust transferability, reliably capturing TABE thermal behavior across diverse climatic conditions. Crucially, the study translates numerical optimization into architectural strategies, identifying differentiated design recommendations for distinct climate zones and confirming the effectiveness of data-driven approaches in climate-responsive envelope design.

Overall, this research establishes an integrated computational design workflow covering physical modeling, data generation, performance prediction, and optimization. It effectively bridges the gap between engineering analysis and early-stage architectural decision-making. The framework provides a scalable and accessible method for architects to integrate active envelope logic into the design process, ensuring that climate adaptability and energy efficiency goals are addressed from the outset. Furthermore, the proposed workflow is designed to be encapsulated into visual scripting environments (e.g., Grasshopper), enabling architects without thermal engineering expertise to conduct real-time optimization. Future work will explore multi-dimensional foundation models to address the limitations of the underlying one-dimensional model, extend the workflow to broader climate regions and operational scenarios, and investigate coupling with spatial design processes to build a more generalizable and interpretable intelligent envelope design framework.

Acknowledgements

This work is supported by the National Natural Science Foundation of China Project (No.51608370).

Attribution Statement

ChatGPT (OpenAI, 2025) was used to improve the language of the manuscript.

References

Ahmed, A. A., Sayed, S., Abdoulhalik, A., & Oyedele, L. (2024). Applications of machine learning to water resources management: A review of present status and future opportunities. *Journal of Cleaner Production*, 441, 140715. https://doi.org/10.1016/j.jclepro.2024.140715

Bahrami, A., Soltanifar, F., Fallahi, P., Meschi, S. S., & Sohani, A. (2024). Energy and Economic Advantages of Using Solar Stills for Renewable Energy-Based Multi-Generation of Power and Hydrogen for Residential Buildings. *Buildings*, 14(4), 1041. https://doi.org/10.3390/buildings14041041

Chen, S., Chang, T., & Yang, Y. (2024). Summer thermal and energy performances assessment of a modular hydronic thermal barrier wall for ultra-low energy buildings - A field

experimental study. *Applied Thermal Engineering*, 236, 121491. https://doi.org/10.1016/j.applthermaleng.2023.121491

Chen, S., Chang, T., Yang, Y., He, C., & Gong, Q. (2024). Experimental and numerical study on thermal performances of a modular hydronic thermal barrier wall with heat injection system in filler cavities. *International Journal of Thermal Sciences*, 196, 108727. https://doi.org/10.1016/j.ijthermalsci.2023.108727

Fawaier, M., & Bokor, B. (2022). Dynamic insulation systems of building envelopes: A review. *Energy and Buildings*, 270, 112268. https://doi.org/10.1016/j.enbuild.2022.112268

Kang, Y., Zhang, D., Cui, Y., Xu, W., Lu, S., Wu, J., & Hu, Y. (2024). Integrated passive design method optimized for carbon emissions, economics, and thermal comfort of zero-carbon buildings. *Energy*, 295, 131048. https://doi.org/10.1016/j.energy.2024.131048

Krajčík, M., Arıcı, M., Šikula, O., & Šimko, M. (2021). Review of water-based wall systems: Heating, cooling, and thermal barriers. *Energy and Buildings*, 253, 111476. https://doi.org/10.1016/j.enbuild.2021.111476

OpenAI. (2025). ChatGPT Model: GPT-4 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/

Sabapathy, K. A., & Gedupudi, S. (2022). On the thermal performance of naturally ventilated room with straw insulation retrofitted envelope for different climatic zones of India. *Journal of Cleaner Production*, 342, 130665. https://doi.org/10.1016/j.jclepro.2022.130665

Schmelas, M., Feldmann, T., & Bollin, E. (2015). Adaptive predictive control of thermo-active building systems (TABS) based on a multiple regression algorithm. *Energy and Buildings*, 103, 14–28. https://doi.org/10.1016/j.enbuild.2015.06.012

Shen, Z., Shrestha, S., Howard, D., Feng, T., Hun, D., & She, B. (2023). Machine learning–assisted prediction of heat fluxes through thermally anisotropic building envelopes. *Building and Environment*, 234, 110157. https://doi.org/10.1016/j.buildenv.2023.110157

Sun, Z., Wang, X., Huang, H., Yang, Y., & Wu, Z. (2024). Predicting compressive strength of fiber-reinforced coral aggregate concrete: Interpretable optimized XGBoost model and experimental validation. *Structures*, 64, 106516. https://doi.org/10.1016/j.istruc.2024.106516

Yang, X., Liu, D., Tian, Z., Deng, N., Wang, R., Jiang, Y., Tang, R., & Zong, Y. (2024). Comparison of different control methods on the thermally activated building system (TABS) with large energy flexibility. *Applied Thermal Engineering*, 254, 123863. https://doi.org/10.1016/j.applthermaleng.2024.123863

Yang, Y., Chen, S., Chang, T., Ma, J., & Sun, Y. (2021). Uncertainty and global sensitivity analysis on thermal performances of pipe-embedded building envelope in the heating season. *Energy Conversion and Management*, 244, 114509. https://doi.org/10.1016/j.enconman.2021.114509

Zhao, T., Zhuang, Z., Chang, T., Grunewald, J., & Yu, Y. (2025). Performance-oriented design of thermal-activated building envelope based on multi-objective optimization: Method and application. *Energy and Buildings*, 328, 115198. https://doi.org/10.1016/j.enbuild.2024.115198

Multi-Agent Large Language Model Framework for Window Optimization

Bridging human design intent and data-driven environmental performance

Reza Taghavifard[1] and Negin Cheraghi[2]
[1]*University of Montreal*
[2]*Iran University of Science and Technology*
[1]*reza.taghavifard@umontreal.ca, 0009-0003-5267-2546*
[2]*negincheraqi@gmail.com, 0009-0006-6902-7422*

Abstract. Optimizing window placement and dimensions is crucial to balancing daylight access, glare control, and energy efficiency. However, reconciling multiple, often conflicting objectives remains challenging, particularly when qualitative user intent is excluded from the optimization loop. This paper presents a multi-agent Large Language Model (LLM) framework that integrates natural language reasoning, computer vision, and environmental simulation to generate context-aware, energy-efficient window configurations. The system comprises three interdependent agents: The first agent interacts with users and processes floor plan drawings via a YOLOv8-based detection model. The second agent performs daylight and glare analyses using Honeybee and Ladybug in Python, feeds result to the first agent, and defines objective functions dynamically based on user needs. The third agent executes NSGA-II multi-objective optimization, periodically analyzing the evolving Pareto front and reporting key trade-offs to the first agent, which can update objectives and constraints based on user feedback and trigger early stopping once satisfactory solutions are reached. By embedding user priorities and data-driven insights, the framework interprets and selects optimal configurations from the Pareto front, generating personalized, context-aware recommendations. Results confirm the system's effectiveness in harmonizing technical performance with subjective comfort, reducing decision complexity and advancing intelligent, user-adaptive window optimization.

Keywords. Window optimization, Multi-agent framework, Large Language Models (LLMs), Multi-objective optimization

1. Introduction

As global energy demands grow, improving building performance has become essential for sustainability and emission reduction (Trząski & Rucińska, 2025). Windows critically influence daylight and solar gain(Chi, 2021). factors such as window-to-wall ratio (WWR) and orientation strongly impact thermal and visual

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 397-406.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

comfort, and poor designs may cause glare or uneven daylight(Liu et al., 2021). Energy optimization in buildings requires balancing often conflicting criteria, including energy efficiency, daylight quality, and visual comfort. This study employs NSGA-II to handle trade-offs among these objectives, while extending its applicability beyond expert-driven workflows. (Zhao & Du, 2020). However, such methods still rely heavily on structured numerical inputs and expert supervision, limiting their accessibility for non-expert users and placing the burden of qualitative interpretation on domain specialists(Diakaki et al., 2008).

Large Language Models (LLMs), capable of translating unstructured natural language into structured, quantifiable parameters can bridge this communication gap. (Jiang et al., 2024; Zhang et al., 2024; Zhang & Chen, 2025). By aligning user intent with performance targets such as daylight autonomy and glare probability, LLMs enable interactive, human-in-the-loop optimization workflows that extend beyond rule-based automation toward intent-driven performance exploration.(Bagasi et al., 2025). Integrated with multi-agent systems, they act as mediators between user goals, regulations, and building physics, rather than operating as isolated language interfaces (Hou et al., 2024; Postle & Salingaros, 2025).

This study focuses on two interdependent performance parameters, daylighting and glare, as key drivers of window-related energy performance(Konstantzos et al., 2015). Building on LLM-based automation and multi-objective optimization (MOO), a multi-agent LLM framework is proposed to translate qualitative design intent into quantitative simulation for window optimization, addressing a gap between user-driven reasoning and performance-based modeling that is often unaddressed in conventional optimization or LLM-only approaches. The framework autonomously derives window

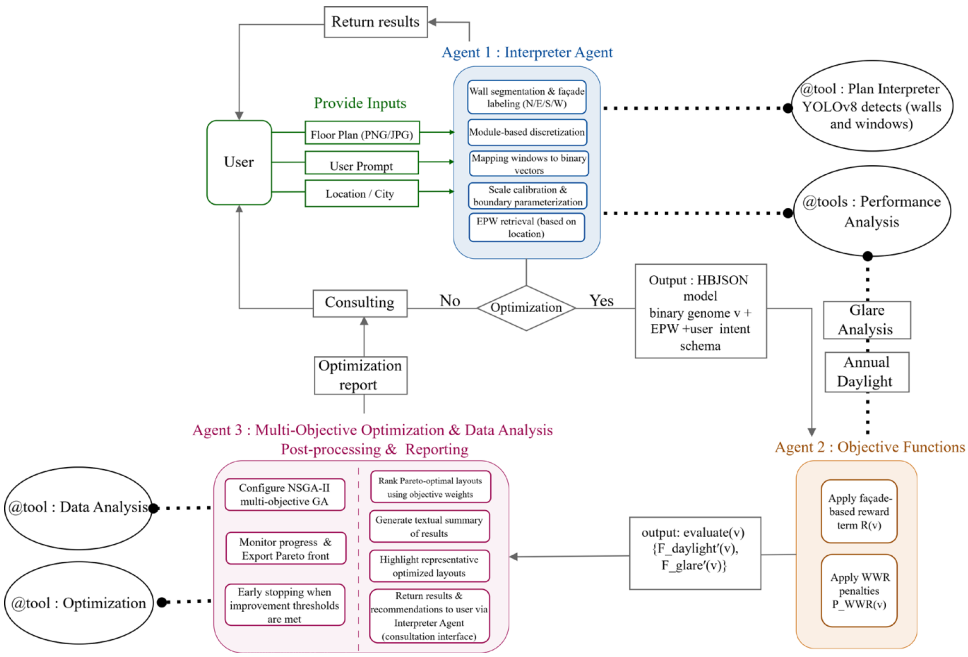


Figure 1.Workflow of the proposed multi-agent LLM framework.

configurations from 2D floor plans, bridging human reasoning and computational intelligence in a way that text-only or parameter-driven optimization approaches cannot. (Figure 1)

2. Methodology

2.1. FRAMEWORK OVERVIEW

This study proposes a three-agent LLM framework that integrates natural-language user intent, computer vision parsing of 2D floor plans, physics-based environmental simulation, and multi-objective optimization, implemented in LangGraph with GPT-4.(LangChain Inc., 2025). Agent A1 captures user intent and converts plans into structured geometry, Agent A2 formulates dynamic objective functions based on performance criteria and user preferences, and Agent A3 performs NSGA-II optimization to generate Pareto-optimal window configurations. The following sections describe the components and present a case study of the full workflow.

2.2. INPUTS

The framework requires a minimal, well-defined set of inputs to ensure robust plan parsing, reproducible simulation, and interpretable optimization. User-provided artifacts include a single-floor architectural raster plan (PNG/JPEG) with clearly defined exterior walls and windows, north alignment, and a known perimeter for scale, along with an EPW weather file that is automatically retrieved based on the user-provided location. Design constraints are specified through global and façade-specific window-to-wall ratio (WWR) bounds (e.g., global WWR $\in$ [0.20, 0.45]; south façade $\in$ [0.15, 0.35]), defining feasible aperture areas, while discrete module or frame sizes (e.g., 0.60 m or 0.90 m with fixed increment steps) reflect fabrication standards (Wright et al., 2014). In addition, qualitative goals are expressed as natural-language statements of comfort or visual objectives, such as maximizing daylight on the south façade or avoiding glare on the west side. These inputs are transformed into a structured schema that guides downstream simulation and optimization.

2.3. A1: INTERPRETER AGENT

The interpreter agent serves as the user-facing interface, capturing design intent and interpreting the floor plan to guide users through analysis or optimization. A GPT-4–based LLM (OpenAI, 2024) conducts multi-turn dialogue to collect inputs, verify scale and orientation, retrieve the EPW file, and elicit façade-specific constraints. When directions are ambiguous (e.g., "exclude the party wall"), the agent requests clarification using cardinal directions (N, E, S, W). Qualitative goals, such as preserving west-facing park views, are then translated into a structured, façade-oriented preference schema, for example:

{"objectives": ["maximize_daylight", "minimize_glare"],
"weights_objectives": {"daylight": 0.6, "glare": 0.4},
"weights_facades": {"N": 0.3, "E": 0.5, "S": 0.2, "W": 0.0},
"wwr_bounds_facades": {"N": [0.15, 0.35], "E": [0.00, 0.2], "S": [0.10, 0.25], "W":

[0.00, 0.00]}}

This schema defines optimization objectives, post-Pareto weighting, façade priorities, hard WWR limits, and discrete module sizes (0.60 m or 0.90 m) for fabrication-based window discretization. Finally, the agent determines whether the user intent corresponds to analysis or optimization mode: in analysis mode, it evaluates the current design and provides natural-language feedback (Section 2.4.2), whereas in optimization mode, it forwards the schema, geometric model, and constraints to downstream agents for multi-objective optimization (Figure 2).

2.4. TOOLS

2.4.1. Plan Interpreter

The Plan Interpreter converts a user-provided raster floor plan into a structured numerical representation of walls and windows for simulation and optimization. Object Detection: A YOLOv8 model(autogyro, 2025) is trained on the CubiCasa5K dataset, detects walls, windows, and doors. Polygon annotations are converted to YOLO-style bounding boxes, retaining only 3 categories. Data are split into training (80%), validation (10%), and test (10%), resized to 1024×1024, and augmented by ±45° rotations(Kalervo Ahtiand Ylioinas, 2019) OCO-pretrained weights, AdamW optimizer, and cosine learning-rate scheduling are used, with performance monitored via COCO-style metrics (mAP@[0.5:0.95], AP_{50})(Lin Tsung-Yiand Maire, 2014). Non-maximum suppression filters detections; wall boxes are reconstructed into polylines, windows snapped to nearest exterior walls, and doors also considered as solid walls. The outer boundary is scaled from the given perimeter and parameterized as a closed curve from a fixed reference corner.

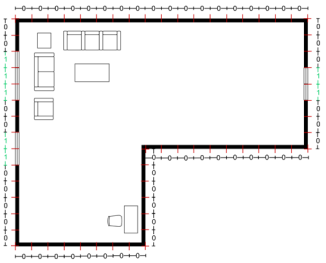
Wall Segmentation & Orientation: Corners split the boundary into wall segments. In this work, we restrict ourselves to orthogonal floor plans. Each segment's outward normal is compared with global axes to assign it to a façade (N, E, S, W).

Module-based discretization: Given a user-defined module size m(e.g., 0.60 m)(Wright et al., 2014), each wall segment i with length $L_i$is discretized into L_i/m one-dimensional slots. For each façade $f \in \{N, E, S, W\}$, this yields a binary vector:

$$\mathbf{v}_f = \left[v_{f,1}, \ldots, v_{f,n_f}\right], v_{f,k} \in \{0,1\},$$

where $v_{f,k} = 0$ denotes opaque wall and $v_{f,k} = 1$ denotes window area at module position $k$on façade f.

Mapping detected windows to modules: Detected windows are mapped onto façade-aligned wall segments as binary vectors (1 = window, 0 = wall), which are concatenated into a single genome used for both optimization and simulation. Window geometry is reconstructed from this encoding to generate Honeybee models and HBJSON files for Ladybug Tools analysis (Figure 2).



Edge	Length (m)	Façade	Binary Vector (window/wall)
1	10.8	N	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
2	8.4	W	[0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0]
3	4.8	S	[0, 0, 0, 0, 0, 0, 0]
4	3.6	E	[0, 0, 0, 0, 0]
5	6.0	S	[0, 0, 0, 0, 0, 0, 0, 0, 0]
6	4.8	E	[0, 0, 0, 1, 1, 0, 0, 0]

Figure 2.Visualized sample floor plan with numerical representation

2.4.2. Performance Analysis

Performance analysis tools are implemented as Python functions that evaluate binary façade encodings using geometric and climatic data to produce scalar performance metrics. The toolbox may combine physics-based simulations and surrogate models to assess daylight (UDI), glare (DGP-based), solar exposure, energy demand, and view indices, with tools selected based on user goals and computational budget. In this study, two physics-based simulations are used: annual daylight performance via UDI Mean (percent of hours with useful daylight) and glare via a DGP-based indicator (percent of hours with DGP > 0.40), both computed using Honeybee and Ladybug Tools (Python–Radiance). A1 reconstructs apertures from façade encodings, runs the simulations, and returns performance metrics, while downstream agents perform optimization.

2.5. A2: OBJECTIVE FUNCTION AGENT

The The A2 Agent receives the binary façade encoding, the HBJSON model, and the user preference schema from A1. Governed by a GPT-4 LLM, it selects the relevant analysis tools and formulates objective functions that combine simulation outputs, arrangement-based rewards, and constraint penalties, and exposes an evaluate(candidate) function for the downstream optimization agent. For a given candidate window layout $\mathbf{v}$, the Objective Function Agent reconstructs the window geometry from the binary vector and matches each entry in the specified objectives to the corresponding performance analysis tool, such as an annual daylight function (UDI Mean) for “maximize_daylight” or a DGP-based indicator for “minimize_glare.” It then calls the selected tools to obtain base performance metrics, for example $D(\mathbf{v})$: representing daylight performance (to be maximized) and $G(\mathbf{v})$: representing glare (to be minimized). At this stage, these metrics reflect only the physical performance of the layout and remain independent of user orientation preferences.

2.5.1. Arrangement-based reward from façade weights

User orientation preferences are encoded in "weights_facades" as non-negative weights $f \in \{N, E, S, W\}$. These weights do not directly scale daylight or glare values; instead, they define a reward term based on window distribution. For a candidate $\mathbf{v}$, the agent computes:

the number of window modules on each façade, $n_f(\mathbf{v})$, the total number of window modules, $n_{\text{tot}}(\mathbf{v})$, and the fraction of windows per façade:

$$q_f(\mathbf{v}) = \frac{n_f(\mathbf{v})}{\max\left(1, n_{\text{tot}}(\mathbf{v})\right)}$$

An alignment score $A(\mathbf{v})$is then defined as:

$$A(\mathbf{v}) = \sum_f w_f \, q_f(\mathbf{v}),$$

which lies in $[0,1]$: it is high when window area is concentrated on high-weight façades and low when windows are placed on façades that the user de-emphasizes or effectively excludes (weight 0). This alignment score is converted into a reward term:

$$R(\mathbf{v}) = \gamma\, A(\mathbf{v})$$

where γ is a scaling factor chosen so that $R(\mathbf{v})$is of comparable magnitude to the main metrics. The reward is then added to the objectives:

Daylight objective (to maximize):

$$F_{\text{daylight}}(\mathbf{v}) = D(\mathbf{v}) + R(\mathbf{v})$$

Glare objective (to minimize):

$$F_{\text{glare}}(\mathbf{v}) = G(\mathbf{v}) - R(\mathbf{v})$$

so that solutions aligning with the user's façade priorities are favored, even when glare values themselves are similar. Façade weights affect optimization via an additional reward/penalty term based on window arrangement, rather than directly scaling physical performance metrics.

2.5.2. WWR-based hard constraints and penalties

The schema also specifies hard WWR bounds per façade via "wwr_bounds_facades". For each façade f, the agent computes the realized WWR, $\text{WWR}_f(\mathbf{v})$, from the binary encoding and checks:

$$\text{WWR}_f^{\min} \le \text{WWR}_f(\mathbf{v}) \le \text{WWR}_f^{\max}$$

Violations are aggregated into a constraint penalty:

$$P_{\text{WWR}}(\mathbf{v}) = \sum_f \big(\max\,(0, \text{WWR}_f^{\min} - \text{WWR}_f(\mathbf{v}))^2 + \max\big(0, \text{WWR}_f(\mathbf{v}) - \text{WWR}_f^{\max}\big)^2\big).$$

The penalized objectives become:

$$F'_{\text{daylight}}(\mathbf{v}) = D(\mathbf{v}) + R(\mathbf{v}) + \lambda P_{\text{WWR}}(\mathbf{v})$$
$$F'_{\text{glare}}(\mathbf{v}) = G(\mathbf{v}) - R(\mathbf{v}) + \lambda P_{\text{WWR}}(\mathbf{v})$$

with λ a large positive factor that discourages WWR violations. Façades declared as non-window façades are therefore strictly enforced. These penalized objectives are encapsulated in a Python evaluation function within a dedicated module, which the Optimization Agent imports and calls as evaluate(v) during the NSGA-II search.

2.6. A3: DATA ANALYSIS AND OPTIMIZATION AGENT

The A3 Agent performs MOO and result post-processing, guided by a GPT-4–based LLM. It receives the binary vector of the plan (design genome), the preference schema (objectives, objective weights, façade weights, WWR), and the evaluation function evaluate(v) constructed by the A2.

2.6.1. Optimization Tool: NSGA-II

The optimization agent applies a standard NSGA-II search on a façade-wise binary genome (0 = wall, 1 = window), with the initial population generated from the current layout and random feasible variants under WWR constraints. New candidates are produced via mutation and crossover, and populations are ranked using non-dominated sorting on penalized daylight and glare objectives, with diversity preserved through crowding distance.

2.6.2. Data analysis

During optimization, the Data Analysis Tool summarizes the evolving design space. It tracks convergence indicators such as hypervolume gains and façade-layout diversity, and groups Pareto-optimal designs into a small number of representative families (e.g., balanced, daylight-oriented, glare-averse). t returns compact tables and ranked shortlists, which A3 packages into short textual reports for A1. A1 then uses these reports in the dialogue with the user (e.g., highlighting that low-glare options darken the core or that most windows cluster on one façade).

2.6.3. Online monitoring and early stopping

For online monitoring, A3 calls the tool every fixed number of generations (e.g., 10) and exports the current Pareto set as a table with façade-level WWRs, daylight and glare metrics, and reward/penalty terms, evaluating improvement and diversity. Instead of automatic stopping, A3 passes the summary to A1, which informs the user and asks whether to continue, stop, or adjust constraints (e.g., shifting emphasis between daylight and glare or excluding façades). A1 updates objective weights, bounds, or filters, while A2/A3 adjust the evaluation function and search; if trade-offs are accepted and further progress is negligible, A3 triggers early stopping and freezes the final Pareto set.

3. Result

3.1. YOLO-BASED DETECTION MODEL PERFORMANCE

The YOLO-based detector achieved strong performance on the validation set, with 74.0% precision, 81.2% recall, and mAP scores of 80.5% (mAP@0.5) and 52.5% (mAP@0.5:0.95). Training curves over 120 epochs show steadily decreasing losses and stabilizing mAP, indicating good convergence without notable overfitting. The normalized confusion matrix further confirms high class-wise accuracy (up to 96% true positives for some classes). Overall, these results demonstrate that the trained YOLO model is efficient and reliable for the target detection task.

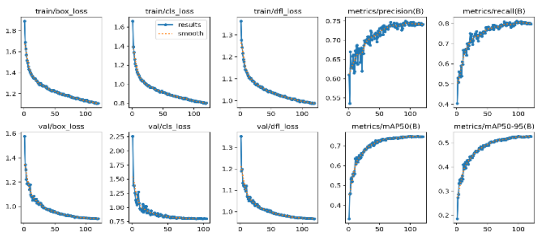


Figure 3 .training process

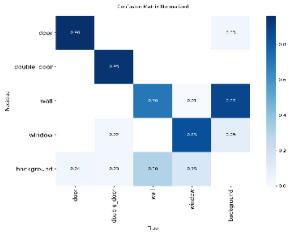


Figure 4 .confusion matrix

3.2. OPTIMIZATION WORKFLOW AND PARETO FRONT GENERATION

To validate the framework, a representative case study was conducted. Agent A1

translated a 2D floor plan and the user query "I need strong daylight for the main area, but I must avoid glare on the west wall where the computer screens are" into a weighted preference schema prioritizing daylight (wd = 0.7) with a façade-specific penalty on west-facing glare. Agent A2 instantiated the evaluation function, and Agent A3 executed NSGA-II while tracking the evolving Pareto front (Figure 5). Unlike conventional workflows that only return a final set of non-dominated solutions, Agent A3 continuously analyzes convergence, diversity, and schema compliance and reports concise summaries to Agent A1, enabling conversational steering. For example, "A3 reports that most low-glare layouts significantly darken the core. Would you accept slightly higher glare on the west wall to brighten the main area?" The user may respond, "Yes, a bit more glare is acceptable if the main area feels brighter overall." Agent A1 then updates the preference schema, prompting Agents A2 and A3 to adjust objectives or trigger early stopping once satisfactory solutions are reached. In the final stage, Agent A3 clusters the Pareto front and recommends a small set of representative designs. Users may further refine results through dialogue (e.g., "Now only show options without windows on the north façade and with compact window patterns"), which is applied as filters to the existing Pareto set, enabling adaptive decision-making rather than static post-processing. To evaluate practical utility, a comparative user study with ten Master's students in Building Energy Engineering compared the proposed framework against a manual MOO workflow and a fixed baseline (UDI = 55%, DGP = 45%) using the same 2D plan and objectives (Figure 6). While the conventional workflow required approximately 450 minutes of computation plus 30 minutes of manual filtering, the proposed framework enabled acceptable solutions in an average of 180 minutes.

Moreover, 40% of participants triggered early stopping and reached satisfactory solutions within 45–115 minutes. Although these solutions occasionally showed slightly lower performance than full optimization, the average time reduction of about 80% (up to 90% for some users) represented a favorable trade-off. Continuous exposure to intermediate Pareto fronts (updated every ~10 seconds) further supported real-time exploration and revealed alternatives unlikely to emerge from static Pareto inspection alone.

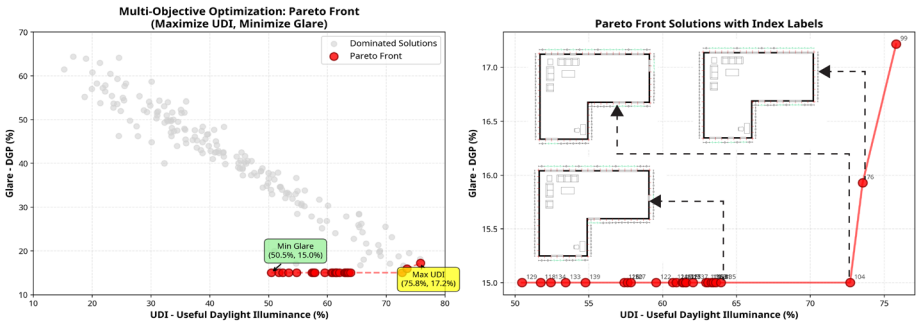


Figure 5 .Pareto front after 50 iterations with three optimized architectural layouts.

4. Discussion, Limitations, and Future Works

The results indicate that the proposed framework substantially outperformed the conventional manual MOO workflow in terms of time efficiency while maintaining competitive performance outcomes. The success of participants U1–U6, who completed the full optimization cycle and often surpassed manual MOO performance, confirms that the preference schema generated by Agent A1 effectively translates qualitative design intent into quantitative optimization weights (wd, wg). The conversational interaction further enabled users to explore solution spaces that are typically inaccessible in static optimization setups. For approximately 20% of participants, dialogue-driven steering redirected the search toward alternatives absent from the original static Pareto set, demonstrating that the framework actively engages users in the optimization process rather than merely post-filtering results.

The current prototype is limited to 2D floor plans, daylight and glare performance metrics, and window layouts discretized into fixed module sizes, which restricts direct applicability to multi-storey, volumetric, or highly irregular façades. Additional limitations include the reliability of LLM reasoning and the potential for error propagation across interacting agents. Future work will extend the framework to 3D geometry and volumetric simulations, integrate additional performance criteria such as thermal comfort and energy consumption, and explore adaptive learning mechanisms that refine preference schemas based on repeated user interactions over time.

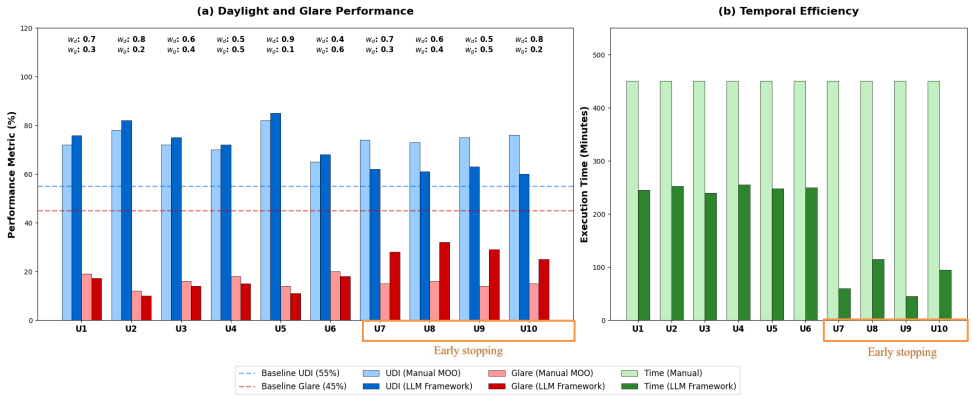


Figure 6 .Quantitative comparison of manual MOO vs. proposed LLM framework.

Attribution Statement

ChatGPT (OpenAI, 2024) was used to refine grammar, wording, and flow.

References

autogyro. (2025). yolo-V8. GitHub. https://github.com/autogyro/yolo-V8

Bagasi, O., Nawari, N. O., & Alsaffar, A. (2025). BIM and AI in Early Design Stage: Advancing Architect–Client Communication. Buildings, 15(12).

Chi, D. (2021). An Approach to Determine Specific Targets of Daylighting Metrics and Solar Gains for Different Climatic Regions. Journal of Daylighting, 8, 1–19.

Diakaki, C., Grigoroudis, E., & Kolokotsa, D. (2008). Towards a multi-objective optimization approach for improving energy efficiency in buildings. Energy and Buildings, 40(9), 1747–1754.

Hou, J., Qiu, K., Zhang, Z., Yu, Y., Wang, K., Capolongo, S., Zhang, J., Li, Z., & Zhang, J. (2024). Wireless-Friendly Window Position Optimization for RIS-Aided Outdoor-to-Indoor Networks based on Multi-Modal Large Language Model. CoRR, abs/2410.20691.

Jiang, G., Ma, Z., Zhang, L., & Chen, J. (2024). EPlus-LLM: A large language model-based computing platform for automated building energy modeling. Applied Energy, 367, 123431.

Kalervo Ahti and Ylioinas, J. and H. M. and K. A. and K. J. (2019). CubiCasa5K: A Dataset and an Improved Multi-task Model for Floorplan Image Analysis. In P.-E. and S. I.-M. and U. J. Felsberg Michael and Forssén (Ed.), Image Analysis (pp. 28–40). Springer International Publishing.

Konstantzos, I., Tzempelikos, A., & Chan, Y.-C. (2015). Experimental and simulation analysis of daylight glare probability in offices with dynamic window shades. Building and Environment, 87, 244–254.

LangChain Inc. (2025). LangGraph. https://github.com/langchain-ai/langgraph

Lin Tsung-Yi and Maire, M. and B. S. and H. J. and P. P. and R. D. and D. P. and Z. C. L. (2014). Microsoft COCO: Common Objects in Context. In T. and S. B. and T. T. Fleet David and Pajdla (Ed.), Computer Vision – ECCV 2014 (pp. 740–755). Springer International Publishing.

Liu, H., Zhang, Z., Ma, X., Lu, W., Li, D., & Kojima, S. (2021). Optimization Analysis of the Residential Window-to-Wall Ratio Based on Numerical Calculation of Energy Consumption in the Hot-Summer and Cold-Winter Zone of China. Sustainability, 13(11).

Postle, B., & Salingaros, N. A. (2025). LLM and Pattern Language Synthesis: A Hybrid Tool for Human-Centered Architectural Design. Buildings, 15(14).

Trząski, A., & Rucińska, J. (2025). Integration of Daylight in Building Design as a Way to Improve the Energy Efficiency of Buildings. Energies, 18(15).

Wright, J. A., Brownlee, A., Mourshed, M. M., & Wang, M. (2014). Multi-objective optimization of cellular fenestration by an evolutionary algorithm. Journal of Building Performance Simulation, 7(1), 33–51.

Zhang, L., & Chen, Z. (2025). Opportunities of applying Large Language Models in building energy sector. Renewable and Sustainable Energy Reviews, 214, 115558.

Zhang, L., Chen, Z., & Ford, V. (2024). Advancing building energy modeling with large language models: Exploration and case studies. Energy and Buildings, 323, 114788.

Zhao, J., & Du, Y. (2020). Multi-objective optimization design for windows and shading configuration considering energy consumption and thermal comfort: A case study for office building in different climatic regions of China. Solar Energy, 206, 997–1017.

Climate-Embedded Building Systems

A Co-Evolutionary Approach to Multi-Objective Optimization in Architecture

Jonathan Wong[1], Puja Bhagat[2], Milad Showkatbakhsh[3], and Mohammed Makki[4]
[1,2,3]*Architectural Association School of Architecture*
[4] *University of Technology Sydney*
[1] *jonathan.wong2@aalumni.org, 0009-0007-7014-0673*
[2] *puja-h.bhagat@aalumni.org, 0009-0002-6286-822X*
[3] *Showkatbakhsh@aaschool.ac.uk, 0000-0001-5440-5175*
[4] *mohammd.makki@uts.edu.au, 0000-0002-8338-6134*

Abstract. Co-evolutionary algorithms (CoEAs) mitigate the limitations of multi-objective evolutionary algorithms (MOEAs) by discretizing a complex and high-dimensional problem into interconnected elements. While the application of CoEAs in other fields is well established, CoEAs find little resonance in the architectural industry. The research compared the capabilities of CoEAs and MOEAs in resolving complex design scenarios through the simultaneous optimization of a tower morphology, its public-private spatial distribution, and its structural system in Hong Kong. The research optimized the design scenario through a CoEA workflow by decomposing the design into discrete, interconnected parts. The workflow employed a novel, cooperative CoEA developed by the authors, which continuously alternated between evaluating each population independently and both collectively. In parallel, a traditional MOEA workflow, where populations evolved as a single system, was employed using the multi-objective optimization engine Wallacei as a baseline comparison. A comparison of the results from each workflow highlighted CoEAs' capability to better rationalize complex optimization challenges and generate balanced solution sets. By decomposing a complex system into its interconnected parts, the research demonstrated how CoEAs facilitate deeper understandings of design problems by emphasizing the relational qualities between components, thereby enabling designers to better frame the problem as compared to MOEAs.

Keywords. co-evolutionary algorithms, evolutionary algorithms, generative design, evolutionary computation, multi-objective optimization

1. Introduction

Architecturally, multi-objective evolutionary algorithms (MOEAs), afford

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 409-418.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

architects the ability to quantitatively reconcile conflicting design criteria to achieve a series of optimized design options. MOEAs computationally mimic Charles Darwin's Theory of Evolution by evolving a design over time in response to specific performance criteria (Popovici et. al. 2012). Through MOEAs, architects efficiently explore the depth and breadth of a design space. Despite these capabilities, MOEAs in architectural applications have several limitations under certain conditions. Firstly, MOEAs experience 'dominance resistance' as the number of objectives in the design problem increases, resulting in the majority (if not all) of the design solutions in the later generations forming the Pareto Front and impacting the algorithm's effectiveness in converging towards the true Pareto Front (Hanne and Moghaddaam, 2025). Furthermore, MOEAs generally fail to develop reasonable design solutions when the complexity of such a search space increases beyond a limit of rationality (Boonstra et. al., 2020). Lastly, MOEAs operate using strictly individualized, quantitative fitness objectives, rather than relational ones, which decontextualizes them from their design space (Potter and Jong, 2000, p. 2). These limitations, which are frequently present in architectural scenarios, hinder MOEAs applications to isolated aspects of a design workflow.

Co-evolutionary algorithms (CoEA) mitigate these issues by simultaneously discretizing a complex, high-dimensional problem into manageable elements. This method mimics natural systems by considering two or more interrelated populations simultaneously. It differs from a traditional MOEA in that the adaptation of each population's individuals is not only informed by its own fitness criteria but also that of other populations (Antonio and Coello Coello, 2018, p. 851). As such, a CoEA places more importance on the relationship between populations as opposed to each in isolation. The relationships between these interrelated populations take four forms: competitive, parasitic, commensalistic, and cooperative (Antonio and Coello Coello, 2018, p. 852). Through these relationships, a CoEA expands the optimization process by contextualizing the individualized performance objectives within a broader interconnected system.

By decomposing a complex system into its interconnected parts, a CoEA empowers architects and designers to have greater agency in formulating a design environment for multi-objective optimization and provide more detailed insights into the independent behaviours of different elements within the design. Thus, CoEAs facilitate a deeper understanding of the design problem by emphasizing the relational qualities between its components, thereby enabling designers to better frame and structure the problem as compared to traditional MOEAs.

2. Evolutionary Algorithms in Architecture

There exists substantive literature on the applications of MOEAs for optimizing architectural elements, including building envelopes, structural systems, etc. (Castrejon-Esparza et. al., 2025). Additionally, architectural research, such as Showkatbakhsh and Makki (2022), examined the use of MOEAs through an emphasis on how the design problem is formulated and what mechanism is used to select a single solution from the evolved population. Thus, the domain of architecture has well adapted MOEAs to solve open-ended design challenges, but

the MOEA limitations discussed previously restrict their application to early-stage design and limit their use in complex scenarios (Boonstra et. al., 2021, p. 3).

Potential solutions to these issues have been explored in computer science such as SANE and CoSOMEA (Ma et. al., p. 434). However, while the field has utilized and advanced CoEAs to solve complex problems for decades, architectural applications of CoEAs have been scarce. As such, the research builds upon the existing frameworks developed within the computer science field and attempts to build a novel workflow applicable in an architectural setting to mitigate the challenges of existing MOEA methodologies.

3. Principles of Co-Evolutionary Algorithms

In 1994, Potter and De Jong proposed the first cooperative CoEA which combined a divide-and-conquer approach with collaboration mechanisms to address certain optimization challenges in computer science. Their proposed algorithm decomposed the problem into interconnected subcomponents which optimized independently and conceptually reassembled into a complete solution for the problem (Potter and De Jong, 2000). The fitness for an individual of a subcomponent was evaluated in relationship to the complete solution (Potter and De Jong, 1994). Since each subcomponent was derived from the same element, their individual optimizations required a level of collaboration to converge on a universal solution. The advantages of their proposed CoEA over MOEAs mainly lied in the robustness of the solutions sets and ability to address the 'curse of dimensionality' that comes with complex problems (Ma et. al., p. 421-422). Particularly, the CoEA mitigates the limitations of MOEAs discussed previously by reducing the overall complexity of the optimization process through localized optimizations with interrelated groupings of objectives.

Extracting from these principles, the research translated the theoretical concepts into a novel computational framework for simulating the basis of co-evolution. Particularly, the authors applied the developed algorithm to the field of architecture to mitigate the design challenges faced by MOEAs.

4. A Novel, Cooperative Co-Evolutionary Algorithm for Architecture

Due to the limited application of CoEAs in an architectural domain, the authors developed a novel, cooperative CoEA as a computational workflow scripted in Grasshopper3D. The CoEA setup assumed three conditions. First, the parametric model defined two interconnected elements of the design, each with their own unique performance objectives. These two elements also shared a common element between them with its own objectives. For example, an office tower and a residential tower positioned within a masterplan would influence each other's performances by sharing the same plot of land and impacting their respective fitness objectives. Next, the generation of each element was defined by a series of independent chromosomes, meaning none are shared directly. Different elements may have the same type of chromosomes, such as the tower height, or influence the chromosomes of another, but the same chromosomes cannot control both elements. Lastly, the fitness criteria of the two related elements were defined as

competitive, parasitic, commensalistic, or mutualistic, while those of the shared design component were neutral, not prioritizing either element. By formulating the design problem in this manner, the design space and optimization process were conceptualized as two major populations with close and evident relationships evolving in tandem. Thus, the evolutionary growth of one group influenced the other, and vice-versa, closely mimicking the principles of co-evolution in nature.

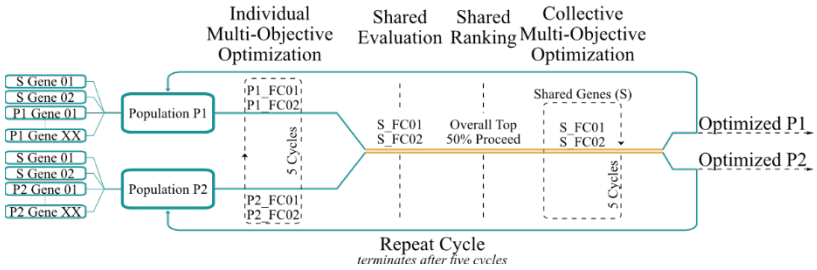


Figure 1. Novel, Cooperative CoEA Workflow

With this setup, the algorithm iteratively evolved the design by cycling between two states: independent optimization of each population and collective optimization of both populations (Figure 1). For a fixed number of generations, each population optimized independently from one another according to their respective fitness criteria and those of the shared component. This optimization process followed traditional MOEA principles of evaluation, selection, crossover, and mutation. Afterwards, the algorithm combined the two populations, evaluating and optimizing the combined population only according to the fitness criteria of the shared element. The collective optimization iterated through a fixed set of generations, again following MOEA principles, to realign the performance of the two populations. The two populations then separated again before starting a new cycle of independent optimization. The algorithm allowed the user to define the number of generations for each state, the number of cycles within the entire simulation, and the number of individuals in each generation. After the simulation terminated, the algorithm produced detailed data about the simulation, including the genes of every generated individual, their evaluated performances for each fitness criteria, and the non-dominated individuals, or Pareto Front members.

5. Experiments

The research explored the architectural applications of CoEAs through the simultaneous optimization of a tower morphology, its public and private space distribution, and its structural system. Two experiments were conducted, an initial exploration of the novel CoEA using a tower segment and a more comprehensive experiment exploring the algorithm's application to a complex, architectural design of the entire tower. To quantify the limitations and benefits of the CoEA, both experiments conducted a MOEA on the same setup as a baseline comparison.

5.1. Exploration of CoEA Principles and Mechanisms

The initial experiment developed, tested, and tuned the novel CoEA by focusing its application on a single tower segment with an exterior structural system and a voxelized distribution of spatial types. The two, interconnected populations for the CoEA were the structural system and public-private distribution. The tower segment acted as the shared element between them. The experiment aimed to generate an optimized design solution which balanced the highly conflicting fitness criteria of the two populations. The authors parametrically generated and analysed all the geometry for the design scenario in Grasshopper3D. Three chromosomes controlled the tower segment form, abstracted as a box: length, width, and height. Its performance was evaluated according to minimized surface area-to-volume ratio, maximized gross floor area, and maximized shading. The structural system, a simplified exoskeleton, was controlled by five chromosomes related to structural member sizing and connection conditions. Its optimization was driven by fitness criteria of minimized embodied carbon and minimized deformation under gravity and wind loads. Finally, the public-private distribution was generated by voxelizing the box and applying two random walker algorithms, one for each space type. Four chromosomes controlled the random walker related to the start locations and growth of each zone. Its performance was evaluated for maximized public space percentage and maximized private space density.

With the parametric model setup, the experiment employed the authors' CoEA workflow to iteratively generate and evaluate various regions of the design space and develop an optimized solution set. The CoEA utilized the described cycle of independent and collective optimizations. The independent optimization iterated for five generations with ten individuals per generation and the collective optimization iterated for five generations with twenty individuals each. Afterward, the two populations separated, and a new cycle was initiated.

During the independent optimization generations, the structural system and public-private distributions were optimized for their respective fitness criteria following a traditional MOEA. When the two populations were optimized collectively, all the individuals were evaluated and optimized according to the shared fitness criteria of the tower segment morphology, rather than their distinct fitness criteria, with the intent of realigning their optimizations with one another. This process again followed a traditional MOEA. The CoEA algorithm completed five cycles of independent and collective optimization, totalling 50 generations for the entire optimization process. In total, the algorithm ran for 4.5 hours. Once completed, the researchers extracted and evaluated the Pareto Front members.

In parallel, a MOEA was employed to optimize the same parametric model and fitness criteria using the Grasshopper3D plugin Wallacei. Instead of handling each element of the design as independent populations, the MOEA workflow attempted to optimize all seven fitness criteria simultaneously for 50 generations with 10 individuals in each generation. By comparison, the MOEA ran for a total of 4 hours. The Pareto Front members were extracted from the solution set to be utilized as a baseline comparison with the CoEA workflow results.

5.2. ARCHITECTURAL APPLICATION OF THE COEA

After completing the initial exploratory experiment, the research built upon its results to broaden the application of the CoEA to a more complex and realistic architectural design challenge. In the same manner as the first experiment, the authors applied the CoEA workflow to an entire tower morphology, its complete structural system, and its internal public-private distribution. The structural system and public-private distribution were considered as the two, interconnected populations for the CoEA, and the tower acted as the shared element between the two populations. The algorithm did not change between experiments, but an emphasis was placed on creating more robust genes and objectives which could facilitate a better navigation of this experiment's more complex design space.

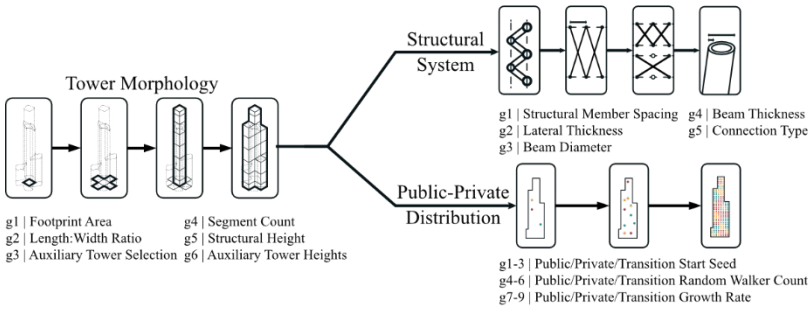


Figure 2. Architectural Application Experiment Setup

The parametric model from the initial experiment was expanded upon by adding additional chromosomes and fitness criteria to more reasonably mirror a typical architectural design environment (Figure 2). In turn, this increased the complexity of the design space by a magnitude of nine. The morphology of the full tower was controlled by seven chromosomes related to its footprint dimensions, segment count, and auxiliary support tower measurements. Its performance was evaluated according to maximized total daylight hours, minimized wind angle deflection, and maximized floor-to-area ratio. The structure, an exoskeleton encasing the tower morphology, was controlled by five chromosomes related to structural member dimensions, spacing, and connectivity. As with the initial experiment, the structural system performance objectives were minimized embodied carbon and minimized deformation under gravity and wind loads. The distribution of public and private spaces throughout the tower were similarly generated according to the initial experiment's voxelization and random walker methodology. Whereas the initial experiment utilized two random walkers, this experiment employed three, one for the public space, one for the private space, and a third for a transitional space between the two. Each random walk algorithm was controlled by three chromosomes, totalling nine for the entire genealogy: the starting seed, the number of starting points, and the growth rate for the voxels. The public-private distribution performance was evaluated according to maximizing the distribution and proximity of each space type to its context.

Utilizing the authors' CoEA computational workflow, the experiment optimized the tower morphology, the structural system, and the public-private distribution for 75 generations with 10 individuals per generation. The

optimization process of the CoEA alternated between 10 independent and five collective optimization states. Taking lessons from the initial experiment results, the experiment placed a higher emphasis on the individual fitness criteria of the public-private distribution and the structural system, as opposed to the shared fitness criteria of the tower morphology, by introducing more generations of independent optimization than the collective. This facilitated a better opportunity to traverse the more complex search space and obtain higher performing solutions for both independent and shared criteria. The cycle of independent and collective optimization continued for five iterations, totalling 15 hours of runtime. After the CoEA completed, the Pareto Front members were extracted from the solution set, and the gene distribution and fitness criteria performance were mapped over time.

As with the initial experiment, a parallel optimization was conducted on the same parametric model solely utilizing a traditional MOEA through the Grasshopper plugin Wallacei, where all ten fitness criteria were optimized at the same time. The MOEA ran for 75 generations with 10 individuals in each generation, reaching a total runtime of 13 hours. The Pareto Front members and corresponding data were identified as a baseline comparison with the CoEA data.

6. Results

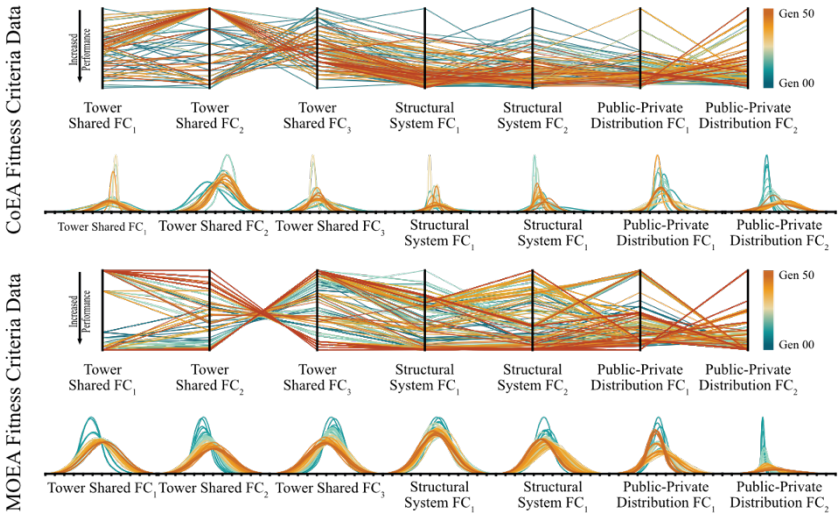


Figure 3. Experiment 01 CoEA and EA Fitness Criteria Data

The experiment results, in comparison with those of the traditional MOEA, emphasized the ability for a cooperative CoEA to be a powerful tool for simultaneously optimizing interconnected parts of a complex system. The CoEA optimizations converged towards an 'ideal team' scenario, where no one fitness criteria converged towards its ideal range nor did any fitness criteria fail to optimize at all. Instead, all fitness criteria optimized moderately together, such that the performance of solution set was balanced across all fitness objectives.

The initial experiment's results showcased the potential viability for

optimizing complex systems using the described CoEA framework. While the results did not show a perfect optimization, there was in fact a convergence of the two populations in relation to the shared fitness objectives. On average, the convergent individuals in the last few generations performed well in accordance with all criteria, including both shared and individual objectives, as seen in the parallel coordinate graphs and mean value graphs (Figure 3). Beyond that, the standard deviation graphs showcased that both population's shared parameters converged to a high performance for all criteria (Figure 3). While these converged populations may not have had the best performance in any one fitness criteria, the fact that all objectives had a relatively high performance within the data set exemplified that both populations engaged in an ideal team scenario.

On the other hand, the MOEA data set revealed none of the fitness criteria, neither individual nor shared, converged meaningfully (Figure 3). While the simulation found some individuals who performed well in one or more criteria, these solutions also performed very poorly for the remaining objectives. This exemplified an individualistic mindset in optimization, which ultimately failed to produce well-fit individuals within the complex search space and time span provided. In comparison, the novel CoEA computational workflow, on average, higher performing individuals across all fitness objectives. Such an occurrence is likely be attributed to the fact the CoEA introduced collectively evolved generations, which re-aligned the design space. This enabled the optimization to identify better performing individuals which were not found during the traditional MOEA due to the scale and complexity of its unfiltered design space (Figure 5).

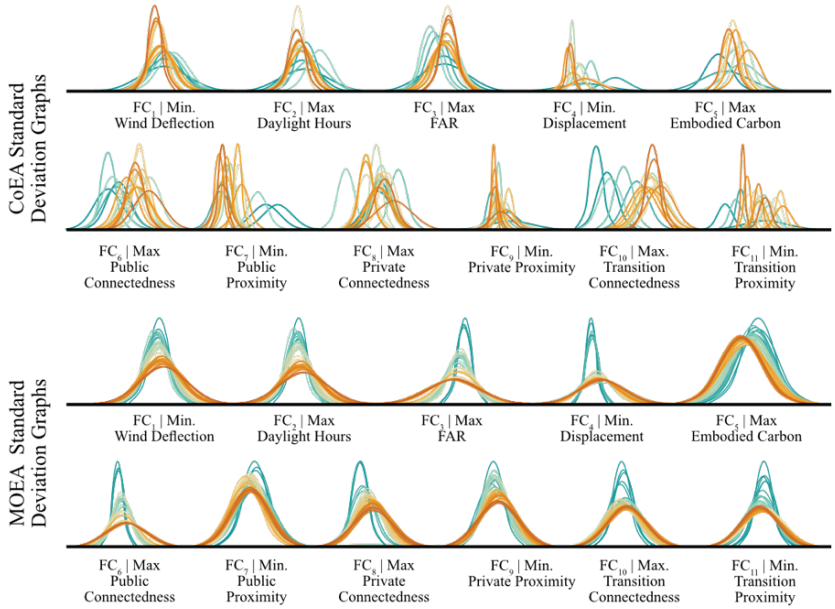


Figure 4. Experiment 02 CoEA and EA Fitness Criteria Data

The research identified a similar phenomenon in the larger architectural scale. Even with a significantly larger search space of nine extra chromosomes and four

additional objectives, the results showcased a similar convergence in the towers evolved using the CoEA workflow and a similar lack of optimization for the MOEA evolved towers. This is clearly showcased in the Pareto Front members alongside the graphs, where the CoEA workflow produced high performing and phenotypically diverse towers (Figure 4 and 6). Thus, the CoEA workflow explored a much broader design space to find optimal solutions which balanced multiple complex and interconnected systems. This convergence exemplified how CoEAs enable discrete yet interconnected systems to gain a cooperative mindset when optimizing complex problems.

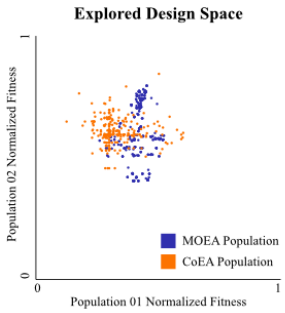


Figure 5. CoEA & MOEA Explored Design Space

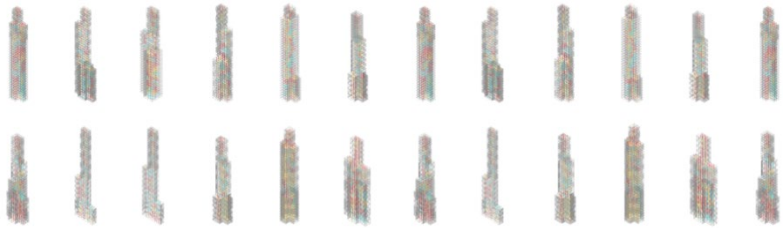


Figure 6. CoEA Pareto Front Members

7. Discussion

By comparing a CoEA and MOEA in the multi-objective optimization of an architectural design scenario, the research highlighted how a CoEA can balance a high degree of conflicting performance criteria within an extremely complex design space. The resultant 'ideal team' scenario of the CoEA emphasized how it is more well suited to addressing the real-world, circumstances of architectural challenges. While the research examined a cooperative, co-evolutionary scenario, future research could investigate other scenarios, such as competitive, parasitic, or commensalistic. Each type of CoEA would require variances and adaptations of the proposed computational workflow due to the changing relationship between the interrelated elements of the design. Additionally, the setup for each CoEA type would require new assumptions, particularly in terms of the relationship between the fitness criteria of each population. Such future research would expand upon the potential for CoEAs to address new design challenges in the built environment that are currently outside the suitable domain of MOEAs.

A comparison of the Pareto Front members and the data set between the CoEA and MOEA optimization results provided a strong basis for understanding the performance of the CoEA, but the research acknowledges the limitations of such an approach. A direct comparison cannot be drawn due to the technical differences in genetic algorithms utilized by the authors' CoEA and Wallacei's MOEA. Further research could pursue a one-to-one comparison using the same, base genetic algorithm. With that said, a comparison with a leading, architectural MOEA tools highlighted the advantages of CoEAs in complex design scenarios.

8. Conclusion

While MOEAs face limitations in highly complex design problems, the research showcased the capabilities of CoEAs to more adequately address these challenges through a relational optimization process. The novel, cooperative CoEA developed by the authors facilitated this contextualization of the optimization through an iterative, back-and-forth workflow between independent and collective populations. Through such a technique, the solution set of the CoEA demonstrated an 'ideal team' scenario, where the optimization of the conflicting objectives converged along a common middle ground. Each fitness criteria optimized moderately, but no single one converged towards their optimal performance ranges nor failed to improve at all. Such a notion of an 'ideal team' scenario acknowledges the complexity of real-world scenarios. Through this research, the authors showcased how a co-evolutionary approach to multi-objective optimization in architecture facilitates a more balanced cooperation between multiple conflicting parameters, allowing more robust solutions to emerge within highly complex design spaces.

References

Boonstra, S., Van Der Blom, K., Hofmeyer, H., & Emmerich, M. T. M. (2021). Hybridization of an Evolutionary Algorithm and Simulations of Co-Evolutionary Design Processes for Early-Stage Building Spatial Design Optimization. Automation in Construction, 124, 103522. https://doi.org/10.1016/j.autcon.2020.103522

Castrejon-Esparza, N. M., González-Trevizo, M. E., Martínez-Torres, K. E., & Santamouris, M. (2025). Optimizing Urban Morphology: Evolutionary Design and Multi-Objective Optimization of Thermal Comfort and Energy Performance-Based City Forms for Microclimate Adaptation. Energy and Buildings, 342, 115750. https://doi.org/10.1016/j.enbuild.2025.115750

Ma, X., Li, X., Zhang, Q., Tang, K., Liang, Z., Xie, W., & Zhu, Z. (2019). A Survey on Cooperative Co-Evolutionary Algorithms. IEEE Transactions on Evolutionary Computation, 23(3), 421–441. https://doi.org/10.1109/TEVC.2018.2868770

Miguel Antonio, L., & Coello Coello, C. A. (2018). Coevolutionary Multiobjective Evolutionary Algorithms: Survey of the State-of-the-Art. IEEE Transactions on Evolutionary Computation, 22(6), 851–865. https://doi.org/10.1109/tevc.2017.2767023

Popovici, E., Bucci, Anthony, Wiegand, R.P. & de Jong, Edwin. (2012). Coevolutionary Principles. In Rozenberg, G., Bäck, T., Kok, J.N. (Eds.), Handbook of Natural Computing. (1st ed., pp. 987-1033). Springer.

Potter, M. A., & Jong, K. A. D. (2000). Cooperative Coevolution: An Architecture for Evolving Coadapted Subcomponents. Evolutionary Computation, 8(1), 1–29. https://doi.org/10.1162/106365600568086

Potter, M.A., De Jong, K.A. (1994). A Cooperative Coevolutionary Approach to Function Optimization. In Davidor, Y., Schwefel, HP., Männer, R. (Eds.), Parallel Problem Solving from Nature — PPSN III. (1st ed., pp. 249-257). Springer.

Showkatbakhsh, M., & Makki, M. (2022). Multi-Objective Optimisation of Urban Form: A Framework for Selecting the Optimal Solution. Buildings, 12(9), 1473. https://doi.org/10.3390/buildings12091473

Hanne, T., & Moghaddam, M. (2025). A Review of the Evolution of Multi-Objective Evolutionary Algorithms. Computers, Materials & Continua, 85(3), 4203–4236. https://doi.org/10.32604/cmc.2025.068087

A Space Habitat

Generative Urban Systems

Ling Kit Cheung[1], Mohammed Makki[2] and Milad Showkatbakhsh[3]
[1,2]University of Technology, Sydney.
[3]Architectural Association
[1]jckira@hotmail.com.hk, 0000-0001-5040-8686
[2]mohammed.makki@uts.edu.au, 0000-0002-8338-6134
[3]showkatbakhsh@aaschool.ac.uk, 0000-0001-5440-5175

Abstract. This research proposes a scalable orbital habitat designed to ease humanity's environmental impact by relocating a portion of human activity into space. The research explores how a self-sustaining system in Earth–Moon orbit could support large populations while remaining closely linked to Earth. Building on life-support principles developed for the International Space Station, the study scales closed-loop cycles for air, water, waste, and food to accommodate up to two hundred thousand inhabitants, supported by solar energy and regenerative ecological systems. The colony's structure is organised through a design database informed by NASA research, categorising modules into residential, public, infrastructural, and life-support classes. These modules are combined through a voxel-based workflow and arranged using a generative system influenced by neighbour rules and urban conditioning. A parametric model, integrated with a multi-objective evolutionary algorithm, optimises life-support performance, energy capture, and spatial qualities, while a weighted selection model evaluates connectivity and feasibility. The results demonstrate the value of combining rule-based growth, performance evaluation, and designer-led selection in shaping large-scale orbital habitats.

Keywords. Space colonisation, Orbital Habitat, Evolutionary algorithm, Voxel-based modelling, Modular urbanism

1. Introduction

The growing vulnerability of humanity in the face of climate change, overpopulation, and ecological collapse has reframed space not just as a frontier for exploration, but as a potential response to these escalating challenges. As Earth's ecosystems edge toward

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 419-428.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

critical tipping points, relocating part of human habitation beyond our planet is emerging as a speculative yet increasingly pragmatic way to reduce pressure on Earth.

Ceylan (2018) situates space colonisation within speculative and cultural narratives, viewing it as an allegory for Earth's ecological anxieties. This research draws on that perspective, framing orbital habitation as both a technological experiment and an ethical reflection on the state of the planet, envisioning a self-sustaining orbital habitat designed to lighten humanity's ecological footprint on Earth. Rather than portraying space colonisation as an escape from Earth, it should be positioned as an ecological redistribution, an auxiliary ecosystem that is intertwine and operating in symbiosis with Earth.

Using the International Space Station (ISS) as a key precedent, the proposed space colony scales its closed-loop life-support principles to support approximately 200,000 inhabitants, reducing dependence on Earth through regenerative air, water, and waste cycles powered by solar energy. Situated within a cis-lunar orbit, the colony maintains proximity to both Earth and the Moon, forming a tri-nodal system for optimum transport and resource transportation. Artificial gravity, achieved through the Coriolis Force of rotation, is used to sustain long-term human health on both physical and physiological levels.

Architecturally, the habitat is composed of modular voxels generating the basis of the colony's spatial structure using data from NASA's research led by study director Gerard K. O'Neill (Billingham et al., 1977). A generative algorithm based on Conway's Game of Life Cellular Automata governs the system's aggregation and growth, while a multi-objective evolutionary algorithm (MOEA) optimises class ratios and environmental performance.

This research explores an integrated generative and evolutionary system to produce adaptive, quantifiable, and ecologically resilient space habitats that serves as a modular, self-sustaining system that can be an extension of Earth's ecology in space, reducing humanity's footprint on the planet while remaining capable of continuous growth and evolution.

2. Background and Context

Models for space colonisation emerged as early as the 1960s, when Gerard K. O'Neill questioned whether Earth was "the best possible place for an expanding technological civilization." In his seminal paper 'The Colonization of Space', O'Neill envisioned large, self-sustaining orbital habitats that could support permanent human settlement beyond Earth (O'Neill, 1974). He proposed that such colonies could relieve pressure on the planet's limited resources while preserving ecological balance through closed life-support systems.

Building on O'Neill's vision, the ISS provides the foundation for the proposed colony's orbital sustainability. Though yet to be tested on a larger scale, its Environmental Control and Life Support System (ECLSS) closed-loop processes allow for life sustaining in microgravity (Seedhouse, 2020). Scaling these systems to a population multiple times the original size transforms technical maintenance into an architectural problem of spatial and ecological integration, which in this experiment, will be overlooked by allowing regular but minimal support from Earth.

Smith (2016) expressed the importance for the human race to adapt to the new lifestyle of space settlement through adaptive evolution, highlighting the eventual split of space settlement from space exploration with increased necessity and urgency. Studies such as 'Kalpana One' (Arora et al., 2006) and Sharma's 'Space settlement design study - An overview' (Sharma & Mahajan, 2013) addressed artificial gravity, mass distribution, and radiation protection but offered limited spatial and urban resolutions. More recent studies (Giang et al., 2023) propose martian habitats that utilise computational systems based on sequential evolutionary algorithms that propose modular habitation for a scalar and progressive population.

This research examines a data-driven computational methodology aimed at generating settlement proposals that integrate essential logistical functions with articulated spatial and architectural qualities.

3. Method

3.1. GENERATIVE LOGIC: CELLULAR AUTOMATA (CA) AS SPATIAL SYSTEM

The research's generative framework is adapted from John Conway's 'Game of Life' (Games, 1970), a cellular-automaton model where local neighbour rules create bottom-up emergent global order. Spatially, the system operates through a three-dimensional voxel grid in which each voxel represents a modular unit belonging to one of four functional classes: residential, public, infrastructural, and life-support.

Each voxel's state (active, inactive, or transformed) is determined by adjacency conditions encoded with programmatic and ecological relationships. Through iterative rule applications, these voxels that are intertwined with neighbouring dependencies generate adaptive growth patterns, forming urban clusters and corridors that resemble ecological networks.

Implemented in Grasshopper, the 3D Cellular Automata algorithm runs as a rule-based generator that evolves voxel populations over multiple cycles. Each iteration produces a new spatial configuration that reflects the balance between density, connectivity, and ecological proximity. By making small changes to the parameters that define the proposed system, each design solution will have drastic variations in the form of the final design and criteria performances. In this way, the Cellular Automata logic becomes a simulation of architectural evolution, producing morphologies capable of continual reorganisation as the urban quality parameters shift.

3.2. MULTI-OBJECTIVE EVOLUTIONARY ALGORITHM

To evaluate and refine these generative results, the NSGA-2 (Deb et al., 2000) multi-objective evolutionary algorithm, through the grasshopper plugin WallaceiX (Makki at el, 2018), is employed. The MOEA operates on populations produced by the CA, selecting and breeding variants according to measured performance against the criteria. Rather than converging on a single optimum, it evolves a spectrum of solutions that balance multiple criteria such as environmental efficiency, spatial accessibility, and structural stability.

Through cycles of selection, crossover, and mutation, the algorithm identifies

parameter configurations that best satisfy competing criteria, creating a feedback loop between generation and optimisation. This iterative process provides the opportunity for intervention during any step of the process, allowing both the design solutions and the design problem itself to evolve, refining rule weights, adjacency priorities, and spatial hierarchies in response to emerging results.

4. Experiment

Drawing from Globus's (2017) argument that "building the first settlements close to home makes everything much easier", suggesting that Equatorial Low Earth Orbit (ELEO) offers dramatically lower transportation costs, shorter communication delays, and reduced radiation exposure, the proposed colony adopts a similar principle of strategic proximity. Positioned midway between Earth and the Moon's orbital paths, the colony is situated close enough to Earth to facilitate the transportation of goods, people, and materials with reasonable fuel efficiency, yet distant enough to integrate the Moon as a secondary station and resource hub (figure 1).

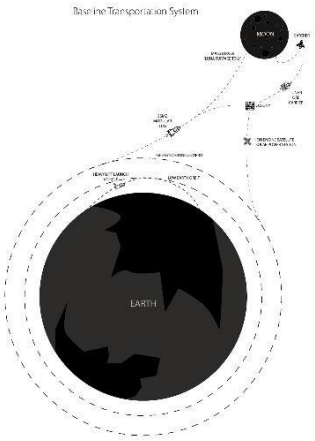


Figure 1. Colony position in earth's orbit and transport loop

Bock et al. (1979) analysed how environmental factors influence the structural mass and overall cost of orbital habitats. This data was scaled to satisfy the goal of a population size of 200,000 people; forming the residents of the first full colony. More importantly, the research provides the specific space functions and the corresponding sizes and connection relationship for each functional program type.

These functions were then simplified and classified into four class systems: residential, public, infrastructural, and life-support, in a set ratio and neighbouring configuration, generating the basis of the colony's spatial structure: 10% residential, 30% public space, 5% infrastructure, 55% life support. The neighbouring configuration will follow Bock's research with agricultural, food, mechanical all grouped into the life support class.

Each voxel represented one 'cell' belonging to a class, with a volume of 17,640

cubic metres and span of 35 metres. For residential class, each cell houses around 60 households, equating to roughly 240 residents. Using the Cellular Automata framework, residential cells would grow into clusters (superblocks) of 25-35 cells (figure 2).

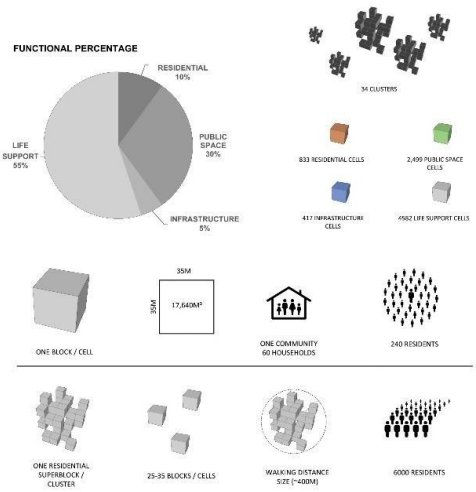


Figure 2. The specification that contribute to the colony cells

4.1. MODULAR DEVELOPMENT

The development of the modular system is split into six stages (figure 3). In Stage 1, the colony boundary is defined by setting the circumference and estimated width of the ring, establishing the spatial limits from which all voxel positions emerge. Stage 2 forms the cluster 'urban' centres, where points are distributed along the ring to establish initial cluster centres. Their spatial proximity determines when centres merge into larger cells, which become geometric anchors for future urban districts, expressed as public space cells such as green spaces, plazas, and recreational zones.

Stage 3 expands neighbourhoods, activating cells adjacent to each cluster centre based on neighbouring conditions and producing the first layer of populated geometry in the form of residential cells. Stage 4 establishes intermediate connections, as cells bordering both public and residential zones activate to create the connective tissue of the ring through infrastructure cells, supporting medical, commercial, educational, transport, and mechanical functions. Stage 5 introduces radial 'rural' linkages, where linear paths between cluster centres activate a secondary system of cells that extend outward as radial arms. These links connect clusters where possible, while failed connections form super-clusters. Functions along these arms vary with proximity—from small-scale oxygen, water, and power hubs and urban farms near the centres, to agricultural, livestock, industrial, and large-scale resource zones farther away. Finally, Stage 6 establishes the vertical anchors and rotational core, where vertical shafts rise from key positions along the ring and converge toward the colony's rotational centre. These elements set the vertical geometry and articulate the relationship between orbit,

rotation, and structural equilibrium, while also supporting essential structural and circulatory functions that complete the generative formation of the phenotype.

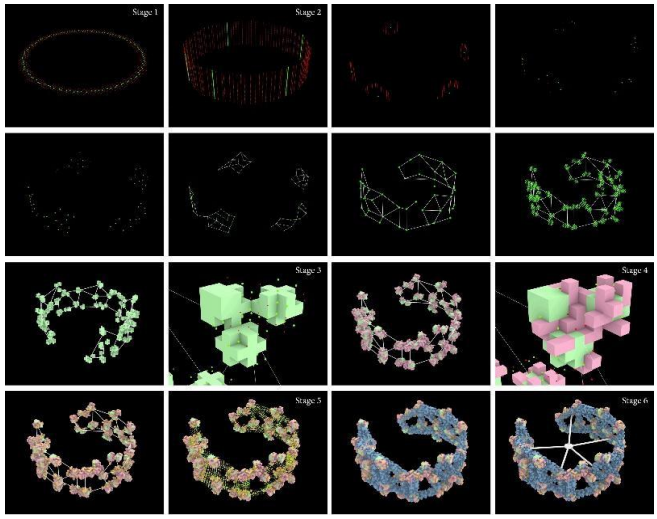


Figure 3. Six stages of the development of each design solution

4.2. FITNESS OBJECTIVES

Four fitness objectives were identified to address the project's functional and logistical requirements, serving as the driving criteria for the MOEA. 1. Maximise residential cell rooftop openings, making sure there are no other cells blocking views and solar gain into the module; 2. Maximising solar exposure on life support modules surfaces; 3. Minimising travel distance between clusters; 4. Minimising centre of gravity deviation from the phenotype's cylindrical centroid (figure 4).

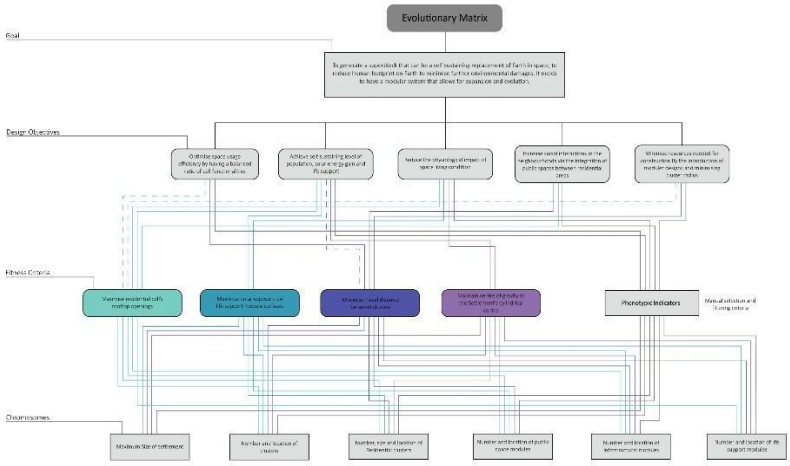


Figure 4. Evolutionary matrix showcasing the relationship between chromosomes and objectives.

5. Experiment results and Selection Process

5.1. ALGORITHM RESULTS

The final simulation produced a population of 4000 results, comprised from 80 generations of 50 individual in each, with a total runtime of 38hrs. Analysis of the phenotype results and graphs output by the simulation indicates that the overall results for all 4 objectives did not cohesively improve throughout the simulation; Despite fitness objectives 2 and 4 (solar exposure on life support cells and centre of gravity) seeing improvement from the earlier generations, fitness objectives 1 and 3 (solar exposure on residential cell roofs and cluster distances) had in fact decreased in their mean value. This highlights the conflicting nature in these fitness objectives; By reducing the distance between clusters, life support cells will also reduce in number due to the colony size decreasing, and minimising blockage to residential cell roofs will also reduce potential activation of life support cells. This is evident in the parallel coordinate plot where fitness objectives 2 and 3 showing high conflict, with 1 and 3 being a less extreme conflict (figure 5).

5.2. SELECTION PROCESS

As the simulation generated 4000 results, it is important to develop a clear methodology in the filtering and selection of the results to further analyse and select the final result. Therefore, a 3-stage selection matrix was developed to aid the process, based on the selection framework presented by Showkatbakhsh & Makki (2022).

5.2.1. Stage 1 - Pareto Fronts Cluster Centers

The simulation produced a total of 284 pareto front solutions. Using K-means clustering, the pareto front was clustered into 15 clusters (figure 5). The cluster centres were identified and selected as the representative solution for each cluster. Additionally, the fittest solution for each fitness objectives was also selected (The outliers - best performing result for fitness objectives 3 and 4 were the same). Finally, result closest to the utopia point (theoretical ideal where all objectives achieve best possible values simultaneously) and the average result (overall balanced across the objectives) were also selected. This resulted in a total of 21 solutions selected for further analysis.

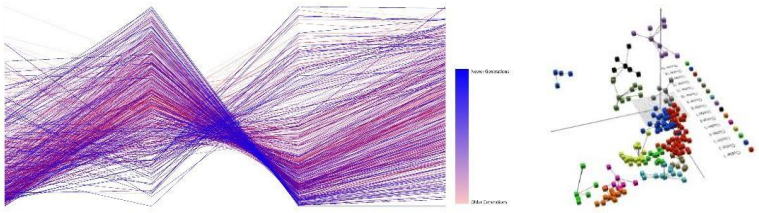
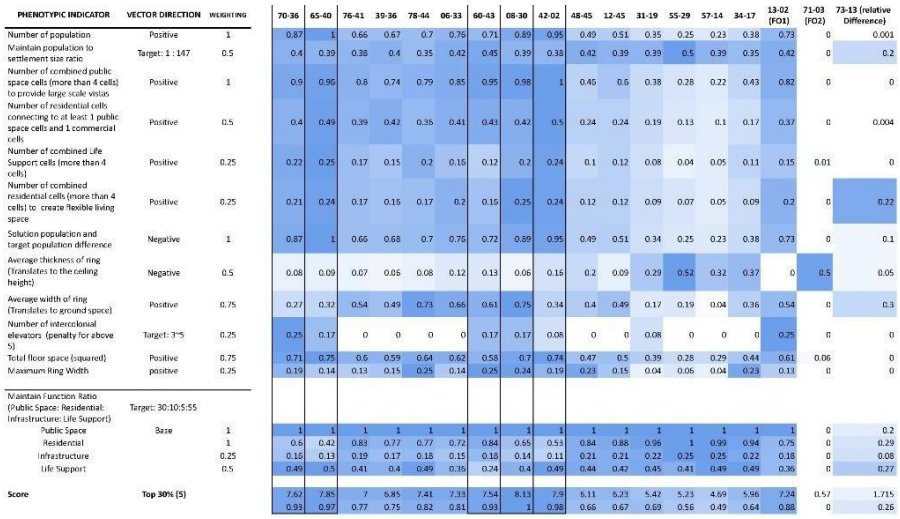


Figure 5. Parallel coordinate plot (left) and k-means clustered solutions (right)

5.2.2. Stage 2 - Phenotypic indicators

Additional objectives (phenotypic indicators) were identified to further analyse the selected solutions, these phenotypic indicators help convert the phenotype's geometric and spatial characteristics into quantifiable data. Each phenotypic indicator has a positive or negative weighting, informing whether the score will be credited or demerited. To allow for a more in-depth decision-based intervention, each indicator has a weight according to their importance to the design. The top 30% of the best scoring (normalised) results, in this case 5 results, are then selected for the next stage (figure 6).



PHENOTYPIC INDICATOR	VECTOR DIRECTION	WEIGHTING	70-36	65-40	76-41	39-36	78-44	06-33	60-43	08-30	42-02	48-45	12-45	31-19	55-29	57-14	34-17	13-02 (FO1)	71-03 (FO2)	73-13 (relative Difference)
Number of population	Positive	1	0.87	1	0.66	0.67	0.7	0.76	0.71	0.89	0.95	0.49	0.51	0.35	0.25	0.23	0.38	0.73	0	0.001
Maintain population to settlement size ratio	Target: 1 : 147	0.5	0.4	0.39	0.38	0.4	0.35	0.42	0.45	0.39	0.38	0.42	0.39	0.39	0.5	0.39	0.35	0.42	0	0.2
Number of combined public space cells (more than 4 cells) to provide large scale vistas	Positive	1	0.9	0.96	0.8	0.74	0.79	0.85	0.95	0.98	1	0.46	0.6	0.38	0.28	0.22	0.43	0.82	0	0
Number of residential cells connecting to at least 1 public space cells and 1 commercial cells	Positive	0.5	0.4	0.49	0.39	0.42	0.35	0.41	0.43	0.42	0.5	0.24	0.24	0.19	0.13	0.1	0.17	0.37	0	0.004
Number of combined Life Support cells (more than 4 cells)	Positive	0.25	0.22	0.25	0.17	0.15	0.2	0.16	0.12	0.2	0.24	0.1	0.12	0.08	0.04	0.05	0.11	0.15	0.01	0
Number of combined residential cells (more than 4 cells) to create flexible living space	Positive	0.25	0.21	0.24	0.17	0.15	0.17	0.2	0.16	0.25	0.24	0.12	0.12	0.09	0.07	0.05	0.09	0.2	0	0.22
Solution population and target population difference	Negative	1	0.87	1	0.66	0.68	0.7	0.76	0.72	0.89	0.95	0.49	0.51	0.34	0.25	0.23	0.38	0.73	0	0.1
Average thickness of ring (Translates to the ceiling height)	Negative	0.5	0.08	0.09	0.07	0.06	0.08	0.12	0.13	0.06	0.16	0.2	0.09	0.29	0.52	0.32	0.37	0	0.5	0.05
Average width of ring (Translates to ground space)	Positive	0.75	0.27	0.32	0.54	0.49	0.73	0.66	0.61	0.75	0.34	0.4	0.49	0.17	0.19	0.04	0.36	0.54	0	0.3
Number of intercolonial elevators (penalty for above 5)	Target: 3~5	0.25	0.25	0.17	0	0	0	0	0.17	0.17	0.08	0	0	0.08	0	0	0	0.25	0	0
Total floor space (squared)	Positive	0.75	0.71	0.75	0.6	0.59	0.64	0.62	0.58	0.7	0.74	0.47	0.5	0.39	0.28	0.29	0.44	0.61	0.06	0
Maximum Ring Width	positive	0.25	0.19	0.14	0.13	0.15	0.25	0.14	0.25	0.24	0.19	0.23	0.15	0.04	0.06	0.04	0.23	0.13	0	0
Maintain Function Ratio (Public Space: Residential: Infrastructure: Life Support)	Target: 30:10:5:55																			
Public Space	Base	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0.2
Residential		1	0.6	0.42	0.83	0.77	0.77	0.72	0.84	0.65	0.53	0.84	0.88	0.96	1	0.99	0.94	0.75	0	0.29
Infrastructure		0.25	0.16	0.13	0.19	0.17	0.18	0.15	0.18	0.14	0.11	0.21	0.21	0.22	0.25	0.25	0.22	0.18	0	0.08
Life Support		0.5	0.49	0.5	0.41	0.4	0.49	0.36	0.74	0.4	0.49	0.44	0.42	0.45	0.41	0.49	0.49	0.36	0	0.27
Score	Top 30% (5)		7.62	7.85	7	6.85	7.41	7.33	7.54	8.13	7.9	6.11	6.23	5.42	5.23	4.69	5.96	7.24	0.57	1.715
			0.93	0.97	0.77	0.75	0.82	0.81	0.93	1	0.98	0.66	0.67	0.69	0.56	0.49	0.64	0.88	0	0.26

Figure 6. The phenotypic indicators and weighted scores for the selected solutions. The 5 top solutions are highlighted.

5.2.3. Stage 3 - Visual Cross Examination

The final 5 results' raw fitness values alongside a visualised representation of each of their 4 fitness objective are analysed. Additional data relevant to the fitness values were also extracted for further assessment. This data was examined both numerically and visually, combining quantified performances, aesthetics and rotational stability. Ultimately, through the process of elimination on the weighting of the results flaws and credits, the solution located in Generation 08 individual 30, was selected as the final result (figure 7).

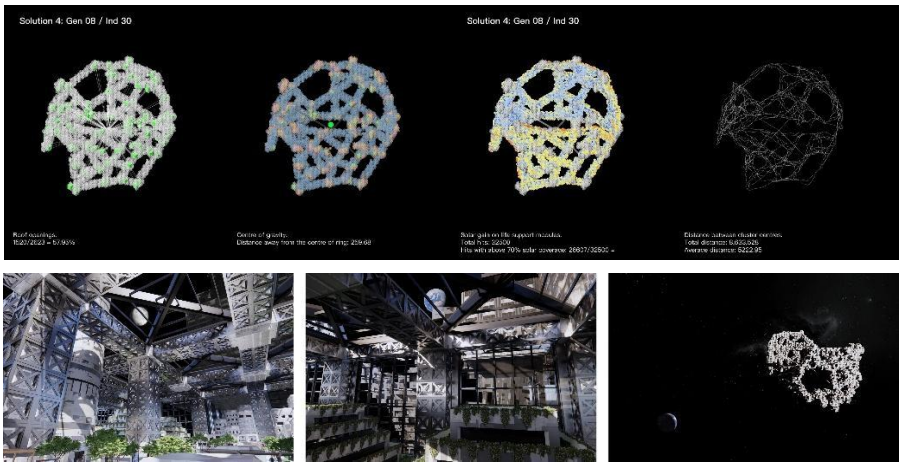


Figure 7. The final selected solution and its visualisation

6. Discussions and Conclusion

The proposed space habitat demonstrates the potential of combining generative and evolutionary algorithms to explore architectural and urban formations in a context beyond Earth. The geometric principles first abstracted and then evolved into morphologies can, in turn, develop into spatial qualities capable of meeting specific logistical demands for orbital habitats. As a proof of concept, this demonstrates the viability of the workflow as a problem-solving system that remains independent of any single context. Its level of abstraction enables scalability and adaptability across diverse scenarios. Through the iterative interplay between CA's generative logic and Wallacei's multi-objective evolutionary framework, the project establishes a workflow matrix that merges morphological generation, quantitative evaluation, and designer-led interpretation. The results highlight both the promise and the current limitations of such an approach.

While the methodology successfully produced a diverse set of morphologies and quantifiable performances, the results also expose the limitation of the computational infrastructure on the simulation. Despite the notable conflict between fitness objectives, the long runtime and consumer-grade hardware constrained population size and the number of generations, which may help the convergence of results when increased. The reliance on voxel-based geometry proved effective for data management and rule-based manipulation, but when combined with the cylindrical colony configuration it revealed a fundamental geometric inconsistency. The cylindrical form was necessary to create artificial gravity, which causes each "floor" to occupy the outermost surface of the rotation, meaning that gravitational orientation is not consistently represented across the habitat as certain cells will have the gravitational force pulling not directly downwards. Future iterations can benefit from parametric systems that identify this inconsistency and deform the voxel geometries to suit the gravitational force. In addition, the project remains highly hypothetical, abstracting away major technical and

engineering challenges related to construction in space. Visualisation of the results provides insight into potential spatial and urban qualities of the colony, though it primarily focuses on the residential and public classes. Further study could expand on this by exploring the life-support and infrastructural zones in greater detail, as well as the human-scale spatial experience.

From a methodological perspective, the research emphasizes the importance of designer involvement throughout the evolutionary process. The proposed workflow methodology enabled continuous interpretation and intervention of fitness data, allowing for informed selection and iterative adjustment of both objectives and generative rules. This combination of rule-based growth simulation, evolutionary selection, and performance analysis can inform sustainable urban morphogenesis not only in space, but also back on earth, particularly in environments requiring self-sufficiency or modular adaptability. Ultimately, this research positions space habitat design not as speculative escapism but as an exploratory framework for rethinking architectural agency, where design, technology, and ecology co-evolve for a self-sustaining future both on and beyond Earth.

References

Arora, N., Bajoria, A., & Globus, A. L. (2006). Kalpana One: Analysis and design of space colony. 47th Structures, Structural Dynamics, and Materials Conference, 2183.

Billingham, J., Gilbreath, W., O'Leary, B., & Gossett, B. (1977). Space resources and space settlements. Ames Research Center.

Bock, E., Lambrou Jr, F., & Simon, M. (1979). Habitat Structural Weight and Cost. Space Resources and Space Settlements: Technical Papers Derived from the 1977 Summer Study at NASA Ames Research Center, Moffett Field, California, 428, 33.

Ceylan, S. (2018). Space, architecture, and science fiction: An architectural interpretation of space colonization. The International Journal of the Constructed Environment, 9(2), 1.

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. International Conference on Parallel Problem Solving From Nature, 849–858.

Games, M. (1970). The fantastic combinations of John Conway's new solitaire game "life" by Martin Gardner. Scientific American, 223(120–123), 11.

Giang, L. A., Tran, M. T., Balossino, S., & Makki, M. (2023). Urbanising Mars: Establishing an Urban Framework for Martian Settlement Through Sequential Evolutionary Simulations. 28th International Conference of the Association for Computer-Aided Architectural Design Research in Asia.

O'Neill, G. K. (1974). The colonization of space. Physics Today, 27(9), 32–40.

Seedhouse, E. (2020). International Space Station life support system. In Life Support Systems for Humans in Space (pp. 151–179). Springer.

Sharma, N., & Mahajan, A. (2013). Space Settlement Design Study: An Overview. Atti Della Fondazione Giorgio Ronchi, 68, 1.

Showkatbakhsh, M., & Makki, M. (2022). Multi-Objective Optimisation of Urban Form: A Framework for Selecting the Optimal Solution. Buildings, 12(9), Article 9.

Smith, C. M. (2016). An adaptive paradigm for human space settlement. Acta Astronautica, 119, 207–217.

Optimising Multifunctional Space Use in Mission-Critical Facilities

A Multi-Objective Reinforcement Learning Approach

Eduard Haiman[1] and Davide Schaumann[2]
[1,2]*Technion – Israel Institute of Technology*
[1]*eduardhaiman@campus.technion.ac.il, 0009-0007-2571-1457*
[2]*d.schaumann@technion.ac.il, 0000-0001-9899-2818*

Abstract. Mission-critical facilities, such as healthcare buildings, often feature multifunctional spaces that support a wide range of activities. Nonlinear interactions among space affordances, staff availability, and activity requirements make optimal space use challenging. Existing methods typically explore limited space-allocation scenarios, focus on a single-objective focus, and consider short time horizons (typically days), thereby hindering longer-term decision-making in real-world settings. This research proposes a multi-objective reinforcement learning framework to generate weekly use strategies for multifunctional spaces that approximate the optimal achievable use under realistic constraints. The algorithm refines allocation policies for assigning staff and activities to spaces, while balancing spatial, social, and operational evaluation criteria. The methodology involves: (a) modelling the environment, (b) identifying optimal resource allocation strategies, and (c) analysing space management scenarios providing trade-offs among competing criteria. The framework was tested in an outpatient clinic with 19 rooms, 43 staff members, and eight treatment types, considering five working days under varied multi-criteria scenarios. Resulting allocation strategies were evaluated for trade-offs among spatial efficiency, user experience, and operational performance.

Keywords. Multifunctional Space, Space-Use Analysis, Building Optimisation, Space Utilisation, Reinforcement Learning

1. Introduction

Mission-critical facilities, such as healthcare buildings, typically feature multifunctional spaces that can accommodate a wide range of activities based on service demand. However, determining a space's function at any given time is challenging due to non-linear interactions among spatial factors such as geometry, affordances, and adjacencies (Tekin et al., 2023), social factors including staff availability and responsibilities (Guerriero & Guido, 2011), and operational factors specifically activity sequences and required resources (Hicks et al., 2015). Ignoring

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 429-438.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

these interdependencies can lead to disrupted workflows, inefficient resource consumption, and decreased occupant satisfaction.

Prior research on space-use analysis examined how occupants use spaces to meet activity requirements (Kim et al., 2013). Other approaches mapped activity demands to available spaces based on spatial affordances and capacity (Chen et al., 2019). Agent-based simulations of occupant behaviour have been employed to evaluate architectural design solutions (Schaumann et al., 2019). Nevertheless, these approaches exhibit three notable limitations. First, they typically assess a small set of predefined scenarios rather than identifying solutions that approximate the optimal achievable use under realistic constraints, such as sequential activities distributed across multifunctional spaces, staff availability and roles, and user journeys. Second, they predominantly adopt single-objective evaluations (e.g., maximising space utilisation) without explicitly addressing trade-offs among multiple criteria, including spatial efficiency, user experience, and operational performance. Third, they restrain analysis to daily time horizons to limit computational complexity, thereby hindering longer-term decision-making.

Operations research approaches focus on scheduling efficiency in mission-critical facilities, such as patient scheduling in healthcare environments (Rezaeiahari & Khasawneh, 2020). However, they overlook the potential for functional adaptation of space use and long-term trade-offs across spatial, social, and operational performance that arise from a multi-day horizon.

Recent advances in reinforcement learning (RL) have demonstrated the potential to optimise resource allocation across various domains, such as building system management (Vázquez-Canteli & Nagy, 2019) and multi-objective scheduling (Ding et al., 2024) . However, its application to multi-objective space-use optimisation under comprehensive real-world constraints remains largely unexplored.

This study introduces a multi-objective RL-based framework for generating weekly use strategies for multifunctional spaces, accounting for constraints and dependencies among spatial and people resources to meet the activity demand. The algorithm iteratively tests and refines allocation policies for assigning staff and activities to spaces based on their functional affordances, while balancing multiple performance criteria. By supporting the adjustment of weights assigned to performance criteria, the framework enables the systematic exploration of trade-offs and non-linear dependencies among strategies, facilitating the identification of optimal space-use solutions that align with an organisation's design and management objectives.

2. Computational Framework

The proposed framework consists of the following components: input, computational environment, training, optimal strategies, and analysis cycle, as illustrated in Figure 1. The input and computational environment components model the interactions among activities, spatial resources, and people resources while accounting for resource availability, allocation constraints, and dependencies between resources and activities. The training component focuses on learning the optimal policy for space use

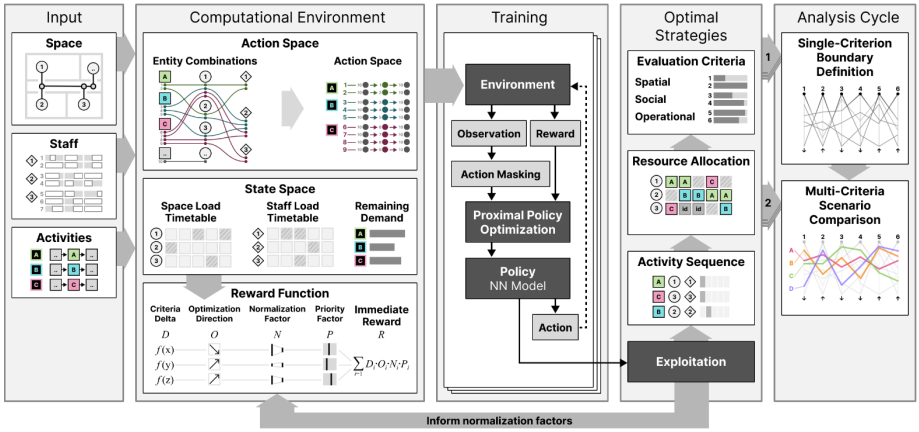


Figure 1. System architecture for space-use analysis

management. The optimal strategies component generates the best resource allocation strategy using a learned policy that balances spatial, social, and operational evaluation criteria. A two-stage analysis cycle first discovers bounds of the feasible optimal values across individual criteria and then explores optimisation strategies that consider trade-offs among multiple criteria, balancing priorities.

2.1. INPUT

The input component defines the environment configuration, including space, staff, and activities. Space is represented as a graph. The nodes of the graph representing multifunctional spaces encode spatial affordances with a degree of fit between spaces and activities. The edges of the graph encode spatial relationships, including circulation paths with distances. Staff data consists of their roles (e.g., physicians, nurses) and availability within the considered timeframe. Activities include duration, revenue they bring to the department, social value to the community, and service demand limits, which set the maximum number of core activities the facility targets (e.g., physicians' treatment) required within the community, thereby limiting the number of activities that can be assigned. Activities are structured into narratives that encompass activity sequences, relative timing, and requirements for space and staff, specifying the type of staff needed to perform each activity. Narrative can also encode spatial constraints, ensuring spatial adjacency among complementary activities in a narrative, or social features to consider group or individual treatments (Schaumann et al., 2019).

2.2. COMPUTATIONAL ENVIRONMENT

This component models the 'world' that the RL agent can observe, the actions it can take, and the relative rewards of these actions. Accordingly, it has three key components: the action space, the state space, and the reward function.

The action space comprises all possible actions that the RL agent can take, with each action representing a specific narrative that encompasses a sequence of activities along with the corresponding space and staff allocated to it. To construct the whole

action space, the system first computes all feasible combinations of spaces and staff required for each activity, considering only spaces with a non-zero degree of fit to a given activity and the necessary staff. It then assembles these combinations into unique sequences that follow the predefined activity order of each narrative.

The state space comprises all possible combinations of resource allocations and corresponding remaining demand. A state includes space and staff timetables, representing their status (e.g., available or occupied) in all time slots. Additionally, it accounts for the remaining service demand across core activities, defined as the difference between the service demand limit for these activities and the number of activities already assigned to perform.

The reward function provides feedback that guides the algorithm in learning effective decision policies. At each step, the algorithm selects actions based on an immediate reward derived from performance criteria. To align rewards with prioritised objectives, a scalarization approach (Roijers et al., 2013) combines criteria of different scales into a single scalar value.

$$R = \sum_{i=0}^{n} D_i \times O_i \times N_i \times P_i$$

Here, D_i is the value change for each criterion i between steps; O_i (± 1) indicates whether improvement requires minimisation (-1) or maximisation (+1); N_i normalises values to $[0,1]$ are defined as shown in Section 2.5; and P_i denotes a user-defined relative importance for each criterion, with the sum of their values equal to 1. This formulation enables the framework to balance spatial, social, and operational criteria within a unified reward structure.

2.3. TRAINING

This component aims to develop a policy for an optimal space use strategy. To achieve this, we employ a single-agent RL framework in which a single agent is responsible for space-use decisions across all resources within the facility. This approach involves space-use optimisation from a single agent's global view, enabling global awareness on the status of the world and coordination in the resource allocation process. The chosen algorithm is the Maskable Proximal Policy Optimisation (PPO) approach (Schulman et al., 2017). As a model-free RL approach (Gosavi, 2015), PPO does not require explicit modelling of the environment's transition dynamics, making it highly adaptable and demonstrating stable, reliable learning. A key feature of Maskable PPO is its support for invalid action masking, which is particularly valuable in discrete action spaces where certain actions may be prohibited in specific states (Huang & Ontañón, 2022). This is particularly relevant in environments with large action spaces and constraints, like our case.

The RL algorithm observes the current state and selects actions based on a probability distribution over the feasible actions in that state, as defined by the learned policy through a reward mechanism. In each episode, the algorithm continually takes a feasible action according to the action mask until the remaining resources are insufficient for any action, at which point it starts a new learning episode. For implementation, we utilised the Stable Baselines 3 Python library (SB3) (Raffin et al., 2021) and its extension, SB3-Contrib, which enables masking invalid actions.

2.4. OPTIMAL STRATEGIES

In this component, the algorithm exploits the learned policy to identify optimal space-management solutions. The outcome is an activity sequence assigned to a particular space and staff. This sequence is translated into resource allocation strategies, represented as space and staff schedules that allocate activities. Indicating functional attribution for the facility resources at a specific point in time determines the values of evaluation criteria across spatial, social, and operational categories.

2.5. ANALYSIS CYCLE

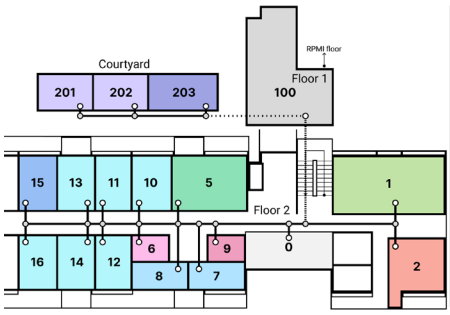
This component includes two stages: single-criterion boundary definition and multi-criteria scenario comparison.

In the single-criterion boundary definition, the algorithm optimises each criterion individually. The immediate reward is computed directly as the delta of the target performance criteria, using its original scale without normalisation, and ignoring the remaining performance criteria. This allows the system to generate optimal solutions for each performance criterion in isolation, without considering trade-offs with other criteria. The total rewards from the first stage are used to calculate normalisation factors for each criterion on the second stage, ensuring that all criterion values in the multi-criteria scenarios are evaluated on a unified scale, enabling fair trade-offs.

In the multi-criteria scenario comparison, the algorithm optimises strategies for different prioritisation scenarios, exploring the trade-offs among criterion values. Priority and normalisation factors are assigned to each criterion, where normalisation is based on the best results obtained in the previous stage. That allows the system to balance multiple performance goals simultaneously and evaluate solutions based on a weighted combination of normalised performance outcomes.

3. Case-study Overview and Experiment Setup

A case study aims to demonstrate the framework's capabilities to generate and compare alternative space management strategies that optimise interdependent spatial, social, and operational efficiency criteria. We tested the proposed approach at the Pain Medicine Institute (RPMI), an outpatient clinic within the Rambam Healthcare Campus in Haifa. RPMI faces growing service demands and adaptation needs to



ID	Type	Afforded activities	Patient capacity
0	Secretaries	**Check in/out**	-
1	Intervention room	**E**·**I**·A·B·C·D	6
2	Imaging room	**F**·**G**·**H**·A·B·c	1
5	Intervention room	**B**·**D**·**E**·**F**·A·C·I·h	6
6	Office	**C**·I	1
7	Clinic	**B**·A·C·I	1
8	Clinic	**B**·A·C·I	1
9	Office	**C**·i	1
10	Clinic	**A**·**B**·**C**	1
11	Clinic	**A**·**B**·**C**	1
12	Clinic	**A**·**B**·**C**	1
13	Clinic	**A**·**B**·**C**	1
14	Clinic	**A**·**B**·**C**	1
15	Clinic	**A**·**B**·**C**·f	1
16	Clinic	**A**·**B**·C	1
100	Waiting space	**Waiting**	-
201	External space S	**C**·a·b	1
202	External space S	**C**·a·b	1
203	External space L	**C**·**D**·a·b	6

#	Activity
A	Physician visit
B	Physical treatment
C	Psychological treatment
D	Group therapy
E	Intravenous medications
F	Ultrasound
G	X-ray
H	Pumps
I	Pre\Post treatment

Degree of fit:
HIGH MID low

Figure 2. Schematic facility layouts with multifunctional spaces and the spatial graph.

fluctuations in patient volume, staff availability, spatial congestion, and space allocation requirements. The RPMI is located on the second floor of its building, with a waiting area one level below, as shown in Figure 2. To accommodate treatments, staff also have access to external spaces in the building's courtyard, which are equipped to meet the clinic's needs.

Data has been collected to model the environment. Spatial data was encoded as a graph based on the facility's floor plans, while staff schedules were used to determine staff availability and roles. Additionally, interviews with the director of RPMI provided critical insights into spatial and staff compositions, affordances and constraints, space-staff-activity dependencies, narrative composition, and facility workdays and hours. Interviews also informed relevant performance criteria for facility analysis.

The experimental setup is based on the RPMI facility, with 19 spaces, including 17 multifunctional rooms, offering a total of 12 functional affordances. The facility is staffed by 43 personnel, divided into five roles: physicians, resident physicians, physical therapists, psychotherapists, and nurses. They are involved in eight treatment narratives: physician visit, physical treatment, psychological treatment, group therapy, intravenous medications, ultrasound, X-ray, and pumps, which correspond to specific steps, spatial requirements, and staff responsibilities. For instance, a narrative for ultrasound treatment comprises six steps: reception at the secretary's desk, waiting in the waiting area, pre-treatment, ultrasound, post-treatment, and checkout at the secretary's desk. Operational constraints posit that pre- and post-treatment should be carried out in a room located within ten meters from the ultrasound room, and the same nurse should be assigned for pre- and post-treatments and ultrasound. The considered time frame in this experiment is a five-day workweek, covering six hours from 8:00 to 14:00. The algorithm estimated critical evaluation criteria in healthcare environments (Hicks et al., 2015), as shown in Table 1.

Table 1. Performance criteria, where ↓ is minimisation, ↑ is maximisation.

Categories	Criteria	Definitions	Units	Goal
Spatial	Space Idle Rate	Unused space time ÷ initial available time	Rate	↓
	Space-Activity Fit	Average match between spaces and activities	Degree	↑
Social	Staff Idle Rate	Unused staff time ÷ initial available time	Rate	↓
	Patient Walking Distance	Average distance walked per patient	Meters	↓
Operational	Activity Revenue	Total revenue generated by activities	Points	↑
	Social Value	Total social value generated by activities	Points	↑

4. Experimentation and Results

The study investigated the effects of RL optimisation on generating resource allocation strategies and the trade-offs that emerge in optimisation scenarios. The experimentation followed a two-stage methodology. In the first stage, six single-criterion optimisations were conducted to define the boundary. The second multi-criteria prioritisation stage included four scenarios. The first one prioritises all the criteria shown in Table 1 equally. The remaining three prioritise criteria equally within specific categories—

spatial, social, or operational—ignoring the remaining factors.

The results produced allocation strategies represented as the 3-way combination of space, staff, and activity allocated at subsequent time steps, as shown in Figure 3. In the five-day schedule, each multifunctional room (represented as a row in Figure 3) is associated with an individual usage outcome, combining specific activity types (coloured bars within a space timeline) and the ID of staff member(s) involved (represented as numbers within the activity bars). In addition to assigned activities, the schedule includes idle time slots caused by planned staff activities, such as regular morning preparations or the Tuesday all-staff meeting, or by a lack of activity allocation due to resource scarcity.

The single-criterion boundary definition stage yielded values for single-criterion optimisation and evaluated trade-offs, as shown using grey lines in Figure 4. The near-optimal criterion values defined the upper bounds of the facility's potential performance. The trade-off analysis of results indicated several correlations. For instance, prioritising the fit spaces to activities, striving to increase people's satisfaction, led to space inefficiency and reduced staff utilisation (the space idle rate increases by 8.5% and the staff idle rate increases by 33.2%). Additionally, optimising the facility's

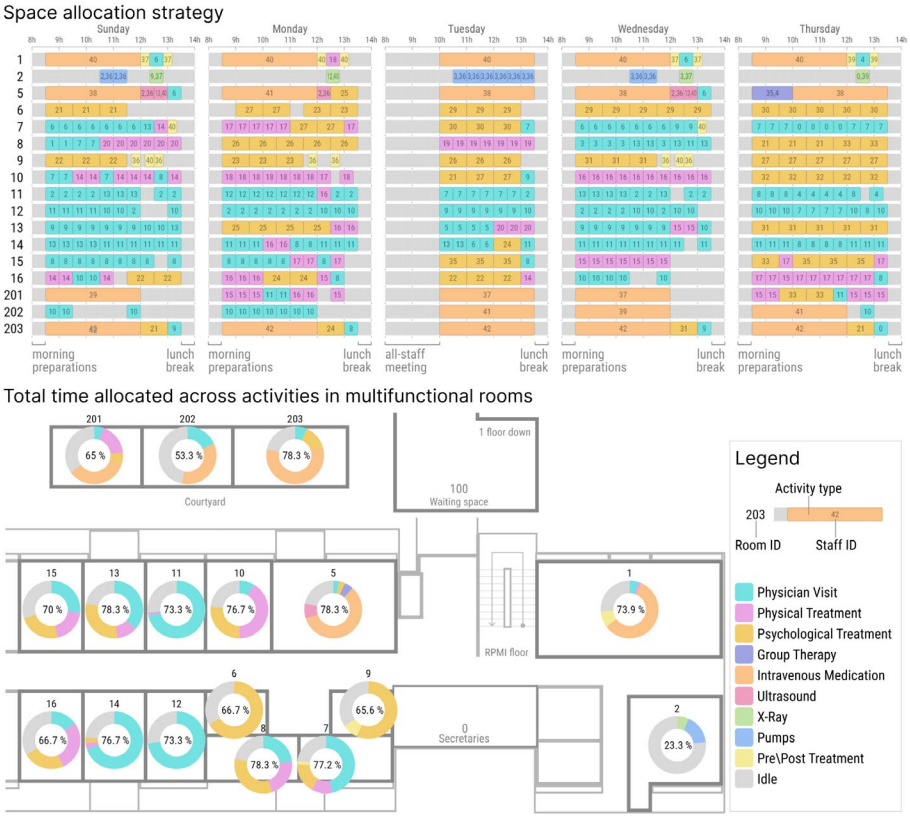


Figure 3. Space-staff allocation strategy balancing all criteria equally.

overall social value led to deterioration across all spatial and social criteria, with the maximum adverse effect on fit spaces to activities (4.3%) and patient walking distance (17.3%). In contrast, optimising activity revenue resulted in only moderate deterioration of these same criteria.

The multi-criteria scenario comparison stage identified trade-offs between the near-optimum values obtained in the first stage and across different prioritisation scenarios, as shown using coloured lines in Figure 4. The scenario balancing all criteria equally yielded equilibration across all criteria, except for a moderate level of space-activity fit. This outcome reflects a strong conflict between two objectives: minimising the space idle rate reduces the ability to maximise the rate of space-activity fit, and vice versa. The scenario focused on spatial criteria optimisation led to a crucial deterioration in operational criteria, while the values of social criteria remained balanced in the middle range. Social and operational prioritisation scenarios exhibited similar behaviour with respect to spatial criteria but yielded opposite results: improving one entitled to deterioration the other, highlighting significant trade-offs.

Training consisted of 7.5 million steps per cycle, conducted on a system equipped with an NVIDIA RTX 5000 Ada GPU (64 GB memory) and 128 GB RAM, with each cycle averaging 15 hours.

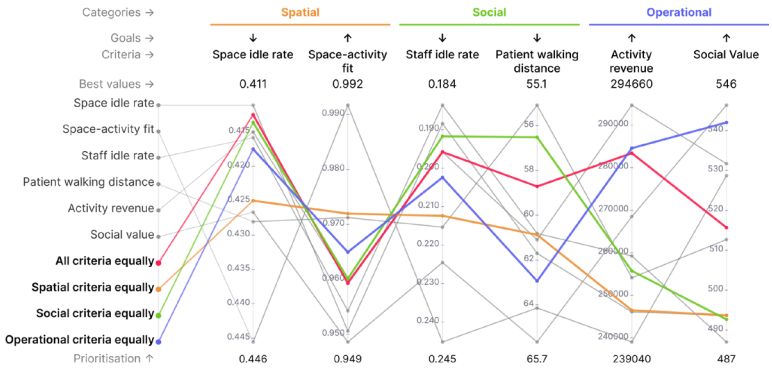


Figure 4. Trade-off analysis for two stages (the top values are the best).

5. Conclusion and Discussion

The proposed RL-based framework advances space-use analysis in three ways. First, it optimises space allocation while capturing the complexity of a healthcare setting, featuring 19 multifunctional rooms, 43 staff members across five roles, and eight treatment narratives, each affecting performance criteria. Second, it evaluates long-term resource allocation across a five-day horizon, reflecting weekly staff availability. Third, its two-stage methodology supports multi-criteria optimisation by (a) identifying near-optimal single-criterion values and (b) balancing multiple criteria to reveal trade-offs arising from non-linear spatial, social, and operational dependencies, which is especially evident in scenarios that prioritise criteria with contradictory behaviour, such

as spatial criteria.

Despite capturing the interactions between spatial configurations, staff numbers and roles, and treatment activities, the study has several limitations. The current approach assumes fixed activity durations, omitting sudden changes and allocating slack time to mitigate their effects. Patient walking movement is simulated to compute patient walking distances without accounting for waiting times due to varying arrival times and possible delays, as well as disruptions from unplanned events such as equipment malfunctions or employee unavailability. Besides patient walking distance, staff walking distance should also be considered, which requires modelling staff members' movements and accounting for changes in rooms and behaviour during idle time. While results are promising for a single facility, testing in varied settings and collecting stakeholder assessments are needed to evaluate the framework's scalability and practical applicability. Further research could also examine integration with real-time operational data and adaptability to dynamic events and changing conditions.

Maskable PPO has shown sensitivity to hyperparameters, requiring scrupulous tuning to achieve efficient learning. Despite the scalarization approach simplifying the prioritisation process in multi-objective optimisation by weighting metrics within reward shaping according to the priorities of the criteria, the outcome approximates the criterion values to the target widths, without precisely matching them. Due to the non-linear relationships between the prior scalarization of priorities and the resulting trade-offs, it remains challenging to search for the full Pareto front, which can support trade-off analysis among allocation strategies when no single objective outperforms all others. Comparison with the baseline resource allocation algorithm can also help to evaluate the framework's performance.

Practically, the RL-based framework integrates spatial, social, and operational criteria to assess overall facility performance. During the generation of an allocation strategy, the metrics for these criteria are calculated for each space or staff and then accumulated to derive the total impact. These features can be used to analyse space use on the level of a single multifunctional room or to analyse patterns among staff roles and availability. Addressing room-level analysis can identify performance bottlenecks and rooms with outstanding performance. Moreover, staff-level analysis can identify discrepancies in people resources for a given spatial configuration, affordances, and service demand. Performing both high-level analyses to evaluate facility performance and low-level analyses of a particular resource entity can be used to assess what-if scenarios during facility adaptation and performance improvement. These capabilities can broaden the applicability of the proposed approach, aiding facility adjustments and long-term adaptations that support resilience, longevity and sustainability.

Acknowledgements

We thank Dr. Amir Minerbi, Director of the Rambam Pain Medicine Institute, for his invaluable assistance and collaboration. This research was supported by the Israel Science Foundation (Grant No. 2351/24).

Attribution Statement

Grammarly (Grammarly Inc., 2025) was used to correct grammar, spelling, and

punctuation. ChatGPT (OpenAI, 2025) was used to improve clarity and flow.

References

Chen, X., Kim, T. W., Chen, J., & Xue, B. (2019). Ontology-Based Representations of User Activity and Flexible Space Information: Towards an Automated Space-Use Analysis in Buildings. Advances in Civil Engineering, 2019(1), 3690419. https://doi.org/10.1155/2019/3690419

Ding, L., Guan, Z., Rauf, M., & Yue, L. (2024). Multi-policy deep reinforcement learning for multi-objective multiplicity flexible job shop scheduling. Swarm and Evolutionary Computation, 87, 101550. https://doi.org/10.1016/j.swevo.2024.101550

Gosavi, A. (2015). Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning (Vol. 55). Springer US. https://doi.org/10.1007/978-1-4899-7491-4

Grammarly Inc. (2025). Grammarly: AI Writing and Grammar Checker App (Version 1.2.209.1783) [Large Language Model]. https://app.grammarly.com/

Guerriero, F., & Guido, R. (2011). Operational research in the management of the operating theatre: A survey. Health Care Management Science, 14(1), 89–114. https://doi.org/10.1007/s10729-010-9143-6

Hicks, C., McGovern, T., Prior, G., & Smith, I. (2015). Applying lean principles to the design of healthcare facilities. International Journal of Production Economics, 170, 677–686. https://doi.org/10.1016/j.ijpe.2015.05.029

Huang, S., & Ontañón, S. (2022). A Closer Look at Invalid Action Masking in Policy Gradient Algorithms (No. arXiv:2006.14171). arXiv. https://doi.org/10.48550/arXiv.2006.14171

Kim, T. W., Rajagopal, R., Fischer, M., & Kam, C. (2013). A knowledge-based framework for automated space-use analysis. Automation in Construction, 32, 165–176. https://doi.org/10.1016/j.autcon.2012.08.002

OpenAI. (2025). ChatGPT Model: GPT-5 [Large Language Model]. https://chatgpt.com

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-baselines3: Reliable reinforcement learning implementations. J. Mach. Learn. Res., 22(1), 268:12348-268:12355.

Rezaeiahari, M., & Khasawneh, M. T. (2020). Simulation optimization approach for patient scheduling at destination medical centers. Expert Systems with Applications, 140, 112881. https://doi.org/10.1016/j.eswa.2019.112881

Roijers, D. M., Vamplew, P., Whiteson, S., & Dazeley, R. (2013). A Survey of Multi-Objective Sequential Decision-Making. Journal of Artificial Intelligence Research, 48, 67–113. https://doi.org/10.1613/jair.3987

Schaumann, D., Putievsky Pilosof, N., Sopher, H., Yahav, J., & Kalay, Y. E. (2019). Simulating multi-agent narratives for pre-occupancy evaluation of architectural designs. Automation in Construction, 106, 102896. https://doi.org/10.1016/j.autcon.2019.102896

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms (No. arXiv:1707.06347). arXiv. https://doi.org/10.48550/arXiv.1707.06347

Tekin, B. H., Corcoran, R., & Gutiérrez, R. U. (2023). A Systematic Review and Conceptual Framework of Biophilic Design Parameters in Clinical Environments. HERD: Health Environments Research & Design Journal, 16(1), 233–250. https://doi.org/10.1177/19375867221118675

Vázquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modeling techniques. Applied Energy, 235, 1072–1089. https://doi.org/10.1016/j.apenergy.2018.11.002

The Multi-objective Optimization of Breeze Blocks

Re-thinking modern architecture design features in a performance-oriented method

Jiajian Chen[1] and Hiroyuki Shinohara[2]
[1,2]*The Chinese University of Hong Kong*.
[1]*ianchenj@link.cuhk.edu.hk, ORCID0009-0004-8683-394X*
[2]*hiroyukishinohara@cuhk.edu.hk, ORCID0000-0003-2644-0592*

Abstract. This paper investigates a design methodology for the applying breeze blocks in semi-outdoor conditions by integrating environmental simulation and multi-objective optimisation. Design parameters of breeze block such as thickness and perforation size affect daylight distribution, thermal comfort, and physical visibility within the semi-outdoor spaces. The relationships between these design parameters and environmental performance can be explores through simulations and statistical analysis. While previous studies have applied computational simulation and optimisation to evaluate the environmental performance of passive building design features, investigations majorly have focused on daylighting and energy performance, with limited attention given to thermal comfort for occupants. The findings of this study indicate that using computational simulation and optimisation in the design of breeze block can enhance designer's understanding of their environmental performance and thus benefit design decisions. This study introduces a practical framework for architects and designers to adopt evidence-based methods for incorporating breeze blocks into building design taking into account specific climatic conditions and site contexts.

Keywords. perforated blocks, computational simulation, parametric design, passive design, multi-objective optimization

1. Introduction

In the modern architecture of Hong Kong and other Southeast Asian regions with tropical and subtropical climates, perforated blocks, commonly known as breeze blocks, are extensively used as passive design features. In Hong Kong during 1950s and 1960s, breeze blocks were introduced to enhance sunlight penetration and ventilation in public spaces in response to the challenges posed by overcrowded living conditions in hot and humid climate (Lai, 2020). However, as Han (2016) notes, there remains ambiguity regarding the relationship between the formal characteristics of breeze blocks and their functional effects, leaving a gap between the intention of formal design and actual functional performance. The study rethinks the potential of this

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 439-448.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

traditional building technology for contemporary architectural applications by analysing its functions and exploring advancement in computational design tools.

To examine the formal-functional relationship, Yumna et al. (2020) investigated the daylight performance of breeze blocks, indicated by Daylight Glare Probability (DGP) through a parametric simulation study. Shinohara (2023) studied the daylight and ventilation performance of breeze blocks from several existing modern buildings in Hong Kong, analysing their morphological characteristics using annual daylight simulation and computational fluid dynamics (CFD). These studies highlighted the effectiveness of breeze blocks as a passive design feature and revealed the connection between formal design and environmental performance. However, they also underscore the needs for further investigation into the interactions among multiple control variables and multiple functional outcomes.

The rise of computational technology has introduced new possibilities for analysing the intricate relationships between form and performance in breeze block design. Yuliana et al. (2023) employed computational simulations and Multi-Objective Optimisation (MOO) to design breeze block that balance daylight, ventilation, and structural rigidity for optimal environmental performance. Susanto et al. (2025) further advanced the application of MOO in the parametric design of breeze blocks, optimising daylight, ventilation, and visibility. These studies provide valuable insights into the quantification of functional effects and the deliberate design of breeze blocks based on their environmental performance. However, the assessments of ventilation and shading performance were conducted only in simplified hypothetical environmental settings. The simulation setups in previous studies represent indoor conditions rather than semi-outdoor scenarios, differing from the common application of breeze block walls as façade components which results in distinct pattern in wind and light environment. Moreover, since the studies analysed ventilation and sun shading performance separately, the relationship between breeze block design and human thermal comfort, which requires investigation into synergetic interaction among multiple environmental variables, remains unclarified in the existing studies.

This research addresses these gaps by developing a design framework that systematically re-evaluates the potential of breeze blocks for contemporary architectural applications. The research questions are: How do design parameters of a breeze block affect its environmental performance in terms of thermal comfort, daylight and social privacy? How can a computational design method facilitate performance-oriented design of breeze blocks through an optimisation process?

2. Methodology

2.1. CASE COLLECTION AND SELECTION

This study combines a four-phase methodology, including case collection and selection, parametric simulation, statistical analysis and design optimisation as illustrated in Figure 1. Firstly, field investigation of breeze block applications in existing modern architecture in Hong Kong was carried out by manual search, resulting in the documentation of 78 cases. For each case, the formal characteristics of the breeze block and its relationship to the deployed site were analysed to establish a profile of

breeze block applications in the city. From the understanding of the profile, a design of square block with two parallel vertical rectangular openings, commonly found on the façades of many public housing estates, as illustrated in Figure 2, was selected as the Typical Breeze Block (TBB). This design represents the most typical and applicable design in Hong Kong, to be investigated in this study.

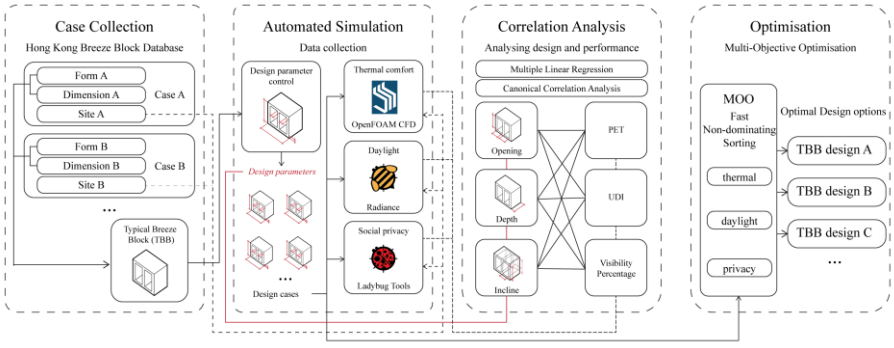


Figure 1. Research methodology

2.2. AUTOMATED ENVIRONMENTAL SIMULATION SYSTEM SETUP

For the design of the TBB, three key design parameters are defined: edge thickness, depth, and the area ratio between front and rear openings. These parameters respectively control the block's wall thickness, overall depth, and the degree of inclination of the internal slope in openings. The detailed configuration of these design parameters, which generates 379 cases of design options. The edge thickness varies from 10mm to 50mm, the depth ranges from 80mm to 180mm and area ratio ranges from 0.7 to 1.3. To automate the environmental performance simulation process, an iterative simulation workflow was developed with an automated control component on design parameters and a simulation component. The automated parameter control component systematically generates TBB design options based on predefined sets of parameters and passes the designed geometries to the simulation component for iterative environmental performance analysis.

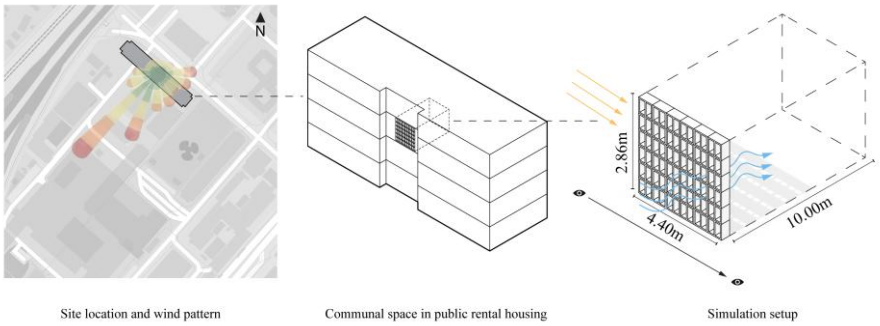


Figure 2. Environmental simulation context

This study evaluates the environmental performance of TBB in three key aspects: human thermal comfort, daylight and social privacy, which are respectively indicated by Physiological Equivalent Temperature (PET), Useful Daylight Illuminance (UDI), and physical visibility percentage (VisiPer). These performance aspects are assessed through computational simulations for a stacked TBB wall in an environment setting simplified from a real-world site featuring breeze block applications. The dimension of the simulated site is 4.40m (w)*2.86m (h)*10.00m (d) as shown in Figure 2. The site is based on Luk Chuen House of Lek Yuen Estate in Hong Kong, where the prevailing summer wind direction is perpendicular to the breeze block wall. Meteorological and microclimatic data for the site were incorporated into the simulation environment using publicly available databases, with the simulation period set to summer (June to August). Human activity was defined as walking in 1m/s with metabolic rate of 2.4, reflecting the site's function as a communal space. Measurements of the performance indicators were taken on an object surface located 2 metres from the TBB wall and elevated 1.5 metres above floor level.

Both automated parameter control component and simulation component were implemented in Grasshopper. CFD analysis was performed using OpenFOAM via Butterfly for the wind environment, while simulations of radiance and calculations of PET, UDI and VisiPer were performed using Ladybug tools. Additionally, a base case without TBB wall, under the identical simulation condition, was established to serve as a baseline of performance for comparative analysis.

2.3. STATISTICAL CORRELATION ANALYSIS

For data pre-processing, the values of performance indicators were averaged and normalised to eliminate the influence of differing data scales on the correlation analysis models. The mean PET on the measured surface is denoted as mPET, so as for mUDI and mVisiPer. The normalised variables were entered as dependent variables in the correlation analyses and optimisation process.

Multiple linear regression (MLR) analysis was performed for each environmental performance indicator as exploratory models to reveal the relationships between the set of design parameters and individual aspects of environmental performance. To ensure that no potentially relevant design parameters were excluded, a simultaneous entry method was adopted preserving all independent variables. Standardised beta coefficients were also calculated to analyse the different contributions to one performance indicator between design parameters.

To further examine cross-relationships between two sets of variables, Canonical Correlation Analysis (CCA) method was employed to model the interdependencies between design parameters and performance indicators. The method derives paired canonical variates from both sets of variables and performs Eigenvalue-based optimisation to maximise the inter-set association modelled (Clark, 1975). Variable contributions to each variate are interpreted through loadings, which constitute the canonical correlation between variables. Unlike MLR, which describes how a single design parameter contributes to a specific performance indicator, CCA reveals how strong individual design parameters are correlated with a specific performance indicator.

2.4. MULTI-OBJECTIVE OPTIMISATION

While a majority of existing research have adopted the SPEA-2 and HypE algorithms implemented through Octopus plugin in Grasshopper for MOO, given the small dataset size with fixed, discrete data points in this study, the optimisation process prioritises efficiency in identifying the non-dominated front from within the predefined dataset. Therefore, this study uses Fast Non-dominated Sorting method from NSGA-II algorithm, excluding additional components of the original algorithm that might increase unnecessary computational complexity for this specific dataset (Wang, 2023).

With the purpose to enhance thermal comfort, daylight performance, and privacy, the MOO process seeks to balance three objectives: minimising the mPET, maximising the mUDI, and minimising mVisiPer. Optimal solutions are obtained by identifying the pareto front, representing optimal design options that cannot be improved in one objective without worsening another. Objective space plot is used to visualise the distribution of optimal design options and the pattern in distribution of parameters.

3. Findings

3.1. SIMULATION RESULTS

Environmental simulations and calculation on indicators were completed for all 379 cases after 26 hours running on a personal computer with AMD Ryzen 7700X CPU utilising 14 cores. The average simulation time per each case was 4.11 minutes. After excluding 55 invalid data points, due to the dimensions of the openings exceeding the defined boundaries of a single TBB unit, and 26 outlying data points with exceptional high mPET (greater than 36°C), due to overly small openings that made the TBB wall perform similar to solid, resulting in significant drop in wind velocity which leads to an unusually high mPET, dataset for 298 valid cases was obtained. The normalised mPET of each case was the mPET divided by 28.8293°C, the mPET value for the base case, with mUDI divided by 66.6542% and mVisiPer divided by 84.6189%.

3.2. CORRELATION ANALYSIS

The correlation between mPET, mUDI and mVisiPer as dependent variables and edge thickness, depth and opening area ratio as independent variables were examined using MLR. A substantial proportion of variance is explained by the models, as indicated by all coefficients of determination (R^2) over 0.7 in all cases. The lower R^2 values in the models for mPET (0.770) and mUDI (0.783) in comparison with that for mVisiPer (0.911) might be attributed to more complex and interdependent relationships in variables for thermal comfort and daylight simulations. As shown in Figure 3, the plots of standardised coefficient betas from MLR models reveal that no single design parameter can have only a positive effect on all aspects of environmental performance. For example, increasing the depth leads to better social privacy and cooler thermal experience but simultaneously reduce comfortable daylight hours. These findings underscore the necessity of adopting an optimisation approach in design practice.

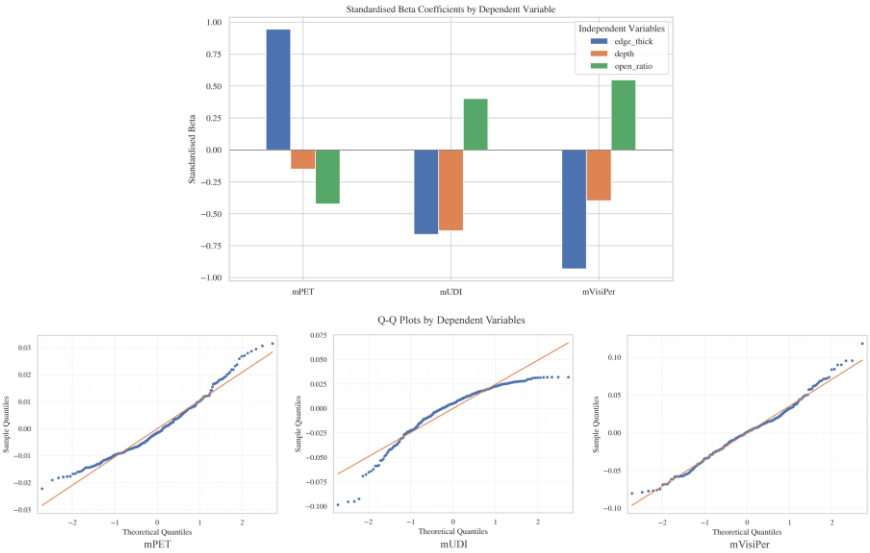


Figure 3. Standardised beta coefficients and Q-Q plots

As shown in Figure 3, the Q-Q plots for all three dependent variables exhibit curved pattern in MLR, with upward curving in mUDI and downward curving in mPET and mVisiPer, indicating that the distribution of data does not strictly follow the theoretical normal distribution. This departure suggests the presence of non-linear relationships between two sets of variables. To examine these potential non-linear relationships between a set of three design parameters and a set of three environmental performance indicators, a Non-Linear Canonical Correlation Analysis (NLCCA) model was employed to accommodate the inadequacies in the assumptions of linear regression models and derive analytical model with more interpretability (Huang et al., 2006).

Compared to the MLR models, increased percentage of explained variance in dependent variables can be observed as 88.18% for mPET, 97.52% for mUDI and 98.85% for mVisiPer, indicating that the relationships between design parameters and performance indicators are better modelled in NLCCA. The correlation strengths among variables are displayed as the correlation heat map in Figure 4, where the directions of correlation between independent and dependent variables are consistent with the correlation pattern in the MLR models. Edge thickness has moderate (-0.5) to strong (0.77, -0.71) correlations with the performance indicators, while depth shows weak (-0.15) to moderate (-0.63, -0.4) correlations. In contrast, the opening area ratio shows only slight (-0.037) to weak (0.13, 0.17) correlations. This pattern of correlation strengths suggests that, when designing breeze blocks for improved environmental performance, the opening size should be prioritises, followed by the block thickness, as key parameters influencing overall performance outcomes.



Figure 4. Correlation heatmap indicating correlation strengths

3.3. MULTI-OBJECTIVE OPTIMISATION

The MOO process found 30 optimal design options out of all 298 evaluated cases, of which 12 achieved a reduction in mPET compared with the base case, as illustrated in Figure 5. The 30 optimal design options compose the pareto front within the objective space. For easier interpretation, the pareto front in the three-dimensional objective space was projected onto planes constructed by a pair of objectives as orthogonal axes, revealing predominant nonlinear trade-offs between mUDI maximization and the minimizations of mPET and mVisiPer, with clustering patterns that indicate preferential regions of optimal design solutions.

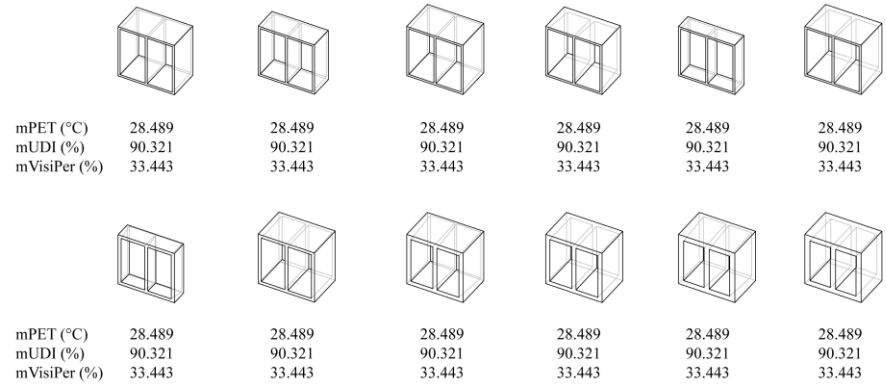


Figure 5. Optimal design options with mPET lower than base case

The two-dimensional projections of the pareto front show the multi-objective trade-offs in optimising environmental performance in Figure 6. In the mUDI-mPET plane, the negative slope of pareto front implies that daylight enhancement is aligned with thermal comfort improvement, following an approximately linear trend. In contrast,

the mUDI-mVisiPer projection reveals a pronounced concave curvature, indicating a nonlinear conflict between daylight and privacy performance. Initial mUDI increases (from 1.10 to 1.30) are associated with modest rises in the level of mVisiPer (up to 0.3), but beyond mUDI level of 1.30, the steeper penalties emerge and significantly deteriorate social privacy performance, as indicated by soaring level of mVisiPer values. This curvature implies trade-offs between confronting daylight and social privacy performance. Similarly, the mVisiPer-mPET plane exhibits a convex curve, suggesting less intense conflicts between thermal comfort and social privacy. Although efforts in lowering level of mPET tend to produce steep increases in mVisiPer, clustering of optimal design options near the level of mPET in base case (from 0.99 to 1.01) allows design options with mVisiPer ranging from 0.1 to 0,3, while further marginal improvement in social privacy will sharply deteriorate thermal comfort.

Notably, among the optimal design options, the one with the lowest mPET, highlighted in Figure 6, exhibits mUDI and mVisiPer levels close to the midpoints between their respective maxima and minima for all cases with mPET values below the base case. The presence of this option indicates the possibility of prioritising thermal comfort without sacrificing either daylight or privacy through MOO process.

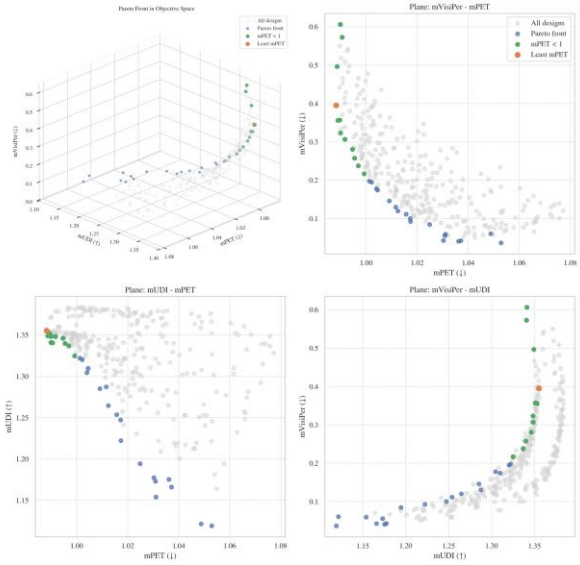


Figure 6. 2D and 3D catter plot for objective space

4. Discussion

The findings from statistical analysis and optimisation process have elicited the relationships between form and environmental performance in breeze block design through analysis in specific design parameters and performance indicators. The revealed relationships not only indicate potential design priorities for improving the environmental effectiveness of breeze blocks but also implies a pathway to understand the multifunctionality, offering insights into how formal variations can simultaneously influence multiple aspects of performance.

The mean mPET in all cases except base case is 29.4°C, which will create slightly warmer thermal experience for occupants. This value is marginally above the neutral PET range for Hong Kong's outdoor environment (27–29 °C) when accounting for acclimatisation effects (Ng & Cheng, 2012). Given that the base case mPET is 28.8 °C, it might raise questions in necessity on deployment of breeze blocks. However, when examining the PET data from the simulations, which have important components as the wind velocity, air temperature and mean radiant temperature (MRT), the effect by deployment of breeze block can be observed in reduction of wind velocity and MRT compared to the base case. While decrease in wind velocity tends to produce a warmer thermal experience, as compensation, the breeze block wall creates shading effect from excessive solar heat and maintains similar thermal environment as the base case. Consequently, the results suggests that breeze blocks function primarily as a shading device rather than a means for promoting airflow, creating a breeze.

The shading effect provided by breeze blocks not only contributes to enhance daylight quality and social privacy but may also exert an indirect positive influence on thermal comfort. Chinazzo et al. (2019) discovered that people tend to perceive cooler environment under warm conditions when daylight levels are within low to medium illuminance ranges. This indicates that daylight exposure can psychologically modulate people's perception on thermal comfort, which cannot be predicted by physiological thermal indices like PET. Based on this understanding of human thermal perception beyond physical prediction models, the shading effect of breeze blocks deployed in semi-outdoor space in Hong Kong is likely to foster a cooler experience. These intricate interrelationships among multiple aspects of environmental performance highlight the needs for a holistic design approach that integrates diverse performance criteria, of which the challenges are well-addressed through the MOO process investigated in this study.

The characteristics observed in the pareto front confirm the presence of a convex pareto surface in three dimensions, dominated by mUDI-mVisiPer tensions, while mPET exerts subtler influence. The continuity and spread of the pareto front demonstrate the effectiveness of the MOO algorithm in sorting for optimal solutions for practical applications in breeze block design, where daylight enhancement must balance with thermal comfort and social privacy. The presence of the lowest-mPET solution illustrates that designers can strategically select either from extreme or intermediate points according to specific performance priorities, highlighting the potential for targeted refinements through adapting objective weighting or constraint integration within the optimisation process.

5. Conclusion

This study investigated the computational design methodology for breeze blocks through the integration of statistical analysis and multi-objective optimisation technique. The findings indicate that the size of the openings had great influence on the environmental performance of the breeze block walls, followed by block depth, while the opening incline only exhibits a minor association. The application of the MOO process has revealed that design options can achieve balanced performance outcomes across thermal comfort, daylight quality, and social privacy, demonstrating the potential of computational optimisation to guide data-informed and performance-

oriented design decisions.

The meteorological and microclimatic condition in this study only include scenarios that sun azimuth angle and prevailing wind direction are in concurrent situation. The performance of breeze block might vary when the relationship between orientation of breeze block wall and sun and wind direction changes and the association between design parameters and performance might differ as well. This study has only included discrete parameters for the convenience in physical production. In future studies, the controlling parameters of breeze blocks can be developed in the format of continuous values and enables evolutionary optimisation processes for more detailed achievement in optimising the performance.

By reimagining breeze blocks through a performance-oriented, computational approach, this study promotes a computational framework for architects and designers to adopt evidence-based methods for integrating the cultural heritage with innovative technologies, fostering a responsive and adaptive architectural practice.

Reference

Chinazzo, G., Wienold, J., & Andersen, M. (2019). Daylight affects human thermal perception. *Scientific Reports*, 9(1), 13690. https://doi.org/10.1038/s41598-019-48963-y.

Clark, D. (1975). *Understanding canonical corelation analysis*. University of East Anglia.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197. https://doi.org/10.1109/4235.996017.

Han, M. (2016). *The characteristics of modern architecture in Hong Kong, 1930s-1970s*. (Thesis) Chinese University of Hong Kong.

Huang, S.-Y., Lee, M.-H., & Hsiao, C. K. (2006). *Kernel canonical correlation analysis and its applications to nonlinear measures of association and test of independence*. Institute of Statistical Science: Academia Sinica, Taiwan.

Lai, C. W. C. (2020). *Shanghai plaster: a history of cement architecture and colonial modernity in Hong Kong, 1920s - 60s*. (Thesis). University of Hong Kong.

Ng, E., & Cheng, V. (2012). Urban human thermal comfort in hot and humid Hong Kong. *Energy and Buildings*, 55, 51–65. https://doi.org/10.1016/j.enbuild.2011.09.025.

Salsabila, S. Y., Hariyadi, A., Afif, N., & Cinderakasih, A. P. (2020). *Parametric Roster Concrete Brick Modules as External Shading Device to Maximize Indoor Visual Comfort*. International Conference on Less is MOji-REboot the City, Fukuoka-Kitakyushu, Japan.

Shinohara, H. (2023). Rethinking breeze blocks: a study of form and performance complexity ìn hong kong's building designs. In *HUMAN-CENTRIC, Proceedings of the 28th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2023*, Volume 2, 633-642.

Susanto, D., Alifa, R. P., Aylien, S., & Alkadri, M. F. (2025). Advancing building façade design: Digital fabrication of an optimized non-conventional roster brick prototype. *Journal of Asian Architecture and Building Engineering*, 1–25. https://doi.org/10.1080/13467581.2025.2505795.

Wang, B. (2023). *NSGA-II-in-python* [Source code]. GitHub repository. Retrieved from https://github.com/bujibujibiuwang/NSGA-II-in-python.

Yuliana, Y., Agung, M. R. C., Rahman, M. A., Widyarko, W., & Alkadri, M. F. (2023). Investigating roster brick based on multi-objective optimization between daylighting, wind velocity and structural displacement. *IOP Conference Series: Earth and Environmental Science*, 1267(1), 012048. https://doi.org/10.1088/1755-1315/1267/1/012048.

Enhancing Urban Ecosystem Services

A computational approach to multispecies habitat distribution on building envelopes

Anastasia Slavina[1], Surayyn Uthaya Selvan[2], Oren Agay Shay[3], Narkis Joffe [4], Lilu Golan[5] and Yasha J. Grobman
[1,2,3,4,5,6] *Technion - Israel Institute of Technology.*
[1]*anastasiias@campus.technion.ac.il, 0009-0007-4751-1296*
[2]*surayyn@campus.technion.ac.il, 0000-0001-9401-1170*
[3]*orenagayshay@campus.technion.ac.il, 0009-0003-1522-1631*
[4] *narkis.joffe@campus.technion.ac.il, 0009-0009-3380-4762*
[5]*lilu.golan@campus.technion.ac.il, 0009-0003-1713-5446*
[6]*yasha@technion.ac.il, 0000-0003-4683-4601*

Abstract. Rapid urbanisation has fragmented natural habitats, displacing species and disrupting ecological balance, which contributes to biodiversity loss and intensifies climate change. While numerous ecological strategies have been proposed to counter biodiversity loss, their integration into architectural design remains limited by the absence of data-driven workflows that link ecological modelling with spatial performance. This research introduces a computational framework for distributing multispecies habitats within different elements of urban structures, focusing particularly on building envelopes. The framework integrates architectural Key Performance Indicators (KPIs) with principles of ecological Habitat Suitability Modelling (HSM). It combines biotic and abiotic predictors, such as vegetation proximity, street lighting, noise, and human activity, to estimate suitability gradients and guide adaptive habitat allocation. A subsequent Multi-Objective Optimisation (MOO) stage balances diverse stakeholder preferences and real-world constraints, such as spatial and maintenance limitations, generating a range of viable design options. The research contributes to biodiversity-inclusive architecture by providing a transparent, data-driven framework for enhancing urban ecosystem services through design.

Keywords. multispecies habitat distribution, cityscape contextual variables, habitat suitability index, key performance indicators, multi-objective optimisation

1. Introduction

Biodiversity plays a pivotal role in regulating ecosystem processes such as carbon sequestration, temperature moderation, and pollination. Yet, rapid urbanisation has resulted in the progressive fragmentation of natural habitats, reducing ecological

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 449-458.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

connectivity and contributing to species decline (Weisser et al., 2022). While traditional solutions such as nature parks and wildlife crossings remain valuable, they are often insufficient in mitigating species extinction. Therefore, there is a pressing need for integrated strategies that synthesise ecological and architectural data into a coherent and transparent design workflow (Grobman et al., 2023). Recent research on multispecies design developed a framework for distribution of various species habitats on building envelopes based on climatic and ecological variables (Selvan et al., 2025). However, these methods employ computationally intensive databases and do not incorporate numerous critical architectural and cityscape parameters that would enable their integration into the design process and be accessible for practitioners.

This research attempts to mitigate these gaps by developing a designer-oriented framework and computational model for the distribution of multispecies habitats on a building envelope in a dense urban environment. The suggested framework integrates architectural key performance indicators (KPIs) with principles of ecological Habitat Suitability Modelling (HSM) (Selvan et al., 2025), which estimates the probability of species occurrence based on environmental factors, traditionally applied at urban or regional scales (Hirzel et al., 2008). This paper contributes to existing knowledge in two domains. The first is a designer-oriented workflow that allows architects to reach ecologically informed multispecies habitats distribution on a building envelope. The second is the integration of architectural, anthropogenic, and cityscape contextual variables, such as noise pollution and street lighting, into distribution scenarios adapted to the building scale. These scenarios reflect context-specific design constraints, including the percentage of multispecies coverage and dedicated areas for habitat placement. The suggested design framework and computational tool are tested in a case study design process, demonstrating its potential to enable designers to integrate a wide array of real-world architectural and ecological criteria, producing realistic species distribution outcomes on building envelopes.

2. Existing Models for Habitat Distribution at Urban and Building Scales

While numerous studies have explored the influence of non-anthropogenic factors on biodiversity at the city scale, practical implementation requires research at the scale of architectural and urban elements, such as facades, roofs, or other environmental interfaces. Addressing this gap, the ECOLOPES project introduced an ecosystem-driven design framework that reconceptualises the building envelope as a multispecies living interface shared by humans, plants, animals, and microbiota (Weisser et al., 2022). Its methodology combines ecological and architectural knowledge within a data-integrated computational design environment employing approaches such as Habitat Suitability Modelling (HSM) to promote cross-disciplinary decision-making (Selvan et al., 2025).

HSM predicts the probability of species occurrence by relating recorded species observations to the environmental characteristics of the locations where they were observed, such as temperature, solar radiation, and topography. This process identifies the most suitable areas for species survival (Hirzel et al., 2008). Recent studies have

further advanced this approach in urban contexts by integrating cityscape parameters that capture anthropogenic structures and spatial heterogeneity (Fung et al., 2024).

Alongside the statistical HSM approach, a rule-based Multi-Criteria Analysis (MCA) provides an analytical framework for creating suitability maps by combining multiple environmental and urban variables with weighted importance, enabling habitat assessment where empirical occurrence data are limited (Muller et al., 2018). This research adopts ranked HSM through MCA framework to produce a gradient of suitability scores that allows finer-scale prioritisation and design adaptation (Bhakti et al., 2024).

Despite extensive applications of HSM at regional and city scales, a significant gap remains when translating these models to the scale of architectural or urban elements, where microclimatic, geometric, and anthropogenic factors strongly influence habitat potential. Moreover, we did not find previous research that examined ranked HSM on the building scale in general and for practice-oriented application specifically. This research aims to address these gaps by developing a workflow that integrates architectural and environmental variables with multispecies stakeholders (humans, plants, and animals) in a practice-oriented approach.

3. Methodology

The proposed designer-oriented framework is comprised of a four-stage hybrid computational workflow for multispecies habitat distribution within urban structures (Figure 1). The first stage focuses on data extraction and processing. The second stage involves spatial and environmental analyses to retrieve biotic, abiotic and cityscape contextual variables intrinsic to a selected urban area and relevant to species survival. The third stage develops habitat placement strategies by mapping species preferences in relation to retrieved variables and assigning ranked Habitat Suitability Indices (HSIs) to potential habitat units. The final stage integrates these strategies into the multi-objective optimization (MOO) model (adapted from Selvan et al., 2025) to balance multiple ecological and architectural objectives.

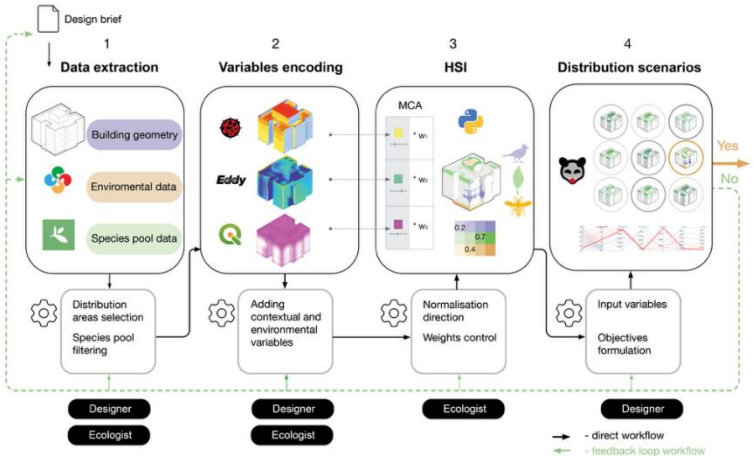


Figure 1. Four-stage computational workflow showing all adjustments and settings on each step.

The process encompasses environmental simulations, ecological data modelling, and multi-objective optimization to generate adaptive and ecosystem-oriented design scenarios. The proposed computational framework is developed within the Rhino-Grasshopper environment, using native, custom-scripted, and third-party plug-in components including Opossum, Eddy 3D and Ladybug tools, along with QGIS platform.

3.1. STAGE 1. DATA EXTRACTION AND PROCESSING

The workflow begins with a building envelope within a defined urban context that needs to be integrated into a multispecies design framework. In this stage, three core data types are assembled:

- Geometric data, including the 3D model of the building envelope and its surrounding context.
- Environmental data, extracted from the EnergyPlus Weather database to characterize climatic conditions of the selected site.
- Ecological data, derived from the Global Biodiversity Information Facility (GBIF), identifying the local species pool relevant to the site.

The input data are aligned with the different stakeholders' objectives, which are formulated as a primary design brief. The architect, ecologist, and engineer collaboratively define initial constraints, such as the target facade area that can be covered with multispecies habitats, species selection from the local pool, and their prioritization based on ecological roles (e.g., pollinators, seed dispersers). By the end of this stage, the system establishes three foundational datasets: the 3D envelope model, the species pool, and the environmental parameters of the study area.

3.2. STAGE 2. ENVIRONMENTAL AND CITYSCAPE CONTEXTUAL DATA ENCODING

The second stage discretizes the building envelope into spatial units, each representing a potential multispecies habitat element. The resolution depends on a minimum habitat size and the level of approximation. Each discretized unit is encoded with multi-layered data.

The dataset encompasses environmental parameters: solar radiation and wind pressure; morphological attributes: surface inclination, height from ground, orientation; anthropogenic influences: proximity to windows, balconies, or electrical systems; abiotic and cityscape contextual factors: proximity to greenery as potential food resources, exposure to noise and light pollution, and other identified site-specific factors. All contextual variables can be derived using QGIS platform, where ecological features such as trees and street lighting are mapped as point data, and Euclidean distances are computed from each facade cell to these points to quantify proximity-based influence; urban noise is similarly represented through distance-to-road metrics, capturing the gradient of anthropogenic disturbance.

This process results in a spatially explicit dataset in which each unit of the analysed building element is represented by a set of environmental, morphological, and

cityscape contextual variables.

3.3. STAGE 3. HABITAT SUITABILITY INDEX (HSI) MAPPING

In this stage, the discretized data are translated into species-specific suitability scores using a ranked HSM approach combined with MCA. Each species in the local pool possesses defined functional traits, which describe ecological tolerances and are converted into parameter thresholds for analysis (e.g., minimum, maximum, and optimal values for sun radiation, noise, or proximity to human activity). For each parameter, suitability is computed using threshold-based piecewise linear or linear functions, producing a normalized score between 0.0 and 1.0. These normalized values are subsequently multiplied by assigned weights representing the relative ecological importance of each parameter for a species under consideration and can be determined through ecological consultation or specified design goals (Figure 2). Weighted parameter scores are aggregated into a composite HSI for each spatial unit, calculated as a weighted mean of the parameter-specific suitability values.

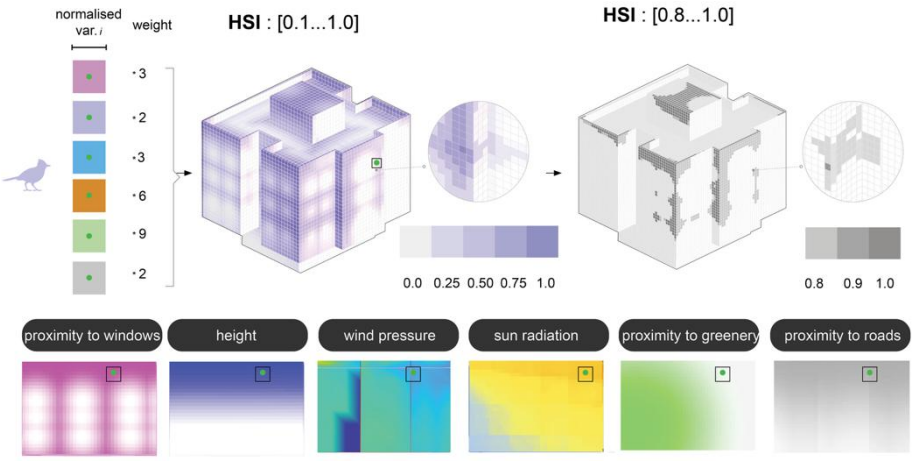


Figure 2. Habitat Suitability Index (HSI) mapping for a single species. On the left, the HSI gradient is displayed, ranging from 0.0 to 1.0. Each cell's HSI value is encoded as a weighted sum of normalized predictor variables, shown at the bottom. On the right, areas with HSI values from 0.8 to 1.0 represent the highest suitability, highlighting the most suitable areas of the building envelope.

Interspecies dependencies are modelled using a proximity coefficient C - the mean HSI of all cells within a defined neighbourhood N (i.e., all cells located within a specified radius around cell j) (Eq. 1), to represent localized ecological interactions and spatial continuity (Eq. 2). For example, the HSI of avian species may be influenced by the surrounding HSI of vegetation, which functions as both trophic support and habitat structure:

$$C_j = \frac{1}{\mathrm{N}} \sum_{i=0}^{N} HSI_{plants} \quad (1)$$

$$HSI'_{species\ A} = HSI_{species\ A} \times C_j \quad (2)$$

3.4. STAGE 4. MULTI-OBJECTIVE OPTIMIZATION

This stage integrates ecological and architectural priorities into a Multi-Objective Optimization (MOO) model based on Selvan et al. (2025). The optimization operates within a nested hierarchical structure of stakeholder objectives, where human, plant, and animal goals are decomposed into measurable Key Performance Indicators (KPIs). Each KPI functions as a directional constraint, guiding the algorithm toward trade-offs among competing criteria.

Within this framework, each cell of the discretized building envelope is treated as an individual design variable, containing both HSI values for each species and a dataset of encoded architectural, ecological and contextual parameters.

The MOO process is implemented with Opossum plug-in, applying the RBFMOpt algorithm. This approach explores interdependent objectives to generate a set of Pareto-optimal configurations, representing a distinct balance between biodiversity and architectural performance.

4. Case study

The developed designer-oriented multispecies habitat distribution framework was tested on a four-storey H-type residential building located in Tel Aviv. The design brief specified 30% coverage of the building envelope with multispecies habitats, distributed to maximize species richness and habitat suitability. Two optimization scenarios were considered: one minimizing occupant disturbance, and the other maximizing thermal performance of the building envelope. In the latter, the optimized distributions were expected to balance biodiversity objectives with architectural performance by approximating the multispecies habitats as a secondary insulating layer that enhances the building's thermal stability (Coma et al., 2017).

Following the design brief, species occurrence data within a 3.5 km radius were extracted from the GBIF database and refined with ecologist input to select four taxa based on ecological importance and adaptability to building-scale habitats. The selected species pool represents ecologically complementary taxa requiring minimal maintenance. The faunal species include Apis mellifera (honeybee, pollinator), Passer domesticus (house sparrow, seed disperser) and Pipistrellus kuhlii (bat, nocturnal insectivore). The vegetative layer was designed to support these functional species groups through drought-tolerant, vertical-growing species Bryonia cretica. The initial species pool was composed to establish an interconnected and resilient ecological network, enabling mutual support between fauna and flora while maintaining compatibility with architectural integration.

4.1. DATA-ENCODING AND HSI MAPPING

The building envelope was discretized in Rhino-Grasshopper environment using a multi-resolution grid, scaled to species-specific habitat sizes, ranging from 0.25 to 1.0-meter cell's side length. Environmental performance simulations were conducted with the aid of Ladybug tools and Eddy3D plug-in to derive solar radiation and building-envelope wind pressure coefficients. HSI values were assigned based on the normalized relationship between each species' tolerance range and the corresponding

environmental variables. Species with high irradiance tolerance, such as Bryonia cretica, followed a direct relationship, while shade-preferring or heat-sensitive taxa, such as the Kuhl's pipistrelle bat, followed an inverse one. For wind pressure, species preferring sheltered conditions (e.g., the honeybee) were assigned higher suitability in areas with low wind-pressure coefficients, whereas airflow-tolerant species (e.g., the bat) were favoured in wind-exposed facade zones.

Proximity to greenery was retrieved from QGIS vegetation layers, using Quick OSM plugins via key words search, imported into Grasshopper via @IT plug-in, and site photography by mapping a point cloud of individual trees and the centroids of green patches. Each discretized unit of the building envelope was assigned a distance value to the nearest mapped point. Artificial light and road noise disturbances were computed following the same logic: lamp locations and road networks were obtained from the QGIS database and verified through on-site photographic mapping, with distances from facade cells encoded according to the assumption that areas less exposed to urban disturbance factors exhibit higher suitability. Weighting coefficients for the final HSI computation were adjusted to account for species-specific behavioural patterns, with nocturnal taxa receiving higher sensitivity to streetlight-related disturbance gradients, while for house sparrows, proximity to the nearest trees received the highest weighting.

To incorporate ecological interdependencies, a hierarchical suitability mapping was implemented. HSI values for vegetation layers (Bryonia cretica) were first computed as baseline habitat resources (Filazzola et al., 2019). Animal taxa were subsequently mapped relative to these vegetative layers through a distance-based proximity function (Eq. 2), with interaction thresholds set according to species-specific behavioural patterns: 1 m for pollinators (Apis mellifera), 1.5 m for nesting birds (Passer domesticus), and 3 m for insectivorous bats (Pipistrellus kuhlii).

Each discretized cell of the building envelope contained six HSI values corresponding to the modelled species, along with six predictor variables: solar radiation, wind pressure coefficient, proximity to human activity, proximity to greenery, street lighting, and noise exposure. Twelve values per cell served as the initial input variables for the Opossum optimization environment.

The primary objective was to maximize the total HSI, representing the overall ecological functionality of the facade, following two defined optimization cases. In the first, a secondary objective aimed to enhance human comfort by maximizing the distance to human activity; only facade cells with HSI > 0.7 were included in the input variables to prioritize habitat suitability and species persistence. In the second, the alternative objective focused on thermal performance, approximated by selecting facade cells with the highest solar radiation. Here, the HSI threshold was lowered to 0.5, shifting the optimization emphasis from ecological suitability toward building thermal regulation, under the assumption that multispecies habitats on highly irradiated areas would act as a secondary insulating layer. Trade-offs arose from differing species functional traits. Vegetative taxa benefit from high irradiance, while associated faunal species that depend on them for trophic and shelter resources generally favour shaded environments. To prevent spatial overlaps between species, the optimization process followed a specific rule-based selection strategy: cells with the highest suitability for the house sparrow were identified first and excluded from the

subsequent search space; vegetation layers were then selected within a defined proximity zone, while bat and bee habitats were selected from the remaining areas beyond this spatial buffer.

The Opossum solver set to RBFMOpt algorithm was used to generate 200 alternative distribution scenarios for each case, from which the Pareto-front solutions were retrieved. The final selected configurations of multispecies habitats distribution represented a balanced trade-off that simultaneously optimized species suitability, human comfort and thermal performance (Figure 3).

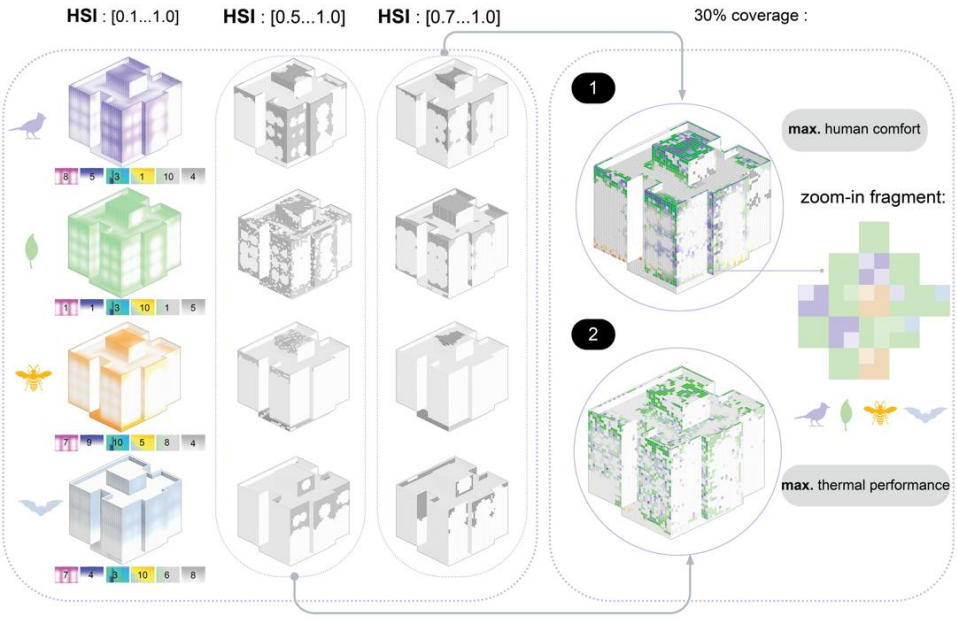


Figure 3. From left to right columns: HSI gradients of four species with suitability from 0.0 to 1.0; areas of the building envelope with HSI from 0.5 to 1.0; selected distribution scenarios for two proposed design cases; a zoomed-in fragment of multispecies distribution.

4.2. RESULTS AND DISCUSSION

Two retrieved distribution scenarios from the Pareto-front solutions, corresponding to two selected design cases, revealed distinct differences in spatial allocation patterns. In the first case, 30% of the total multispecies habitat coverage was concentrated on the northeastern corner of the facade. This clustering can be explained by the high weighting assigned to proximity to external greenery for the Passer domesticus HSI mapping. The total area of these 30% amounts to 1,749 m², within which Passer domesticus nests constitute 32% (average HSI = 0.7), vegetation covers 38% (average HSI = 0.6), Apis mellifera occupies 16% (average HSI = 0.8), and Pipistrellus kuhlii 14% (average HSI = 0.6). In the second design case, the same 30% coverage was more evenly distributed, primarily across the eastern and southern facades, receiving the highest solar radiation. Here, 60% of the area corresponded to vegetation, prioritized

in the MOO input variable selection, with an average HSI of 0.7; birds accounted for 22% (HSI = 0.5), while Apis mellifera and Pipistrellus kuhlii represented 10% and 8%, with HSI values of 0.5.

Distribution scenarios were additionally analysed in the Design Explorer environment to examine how different species' functional traits interact with design-brief objectives. For instance, in the second design case, it becomes evident that achieving maximum thermal comfort is not possible without compromising the HSI of certain species, as scenarios with the highest HSI for faunal species are less effective in terms of thermal performance of the building envelope (Figure 4). These findings underscore the potential of hybrid computational–ecological models to inform data-driven architectural design.

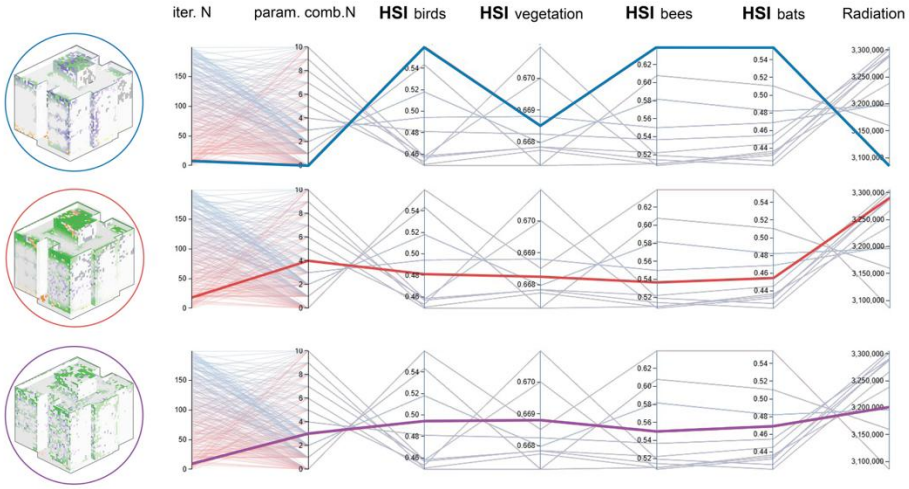


Figure 4. Species HSI and objectives data relation generated in Design Explorer online platform. On top: option with the highest fauna species HSI; in the middle: scenario with the highest thermal comfort potential; on the bottom: selected option with the trade-off of all parameters.

5. Conclusion and Future Development

This study presents a computational data-driven framework that integrates ecological modelling and architectural design workflow to guide multispecies habitat placement on building envelopes. The framework couples GIS-based contextual encoding, Habitat Suitability Index (HSI) mapping with a KPI-based MOO model to evaluate trade-offs between ecological and human-centered design objectives that balance biodiversity support with human and environmental performance.

The proposed computational framework is highly flexible, allowing control over data flows and accommodating all stakeholders in the design process: architects, ecologists and engineers, supporting collaborative biodiversity development strategies, including targeted habitat allocation and ecological network enhancement.

While the current study focused on a single building envelope, the framework is inherently scalable and can be extended to networks of urban structures. However, current HSI formulations remain constrained by their temporally fixed

structure and limited representation of behavioural dynamics.

Future development could integrate statistical validation methods to strengthen predictive reliability, enabling adaptive, feedback-based design environments that simulate seasonal dynamics, behavioural interactions, and evolving ecological niches. This approach advances data-driven ecological design towards more designer-oriented methodologies, positioning building envelopes as active mediators of urban biodiversity, and provides a versatile tool for multispecies envelopes development strategies.

Acknowledgements

This research was supported by the Israel Science Foundation (ISF) (Grant No. 485/25). The author(s) would also like to thank Victoria Culshaw for her ecological consultation and valuable insights.

Attribution Statement

Language editing assistance was provided using ChatGPT (OpenAI). The author(s) are responsible for the final content and interpretation of the manuscript.

References

Weisser, W., Hensel, M., Barath, S., Culshaw, V., Grobman, J., Hauck, T., Joschinski, J., Ludwig, F., Mimet, A., Perini, K., Roccotiello, E., Schloeter, M., Shwartz, A., Sunguroğlu Hensel, D., and Vogler, V. Creating ecologically sound buildings by integrating ecology, architecture and computational design. People and Nature, 4, 2022.

Grobman, Y. J., Weisser, W., Shwartz, A., Ludwig, F., Kozlovsky, R., Ferdman, A., Perini, K., Hauck, T. E., Selvan, S. U., Saroglou, S., Barath, S., Schloeter, M., and Windorfer, L. Architectural multispecies building design: Concepts, challenges, and design process. Sustainability, 15(21), 2023.

Uthaya Selvan, S., Windorfer, L., Saroglou, T., Calbi, M., Culshaw, V., Roccotiello, E., Grobman, Y. J., and Barath, S. Multispecies design decision-making: A hybrid computational model for ecological building envelope optimization and evaluation. Journal of Building Engineering, 108, 2025.

Hirzel, A. H., and Le Lay, G. Habitat suitability modelling and niche theory. Journal of Applied Ecology, 45(5): 1372–1381, 2008.

Fung, Uva Yu Yan, Chris Carbone, Kate Scott-Gatty, Robin Freeman, Robert M. Ewers, and Jessica Turner. "Habitat Suitability as an Indicator of Urbanisation Potential in Four UK Mammals." Mammal Review 54, no. 2, 2024.

Müller, Anke, Peder K. Bøcher, Christina Fischer, and Jens-Christian Svenning. "'Wild' in the City Context: Do Relative Wild Areas Offer Opportunities for Urban Biodiversity?" Landscape and Urban Planning 170, 2018.

Bhakti, T. J. C. P., Mourão Moura, A. C., Puijoni, D., Saliba, L., and Rodrigues, M. Urban biodiversity suitability index: Decoding the relationships between cities and birds. Urban Ecosystems, 27(1): 305–319, 2024.

Coma, Julià, Gabriel Pérez, Alvaro De Gracia, Silvia Burés, Miguel Urrestarazu, and Luisa F. Cabeza. "Vertical Greenery Systems for Energy Savings in Buildings: A Comparative Study between Green Walls and Green Facades." Building and Environment 111, 2017.

Filazzola, A., Shrestha, N., and MacIvor, J. S. The contribution of constructed green infrastructure to urban biodiversity: A synthesis and meta-analysis. Journal of Applied Ecology, 56(9), 2131–2143, 2019.

Climate-Aware Adaptive Regeneration of Ageing Housing Stock

An Agent-Based Framework for Residential Layout Design and Optimisation

Zhang Yin[1] and Frederick Chando Kim[2]
[1,2] *National University of Singapore.*
[1] *zhangyin@u.nus.edu, https://orcid.org/0009-0002-3778-506X*
[2] *fredkim@nus.edu.sg, https://orcid.org/0000-0001-9311-5322*

Abstract. The climate-aware regeneration of residential layouts is essential for meeting global climate goals while addressing functional requirements arising from evolving household needs. However, regeneration methods for residential layouts remain underexplored and rarely integrate environmental performance into the design process. To address this, our research proposes a climate-aware framework for the regeneration of existing housing layouts that adapts to diverse functional needs and supports long-term sustainability. Specifically, an agent-based modelling approach is used to regenerate feasible layouts while preserving the existing structural system. Then, a multi-objective optimisation approach is applied to optimise daylight and energy performance of the generated layouts. To validate the effectiveness of the proposed framework, we use an ageing residential building in Shanghai as a case study. Results indicate that the method generates diverse regeneration options with substantially improved thermal performance in annual heat and cooling load reductions, while maintaining daylight performance comparable to the original layout. Overall, our approach demonstrates robust adaptability in regenerating residential layouts across diverse functional needs. It supports climate-aware regeneration of ageing housing stock and serves as an effective decision-support tool, with the substantial potential to advance sustainable urban transformation.

Keywords. Adaptive regeneration, climate-aware, residential layout renovation, agent-based modelling, multi-objective optimisation, ageing housing stock

1. Introduction

In the past few decades, shifting household needs, evolving building regulations, and rising energy performance expectations have revealed the growing inadequacy of ageing residential buildings (Liang et al., 2025). At the same time, the building sector contributes nearly 40% of global greenhouse gas emissions, making it a critical area for climate change mitigation (International Energy Agency & United Nations Environment Programme, 2018). Regenerating housing, which refers to upgrading

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 459-468.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

existing dwellings to improve spatial functionality, living quality, and environmental performance, has become critical for achieving decarbonization and aligning with global climate goals (Benedetti et al., 2025). In response, China has proposed regenerating about 219,000 urban residential neighbourhoods built before 2000, representing over 31% of its housing stock (NDRC, 2023). This highlights the urgent need for climate-aware housing regeneration strategies that enhance energy efficiency and adapt to evolving household needs.

Existing research on housing regeneration has predominantly focused on upgrading building envelopes (H. Zhang et al., 2025), structural systems (Theilig et al., 2025), and heating, ventilation and air-conditioning (HVAC) systems (Krajčík et al., 2023). However, these interventions primarily aim to enhance environmental performance and do not directly address changing household and functional needs. Layout regeneration provides a critical pathway to accommodate evolving household needs through spatial reorganisation and can also extend the service life of existing buildings by improving performance. Although spatial renovation strategies have been explored in industrial and hospital buildings (Vardouli & Theodore, 2021; Z. Zhang et al., 2024), research on residential buildings remains limited and has seldom incorporated performance evaluation or climate-responsive strategies into the regeneration process. These gaps underscore the pressing need for a digital decision-support framework that integrates spatial reasoning and environmental performance to guide the regeneration of existing ageing dwellings.

In addition, in recent years, automated methods for climate-aware spatial layout generation have rapidly developed (Du et al., 2020), focusing on improving building performance indicators such as energy efficiency, daylighting, and ventilation. However, most existing approaches have been developed for new construction, primarily addressing geometric and topological constraints (Ko et al., 2023). For regeneration, more constraints must be considered, such as existing structural systems and fixed service cores, which restrict design flexibility and significantly increase computational complexity. Machine learning methods and rule-based algorithms have been widely applied to layout generation (Weber et al., 2022). However, the former often lack precise control over functional and geometric variables, while the latter tend to be rigid and difficult to adapt to complex renovation constraints. To address these limitations, agent-based modelling has demonstrated robust performance while effectively reducing computational complexity, and it has been successfully applied to multi-storey and urban block layouts (Guo & Li, 2017; Huang et al., 2025), making it well-suited for renovation contexts with diverse functional needs and complex constraints.

Consequently, this study proposes a climate-aware approach for the adaptive regeneration of existing residential layouts, addressing diverse functional and spatial demands while integrating daylighting and energy performance into the regeneration process. By developing an efficient decision-making tool, the method aims to support large-scale renovation strategies and advance sustainable urban transformation.

2. Method

2.1. AGENT-BASED FRAMEWORK FOR LAYOUT REGENERATION

We propose a climate-aware framework for the automated regeneration of residential layouts, as shown in Figure 1. The workflow consists of three main stages: constraints input, layout generation, and layout optimisation. First, spatial requirements are established based on household needs, and structural constraints are extracted from the current layout. Second, an agent-based method is employed to establish constraints and generate feasible layouts. Finally, a multi-objective optimisation approach based on non-dominated sorting genetic algorithm II (NSGA-II) is employed to assess the daylight access and energy performance of the generated layouts, using spatial daylight autonomy (sDA), annual cooling loads, and annual heating loads as key metrics.

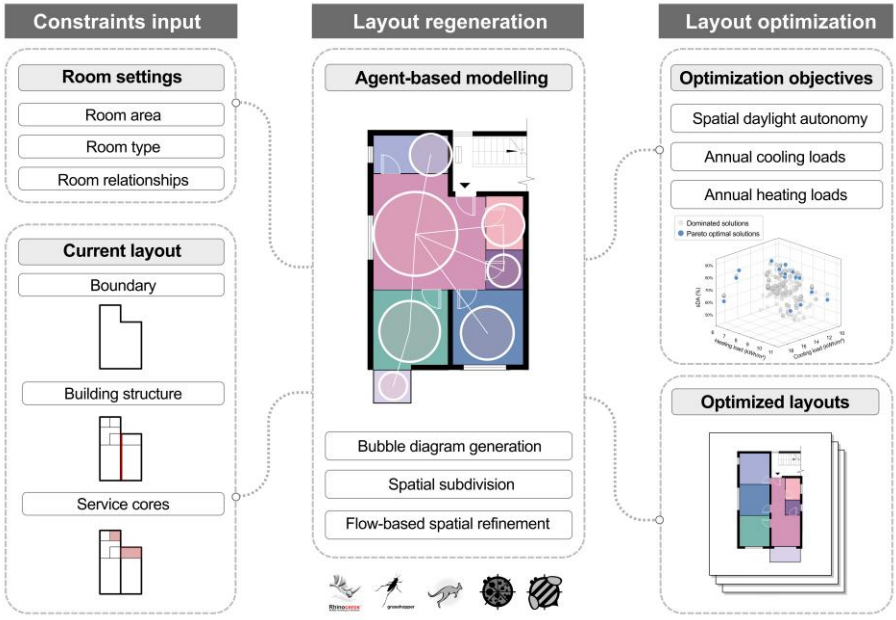


Figure 1. Agent-based framework for residential layout regeneration

2.2. CONSTRAINTS INPUT

This section defines the input parameters for residential layout regeneration. Functional requirements include room types, room areas, and topological relationships, all of which are stored in a CSV file for subsequent reading and processing. The model also extracts the current structural system from the current layout, including the layout boundary, fixed structural elements, and immovable service cores. Fixed structural elements comprise load-bearing walls and columns. Immovable service cores, including kitchens and bathrooms, cannot be relocated due to embedded Mechanical, Electrical, and Plumbing (MEP) systems. All these inputs are used to define layout constraints during the layout regeneration stage.

2.3. LAYOUT GENERATION

This section establishes an agent-based modelling method for residential layout

regeneration that accommodates diverse functional needs while preserving the existing structural system. It comprises three main parts: bubble diagram generation, spatial subdivision, and flow-based spatial refinement.

2.3.1. Bubble Diagram Generation

Bubble diagrams are widely used to represent room functions, topological relationships, and relative positions (Zheng & Petzold, 2023). Here, we model rooms as circular agents sized by target area, with spring-like links encoding topology. To model the room agents, we employ Kangaroo, a physics-based simulation plugin in Grasshopper. We define three constraint types to control the agents' movements and interactions:

1) Attraction and repulsion constraints. Attraction forces are applied to keep adjacent rooms connected and to pull daylight-required rooms toward exterior walls. Repulsive forces are applied to prevent overlap between room agents.

2) Boundary constraints. Rigid-body collisions between room agents and the layout boundary are applied so that all rooms are kept within the layout.

3) Structural constraints. To preserve the existing structure, load-bearing walls and columns are modelled as fixed obstacles, and repulsive forces are applied to prevent room agents from occupying these zones. Additionally, kitchens and bathrooms are treated as immovable service cores, with their locations fixed.

Under the above constraints, room agents interact dynamically until the system becomes stable. Figure 2 shows the process for generating a bubble diagram. The output is a feasible bubble diagram that satisfies functional requirements, topological relationships, and structural constraints.

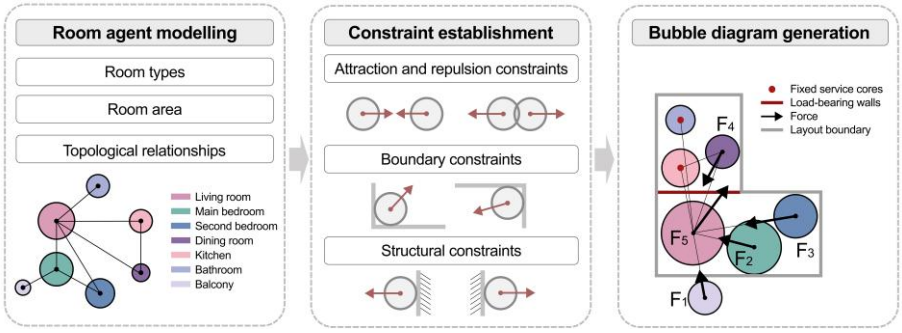


Figure 2. The process of bubble diagram generation

2.3.2. Spatial Subdivision

After generating the bubble diagram, we convert it into a spatial layout using a weighted Voronoi diagram (WVD) (Figure 3). By assigning each room a weight based on its target area, rooms with larger weights receive larger spatial allocations while preserving relative agent positions. A grid system is established to generate regular room layouts, with each grid cell assigned to a room if its centroid lies within that room's boundary. By classifying cells in this way, an initial functional layout is obtained. Because the bubble diagram does not overlap structural walls, the resulting

room layout maximises preservation of the existing structural system. Finally, to ensure room comfort and usability, we set a maximum aspect ratio of 1:2. If any room exceeds this limit, the layout is deemed infeasible and the bubble diagram is regenerated.

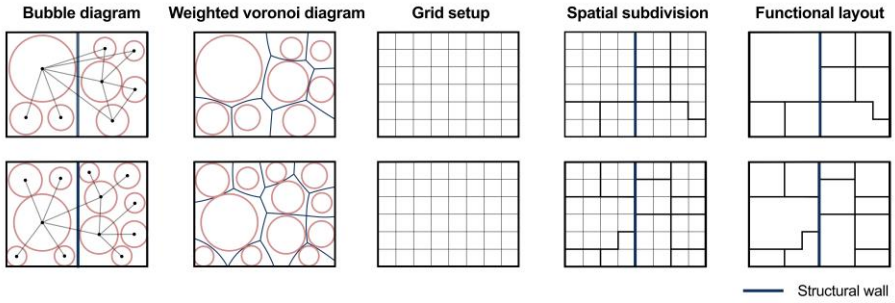


Figure 3. Spatial subdivision based on WVD

2.3.3. Flow-Based Spatial Refinement

After generating the functional layout, an overall circulation network is established to ensure overall network efficiency. We first set the main entrance according to the user's needs. For each space adjacent to the entrance, feasible door locations are identified and the shortest path from the entrance to that space is computed. The shortest paths between connected rooms are then computed without effective circulation. If multiple paths have equal length, the system selects the path closest to the existing circulation network to improve overall network efficiency. All cells along the selected paths are reclassified as circulation. Since the living room is the primary public and circulation space, the circulation area is assigned to it. After establishing the circulation network, the layout is finalised to ensure efficient flows (Figure 4).

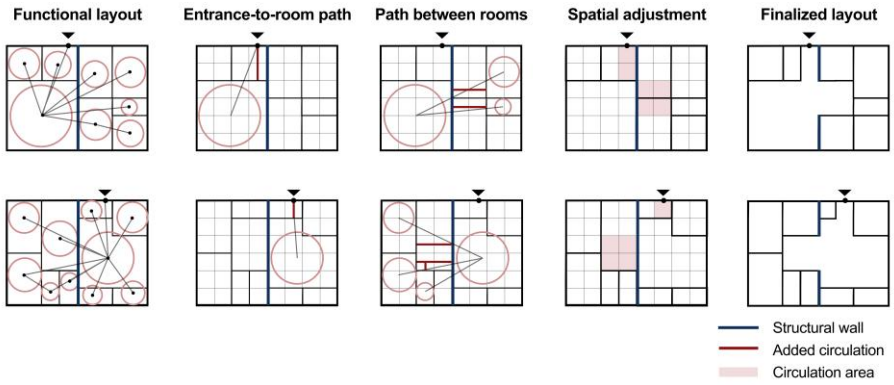


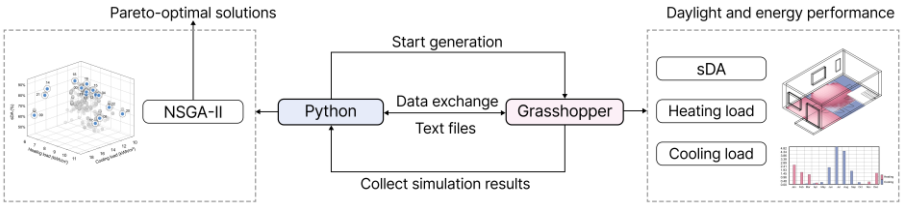
Figure 4. The process of flow-based spatial refinement

2.4. LAYOUT OPTIMIZATION

To obtain climate-aware regeneration layouts, we apply multi-objective optimisation to improve daylight and energy performance of the regenerated layouts. We use three

metrics simulated in Rhino and Grasshopper with Ladybug and Honeybee. Specifically, sDA is used to assess daylight access, defined as the percentage of annual hours when indoor illuminance is at least 300 lx. Annual cooling and heating loads are used to evaluate cooling and heating demand, respectively. For optimisation, we implement NSGA-II in Python, which effectively balances convergence and diversity for complex, nonlinear problems (Deb, 2011). To ensure real-time data exchange between Rhino and Python, Grasshopper writes simulation results to text files after each generation, and Python reads them to drive the optimisation. After optimisation, the optimised layouts and performance outputs are exported as CSV files for subsequent analysis. The workflow of the optimisation process is shown in Figure 5.

Figure 5. The workflow of the multi-objective optimization process



3. Case Study

This study validates the framework using a residential floor plan from Tongji Xincun, Shanghai. The building completed in 1987, has a brick-concrete structure with a 3.3 m storey height. Its layout exhibits the typical plan characteristics of Danwei Compound housing before 2000 (Figure 6). Originally built for students and staff, it no longer meets current household needs, building regulations and performance standards. In our experiments, we merge two adjacent units to accommodate diverse residential needs, while maximising the preservation of the load-bearing wall between them.

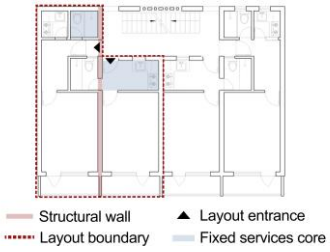


Figure 6. Floor plan of the residential building in Tongji Xincun

In the building modelling stage, the window-to-wall ratios (WWRs) are set to 0.2 (north) and 0.3 (south), with a window height of 1.5 m and a sill height of 0.9 m. The room types, area ranges, and topological constraints are given in Table 1. Simulations use the Shanghai EPW climate file. Lighting, occupancy, equipment, and heating-cooling schedules are set in accordance with the Code for Thermal Design of Civil Buildings (GB 50176-2016). The K-value is specified as 0.6 W/(m²·K) for external walls, 0.4 W/(m²·K) for the roof, 1.5 W/(m²·K) for internal walls, and 1.2 W/(m²·K) for the ground. Windows are modelled with a U-value of 2.2 W/(m²·K), a solar heat gain coefficient (SHGC) of 0.35, and a visible transmittance of 0.66. Natural ventilation

is enabled by opening windows only when the outdoor air temperature is between 20-26 °C. HVAC efficiency is represented by fixed coefficients of performance (COPs) of 4.4 for cooling and 3.5 for heating. Doors are assumed to remain closed due to the minor impact on energy load results.

Table 1. Room constraint settings

Room Number	Room Type	Area (m²)	Adjacent room (s)
0	Living room	[6, 15]	1, 2, 3, 4, 5
1	Bedroom	[4, 12]	0
2	Dining room	[3, 6]	0, 3
3	Kitchen	[4, 6]	0, 2
4	Bathroom	[3, 6]	0
5	Storage room	[3, 6]	0
6	Balcony	[2, 8]	0 or 1

4. Results

For layout generation, we evaluate the effectiveness and flexibility of the proposed agent-based regeneration method by testing four functional requirement sets that cover diverse household needs. These sets span household structures from single occupant to co-living households. The generated layouts are shown in Figure 7.

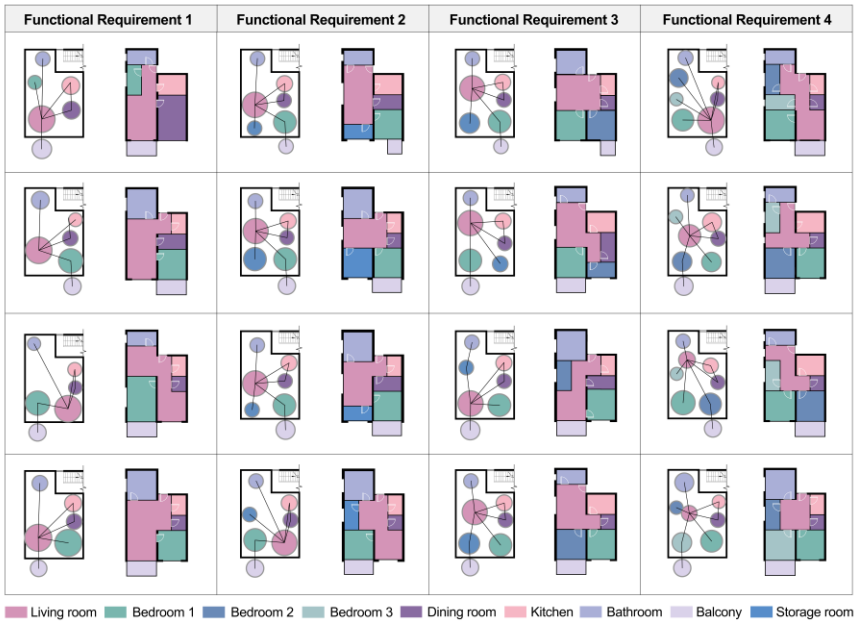


Figure 7. Layout generation results under 4 different functional requirements

The generated layouts exhibit robust adaptability to varying functional requirement sets, satisfying room-type and area requirements. They also exhibit efficient circulation

and meet topological constraints effectively. In addition, the resulting layouts effectively preserve the existing structural walls, with only partial interventions, thereby reducing demolition volume and renovation costs.

For optimisation, we use functional requirement set 3 as the test case and perform multi-objective optimisation. The algorithm ran with a population size of 50 for 30 generations. The optimisation process took about 24.25 hours, resulting in 23 Pareto-optimal solutions. Figure 8 illustrates the spatial distribution of all feasible solutions.

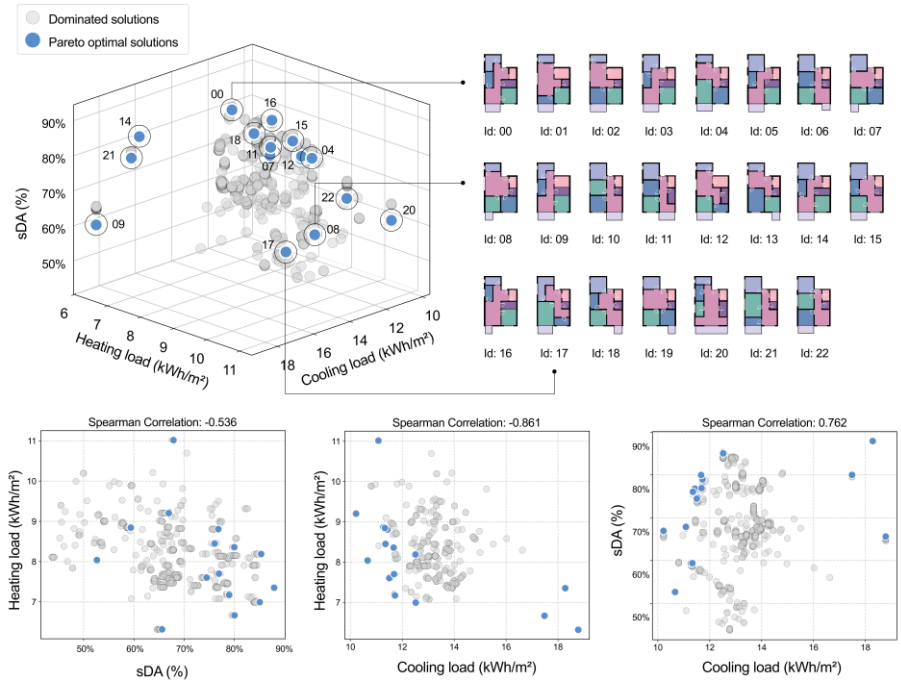


Figure 8. Distribution of optimization solutions

The optimisation results show a wide distribution of solutions across the objective space, demonstrating that the method can generate diverse layouts according to specific functional requirements. Compared with the original layout, Pareto-optimal solutions reduce the annual heating load by 21–57% and the annual cooling load by 20–56%. Daylight performance remains comparable to or higher than the original layout, with sDA values between 53–88%. In addition, the Pareto front illustrates a balanced compromise between daylight and energy performance. For instance, solution 0 achieves high sDA performance but pays with higher cooling loads due to increased solar heat gain. In contrast, Solutions 8 and 17 maintain lower energy demand, but sDA decreases, reflecting a compromise in daylight availability. Overall, these results indicate that the proposed approach can substantially reduce energy demand while preserving daylight performance compared with the original layout, offering decision-makers diverse regeneration options.

5. Conclusion and Discussion

To adapt to evolving functional needs, this study establishes a climate-aware framework for the regeneration of ageing residential layouts, offering a practical pathway for the sustainable transformation of ageing housing stock. The framework integrates functional requirements and building performance while considering structural elements and building service systems in renovation contexts. The case study of a residential building in Shanghai effectively demonstrated the framework's flexibility in meeting diverse functional needs, offering options for climate-aware regeneration. This approach enables rapid design and iterative optimisation in the early stages of design, providing decision support for policymakers and designers. The resulting layouts can be developed into detailed floor plans, demonstrating practical applicability (Figure 9). The method also has the potential to support a dynamically adaptable paradigm for housing regeneration, supporting renovation strategies across different periods and contributing toward carbon-neutral goals.

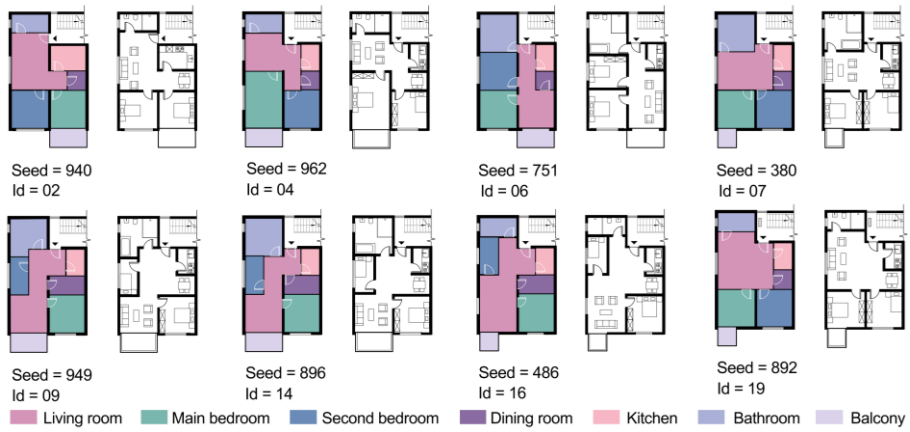


Figure 9. Partial optimisation results and detailed layouts

Although the method proves effective for residential layout regeneration, future work will extend it to more building types and climate zones to test robustness and incorporate sensitivity analysis to quantify environmental performance trade-offs as decision preferences. Moreover, future work will enable 3D component-level interventions (e.g., façade, window, and envelope) to develop an integrated building repair framework spanning spatial, structural, and environmental dimensions.

Attribution Statement

ChatGPT (OpenAI, 2025) was used to translate portions of the manuscript and to improve flow. Grammarly (Grammarly Inc., 2025) was used to correct errors in spelling and grammar.

References

Benedetti, A. C., Costantino, C., Lobosco, R., Predari, G., & Gulli, R. (2025). Comparative LCA Scenarios for Urban Regeneration of Residential Building Stock. Application to an

Existing High-density Urban Block in Bologna. Energy and Buildings, 329, 115270. https://doi.org/10.1016/j.enbuild.2024.115270
Deb, K. (2011). Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction. In L. Wang, A. H. C. Ng, & K. Deb (Eds), Multi-objective Evolutionary Optimisation for Product Design and Manufacturing (pp. 3–34). Springer. https://doi.org/10.1007/978-0-85729-652-8_1
Du, T., Turrin, M., Jansen, S., van den Dobbelsteen, A., & Fang, J. (2020). Gaps and requirements for automatic generation of space layouts with optimised energy performance. Automation in Construction, 116, 103132. https://doi.org/10.1016/j.autcon.2020.103132
Guo, Z., & Li, B. (2017). Evolutionary approach for spatial architecture layout design enhanced by an agent-based topology finding system. Frontiers of Architectural Research, 6(1), 53–62. https://doi.org/10.1016/j.foar.2016.11.003
Huang, Y., Yang, J., Zhang, X., Chen, D., Zhang, Z., & Chen, X. (2025). Intelligent optimization of building complex layouts under multi-constraint conditions: A multi-agent modeling approach. Sustainable Cities and Society, 131, 106734. https://doi.org/10.1016/j.scs.2025.106734
International Energy Agency, & United Nations Environment Programme. (2018). Efficient and Resilient Buildings and Construction Sector. https://wedocs.unep.org/xmlui/handle/20.500.11822/27140
Ko, J., Ennemoser, B., Yoo, W., Yan, W., & Clayton, M. J. (2023). Architectural spatial layout planning using artificial intelligence. Automation in Construction, 154, 105019. https://doi.org/10.1016/j.autcon.2023.105019
Krajčík, M., Arıcı, M., & Ma, Z. (2023). Trends in research of heating, ventilation and air conditioning and hot water systems in building retrofits: Integration of review studies. Journal of Building Engineering, 76, 107426. https://doi.org/10.1016/j.jobe.2023.107426
Liang, Y., Qian, Q. K., Li, B., An, Y., & Shi, L. (2025). A critical assessment on China's old neighborhood renovation: Barriers analysis, solutions and future research prospects. Energy and Buildings, 332, 115407. https://doi.org/10.1016/j.enbuild.2025.115407
NDRC. (2023). Policies-National Development and Reform Commission. https://en.ndrc.gov.cn/policies/index_2.html
Theilig, K., Vollmer, M., Lang, W., & Albus, J. (2025). Multi-criteria decision-making for energy building renovation: Comparing exterior wall structures with the AHP, ANP, utility analysis, and TOPSIS. Building and Environment, 280, 113075. https://doi.org/10.1016/j.buildenv.2025.113075
Vardouli, T., & Theodore, D. (2021). Walking Instead of Working: Space Allocation, Automatic Architecture, and the Abstraction of Hospital Labor. IEEE Annals of the History of Computing, 43(2), 6–17. https://doi.org/10.1109/MAHC.2020.2990111
Weber, R. E., Mueller, C., & Reinhart, C. (2022). Automated floorplan generation in architectural design: A review of methods and applications. Automation in Construction, 140, 104385. https://doi.org/10.1016/j.autcon.2022.104385
Zhang, H., Hewage, K., Bakhtavar, E., Sun, Q., & Sadiq, R. (2025). Robust building energy retrofit evaluation under uncertainty: An interpretable machine learning approach. Energy Conversion and Management, 345, 120362. https://doi.org/10.1016/j.enconman.2025.120362
Zhang, Z., Guo, Z., Zheng, H., Li, Z., & Yuan, P. F. (2024). Automated architectural spatial composition via multi-agent deep reinforcement learning for building renovation. Automation in Construction, 167, 105702. https://doi.org/10.1016/j.autcon.2024.105702
Zheng, Z., & Petzold, F. (2023). Neural-guided room layout generation with bubble diagram constraints. Automation in Construction, 154, 104962. https://doi.org/10.1016/j.autcon.2023.104962

Optimising Pareto Front Selection Through the Use of Artificial Intelligence

A Generative Revitalisation of Rabdan's Pedestrian Landscape

JAMES GOMEZ[1], JOSEF SFARA[2] and MOHAMMED MAKKI[3]
[1,2,3]*University of Technology Sydney.*
[1]*james.gomez@student.uts.edu.au, 0009-0006-1833-2195*
[2]*josef.sfara@student.uts.edu.au, 0009-0008-4375-1863*
[3]*mohammed.makki@uts.edu.au, 0000-0002-8338-6134*

Abstract. Selecting an optimal solution from large Pareto front sets remains a central challenge within multi-objective evolutionary algorithm (MOEA) workflows, particularly in large-scale urban design problems with conflicting objectives. Conventional selection methods, including clustering and phenotypic indicators, are often data intensive and rely on manual interpretation. This paper proposes a reasoning-based Pareto front selection methodology that integrates artificial intelligence into sequential evolutionary simulations. The GPT-4.1 mini model is used to generate context-specific weightings for fitness objectives from a design prompt, within a Rhino/Grasshopper and Wallacei X workflow. The method is tested through an urban redevelopment case study of the Al Qana precinct in Abu Dhabi. Results demonstrate that AI-assisted weighting can streamline selection while preserving designer intent.

Keywords. Abu Dhabi, Multi-Objective Evolutionary Algorithm, Urban Redevelopment, Artificial Intelligence, Chat GPT

1. Introduction

The analysis of Pareto front solutions for the purpose of selecting the most suitable solution for the design problem being investigated remains a critical challenge within Multi-Objective Evolutionary Algorithm (MOEA) workflows, particularly when applied to large-scale urban projects. Complications arise when the design problem is comprised from multiple conflicting objectives, and the MOEA generates a large number of Pareto front solutions, leaving the designer with many viable design solutions, each one optimised for a different fitness objective (Petchrompo et al., 2022). As it stands, there are various ways to narrow down the Pareto front solutions generated by a MOEA's, such as clustering the pareto front for manual assessment of cluster centres or the application of 'phenotypic indicators' to differentiate between pareto front solutions based on metrics outside of the ones used by the fitness objectives (Showkatbakhsh & Makki, 2022). Although such approaches provide the user with a suitable method of assessing the pareto front for the purpose of selection, they are time

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 469-478.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

consuming and ‘data heavy’, necessitating manual tasks that are seldom automated.

Alternative methods for narrowing down the pareto front for selection is through the assignment of weightings to individual fitness objectives. While this approach allows designers to prioritise the Pareto front solutions of specific fitness objectives, the reasoning for prioritising certain fitness objectives tends to simply be based on which objective is deemed the most important, rather than a means of selection specific to the context of the outputted phenotype.

This research proposes a new methodology for Pareto front selection that intends to utilise AI, specifically the [gpt-4.1-mini] OpenAI model, to develop a reasoning-based means for weighting solutions, within existing design tools, primarily Rhino/Grasshopper and Wallacei (Makki et al., 2018) This process would utilise a prompt specific to the context of the design problem being investigated, with the AI model then returning a new weighting specific to the problem’s design objectives.

To test this methodology, an urban redevelopment project in Abu Dhabi is used as a case study, focusing on the Al Qana precinct. Al Qana provides a strong basis for this investigation due to the presence of multiple interdependent factors whose improvement produces inherently conflicting fitness objectives. The precinct exhibits several characteristics associated with car-centric urban design, including fragmented pedestrian networks, limited spatial continuity, and reduced walkability, making it a suitable context for testing multi-objective optimisation and selection strategies.

2. Background and context

2.1. PARETO FRONT ANALYSIS AND SELECTION

The advantage of employing an MOEA is primarily due to the algorithm’s ability to solve multiple conflicting objectives without the requirement for the designer to make trade-off decisions; however, in doing so, the algorithm presents a primary challenge to the designer. Optimising for multiple conflicting objectives naturally leads to multiple optimal solutions, where there is no single ‘best’ solution for the designer to select (Deb, 2001). What is usually generated is a number of optimised trade-off solutions that form the Pareto front; the larger the Pareto front, the more challenging it is for the user to select the ‘optimal solution’ (Luke, 2013). Therefore, the process of selection plays a key role in how the designer rationalises the algorithm’s output, one that is directly correlated with the complexity of the design problem.

There have been various intervention strategies to extract a single design solution from the algorithm’s output, such as merging the different fitness objectives into one fitness objective, resulting in a single objective problem with a single design solution (Zhang et al., 2024), or through directing the algorithm to optimise for one fitness function over the other, primarily through defining weights to the fitness objectives prior to running the algorithm (Marsault, 2013), or through the analysis and filtration of the pareto front after the algorithm has concluded based on metrics independent from the fitness objectives (Makki & Melsom, 2024).

Most recently, research has examined the integration of Large Language Models (LLMs) within MOEAs as a means of guiding evolutionary search, supporting decision-making, and optimisation guidance (Liu et al., 2025; Wang et al., 2024),

however their role in post optimisation analysis of the Pareto front for selection remains underexplored, despite its potential as a valuable alternative to the analysis and selection strategies outlined above.

2.2. URBAN SEGMENTATION - ABU DHABI'S CAR-CENTRIC MASTER PLANNING

The urban development of Abu Dhabi has been highly turbulent, with the city's master plans primarily coming from foreign architects and urban planners, changing hands on multiple occasions. This reliance on international consultants has contributed to Abu Dhabi's car-centric culture, employing design principles that do not address the city's context. In 1988, following a period of economic growth and mass urbanisation due to the increased price of oil, Abu Dhabi's Executive Council identified the need for a city-wide planning framework to better control and coordinate urban growth. To address this, W.S. Atkins & Partners Overseas was commissioned to prepare 'The Master Directive Plan for Abu Dhabi and its Environs' (MDP), a document that would guide development between 1990 and 2010 (Elsheshtawy, 2008). While the MDP was never formally adopted, elements of the plan were applied by municipal authorities, in particular, the recommendation to shift large-scale development towards the mainland.

These plans are one of the primary contributors to the car-centric design of Abu Dhabi, segmenting single-family housing in peripheral suburbs and densifying high-rise commercial and residential towers in the city core. The resulting urban segmentation that occurred due to these plans has been detrimental to the potential for pedestrian activity in Abu Dhabi, as by isolating suburbs, the city's harsh climate deters any pedestrian ventures between homes and amenities given the sheer resulting distance (Scoppa et al., 2018). This again highlights the poor application of international design principles that are unresponsive to the city's own environmental conditions, with many of the MDP characteristics influenced by North American suburban planning principles, a culture which does not contend with a similar climatic context (Elsheshtawy, 2008). As demonstrated by a space syntax study of Abu Dhabi in 2023 (Gomes et al., 2023), the network of highways that interconnect the city, in the absence of a car, isolates its neighbourhoods with a lack of consideration as to pedestrian traversal and access both in and between neighbourhoods. The result is a city comprised of highly segmented program and facilities, with the urban landscape, especially in suburban areas, providing no incentive for pedestrian activity.

2.3. THE PEDESTRIAN EXPERIENCE OF AL QANA AND RABDAN

The relationship between Al Qana and the suburb of Rabdan exemplifies the broader failures of pedestrian accessibility in much of Abu Dhabi's suburban environment. Al Qana, an artificial canal and entertainment precinct directly south of Rabdan, attracts high levels of pedestrian activity relative to typical public spaces in the city, owing to the integration of mixed use typologies, as well as the creation of overshadowed spaces as a result of the precinct's urban configuration. Despite this, the built environment fails to facilitate direct pedestrian connectivity between Rabdan and Al Qana. The urban fabric separating these areas is dominated by a combined main road and parking strip exceeding 30 m in width, almost entirely unshaded, with no clear or legible means of

pedestrian access. This condition reflects the prioritisation of vehicular transportation and the fragmentation of pedestrian networks that characterises much of Abu Dhabi's superblock-based planning, which impedes walkability and reduces route directness for pedestrians in low-density suburban contexts.

A potential solution to this lack of connectivity lies within Rabdan itself, this being the network of 'Sikkak'; narrow pedestrian alleyways that separate residential structures and connect to the main streets. The Sikkak network has been shown through space syntax and network analyses to contribute substantially to the efficiency and directness of street networks by providing alternative, shorter routes that complement primary vehicular streets (Scoppa et al., 2018). The role of the Sikka in enhancing pedestrian circulation is recognised in Abu Dhabi's planning manuals, where they are described as key pedestrian routes that should link residential areas to local services and be shaded by surrounding buildings (Abu Dhabi Public Realm Design Manual, 2011). Compared to surrounding main roads, Sikkak typically offer more favourable walking conditions, providing partial shading for the majority of the day, developing permeable urban networks and human-scaled spatial configurations.

Cultural norms and climate have conditioned residents to favour car use over walking (Ochieng & Jama, 2015), an outcome of decades of car-oriented planning and automotive normalisation. As a result, the spatial and socio-cultural conditions that could support pedestrian movement are undercut, reinforcing auto-dependence even where attractive pedestrian environments could exist within residential neighbourhoods.

3. Methodology

The presented experiment utilises Abu Dhabi's Al Qana suburb as a case study to examine the role of LLMs to analyse the pareto front of an MOEA for the purpose of selection. OpenAI's [gpt-4.1 mini] LLM was used to read '.csv' data file output from Wallacei X; generating a method to rank solutions from Pareto front. The workflow of the script follows as such:

Grasshopper → Python Server → OpenAI API → Python Server → Grasshopper

The experiment develops possible superblock solutions that address the innate issues that an urban fabric tailored to being car-centric has on different scales. By breaking down design decisions influenced by vehicles, the research introduces sustainable and pedestrian-oriented landscapes that promote economic growth and create 'third spaces' that allow informal activities and social interaction to strive.

Sequential simulations have been utilised in previous research to address complex design problems with a large number of design goals or problems with varying scales (Randall et al., 2020; Giang et al., 2023). The MOEA used for the experiment is the NSGA-2 algorithm developed by Kalyanmoy Deb (Deb et al., 2000) within Wallacei X (Makki et al., 2018), a plugin for Grasshopper 3D. The simulation was performed on a mid-end consumer-grade PC with the following specifications: OS. Windows 11 Pro 64-bit (ver.24H2); CPU. 12th Gen Intel i7-12700KF (12 Cores) @ 3.60 GHz; RAM. 32 GB Corsair DDR5 6000 MHz; GPU. Nvidia RTX 4070 TI SUPER 16 GB.

4. Experiment Setup

The experiment is segmented into two simulations to enable the algorithms to investigate the site through a unique set of criteria. The first simulation optimises the urban network of the site against its surrounding context. Space syntax theory (Hillier et al., 1984) is integrated as a metric for accessibility and spatial configuration to help guide the algorithm develop solutions that respond to pedestrian integration, route choice and centrality. The main intent of this stage is to evolve a solution set that introduces radical design proposals, injecting man-made canals as well as dedicated green spaces throughout the site while maximising pedestrian integration between points of interest within and outside the superblock. The second simulation optimises the topology of built forms based on the resulting phenotypes from the first simulation. This stage analyses criteria that involve pedestrian metrics of comfort and experiences.

4.1. DESIGN PROBLEM SETUP

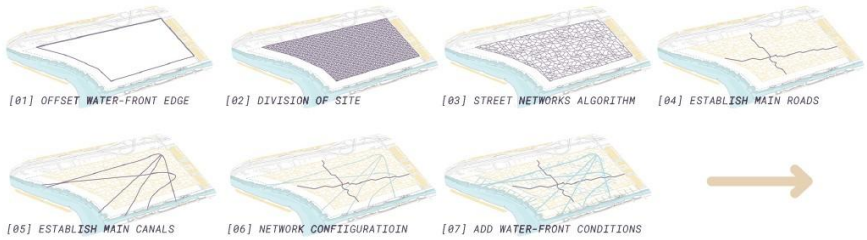


Figure 1. Road Network Construction Method

Figure 1 presents the chromosomes and the step-by-step process to construct the phenotype for the first stage. The waterfront edges are offset to create two sections of the site; the waterfront edge condition is only pedestrian accessible. The latter section of the site is contoured along the existing grain of the surrounding road networks along with an additional diagonal contour to distinguish the network from the common grid network. The algorithm then chooses points along the waterfront in conjunction with random points within the site and designates them as points of interest. An A* search algorithm (Hart et al., 1968) is used to find the shortest walk between the points which becomes the basis for the urban network of the phenotype. Main roads are assigned to allow vehicular access through the site, this decision segments the site into four in which the algorithm randomly selects a segment and a point within to connect with the water's edge, thereby establishing the canals that run throughout the site.

Green spaces are identified within each solution while road and canal networks are introduced into the site based on measures of betweenness centrality; plots are filtered based in order of the following criteria: adjacent to main road intersection point; adjacent to the canal network (an area threshold is introduced to separate mangrove dedicated areas and built spaces); adjacent to nodes with high valency; adjacent to inner main roads; adjacent to superblock boundary edge; then the remaining plots are infilled with residential separated into low, medium, high-rise based on plot area (Figure 2).



Figure 2. Plot Filtering Process

For the first simulation the algorithm optimises for six fitness objectives: 1. Maintain 1:4 Green-to-Built Ratio; 2. Maximise Reach of Canals; 3. Maximise Distance Between Mosques; 4. Minimise Turns on Main Road; 5. Minimise Travel Time to Points of Interest; 6. Maximise Mangrove Area.

Stage 2 utilises an output from Stage 1 as a basis for creating the built forms of the superblock. The algorithm filters the groups further by area size to cull the small land packages and return them as green spaces. A collection of methods and techniques are utilised to create unique topologies that represent each filtered group. In this context, the following fitness objectives are used in the second simulation: 1. 1500 / km2 Population Density; 2. Maximise Evaporative Cooling; 3. Maximise Overshadowing; 4. Maximise Pedestrian Experience; 5. Maximise Views to Points of Interest; 6. Minimise South-East Winds.

5. Selection Process — AI Analysis Proposal

The focus of this experiment is to explore alternative strategies to analysing and selecting solutions that meet the criteria of the design brief. Rather than arbitrarily searching for a Pareto front solution through manually extracting each dataset and comparing it to other solutions, this experiment integrates AI into the selection process to enable quicker decision making across large complex datasets.

5.1. ANALYSIS OF PHENOTYPIC INDICATORS, URBAN METRICS & FITNESS OBJECTIVES

Post-optimisation analysis of Phenotypic Indicators and Urban Metrics can be included within this phase of the selection process. These would be metrics that are not captured by the fitness objectives of the MOEAs and are therefore excluded. Example

phenotypic indicators include: total area of plots (km^2); average area per plot (m^2); number of buildings.. Whereas examples of urban metric analysis involve: average distance between green spaces; connectivity to waterfront; sun-shaded sides of the building. These metrics can be combined with the fitness objective scores in one '.csv' file.

5.2. NORMALISED & WEIGHTED SCORE RANKING SYSTEM

To allow for the LLM to read the compiled '.csv' file, the Pareto front solutions were clustered using a K-means algorithm, with a K-value of 20 to output a list of 20 cluster centres. A prompt was created, based on the original experiment design goals, which the LLM used to normalise the phenotypic indicators, urban metric and fitness objective scores and gives a weight to the items, this creates a scoring system based on the given contextual prompt. The experiment opted into OpenAI's [gpt-4.1 mini] due to the nature of the case study. A large Context Window is needed so that the LLM can read the entire dataset output from Wallacei X. This is due to the fact that all text given to AI is measured in 'tokens'; [gpt-4.1 mini] offered a basis of 1 million tokens compared to the newer [gpt-5.2] model, only offering 400k tokens.

5.3. SELECTION PROCESS WALKTHROUGH

The following presents one example of the generated prompt, and the resulting normalised values for the selected phenotype, Gen. 189 Ind. 35.

" *The site now has several man-made canals to allow for mangroves to be scattered around the site, it now has several 'city centres' that allow for pedestrian focused zones. When planning for the buildings there needs to be focus on the pedestrian experience. If pedestrians are able to utilise this suburb despite the harsh climate this project would be deemed a success.* "

The fitness objectives are normalised and have AI-generated weights based on the prompt applied, then the sum of these scores become the basis of the ranking system (Table 1). The LLM also provides a summarised analysis of the phenotype:

"*This solution balances high mangrove coverage with short pedestrian travel times, reinforcing the pedestrian-first planning intent under climatic stress.* " Gen. 189 Ind. 35 LLM Rationale Statement

The method is applied to relevant phenotypic indicators and urban metrics. The AI model returns the following: a list of how the items are weighted; the 20 cluster centres based on this ranking system; rationale on why it thinks the solution is ranked this way; and general design considerations moving forward:

" *Prioritize designs that maximize canal reach and mangrove coverage to enhance ecological and waterfront benefits, as seen in top clusters 11 and 16…Overall, integrate ecological features with efficient transport and urban form to maximize both environmental and social benefits…* " LLM General Design Considerations

Fitness Objectives	Raw Values	Normalised	Weight	Adjusted
FO1 Maintain 1:4 Green-to-Built Ratio	20.68 %	0.81	1.0	0.81
FO2 Maximise Reach of Canals	96.13 %	0.64	0.8	0.51
FO3 Maximise Distance Between Mosques	566.7 m	0.72	0.2	0.14
FO4 Minimise Turns on Main Road	28 turns	0.54	0.3	0.36
FO5 Minimise Travel Time to Points of Interest	8 mins	0.78	0.6	0.47
FO6 Maximise Mangrove Area	12.62 %	0.88	0.9	0.79
Total Score				**3.08**

Table 1. Gen. 189 Ind. 35 LLM Scoring System

Gen. 189 Ind. 35 was chosen for the urban network and the basis for the second simulation. This solution was ranked 3rd overall despite not achieving the highest rank for any single fitness function. The LLM noted that this solution was consistent across pedestrian metrics, revealing the robustness of the selection process rather than selection based on extremities (figure 3).

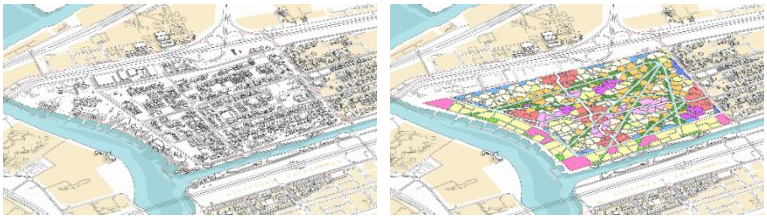


Figure 3. Comparison of Original Suburb and Selected Solution (Gen. 189 Ind. 35)

The same process was completed for the second simulation using the same prompt as a control variable, to verify if the reasoning of the LLM stayed consistent throughout the experiment. The returned weighting reflected the same rationale for the first simulation prioritising pedestrian experiences and climatic comfort (Table 2).

Fitness Objectives	Raw Values	Normalised	Weight	Adjusted
FO1 1,500/km^2 Population Density	2,363 / km^2	0.49	0.4	0.20
FO2 Maximise Overshadowing	46.06 %	0.84	0.9	0.76
FO3 Maximise Evaporative Cooling	36.89 %	0.62	0.8	0.50
FO4 Maximise Pedestrian Experience	13.83 %	0.60	1.0	0.60
FO5 Maximise Views to Points of Interest	7.61 %	0.53	0.2	0.06
FO6 Minimise South-East Winds	37.76 %	0.73	0.7	0.51
Total Score				**2.63**

Table 2. Gen. 86 Ind. 20 LLM Scoring System

"*Cluster 8 (Gen. 86 Ind. 20) leads with the highest score (2.63), excelling in overshadowing and wind minimisation while maintaining strong pedestrian experience and moderate population density…* " Gen. 86 Ind. 20 LLM Rationale

Statement.

The ranked Pareto-front solutions from the LLM are evaluated against independent criteria that are detached from the selection process; these solutions are extracted from the LLM into the modelling environment to assess whether their quantitative performance translates into spatial and urban qualities required by the design goals. This assessment provides validation that the LLM selection process is not arbitrarily ranking solutions, but rather correctly identifying patterns in quantitative data within the Pareto-front set in relation to the design brief.

6. Discussions and Conclusions

The experiment displays the potential of AI within the realms of grasshopper but more importantly an AI-assisted post-optimisation selection framework that is replicable with other MOEAs. Optimisation of the '.csv' file would allow more flexibility for different LLMs to be utilised, in which complex MOEAs, with a large number of Pareto front solutions, are able to be instantaneously ranked based on a contextual prompt of the design problem.

This experiment demonstrates the capabilities of an AI-assisted selection process for the Pareto front solutions. When the selection process was repeated using identical prompts, three of the top 5 phenotypes stayed consistent which indicates stability and robustness of the system. The LLM selection process does not replace existing manual explorations of the Pareto-frontier but provides an option that helps designers operate MOEAs with more than 20,000 solutions. Opting into this process over manual workflows allows designers to immediately rank and identify a solution, thereby enabling designers to focus on the manual design stage rather than allocating substantial amounts of time to analyse large datasets. The final selection of the 3rd ranked solution (Gen. 189 Ind. 35) was a result of the prompt devaluing a fitness function and phenotypic indicator that was particularly important for the design brief. Hence, when utilising this selection method, sensitivity and caution must be exerted when writing the contextual prompt in order to get the desired rationale for selection.

Further research is needed to examine why the AI model selected a specific solution over another and ranked it the best. This experiment lays the foundations to be able to extrapolate what different LLMs read and ignore from the contextual prompt to give the weights as well as how the specified AI model interacts with problems within architecture and urban design. If trained efficiently, AI can be capable of interpreting a designer's intentions based on a project's design brief, facilitating new methods for selecting the optimal design solution that responds to complex design problems.

ChatGPT [gpt-5.1] (OpenAI, 2025) was used to write Python 3 code to be integrated into a Grasshopper GhPython Script component.

References

Abu Dhabi Public Realm Design Manual. (2011). *Abu Dhabi Urban Planning Council.* Abu Dhabi.

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *In International conference on parallel problem solving from nature* (pp. 849-858). Springer, Berlin, Heidelberg.

Deb, K. (2001). *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley & Sons.

Elsheshtawy, Y. (2008). What happened to Abu Dhabi's urbanism? *The question of regional integration. Built Environment*, 34(3), 352–364.https://doi.org/10.2148/benv.34.3.352

Giang, L. A., Tran, M. T., Balossino, S., & Makki, M. (2023). Urbanising Mars: Establishing an Urban Framework for Martian Settlement Through Sequential Evolutionary Simulations. Proceedings of the *28th CAADRIA Conference*, 1, 291–300.

Gomes, A., Kyriazis, A., Schwinger, P., Montagne, C., & Zhand, S. (2023). Middle East Centre Roads (Dis) Connecting Cities and Neighbourhoods A Socio-Spatial Study of Abu Dhabi. *LSE Middle East Centre Paper Series* , 80.

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2), 100–107. https://doi.org/10.1109/TSSC.1968.300136

Hashim, A. R. B. (2018). *Planning Abu Dhabi: An Urban History*. Routledge.

Hillier, B., & Hanson, J. (1984). *The Social Logic of Space*. Cambridge University Press.

Liu, F., Lin, X., Yao, S., Wang, Z., Tong, X., Yuan, M., & Zhang, Q. (2025). Large language model for multiobjective evolutionary optimization. *International Conference on Evolutionary Multi-Criterion Optimization*, 178–191.

Luke, S. (2013). *Essentials of Metaheuristics (Second edition).* Lulu. http://cs.gmu.edu/sean/book/metaheuristics/

Makki, M., & Melsom, J. (2024). A Modular Greening Intervention in Sydney CBD: Variation and Optimization through Evolutionary Computation. *In The Routledge Handbook on Greening High-Density Cities*. Routledge.

Makki, M., Showkatbakhsh, M., & Song, Y. (2018). *Wallacei: An evolutionary and Analytic Engine for Grasshopper 3D.* Wallacei. https://www.wallacei.com/

Marsault, X. (2013). A multiobjective and interactive genetic algorithm to optimize the building form in early design stages. *Building Simulation 2013*, 13, 809–816.

Ochieng, M., & Jama, M. (2015). *The implications of automobile dependency in Abu Dhabi city*.

OpenAI. (2025). *ChatGPT Model: GPT-5.1 [LLM]*. https://openai.com/chatgpt/

Petchrompo, S., Coit, D. W., Brintrup, A., Wannakrairot, A., & Parlikad, A. K. (2022). A review of Pareto pruning methods for multi-objective optimization. *Computers & Industrial Engineering*, 108022.

Randall, M., Kordrostami, T., & Makki, M. (2020). The Taikoo Shing Superblock: Addressing urban stresses through sequential evolutionary simulations. *Proceedings of the 25th CAADRIA Conference*, 1, 415–424.

Scoppa, M., Bawazir, K., & Alawadi, K. (2018). Walking the superblocks: Street layout efficiency and the sikkak system in Abu Dhabi. *Sustainable Cities and Society*, 38, 359–369.

Showkatbakhsh, M., & Makki, M. (2022). Multi-Objective Optimisation of Urban Form: A Framework for Selecting the Optimal Solution. *Buildings*, 12(9), Article 9.

Wang, Z., Liu, S., Chen, J., & Tan, K. C. (2024). Large language model-aided evolutionary search for constrained multiobjective optimization. *International Conference on Intelligent Computing*, 218–230.

Zhang, R., Xu, X., Liu, K., Kong, L., Wang, X., Zhao, L., & Abuduwayiti, A. (2024). Does architectural design require single-objective or multi-objective optimisation? A critical choice with a comparative study between model-based algorithms and genetic algorithms. *Frontiers of Architectural Research*, 13(5), 1079–1094.

Urban Cohabitation within Mangrove Ecosystems through Evolutionary Computation

A Case Study of Abu Dhabi

Laiba Jamil[1], Nancy Sun[2], Mohammed Makki[3] and Apostolos Kyriazis[4]
[1,2,3] *University of Technology Sydney*
[4] *Abu Dhabi University*
[1]*laiba.jamil@student.uts.edu.au, 0009-0002-9633-1471*
[2]*nancy.sun@student.uts.edu.au, 0009-0002-2092-8626*
[3]*mohammed.makki@uts.edu.au, 0000-0002-8338-6134*
[4]*apostolos.kyriazis@adu.ac.ae, 0000-0001-6052-6782*

Abstract. Contemporary urban developments are often delivered at the expense of existing ecological systems, with landscapes cleared or stabilised to accommodate for standardised, imported urban models and globalised architectural languages. In regions rapidly developing with foreign planning logics with the desire for international visibility, the result has often been a built environment ill-suited to the local climate and its ecologies. This paper situates this discussion through the development of Shams in Al Reem Island, Abu Dhabi, where large-scale mangrove removal facilitated land reclamation occurs and hence high-density urbanisation. The study examines how urban development could have evolved if ecological systems were treated as generative forces and not as obstacles. Through counterfactual reconstruction, the present-day site is resituated within a hypothetical pre-development condition shaped entirely by mangrove ecologies. An evolutionary generative framework employing multi-objective optimisation evaluates urban morphologies shaped by ecological continuity, density and connectivity priorities. The findings demonstrate that urban form can emerge through cohabitation with ecological systems rather than erasure, offering a critique of extractive models of urbanisation while proposing a climate-responsive alternative.

Keywords. Abu Dhabi, Mangroves, Ecosystem, Evolutionary Computation, Urban

1. Introduction

Abu Dhabi is currently dealing with a conflict along its shoreline, driven by a construction sector that favours high-end waterfront developments and by the constant decay and erasure of its coastal frontage (Kyriazis et al., 2022). The city's massive land

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 479-488.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

reclamation projects over the decades reveal the prioritisation of the construction sector of the preservation of local ecological conditions. A primary example of this is the erasure of Abu Dhabi's mangrove habitats to make way for urban and building typologies that are incompatible with the surrounding climatic conditions, ultimately revealing an approach to development that is unsustainable (Paleologos et al., 2019). It is projected that by 2030, more than half of the mangrove habitat will be at risk of destruction, a direct consequence of anthropogenic interventions. It undermines the mangrove ecosystem's ability to sustain extreme weather conditions, in addition to carbon sequestration (Environment Agency - Abu Dhabi et al., 2024). As described in the Plan Abu Dhabi 2030 Urban Structure Framework (Abu Dhabi Urban Planning Council, 2007), the city's islands, sea, and coastlines blend to create Abu Dhabi's unique environment, which co-existed with the people living within it for thousands of years, and these resources preserve the connection between humans and the surrounding environment (Abu Dhabi Urban Planning Council, 2007). This aligns with the visions of Sheikh Zayed, Abu Dhabi's founding father, who reiterated, "*On land and in the sea, our fore-fathers lived and survived in this environment. They were able to do so because they recognised the need to conserve it, to take from it only what they needed to live, and to preserve it for succeeding generations*" (Sheikh Zayed bin Sultan Al Nahyan, 1998).

This ongoing depletion of ecologies not only represents environmental loss but also a continuous detachment from the city's cultural and historical identity. Alongside this, the city's state of perpetual construction, its homogeneous urban structure and heavy reliance on car infrastructure (Gomes et al., 2023) result in a transient urban condition disconnected from place (Elsheshtawy, 2019). This disassociation is best represented by the development in Shams, Al Reem Island, an artificial landform with protruding isolated high-rises that fail to acclimate to its surroundings and demonstrate poor social integration. This study investigates how alternative approaches to urban typologies can foster cohabitation with mangrove habitats. By re-introducing the site's original mangrove ecologies, the project reimagines Shams' urban fabric through the employment of a multi-objective evolutionary algorithm, concluding in an unconventional model that challenges population density while minimising ecological degradation.

2. Background and Context

2.1. URBAN FORMATION OF ABU DHABI

Abu Dhabi did not urbanise through an evolution of growth but instead was produced through rapid, systematic development. In less than five decades, subsidised by oil revenue, it was able to abandon its identity as a rural fishing and pearling village in order to adopt the title of a modernised capital city (Kyriazis et al., 2017). Since the catalytic discovery of oil in 1958, the city underwent several stages of building reform that progressively developed Abu Dhabi's current urban fabric (Bani Hashim, 2018). Early resistance to rapid urbanisation under Sheikh Shakhbut was motivated by his concerns about losing the traditional Bedouin way of life. This reluctance was replaced by a governance shift, where Sheikh Zayed's modernisation agenda leveraged this newfound resource to accelerate redevelopment. This transition marked the departure

from locally rooted building practices as it led to the demolition of traditional barasti houses and replaced them with cement buildings, devoid of ornamentation (Bani Hashim, 2018). Under these conditions, international style became dominant, leading to the claim that modernism became the vernacular architecture of Abu Dhabi (Bani Hashim, 2018). Accompanied by standardisation, suburbs came to be architecturally homogeneous and lacked distinct identity (Bani Hashim, 2018). Regulatory shifts in the 1980s allowed building heights to increase to 20 storeys; and subsequent reforms in 2004 intensified these limits, leading to rapid vertical growth and the emergence of the glass tower and shopping mall typologies during this period of urban expansion (Bani Hashim, 2018). Simultaneously, Abu Dhabi also expanded outwards through its active undertaking of large-scale land reclamation projects along its waterfronts, where between the late 1980s and 2021, it reclaimed an area of 35km² (Subraelu et al., 2022). Driven by financial strategies and technological development, these projects cast urban waterfront developments as the new frontier, attracting major real-estate ventures and remodelling civic coastal spaces into lucrative, desired and privatised assets (Kyriazis et al., 2022). Across waterfront regions such as Saadiyat, Reem, Yas, Maryah, and Hudariyat islands, land modifications created consequential extensive damage to mangrove-protected areas, removing existing ecosystems to make way for urban growth (Kyriazis et al., 2022).

2.2. URBAN FORMATION OF ABU DHABI

Situated northeast of the mainland, Al Reem Island exemplifies the spatial consequences of Abu Dhabi's method of accelerated urban development. A landscape once composed of sabkha sediments, salt marshes and mangrove forests is now occupied by isolated clusters of high-rise towers surrounded by leftover, underdeveloped plots. Reclamation began before the establishment of the Environment Agency (EAD) in 1996 and continued until its development-ready stage in 2007, where the assigned developers established separate portions of the island (Dar Al-Dhabi Holding Co., 2016). The Island's purpose was to accommodate roughly 180,000 inhabitants (Dar Al-Dhabi Holding Co., 2016) and serve as one of the three investment-free zones available for foreigners and expatriates on a leasehold basis (Louvre Real Estate, 2026). The island has been marketed as an investment hub, described as a natural island environment offering extensive waterfront views and a large collection of high-rise residential towers. This branding strategy reflects the island's intended purpose, with zoning alterations made at the very beginning to reduce the amount of commercial property, favouring residential development. However, due to the Global Financial Crisis at the time, Al Reem's future was uncertain, and commitments to large-scale commercial masterplans fell short (Barnard, 2016), ultimately resulting in community clusters that are fragmented and isolated.

2.3. SHAMS ISLAND AND THE ERASURE OF MANGROVES

In the northern region of Al Reem Island lies Shams Island, occupying an area of approximately 1 km by 2 km. Shams operates under the same development framework as the rest of Al Reem Island. Ecological considerations were largely disregarded, with most of the mangrove ecosystem removed and replaced with paved hardscape (Gomes

et al., 2023). This resulted in an urban configuration that is poorly suited to its climatic context, as the terrain does not accommodate excess flooding. More importantly however, it exemplifies a strong disconnect between the constructed urban fabric and the ecological system on which it was built.

The disregard of ecological systems fundamentally contradicts the official Plan Abu Dhabi 2030: Urban Structure Framework, which, in its overarching principles, states that "Abu Dhabi will respect, be scaled to, and shaped by the natural environment of sensitive coastal and desert ecologies" (Abu Dhabi Urban Planning Council, 2007). Central to this are Mangroves, which play a vital role in shaping the City's history, as well as being one of the few ecosystems that thrive in the region's extreme climatic conditions, supporting flora and fauna that can only survive in the conditions created by these systems (Paleologos et al., 2019). By 2050, within Abu Dhabi, more than 50% of mangroves are at risk of collapse due to human-driven pressures (Environment Agency - Abu Dhabi et al., 2024). Land reclamation and river modifications through the construction of artificial channels, dams, and levees disrupt sediment erosion and wave balance, causing flooding in deltas and ultimately affecting mangrove regeneration (Paleologos et al., 2019). These damages undermine the role mangroves play in providing coastal stability and climate resilience, and highlight the need for greater mangrove protection to combat climate change (Environment Agency - Abu Dhabi et al., 2024).

During the period of expansive land reclamation in Abu Dhabi, a general mangrove compensation policy was enforced, requiring the afforestation or reforestation of two trees for every mangrove tree destroyed (International Union for Conservation of Nature (IUCN), 2016), with many of these reforestation efforts concentrated in other areas separate from where the deforestation is occurring. Despite this, soil health, proper water flows and integration with other coastal ecosystems remain more effective than simply mass planting seeds (Environment Agency - Abu Dhabi et al., 2024).

2.4. CO-HABITATION WITH MANGROVES AND BUILDING CONSTRUCTION

Co-habitation with mangrove ecologies constitutes the implementation of non-conventional building construction. Stilt construction is heavily favoured as it can potentially reduce the environmental impact of the mangrove area, and its construction method has been used for long-term sustainability (Zainal & Raja Abdul Kadir, 2022). This is evident in communities in Sabah, Malaysia, where timber-stilt construction is utilised for flood resistance and functionality to preserve shrimp farming (Zainal & Raja Abdul Kadir, 2022). Alongside this, in Shan, Myanmar, coastal communities built bamboo stilt dwellings after experiencing severe floods. Thereafter, conclusions were made that the most suitable construction for this area was a blend of indigenous knowledge and new technologies (Zainal & Raja Abdul Kadir, 2022). Co-habitation with mangroves encompasses ongoing care and stewardship through the restoration of depleted mangroves. In the north coast of Central Java, Indonesia, near the city of Semarang, locals gather wood and bamboo poles to form a permeable barrier to catch sediment that leaves a nutrient base for mangrove growth (Didde & Sujana, 2017). This restoration project achieves this balance of indigenous knowledge and new technology, as the simple technique is integrated with thorough habitat research (Didde & Sujana,

2017). Although situated within another regional context, these examples provide valuable insights into how alternative ways of building and restoration of mangroves could inform models of cohabitation within the Abu Dhabi urban landscape.

3. Methodology

The research uses a multi-objective evolutionary algorithm (MOEA) to explore an alternative urban future for the Shams island in Abu Dhabi. Rather than treating the current site condition as fixed, the study reconstructs the site based on its pre-development mangrove condition and uses this ecological state as the baseline for design. This approach allows the proposal to examine how urban development could have been integrated within the mangrove system instead of displacing it. Serving as a model for future urban development planned for mangrove sites that are still in existence today.

The MOEA has been selected as a model due to the inherent complexity of the design problem, one that comprises multiple conflicting objectives, including urban density, environmental preservation, accessibility, and spatial performance; allowing the different objectives to be addressed without necessitating trade-off decisions between the design goals. The research employs the NSGA-II algorithm (Deb et al., 2000) using the Wallacei plugin (Makki et al., 2018) within the Grasshopper 3D environment. The method follows three steps: formulating the design problem, running the MOEA, evaluating and selecting solutions from the evolved Pareto front.

4. Experiment Setup

The proposed model is situated in a pre-development scenario with a counterfactual reconstruction of the Shams Al-Reem Island site as an unaltered 1960s mangrove field. The site is partitioned into two: dense mangrove belts and tidal channels established as minimal-build zones, and sparser, low tree areas defined as build zones, which ecologically determine the extent and intensity of the urban configurations, while prioritising the preservation of the mangrove and its trees. The phenotype is evolved by the algorithm through a sequential construction process starting from this hypothetical baseline (Figure 1.A). Build zones are further tiered, with plots less than 10,000m² supporting Regional civic infrastructure, while larger plots house Local clusters, hence concentrating activity hierarchically (Figure 1.B). Each Local plot defines a productive anchor through the placement of the only ground-contact elements, the Dock, at the closest water edge, paired with an adjacent raised Aquaculture Tower set within a tight catchment to ensure there is operational viability. A Communal Courtyard (CCY) emerges from the interior of each Local plot, with its placement selected from within a bounded central region of the buildable surface. Residential Buildings are plotted and distributed within the area between the CCY and plot boundary and are oriented inwardly, facing the communal core. Overlapping footprints are reclassified as shared spaces, whereas those structures intersecting with the CCY are converted into Communal Buildings. Residential roof plates are treated as Agricultural Terraces, allowing for the extension of productive capacity within the housing fabric without increasing disturbance to the ground. An elevated circulation network, inspired by sikkas, an urban characteristic that is unique to the region, weaves

the dwellings together, converting the attached private levels into spaces of communality while connecting to the CCY and creating a continuous pedestrian system (Figure 1.C). At the scale of each mangrove field, CCYs across Local clusters are linked through proximity-based parameters. Centralised Bazaar platforms with rings of Civic buildings positioned within the Regional plot, further connect to the midpoints of these links, forming an interconnected region-wide network (Figure 1.D). Through bridging the Bazaars, another higher level of connectivity is produced, which creates potential sites for Global platforms, constrained by rules that prioritise the productive elements while trivialising residential structures (Figure 1.E). All built form is elevated on stilts to factor in ground continuity and tidal flooding, while preserving mangrove regeneration. The evolutionary algorithm (EA) selects parameters for size, distance, height, rise, distribution and positioning from bounded ranges, forming a simplified and abstract output geometry to support fast, iterative evaluations of connectivity, density and ecological impact.

The EA optimises the phenotype in response to 5 fitness objectives, developed with the primary purpose of producing urban conditions that demonstrate the possibility of cohabitation within the landscape. The objectives are as follows: (FO1) maximise dwelling capacity;(FO2) maximise solar exposure to mangrove tree canopies; (FO3) maximise ecologically productive area; (FO4) minimise mangrove clearance and tree disturbance; and (FO5) maximise bridge network efficiency, concentrating movement into key bridges (high betweenness) with minimal bridge infrastructure. Collectively, these objectives aid in the establishment of an urban tissue that limits the cumulative impact of the development on the mangrove's sensitive system.

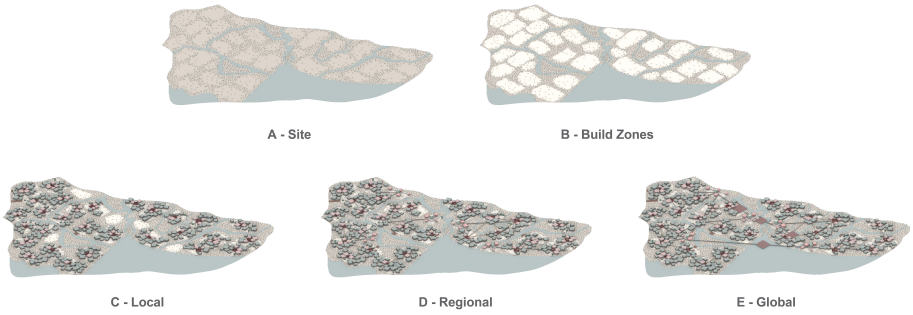


Figure 1A-E. Phenotype Construction

5. Experiment Results / Analysis

The algorithm evolved 5000 solutions (100 generations with 50 solutions per generation) with a simulation runtime of 6 hours and 3 minutes. 889 Pareto front solutions were output by the algorithm. The results demonstrate that there is conflict in the urban and ecological objectives, where FO1 (residential density), FO3 (productive structure area), and FO5 (bridge connections) depict improvement in fitness over the algorithmic run, indicating that the algorithm was more effective at refining urban variables which were in direct control of the phenotype. In contrast, due to this inverse

relationship, there is an inherent trade off where the mangrove-centred objectives, FO2 (solar access on trees) and FO4 (tree destruction) do not depict comparable improvement, implying that these targets were inevitably compromised as built form, so density, productive area and connectivity were favoured.

5.1. SELECTION PROCESS

From the 889 Pareto front solutions, K-means clustering was applied to extract 20 representative outcomes, which helped in preserving variation while aiding in reducing redundancy. Selection then moved through three filtration stages, structured as a progressive elimination framework. Stage one applied the maximum mangrove tree destruction threshold of 20%, where non-compliant solutions were excluded to ensure minimal degradation of ecology. This resulted in a reduced candidate set that met the project's initial environmental criteria.

The second filtration stage introduced a comparative assessment through the employment of a weighted scoring matrix including phenotypic indicator measures as well as fitness objectives (Figure 2). The phenotypic indicators were defined to quantify the impact of structure on each solution in the mangrove field. These included the number of tidal channel crossings and stilts, as they assess interference with the mangrove ecology and its hydrological processes; the residential footprint efficiency, which favoured vertical expansion rather than horizontal sprawl. Additional indicators were quantifying the distribution of different program structures with variable counts, including the total number of Residential and Communal buildings as well as Global platforms. Connectivity was further measured through total bridge lengths alongside average travel distance between Residential buildings and Aquaculture towers, Local clusters to Bazaar and the Bazaar to Global platforms. Environmentally focused criteria, including minimisation of tree disturbance and ground interference alongside maximising solar access, were prioritised through higher weighting in the matrix, ensuring that the top five solutions selected for subsequent testing were ecologically high performing.

PHENOTYPE INDICATORS	WEIGHTING	91|45	91|46	95|40	51|47	70|21	71|48	87|12	27|18	30|38	32|21	03|30	8|36	14|49	20|34
Number of tidal channel crossings	1.00	1.00	0.57	0.14	0.71	0.29	0.57	0.14	0.29	0.86	0.43	0.86	0.71	0.00	0.29
Number of stilts (ground-contact points / piles)	1.00	0.24	0.49	0.96	0.70	0.25	1.00	0.18	0.03	0.25	0.71	0.46	0.95	0.00	0.98
Number of residential buildings	0.50	0.15	0.28	0.49	0.30	0.20	0.50	0.14	0.10	0.20	0.38	0.28	0.37	0.00	0.43
Residential footprint efficiency	0.75	0.46	0.46	0.22	0.56	0.74	0.37	0.17	0.35	0.60	0.43	0.33	0.00	0.41	1.00
Number of communal buildings	0.75	0.35	0.57	0.26	0.12	0.72	0.05	0.54	0.75	0.72	0.30	0.59	0.00	0.33	0.12
Average travel distance between housing and aquaculture	0.25	0.00	0.03	0.25	0.01	0.05	0.25	0.10	0.18	0.20	0.09	0.11	0.24	0.01	0.05
Average distance from residential clusters to nearest bazaar	0.25	0.20	0.14	0.15	0.22	0.11	0.22	0.25	0.00	0.12	0.18	0.17	0.22	0.22	0.16
Average distance from bazaar to global platform	0.25	0.12	0.25	0.02	0.24	0.14	0.03	0.00	0.04	0.17	0.14	0.25	0.14	0.05	0.06
Total bridge length	1.00	0.11	0.87	0.90	0.56	0.80	1.00	0.25	0.37	0.81	0.95	0.91	0.61	0.00	0.87
Number of global platforms (global access points)	0.25	0.08	0.08	0.08	0.17	0.17	0.17	0.17	0.17	0.00	0.08	0.17	0.17	0.17	0.25
FO1 - MAXIMISE DWELLING DENSITY ACROSS THE MANGROVE FIELD	2.00	1.01	1.53	0.36	0.65	2.00	0.00	1.38	1.19	1.65	1.04	1.22	0.21	1.32	0.68
FO2 - MAXIMISE SUN ON MANGROVES	2.00	0.59	0.59	1.50	1.28	0.28	1.87	0.00	0.67	0.47	1.05	1.20	2.00	0.17	1.59
FO3 - MAXIMISE TOTAL ECOLOGICALLY ACTIVE AND PRODUCTIVE AREA	2.00	1.06	1.55	0.34	0.68	2.00	0.00	1.47	1.27	1.61	1.03	1.28	0.30	1.39	0.73
FO4 - MINIMISE MANGROVE CLEARANCE AND TREE DISTURBANCE	3.00	0.42	0.56	2.15	1.73	0.00	3.00	0.35	1.45	0.63	1.25	0.25	2.13	0.26	2.07
FO5 - MAXIMISE CONNECTIVITY EFFICIENCY	1.00	0.61	0.28	0.34	0.36	0.00	0.23	1.00	0.07	0.11	0.28	0.07	0.18	0.46	0.16
TOTAL	16.00	6.41	8.25	8.17	8.29	7.74	9.25	6.14	6.92	8.40	8.33	8.14	8.22	4.79	9.45
RANKS					5		2			3	4				1

Figure 2. Second filtration step for selection

The third and final filtration stage reassessed the five shortlisted solutions through a refined ranking matrix (Figure 3). Average building height and total built footprint within allocated minimal-build zones were evaluated as they impact overshadowing

on mangroves and ecological encroachment. Overlaps converted to voids or shared spaces were quantified due to their influence on solar access to mangroves, and the opportunities it provides for generating micro-communal areas and, as such, an integrated neighbourhood. The matrix was further tested for solar access on productive areas, as it impacts the viability of the agro-aquacultural systems. Furthermore, closeness centrality of the connectivity network is used to test for and pick the most efficient circulation, hence favouring the option with the least structural footprint and most insignificant environmental impact. For each of these ecologically oriented metrics, solutions were ranked on a 1-5 scale, where the most desirable was indicated by 5 and the least, 1. The solution with the highest sum was selected.

PHENOTYPE INDICATORS	20|34	71|48	30|38	32|21	51|47
Minimal average overall height of building	2	3	1	4	5
Minimal total built fooprint on allocated mangrove no-build zones	4	5	1	2	3
Maximal amount of building overlap between the residentials and residential-communal	2	1	5	4	3
Maximal footprint of building overlap between residential and residential-communal	2	1	5	4	3
Optimised sun on active and ecologically productive areas	4	5	1	2	3
Optimised closeness centrality of connectivity network	3	5	1	2	4
TOTAL	17.00	20.00	14.00	18.00	21.00
RANKS	4	2	5	3	1

Figure 3. Third filtration step for selection

5.2. SELECTED PHENOTYPE

Following the filtration process, Generation 51, Individual 47 was selected, which was further developed to translate the algorithmically generated phenotype into an urban proposal that showcases the original intent of the research. This was conducted on two clusters from the selected solution, the local residential cluster and the regional civic zone, which depicts how spatial systems in this city operate across scales (Figure 4).



Figure 4. Selected Phenotype with Focal Points

The local clustering focal point examines everyday inhabitation, exploring residential density, communal life and how they coexist alongside mangrove systems. This focuses on the intersection between housing, communal structures, sikka-inspired bridges and roofscapes. Elevated, inward-facing dwellings wrap around the communal core above the tidal ground layer. Overlaps caused by intersecting residential and communal buildings produce either micro-commons or voids, which allow for light and air while also reducing the disturbance to the mangrove ecosystem below. These act as extensions of the dedicated communal spaces, including the CCY, Communal buildings, and Agriculture terraces, distributing collective use vertically instead of

concentrating it on the ground. Through the placement of cultivation onto roofs, productive capacity is increased without further disrupting ecological continuity.

The regional zone centres on how productive infrastructure and mangrove regeneration interact at a broader communal scale. The focus is organised around the bazaar housing the mangrove nursery wind tower, and adjacent water zones which permit new mangrove growth. The lifted Bazaar operates as the regional exchange hub where it consolidates human activity while leaving the ecological layer that exists below intact. The mangrove nursery wind tower and the woven sediment-trapping bamboo fencing in the surrounding water zones allow for gradual mangrove expansion and hence, future opportunities for cohabitation. Together, the two focal points reveal a plausible symbiotic relationship between urban life and mangrove ecologies.

6. Discussions and Conclusions

This paper presents an alternative trajectory for urbanisation within ecologies, through its selected phenotype, challenging standardised, contemporary development models built through acts of reclamation, erasure and stabilisation. In place of the imposing, gridded, globalised cities, an inhabitation that co-exists with tidal dynamics and existing place-specific ecologies is proposed. Through grounding the site, Shams in Al Reem Island, Abu Dhabi, in its unbuilt 1960s condition of mangrove fields, the study reframes the degradation and displacement of ecologies as a design decision rather than an inevitable consequence (Figure 5).



Figure 5. Current Site vs Proposed Design

Generated by a multi-objective evolutionary framework, the derived, representative morphology negotiates and reconciles trade-offs between mangrove preservation, dwelling density, agro-aquaculture intensity and connective efficiency. The decentralised, layered urban form consisting of clustered residential formations, distributed civic commons and elevated circulation networks demonstrates how alternative priorities when planning can support inhabitation that preserves the ecological processes and continuity of the mangrove forests. In doing so, the paper opposes environmentally exploitative urbanisation, constructing a transferable framework for future research in urban systems for other vulnerable ecologies, such as wetlands and floodplains.

There are certain limitations to the research that can be developed in future research; the model relies on the simplification and abstraction of both urban structures and the mangrove ecosystem for purposes of computation, which enables optimisation but does not fully incorporate the complexities of the ecology. Temporality is also not accounted for, as long-term ecological changes are not simulated. In the instance of

population increase, the mangrove landscapes have limited tolerance for structural occupation before detrimental degradation, so additional growth would need to be accommodated elsewhere. Thus, the study reframes how urban futures should be evaluated, not by their capacity to grow and expand but instead through their ability to coexist with ecological bounds.

References

Abu Dhabi Urban Planning Council. (2007). Plan Abu Dhabi 2030—Urban Structure Framework Plan.

Bani Hashim, A. R. (2018). Planning Abu Dhabi: An Urban History (1st ed.). Routledge.

Barnard, L. (2016, April 12). Abu Dhabi's urban planners working on ways to restart stalled Reem Island projects. The National.

Dar Al-Dhabi Holding Co. (2016). Najmat Abu Dhabi Project.

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. International Conference on Parallel Problem Solving From Nature, 849–858.

Didde, R., & Sujana, N. (2017). Building with Nature on Java: Weaving dams to defend mangroves. COASTAL CONSERVATION.

Elsheshtawy, Y. (2019). Temporary Cities: Resisting Transience in Arabia (1st ed.). Routledge. https://doi.org/10.4324/9780429457838

Environment Agency - Abu Dhabi, Ministry of Climate Change, & Environment and Emirate Nature - WWF. (2024). Mangrove Restoration Guidelines for the United Arab Emirates.

Gomes, A., Kyriazis, A., Montagne, C., Zhand, S., & Schwinger, P. (2023). Roads (dis)connecting cities and neighbourhoods: A socio-spatial study of Abu Dhabi.

International Union for Conservation of Nature (IUCN), A. D. G. E. D. I. (AGEDI). (2016). National Blue Carbon Policy Assessment. UAE: Summary (p. 9) [Policy Assessment Summary].

Kyriazis, A., Almehairi, A., & Seri, I. (2017, June 5). The evolution of Abu Dhabi's Superblocks: From modernist pragmatism to regeneration efforts.

Kyriazis, A., Enab, D. A. N., & Nuaimi, R. T. A. (2022). Comparing Waterfront Land-use Dynamics in Abu Dhabi and Dubai: A Quantitative Analysis and Correlation with the Developing Urban Environment. In Handbook of Waterfront Cities and Urbanism. Routledge.

Louvre Real Estate (2026). Al Reem Island Comprehensive Area Guide. Retrieved January 16, 2026.

Makki, M., Showkatbakhsh, M., & Song, Y. (2018). Wallacei: An evolutionary and Analytic Engine for Grasshopper 3D. Wallacei. https://www.wallacei.com/

Paleologos, E. K., Welling, B. A., Amrousi, M. E., & Masalmeh, H. A. (2019). Coastal development and mangroves in Abu Dhabi, UAE. IOP Conference Series. Earth and Environmental Science, 344(1).

Sheikh Zayed bin Sultan Al Nahyan (1998). To Our Roots Exhibition 2025, Zayed National Museum, AbuDhabi, UAE.

Showkatbakhsh, M., & Makki, M. (2022). Multi-Objective Optimisation of Urban Form: A Framework for Selecting the Optimal Solution. Buildings, 12(9), Article 9.

Subraelu, P., Ebraheem, A. A., Sherif, M., Sefelnasr, A., Yagoub, M. M., & Rao, K. N. (2022). Land in Water: The Study of Land Reclamation and Artificial Islands Formation in the UAE Coastal Zone: A Remote Sensing and GIS Perspective. Land, 11(11), 2024.

Zainal, S., & Raja Abdul Kadir, T. A. Q. (2022). Sustainable On-Stilt Construction Technology for Mangrove Land in Malaysia. 4, 43–50.

Adaptive Environmental Control

Reclaiming Agency in Building Performance Prediction

Enhancing Occupant-Centered Simulation Accuracy through Data Assimilation and Explainable Machine Learning

Yuxin Zhou[1], Yudai Tanino[2], Tomohiro Fukuda[3], and Minoru Etoh[4]
[1,3]*Division of Sustainable Energy and Environmental Engineering, The University of Osaka.*
[2,4]*Institute for Open and Transdisciplinary Research Initiatives, The University of Osaka.*
[1]*yuxin@it.see.eng.osaka-u.ac.jp, 0000-0002-7201-4626*
[2]*tanino.yudai@tsu-x.com, 0009-0003-9543-1486*
[3]*fukuda.tomohiro.see.eng@osaka-u.ac.jp, 0000-0002-4271-4445*
[4]*etoh@ieee.org, 0009-0002-9868-0464*

Abstract. Under global climate change and rapid urbanization, improving the accuracy of building performance simulation (BPS) is essential for low-carbon built environments. However, occupant-related inputs are often uncertain and prediction results can be hard to interpret, limiting their actionable value. This paper proposes an occupant-centered framework that combines data assimilation with explainable machine learning to improve both simulation fidelity and interpretability. A university dining hall was selected as the case study and instrumented with a multi-sensor system to collect year-long indoor environmental data. With the data assimilation workflow, relative errors in daylight and energy simulation results decreased by 2.99% and 38.29%, and predicted mean vote shifted from −0.24 to 0.24. Leveraging these assimilated datasets, a LightGBM predictor achieved strong performance ($R^2 = 0.919$), with SHapley Additive exPlanations analysis identifying the key determinants: daylight was driven by instantaneous weather, energy by thermal and occupancy factors, and comfort by temperature and time. Overall, the framework improves BPS and prediction stages while lowering data-use barriers, supporting more adaptive, occupant-centered building simulation practices.

Keywords. data assimilation, LightGBM, building performance simulation, explainable machine learning, energy efficiency

1. Introduction

Under the combined pressures of climate-induced extreme weather and rapid urbanization, the construction industry stands at a critical point in its low-carbon transition. In this context, the industry is increasingly adopting performance-driven

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 491-500.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

approaches, making accurate meteorological data essential for building performance simulation (BPS) (Kočí et al., 2019). Typical meteorological year (TMY) data is widely used in BPS. However, as global warming enlarges climate variability, discrepancies between TMY data and actual weather are widening (Crawley & Lawrie, 2019), challenging the applicability of BPS methodologies in complex environments.

Most EnergyPlus weather (EPW) files are derived from airports or regional weather stations, which often fail to reflect building-level microclimates (Yuan et al., 2025). Recent studies have explored data assimilation (DA) using real-time monitoring data to reconstruct EPW file, reducing gaps between actual and TMY datasets (Figueiredo et al., 2018). However, most work focuses on improving energy use predictions, with limited integration of multidimensional aspects such as daylight and comfort (Yüceses & Selcuk, 2023). Crucially, the accurate calibration of thermal comfort and occupant sensation remains largely unaddressed (Yuan et al., 2024). Uncalibrated weather data distorts indoor predictions, creating a misalignment between simulated metrics and actual occupant sensation. This discrepancy compromises the reliability of human-centric BPS.

Meanwhile, machine learning (ML) approach has shown strong predictive accuracy in building performance analysis (Rahmani Asl et al., 2017). Yet most ML models operate as "black boxes," offering limited insight into how input variables shape predictions (Zhou et al., 2026). This lack of explainability reduces their value for practice applications. Explainable artificial intelligence (XAI) methods have begun to address this gap in energy prediction and environmental modelling (Machlev et al., 2022), but their use across other performance dimensions remains limited.

To address these challenges, this study assimilates indoor monitoring data into BPS process to improve the prediction of daylight, energy, and comfort. The approach narrows the gap between TMY and real-world weather conditions. It enhances BPS accuracy and provides explainable evidence to support building design and operation. This approach can enable smart, low-carbon, and occupant-centered built environment.

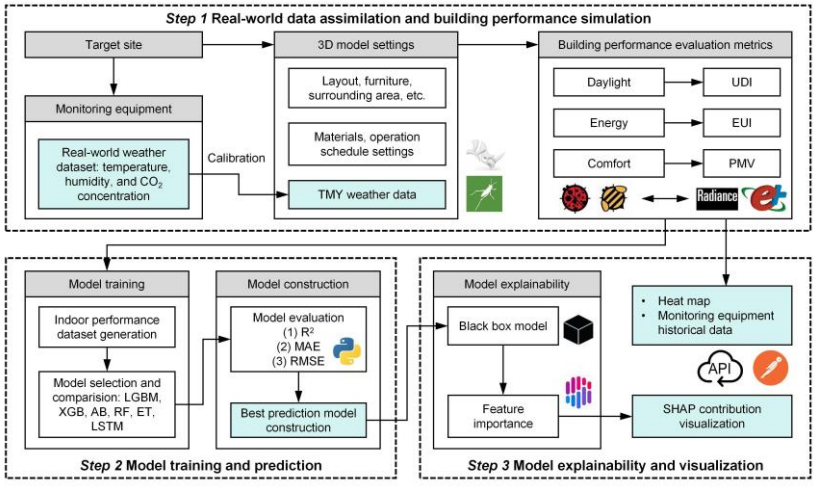


Figure 1. General research workflow.

2. Methodology

This study presents an occupant-centered framework integrating DA with XAI (Fig. 1). Real-time sensor data (temperature, humidity, CO_2 concentration) were used to dynamically calibrate the TMY dataset. Building on this foundation, predictive models for daylight, energy, and comfort were developed, utilizing SHapley Additive exPlanations (SHAP) analysis to interpret key variable contributions.

2.1. DATA ASSIMILATION FOR SIMULATION

2.1.1. Principles of data assimilation

DA is a computational technique that combines observational data with numerical models to optimize system state estimation and enhance prediction accuracy (Bocquet et al., 2015). This method constructs a precise representation of the time-evolving state by integrating diverse information sources including sensor measurements and simulation outputs.

In the context of the built environment, DA bridges the BPS gap to mitigate simulation discrepancies arising from the reliance on static TMY data. Implementing such a framework requires a robust data acquisition infrastructure to capture spatial and temporal environmental variations. Typically, this involves deploying multi-sensor networks to monitor indoor conditions over extended periods, accounting for seasonal dynamics. In modern workflows, application programming interfaces (APIs) are utilized for real-time telemetry: raw sensor data are often processed locally, structured into lightweight formats like JSON, and transmitted to central databases via HTTP request-response mechanisms (Kim et al., 2022). This architecture facilitates both real-time tracking and historical batch retrieval, establishing a high-fidelity data foundation for the assimilation process.

2.1.2. Metrics for indoor performance evaluation

An indoor performance evaluation framework was established based on three dimensions of building performance: daylight, energy, and comfort. (1) Useful daylight illuminance (UDI) represented the percentage of time that work-plane illuminance (0.75 m height) remained within 300-2000 lx, capturing the adequacy of natural light. (2) Energy use intensity (EUI) quantified annual energy consumption per unit area ($kWh/m^2 \cdot year$), reflecting operational efficiency and alignment with carbon reduction goals. (3) Comfort was assessed using the predicted mean vote (PMV) index on a −3 to +3 (cold to hot) scale, considering temperature, humidity, metabolic rate, etc. Together, these indicators provided an integrated view of visual, energy performance, and thermal improvements achieved through DA workflow.

2.2. MODEL TRAINING AND EVALUATION

To capture complex, non-linear interactions, this study employs the Light Gradient Boosting Machine (LGBM). As an efficient implementation of gradient boosting decision trees, LGBM constructs a robust ensemble model by sequentially combining weak learners to correct previous errors, offering superior speed and scalability over

traditional regression (Ke et al., 2017). The objective function is defined as:

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t)$$

Where l denotes the loss function, y_i is the true value, $\hat{y}_i^{(t-1)}$ is the prediction from the $(t-1)$ -th iteration, f_t is the newly added regression tree, and $\Omega(f_t)$ is the regularization term.

To ensure generalization, the dataset was split into 80% training and 20% testing. 5-fold cross-validation and Bayesian optimization were applied to systematically tune key hyperparameters (e.g., learning rate, tree depth, regularization λ, γ) for stability. This approach enhanced prediction accuracy and model stability while maintaining computational efficiency. Performance was assessed using coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE).

2.3. EXPLAINABLE MACHINE LEARNING

ML models frequently operate as "black boxes," obscuring the physical relationships between design variables and performance outcomes. XAI aims to bridge this transparency gap by deconstructing the internal logic behind algorithmic predictions. XAI serves as a vital tool for translating quantitative predictions into practical design insights, allowing stakeholders to understand why a specific performance outcome is predicted rather than just what is predicted.

To achieve this explainability, this study employs the SHAP method (Lundberg & Lee, 2017). Rooted in cooperative game theory, SHAP decomposes complex model predictions into the marginal contributions of individual features. This approach reveals both the direction and magnitude of each variable's influence on daylight, energy, and comfort. By quantifying these impacts, the framework supports reliable, evidence-based decision-making in the design process.

3. Case Study and Results

3.1. CASE STUDY

To validate the proposed framework, a building on the Minoh Campus, The University of Osaka, Japan, was selected as the case study (Fig. 2). Conventional meteorological data from airports or city weather stations are distant from the site and fail to capture its microclimate. Therefore, site-specific monitoring and DA are required to ensure accurate local representation.

A dining hall on the third floor was chosen as the measurement area. This north-facing space, enclosed on three sides by floor-to-ceiling glazing, serves as a high-occupancy zone with strong environmental fluctuations. 22 temperature-humidity and 7 CO_2 concentration sensors were installed across corridors and dining areas to capture spatial variation (Fig. 3). Each sensor had unique coordinates and identifiers for efficient management. Data were recorded every three minutes throughout 2024 and uploaded in real time to a cloud database. The Postman interface enabled live monitoring and anomaly detection, while historical data were retrieved via API for batch analysis. Missing data were interpolated using spatial neighboring averages, and

minute-level data were aggregated into hourly means to match EPW data resolution. Finally, hourly averages across all sensors were used to represent the room's overall conditions, yielding a robust 8,760-hour dataset for analysis.

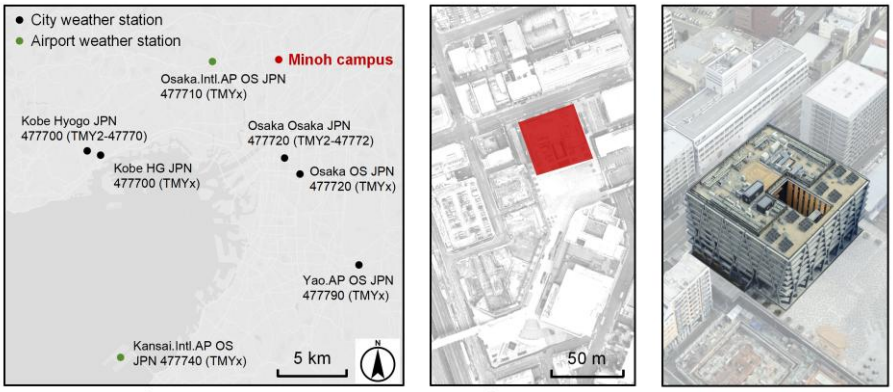


Figure 2. EPW data sources in Osaka and location of the target building.

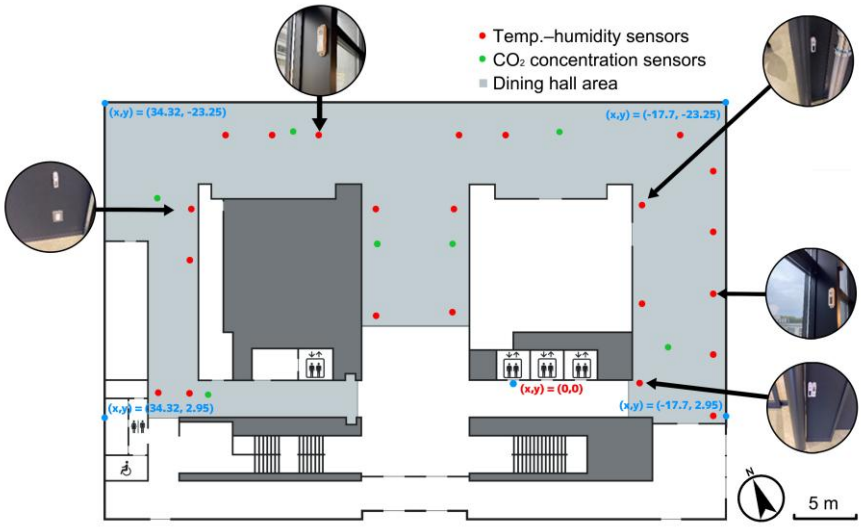


Figure 3. Sensor deployment layout within the dining hall.

3.2. RESULTS

3.2.1. Results of data assimilation and simulation calibration

The processed dataset was assimilated into the EPW file and compared with the TMY dataset. Temperature errors yielded RMSE and MAE values of 7.5 and 6.0, while humidity errors were 21.5 and 17.6, respectively. Table 1 and Fig. 4 summarize the BPS results using the calibrated data (Assimilated Data) versus Original Data.

Daylight differences were minimal (RE = 2.99%). In contrast, energy use showed

a substantial reduction (RE = 38.29%). Notably, heating demand dropped by nearly 90% (20.10 to 2.20 kWh/m²). This sharp decrease is attributed to the divergence between the historical TMY data and the actual warmer winter conditions measured on-site, which naturally reduced the need for active heating. Meanwhile, cooling loads remained stable. Comfort metrics reflected this thermal shift: the annual PMV increased from −0.24 to 0.24, confirming that the original data underestimated indoor warmth. Overall, DA provided a more accurate representation of the building's microclimate and improved BPS reliability.

Table 1. Comparison of simulated results before and after data assimilation.

Evaluation metrics	Assimilated Data	Original Data
UDI (Daylight evaluation, unit: %)	53.50	55.10
Total EUI (Energy evaluation, unit: kWh/m²)	61.28	84.74
Heating EUI	2.20	20.10
Cooling EUI	30.84	32.67
Lighting EUI	4.82	8.54
PMV (Comfort evaluation, unit: N/A)	0.24	-0.24

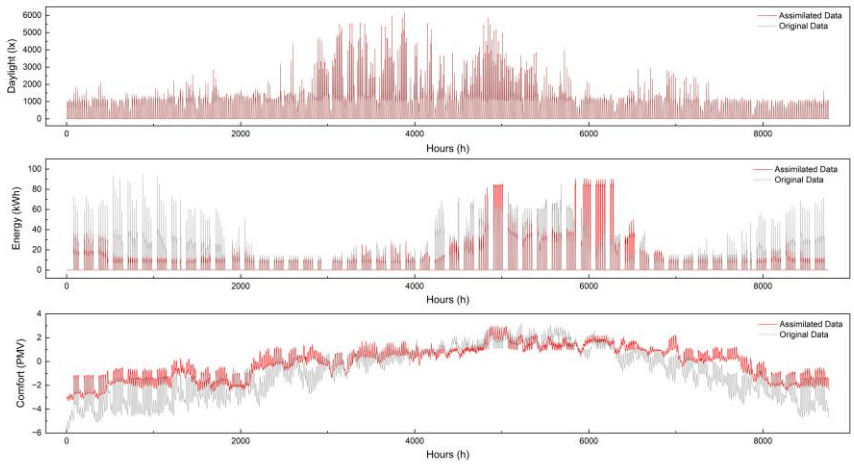


Figure 4. Comparison of hourly results before and after data assimilation workflow.

3.2.2. Results of model training

Fig. 5 and Table 2 illustrate model's predictive performance. Six ML models were compared: LGBM, XGBoost (XGB), AdaBoost (AB), Random Forest (RF), Extra Trees (ET), and Long Short-Term Memory (LSTM). LGBM performed particularly well in energy and comfort tasks, with test set R^2 values of 0.942 and 0.988, respectively, and the lowest RMSE and MAE among all models. Especially for comfort, predicted values nearly overlapped (R^2 = 0.988, RMSE = 0.248, MAE = 0.179), demonstrating excellent robustness. Daylight predictions showed slightly

higher errors due, yet LGBM still outperformed others ($R^2 = 0.827$). The average R^2 across all targets reached 0.919, confirming LGBM's ability to capture complex nonlinear relationships while maintaining high accuracy and generalization.

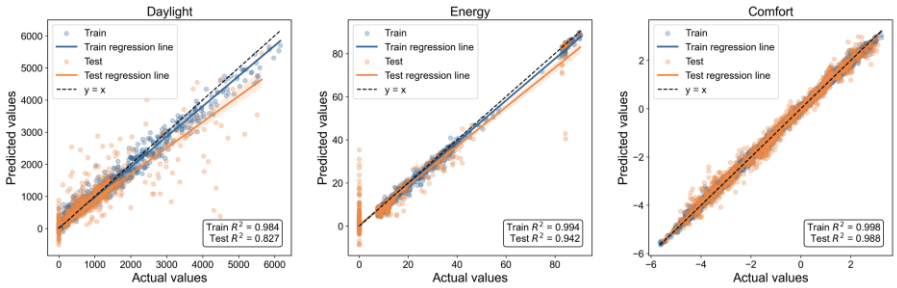


Figure 5. Predictive performance of the LGBM model.

Table 2. Performance results of different predictive models.

Models	Training set			Test set		
	RMSE ↓	R^2 ↑	MAE ↓	RMSE ↓	R^2 ↑	MAE ↓
Daylight performance training results						
LGBM	92.347	0.984	49.822	315.332	0.827	130.527
XGB	52.013	0.995	27.924	326.263	0.814	132.626
AB	396.371	0.709	258.621	503.234	0.558	302.369
RF	286.842	0.848	125.94	384.172	0.742	164.558
ET	273.134	0.862	125.584	380.929	0.747	165.272
LSTM	501.905	0.597	162.352	364.559	0.806	134.382
Energy performance training results						
LGBM	1.019	0.994	0.677	3.213	0.942	1.410
XGB	1.284	0.991	0.836	3.578	0.927	1.822
AB	5.633	0.821	2.857	5.737	0.814	2.839
RF	3.843	0.917	1.773	4.672	0.876	2.272
ET	1.773	0.982	0.929	4.241	0.898	1.938
LSTM	10.261	0.405	5.842	10.406	0.687	5.950
PMV performance training results						
LGBM	0.142	0.998	0.106	0.248	0.988	0.179
XGB	0.209	0.989	0.159	0.291	0.979	0.218
AB	0.315	0.974	0.247	0.366	0.966	0.293
RF	0.32	0.973	0.226	0.402	0.959	0.288
ET	0.172	0.992	0.124	0.298	0.978	0.214
LSTM	0.498	0.935	0.368	0.553	0.923	0.414

3.2.3. Results of model explainability

SHAP analysis clarified the distinct contribution patterns of input variables for each performance index (Fig. 6):

(1) Daylight. Temperature (30.9%) and humidity (27.3%) emerged as the dominant features, together accounting for 58.2% of the model's contribution. Crucially, this high attribution reflects a statistical correlation where these thermodynamic variables act as proxies for solar radiation and sky conditions, rather than functioning as direct physical drivers of illuminance. Specifically, high temperature correlated with positive SHAP values (reflecting clear, sunny conditions), while high humidity showed negative contributions (consistent with overcast or rainy weather).

(2) Energy. Energy use was governed by temperature, hour, and humidity (71.9% total contribution). Temperature exerted the largest impact (26.8%), physically corresponding to increased cooling loads. Hour (22.7%) captured daily operational patterns and occupancy schedules. Humidity (22.4%) showed a symmetric influence, reflecting dehumidification demands.

(3) Comfort. Thermal comfort was dominated by temporal factors, with hour contributing 56.2%, highlighting its strong correlation with daily occupancy rhythms. Temperature (20.4%) ranked second, directly affecting PMV calculations. Beyond these, other features showed limited influence.

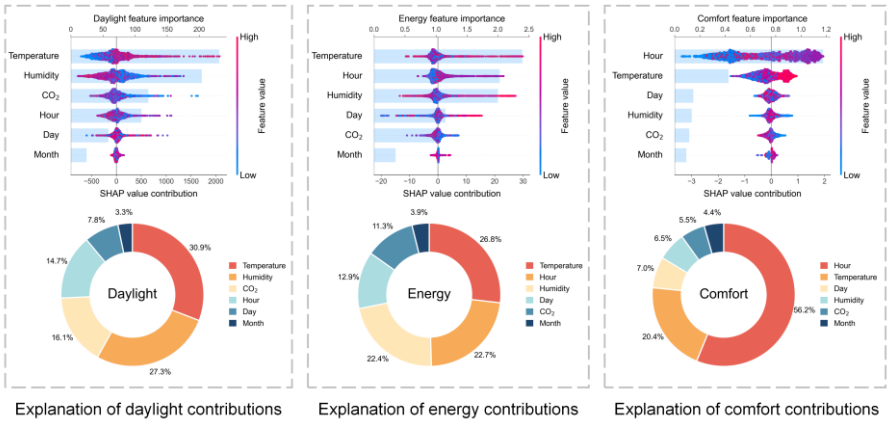


Figure 6. Feature contribution results based on SHAP values.

4. Discussion

This study validates the integration of DA and XAI to improve the BPS accuracy. A critical finding was the ~90% reduction in heating load. This significant decrease highlights the efficacy of the DA framework in correcting TMY-based underestimates of internal heat gains and microclimatic warming. LGBM achieved the best predictive performance and high stability across multiple indicators. It outperformed other models, such as LSTM. While LSTM typically excels with larger datasets, it showed no advantage here, likely due to the limited sample size and high variance in environmental factors. These findings suggest that ML models, like LGBM and XGB, are better suited for this task.

However, daylight prediction showed lower R^2 values (0.827) despite low relative error (2.99%). This reflects the physical difference between thermal inertia and the stochastic nature of daylight, which is highly sensitive to instantaneous cloud movements. Specifically, the dining hall's extensive glazing amplifies the impact of these external sky variations, resulting in high-frequency fluctuations that are inherently difficult to capture compared to spaces with lower window-to-wall ratios. Additionally, dynamic urban changes, such as the high-rise construction on the east side (2023–2024), altered real-world shading patterns that the static simulation model could not fully capture. These irregularities underscore the need for more targeted calibration. Finally, while the single-building scope limits direct numerical extrapolation, the proposed methodological framework is designed to be transferable. Future work should extend the DA workflow to diverse building typologies and incorporate a wider range of indicators, such as air velocity and solar radiation, to further refine the fidelity of occupant-centered simulations.

5. Conclusions

This study presents an occupant-centered framework that integrates data assimilation with explainable machine learning to improve simulation accuracy and predictive model transparency. In the case study, the calibrated EnergyPlus weather file markedly enhanced accuracy: daylight relative error (RE) was 2.99%, energy error decreased by 38.29%, and comfort value rose from −0.24 to 0.24 (RE = 200%). LightGBM achieved the best predictive accuracy (R^2 = 0.919), while SHapley Additive exPlanations analysis identified dominant drivers: daylight was mainly affected by instantaneous weather, energy by temperature and hour, and comfort by temperature and time. Besides, real-time data visualization through application programming interface lowered technical barriers, demonstrating practical applicability.

Overall, this framework demonstrates strong potential for enhancing simulation accuracy and calls for broader validation across climate zones and building types to support its generalization capabilities.

Acknowledgements

This research was supported by the DiCHiLD Project (DAIKIN special project for The University of Osaka) from September, 2024 to June, 2025.

Attribution statement

ChatGPT (OpenAI, 2025) was used to improve the flow and clarity of the text. The use of AI tool was limited to enhancing the writing process and does not replace the author's original work or scientific contributions.

References

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., Flemming, J., Inness, A., Pagowski, M., Pérez Camaño, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., & Seigneur, C. (2015). Data assimilation in atmospheric chemistry models: Current status and future prospects for coupled

chemistry meteorology models. *Atmospheric Chemistry and Physics*, 15(10), 5325–5358. https://doi.org/10.5194/acp-15-5325-2015

Crawley, D., & Lawrie, L. (2019). Should we be using just 'typical'weather data in building performance simulation? *Building Simulation* 2019, 16, 4801–4808. https://doi.org/10.26868/25222708.2019.210594.

Figueiredo, A., Kämpf, J., Vicente, R., Oliveira, R., & Silva, T. (2018). Comparison between monitored and simulated data using evolutionary algorithms: Reducing the performance gap in dynamic building simulation. *Journal of Building Engineering*, 17, 96–106. https://doi.org/10.1016/j.jobe.2018.02.003

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *31st Conference on Neural Information Processing Systems, NIPS 2017* (Vol. 30). The Association for Neural Information Processing Systems (NIPS).

Kim, M., Xin, Q., Sinha, S., & Orso, A. (2022). Automated test generation for REST APIs: No time to rest yet. In *proceedings of the 31st ACM SIGSOFT International Symposium on Software Testing and Analysis, ISSTA2022* (pp. 289–301). The Association for International Symposium on Software Testing and Analysis (ISSTA).

Kočí, J., Kočí, V., Maděra, J., & Černỳ, R. (2019). Effect of applied weather data sets in simulation of building energy demands: Comparison of design years with recent weather data. *Renewable and Sustainable Energy Reviews*, 100, 22–32. https://doi.org/10.1016/j.rser.2018.10.022.

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *31st Conference on Neural Information Processing Systems, NIPS 2017* (Vol. 30). The Association for Neural Information Processing Systems (NIPS).

Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. *Energy and AI*, 9, 100169.

Rahmani Asl, M., Das, S., Tsai, B., Molloy, I., & Hauck, A. (2017). Energy Model Machine (EMM)—Instant Building Energy Prediction using Machine Learning. In *proceedings of the 35th International Conference on Education and Research in Computer Aided Architectural Design in Europe, eCAADE 2017* (pp. 277–286). International Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADE).

Yuan, L., Wang, K., Xue, S., Yin, S., Zhan, Q., & Chao, X. (2024). Assessing and enhancing the applicability of standard effective temperature in outdoor thermal sensation prediction for continental climates. *Energy and Buildings*, 305, 113894.

Yuan, L., Zhou, Y., Wang, K., Wei, L., Huang, C., Zhang, T., Mo, H., Yin, S., Jian, Y., & Wang, Y. (2025). Investigation of the relationships between enclosed architectural spatial forms and indoor–outdoor thermal comfort based on explainable machine learning and optimization algorithms. *Energy and Buildings*, 116465. https://doi.org/10.1016/j.enbuild.2025.116465

Yüceses, B., & Selcuk, S. A. (2023). State-of-the-Art on Research and Applications of Machine Learning in Building Energy Performance Prediction. In *proceedings of the 11th International Conference of the Arab Society for Computation in Architecture, Art and Design, ASCAAD 2023* (pp. 275–292). The Arab Society for Computation in Architecture, Art and Design (ASCAAD).

Zhou, Y., Fukuda, T., & Yabuki, N. (2026). Automatic evaluation and analysis of indoor visual comfort for sustainable building design using interpretable ensemble learning. *Automation in Construction*, 181, 106582. https://doi.org/10.1016/j.autcon.2025.106582

Digital Twins for Urban Public Spaces:

A Virtual Reality-based Method for Analysing Micro-scale Environment-Behaviour Relationships

Wenting Zhang[1], Sifan Cheng[2], and Jeroen van Ameijde[3]
[1,2,3]*The Chinese University of Hong Kong.*
[1]*wenting.zhang@link.cuhk.edu.hk, 0009-0000-5165-9479*
[2]*sifancheng@link.cuhk.edu.hk, 0000-0002-2303-7151*
[3]*jeroen.vanameijde@cuhk.edu.hk, 0000-0002-3635-3305*

Abstract. As primary venues for daily activities, urban public spaces significantly influence individuals' mental health and well-being. They play a crucial role in fostering social interactions, community formation, and the creation of liveable cities. However, our understanding of the spatial patterns of social behaviours in public spaces and their influencing factors remains limited. Virtual reality (VR) technology offers new opportunities for behavioural research, enabling systematic testing on physical and social environment variables. This study developed a VR-based framework to investigate individuals' spatial perceptions and social interaction preferences within urban settings. Through documenting participants' VR experiences, the study analysed their subjective perceptions of spatial environment, encompassing psychological perceptions such as perceived comfort and safety, as well as willingness to engage in social interactions. The study found that moderate environmental features and human presence promote social willingness, while excessive density inhibits it, with gender and personality further moderating these effects. This approach bridges empirical behavioural studies with practical urban design decision-making, aiding in understanding residents' daily experiences and informing strategies for enhancing human-centric urban environments.

Keywords. Public space, social behaviour, virtual reality, human-centred urban design.

1. Introduction

1.1. URBAN PUBLIC SPACES AND SOCIAL INTERACTION

Urban public spaces serve as vital platforms to facilitate and promote social interaction (McGrath et al., 2025). In public spaces, interactions are a constant factor and can be perceived both positive or negatively, with the presence of others acting as a source of

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 501-510.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

stress or comfort. Research indicates that positive social interactions in public spaces can significantly alleviate psychological stress among urban residents (Askarizad and Safari, 2020; Hematian and Ranjbar, 2022). Positive interactions also play a crucial role in enhancing social capital, strengthening community cohesion, and promoting the sustainability and well-being of urban life (Aptekar, 2015; Francis et al., 2012; Holland et al., 2017; Wood et al., 2017).

However, empirical research examining the social functions of public spaces from the user perspective still faces significant challenges (Sones et al., 2025). First, the physical attributes of the built environment possess inherent immutability and complexity, making it difficult for researchers to manipulate them in the real-world (Ibrahim et al., 2025; Struzek et al., 2025). Second, traditional field research methods still face limitations in dynamically capturing behavioural data (Su and Sun, 2025). These approaches often struggle to covertly and synchronously record users' real-time perceptions, which limits the development of detailed understandings of the perceptual and emotional mechanisms underlying social behaviours.

1.2. VIRTUAL REALITY AS A RESEARCH SOLUTION

The rapid advancement of virtual reality (VR) technologies in recent years has introduced transformative experimental possibilities to studying urban public spaces. By generating interactive and computer-simulated three-dimensional environments, VR enables individuals to enter a virtual space and experience physical environments that closely resemble reality (White et al., 2021; Wong et al., 2023). Its immersive and interactive nature allows researchers to simulate realistic spatial experiences under highly controlled conditions. By systematically manipulating environmental variables while keeping other factors constant, researchers can examine how specific spatial characteristics influence perception, social interaction tendencies, group aggregation patterns, and avoidance behaviours (Zhao et al., 2025). This parameterized experimental design effectively overcomes the constraints of variable entanglement in real-world contexts, offering a novel and rigorous pathway for exploring the causal mechanisms linking built environments and social behaviour.

1.3. RESEARCH FRAMEWORK

Based on these opportunities, this study employs a VR-based experimental approach to construct controlled virtual public space scenarios and quantitatively examine the mechanisms through which key objective attribute influence social interaction behaviour (Figure 1). By overcoming the limitations of traditional methods in environmental manipulation and behavioural document, the project provides both a methodological and empirical framework to inform the design of urban public spaces that promote social interaction.

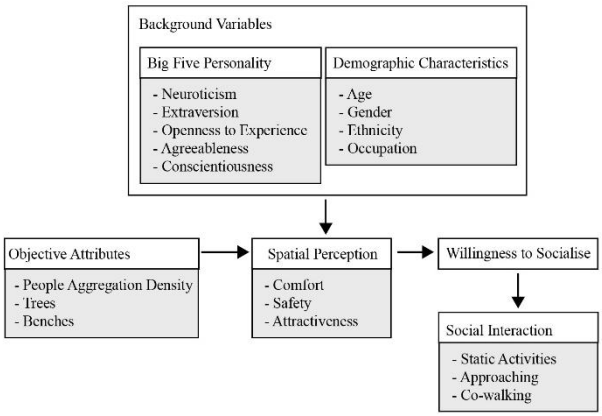


Figure 1. Framework of this study.

2. Methodology

2.1. STUDY AREA

The study was conducted at the "University Mall" on the campus of The Chinese University of Hong Kong (CUHK). This site serves dual functions as both a circulatory and social hub, characterized by high pedestrian volumes and diverse patterns of use. The selection of this site was guided by two main considerations. First, its clearly defined boundaries and moderate spatial scale provide high observability and experimental controllability. Second, this site was used in our previous pedestrian tracking study, which has yielded detailed behavioural data used to calibrate and contrast the current study's findings.

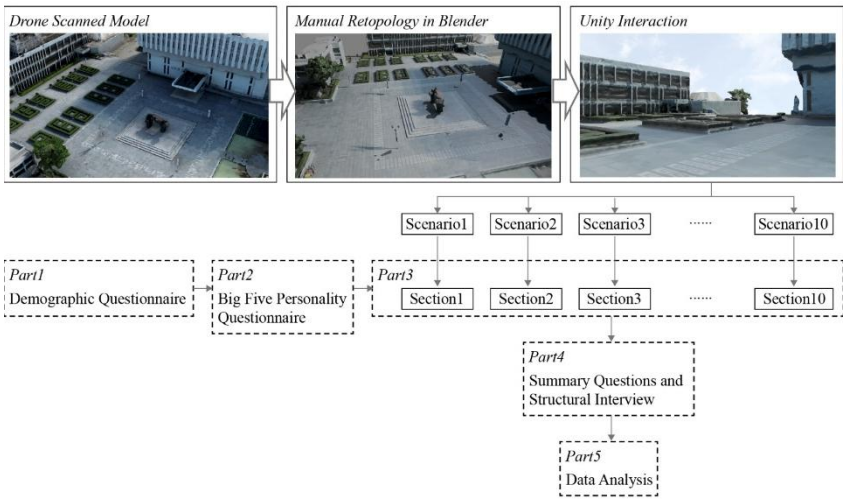


Figure 2. VR experimental workflow.

2.2. CONSTRUCTION OF THE VIRTUAL SCENE PROTOTYPE

The complete workflow is shown in Figure 2. To balance realism with experimental controllability, a high-fidelity digital twin of CUHK Mall was reconstructed as the foundation of the virtual experiment.

During the data acquisition phase, a drone equipped with a LiDAR module was used to perform high-precision scanning of the study area, generating point cloud data that captured the site's topography and spatial morphology. This ensured geometric accuracy at the macro scale and provided a reliable basis for subsequent 3D modelling.

During the model optimisation phase, due to the limited resolution of aerial point clouds at human eye level, the automatically generated meshes were manually refined and reconstructed in Blender. The process involved removing noise and geometric errors, simplifying complex meshes to improve real-time rendering performance, and applying high-resolution texture mapping to enhance visual realism and create an immersive experience.

After optimisation, the virtual model was imported into Unity 2022 LTS to develop an interactive VR experimental platform. The lighting conditions were calibrated to replicate the real environment at 4:00 p.m., ensuring realistic shadow direction and intensity. The final prototype was deployed on a Meta Quest 3 standalone VR headset, leveraging its inside-out tracking system to deliver a six-degrees-of-freedom (6DoF) immersive experience.

2.3. COMPARATIVE EXPERIMENTAL DESIGN

A within-subjects experimental design was adopted to systematically assess how different spatial configurations influence participants' perceptions and social interaction tendencies.

Building upon the research team's previous studies (van Ameijde et al., 2025), three key environmental variables were selected as independent factors: trees, benches, and people. Trees were chosen for their role in modulating thermal and lighting comfort through canopy shading. Seating represented opportunities for rest and lingering, and people were introduced to reflect the level of social vibrancy in the space. Trees and seating were each set at three levels: none, low, and high. The human presence variable included two levels: low aggregation density and high aggregation density. A full factorial combination of these variable levels generated ten experimental scenarios, which listed in Table 1.

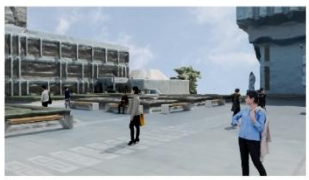
2.4. EXPERIMENTAL PROCEDURE

Before the VR experiment, each participant completed a brief demographic questionnaire and the Big Five Inventory - 10 (BFI-10) to account for potential influences of personalities on social tendencies (Gosling et al., 2003).

Participants were then asked to wear the Meta Quest 3 headset and enter the different virtual environments for short periods of time each. The ten experimental scenarios were presented in randomized order to minimize sequence effects. Each scenario lasted 60 seconds, during which participants were free to look around and move within the virtual space.

After each scenario, participants responded to ten Likert-scale questions that assessed spatial perception and social interaction tendencies. Upon completing all ten scenarios, participants filled out a final summary questionnaire, ranking their overall preferences among the scenes and answering several items regarding the VR experience itself to ensure data validity.

Table 1. Design matrix of virtual scenarios.

Objective Attributes	Levels	Scenario	
		People Aggregation Density - Low (1)	People Aggregation Density - High (2)
Drone scanned model	None (0)		
The presence of trees	Few (1)		
	More (2)		
The presence of benches	Few (1)		
	More (2)		

2.5. STATISTICAL ANALYSIS

First, for each question and scenario, we calculated the average score across all participants involved. To facilitate an in-depth comparison, the ten questionnaire items were condensed into two higher-level factors. Factor 1 (F1) represents spatial perception, capturing perceived safety, comfort, attractiveness, openness, privacy, and related qualities. Factor 2 (F2) represents behavioural preference, indicating the extent

to which a scenario facilitates interaction or encourages individuals to stay in place. F1 and F2 were calculated as the mean of their constituent items. The distributions of participant ratings across scenarios were visualized with box plots. To examine gender differences, box plots were additionally drawn by group (Male, Female, and All). To examine personality influences, an extraversion tendency score was computed for each participant; participants in the middle 10% (between the 45th and 55th percentiles) were excluded, and the remainder were split into Low (≤45th percentile) and High (≥55th percentile) extraversion groups. Box plots analogous to the gender visualization were then used to display within-scene differences among All, Low, and High extraversion groups.

3. Results

3.1. EXPERIMENT OVERVIEW

Participants were recruited using a convenience sampling method. Eligibility criteria included: (1) normal or corrected-to-normal vision (contact lenses permitted) to ensure accurate visual perception during the VR experience; and (2) current enrolment or employment at The Chinese University of Hong Kong, representing potential or actual users of the studied public space. The rationale for selecting this group was to obtain contextually relevant feedback from stakeholders most closely related to the space, thereby enhancing the ecological and applied validity of the findings (Wong et al., 2023).

Prior to the experiment, all participants received an informed consent form detailing the study's purpose, procedure, and potential risks. Participation was voluntary, and only individuals who provided written consent were included in the study. A total of 20 participants completed the study, comprising 8 males and 12 females, aged between 22 and 48 years (mean = 27.2, SD = 5.3). To minimize cognitive bias, participants were not informed of the research hypotheses before the experiment.

Each session lasted approximately 30 minutes, including pre-test questionnaires, VR trials, and post-experiment debriefing. As most participants had no prior VR experience, a 5-minute introductory session was conducted to familiarize them with the Meta Quest 3 headset and its interaction controls. Throughout the experiment, researchers monitored participants' physiological and behavioural responses to ensure safety and data integrity.

Post-experiment feedback indicated that most participants experienced mild to moderate symptoms of simulator sickness, such as light dizziness, nausea, or eye fatigue. Two participants reported temporary spatial disorientation, but no severe discomfort or early termination occurred. Based on subjective feedback and experimental logs, the overall acceptability of the VR experiment was high. Participants demonstrated sustained engagement, and data completeness and reliability met analytical requirements, supporting the robustness and internal consistency of the results.

3.2. FINDINGS

Across all participants, Table 2 shows that scenario 9 (bench (0), tree (0), people (2))

demonstrated relatively poor results. Both spatial perception and behavioural tendency scores were low. Scenario 2 (bench (0), tree (2), people (1)) performed best overall, most effectively attracting participants to engage in social interaction. In Figure 3, scores for F1 were relatively balanced across the ten scenarios, showing no significant overall differences. This suggests that the various combinations of environmental elements offered comparable levels of perceived attractiveness. However, clearer variations emerged in F2 across scenarios. The condition "bench (0), tree (0), people (2)" yielded the lowest social inclination score, indicating that a lack of seating and vegetation combined with high crowd density significantly reduced participants' willingness to engage socially. In contrast, "bench (2), tree (0), people (1)" and "bench (0), tree (2), people (1)" produced the highest social preference scores, suggesting that either an increase in seating facilities or vegetation can effectively enhance social tendencies under moderate crowd density.

Table 2. Mean perceptual and behavioural ratings across experimental scenarios

Scenario	Comfort	Safety	Attractiveness	Openness	Privacy	Interaction	Static	Approaching	Co-walking
1	3.05	3.25	3.15	4.25	1.7	3.15	3.05	3.1	3.9
2	3.9	3.65	3.3	3.7	2.8	3.8	3.6	3.6	3.7
3	3.3	3.4	3.1	3.45	2.3	3.5	2.95	3.2	3.45
4	3.35	3.15	2.7	3.95	2	3.1	2.55	3.05	3.65
5	3.1	3.2	2.85	3.5	2.05	3.15	2.95	3.1	2.9
6	3.15	3.15	2.7	3.4	1.8	3.25	2.9	3	3.2
7	3.6	3.4	3.5	3.15	2.95	3.6	3.1	3.4	3.4
8	3.55	3.3	3.3	3.75	2.7	3.35	3.2	3.3	3.75
9	2.85	2.8	2.75	3.65	1.85	2.9	2.25	2.6	3.05
10	3.55	3.25	3	3.75	2.25	3.5	3.4	3.15	3.55

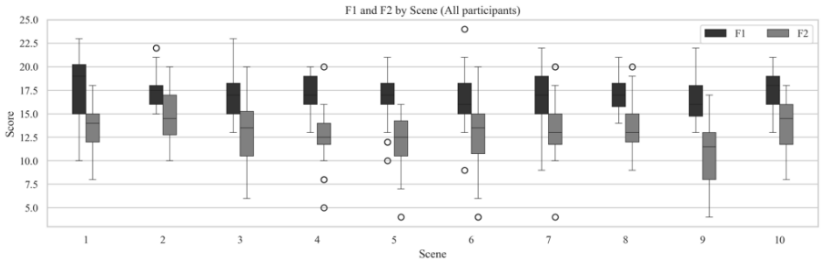


Figure 3. Perceptual and behavioural ratings across scenarios.

From a gender perspective in Figure 4, male and female participants reported similar overall perceptions of these 10 scenarios, indicating general consistency in spatial appraisal. Nonetheless, gender differences emerged in social preferences under certain conditions. In particular, within the "tree (1), bench (0), people (2)" scenario, females exhibited notably higher social inclination than males, suggesting a stronger positive response to moderate vegetation and dense social settings. In "bench (2), tree (0), people (1)" and "bench (0), tree (2), people (1)" scenarios, both genders demonstrated relatively high social preference, while "bench (0), tree (0), people (2)" remained the lowest for both groups. This pattern reinforces that overcrowded environments lacking supportive spatial features can suppress social willingness across genders.

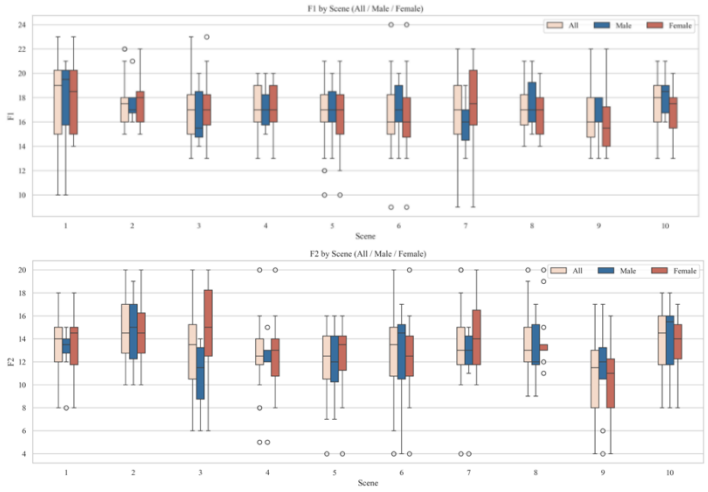


Figure 4. Gender differences in perceptual and behavioural ratings.

Regarding personalities in Figure 5, introverted participants reported significantly lower ratings of social perception in several scenarios (e.g., "bench (1), tree (2), people (2)," "bench (2), tree (0), people (2)"), whereas extraverts maintained relatively stable and higher ratings across environments. This shows that extraverts exhibit greater environmental adaptability and introverts show higher sensitivity to spatial attributes. A similar trend was observed in social behavioural inclination: introverts exhibited lower social tendencies in scenarios such as "tree (1), bench (0), people (2)" and "bench (1), tree (2), people (2)," where aggregation density was high. This pattern indicates that introverted individuals are more likely to exhibit social avoidance in complex or crowded spatial conditions.

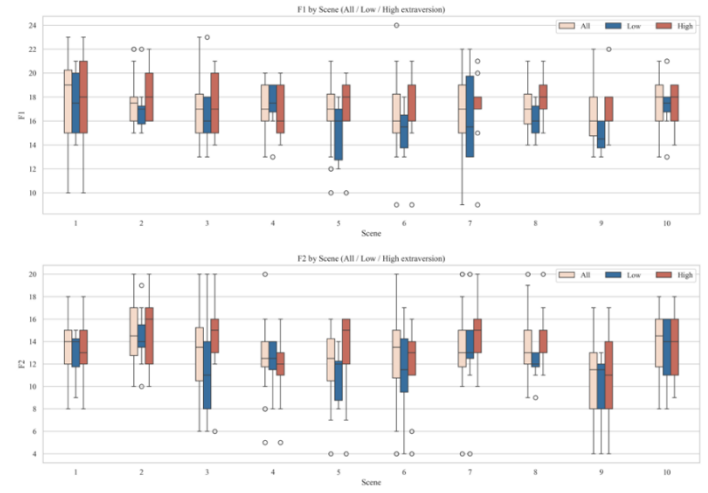


Figure 5. Extraversion differences in perceptual and behavioural ratings.

4. Limitations and Future Work

Although this study effectively revealed the relationship between urban public space objective attributes and spatial perception and social interaction tendencies through virtual reality experiments, several limitations remain.

First, this study selected students from The Chinese University of Hong Kong as the primary participants. While they represent a segment of stakeholders, the sample is relatively homogeneous in terms of age, occupation, and cultural background. The research overlooked other groups such as employees and visitors. Therefore, future studies should consider expanding the sample sources to test the experiment's generalizability. Second, while this study maximised the realism of the VR environment through material simulations and lighting effects, the virtual space still lacks certain sensory elements like temperature and smell. Future research could explore incorporating multimodal immersion technologies to enhance the sensory richness of the experiment. Moreover, as no low-fidelity or abstract VR baseline was included, this study cannot fully disentangle whether the observed behavioural responses are driven by photorealistic visual detail or primarily by spatial configuration itself. Future studies could therefore employ comparative VR conditions with varying levels of visual fidelity to systematically examine the specific contribution of visual realism to behavioural responses.

5. Conclusion

This study employs virtual reality as an experimental platform to investigate how key spatial objective attributes influence individuals' perceptions and social interaction tendencies in public open spaces. By systematically manipulating environmental configurations within a high-fidelity virtual environment, the study created diverse interactive scenarios for participants to experience likely in the real world. Findings indicate that moderate environmental features and human presence effectively promote social willingness, whereas excessive density inhibits interaction intent. Gender and personality analyses further elucidate these results. Female participants exhibited higher social willingness in environments conveying safety and social cues. Extraversion emerged as a stable moderator of spatial experience. Extraverted participants maintained positive evaluations across conditions, whereas introverted participants demonstrated greater sensitivity to environmental variations. These findings underscore the intertwined influence of spatial configuration and individual traits on social behaviour within urban settings.

By integrating behavioural science and immersive technology, this study demonstrates the methodological value of virtual reality in quantifying the social functions of urban design. It transforms traditionally subjective and qualitative design evaluations into quantifiable and comparable behavioural responses. Through simulating variables in spatial configurations, designers can interactively optimise the potential sociability of public open spaces, their perceived safety and comfort for lingering, before a project is implemented. This is particularly valuable in high-density urban settings where high usage levels and varied user needs require extensive consideration. The study not only offers theoretical insights but also delivers evidence-based design guidelines. It specifies density thresholds that can guide design

modifications to encourage social interaction, thereby directly informing the creation of built environments that foster greater social inclusion and mental well-being. Ultimately, this study advances VR as a transdisciplinary platform that bridges behavioural science and urban design, supporting human-centred urban development through empirically informed intervention.

References

Aptekar, S. (2015). Visions of public space: Reproducing and resisting social hierarchies in a community garden. In Sociological Forum (Vol. 30, No. 1, pp. 209-227).

Askarizad, R., & Safari, H. (2020). The influence of social interactions on the behavioral patterns of the people in urban spaces (case study: The pedestrian zone of Rasht Municipality Square, Iran). Cities, 101, 102687.

Francis, J., Giles-Corti, B., Wood, L., & Knuiman, M. (2012). Creating sense of community: The role of public space. Journal of environmental psychology, 32(4), 401-409.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in personality, 37(6), 504-528.

Hematian, H., & Ranjbar, E. (2022). Evaluating urban public spaces from mental health point of view: Comparing pedestrian and car-dominated streets. Journal of transport & health, 27, 101532.

Holland, C., Clark, A., Katz, J., & Peace, S. (2007). Social interactions in urban public places. Policy Press.

Ibrahim, A. F., Taha, M. N. A., & Shqeirat, D. D. H. (2025). The Role of Social Representation in Producing Public Spaces: Amman as a case study. Urban Governance.

McGrath, L., Bower, M., Valentine, K., Donohoe-Bales, A., Fearn-Smith, E., Buckle, C., ... & Wolifson, P. (2025). Active public spaces can protect against multiple dimensions of loneliness: Qualitative evidence from an Australian sample in the COVID-19 pandemic. Health & Place, 96, 103567.

Sones, M., Fuller, D., Kestens, Y., Thierry, B., & Winters, M. (2025). Evaluating the protective effect of public open space on social connectedness: evidence from a natural experiment cohort study in three Canadian cities. Health & Place, 96, 103541.

Struzek, D., Elshawa, H., Müller, L., Randall, D., Tolmie, P., & Müller, C. (2025). Designing for social impact: Understanding the social dynamics of interactive projections for public well-being. International Journal of Human-Computer Studies, 103596.

Su, T., & Sun, M. (2025). Understanding park-based health-promoting behavior and emotion with large-scale social media data: The case of Tianjin, China. Cities, 162, 105987.

van Ameijde, J., Cheng, S., Leung, K. S., Zhang, W., & Dong, J. (2025). Decoding micro-social interactions in public space: a computer-vision-based method. Computational Urban Science, 5(1), 55.

White, G., Zink, A., Codecá, L., & Clarke, S. (2021). A digital twin smart city for citizen feedback. Cities, 110, 103064.

Wong, M. O., Zheng, Z., Liang, H., Du, J., Zhou, S., & Lee, S. (2023). How does VR differ from renderings in human perception of office layout design? A quantitative evaluation of a full-scale immersive VR. Journal of Environmental Psychology, 89, 102043.

Wood, L., Hooper, P., Foster, S., & Bull, F. (2017). Public green spaces and positive mental health–investigating the relationship between access, quantity and types of parks and mental wellbeing. Health & place, 48, 63-71.

Zhao, Y., van den Berg, P., Schulten, B., & Kemperman, A. (2025). Public space and loneliness in young adults in the Netherlands: An experiment with virtual environments. Health & Place, 94, 103480.

Data - Driven Prediction Framework for Natural Ventilation Performance:

A Case Study of Vietnamese Modern Tube Houses

Trong-Tin Tran[1] and Yen-Khang Nguyen-Tran[2]
[1]*Yersin University of Da Lat.*
[2]*Shimane University.*
[1]*tintt@yersin.edu.vn, 0009-0001-7849-9839*
[2]*khang.ntr@riko.shimane-u.ac.jp, 0000-0002-0535-0179*

Abstract. Natural ventilation is a key strategy in traditional architecture, playing a critical role in improving thermal comfort and energy efficiency in hot and humid regions. In dense urban contexts such as Vietnamese tube houses, a narrow dwelling with limited façade exposure, this potential is often restricted, leading to excessive reliance on air conditioning. Besides, existing computation predictions by airflow simulation are complex and time-consuming, challenging their use in early design. This study examines the reliability of Machine Learning (ML) models to support early-stage ventilation prediction and to identify the key architectural and spatial factors influencing ventilation effectiveness. Using a data-driven framework based on architecture form and environmental characteristics, the research aims to provide an accessible alternative to computational simulations of optimization housing ventilation. Based on data collected from 47 tube houses in Ho Chi Minh City, Vietnam, the ML model achieved MAE = 16.77, RMSE = 19.36, and R^2 = 0.661 while identifying spatial configuration such as room pattern as a key parameter while orientation is found to be minor factor in this urban context. The framework offers insights for updating current Vietnamese construction laws and ventilation requirements for tube houses, supporting professional and non-professional designers in integrating natural ventilation strategies. Furthermore, the approach could be studied in similar climatic regions, contributing to tackling the global housing challenges using climate-responsive design.

Keywords. Data-driven design, predictive ventilation model, urban housing, building performance assessment

1. Introduction

In the era of global urbanisation and climate change, designing comfortable, low-energy housing in hot and humid regions remains a complex challenge. Natural ventilation, a principle of vernacular architecture, is believed to provide a sustainable

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 511-520.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

approach to achieving thermal comfort and energy efficiency (Piesik, 2023). However, according to Turkbeyler et al. (2012), its effectiveness is unpredictable in dense urban environments. Such case is pronounced in the Vietnamese "tube house", a housing typology with narrow frontages, deep plans, and restricted exposure to external winds. These characteristics result in poor air quality, while reliance on mechanical cooling also increases energy consumption. Besides, current methods for assessing natural ventilation using Computational Fluid Dynamics (CFD) are valuable for detailed analysis but require intensive workload and specialized expertise. As shown in studies of Nguyen-Tran et al. (2023), their complexity restricts use in early-stage design when decisions have the greatest impact. Additionally, ML model has proven effective in predicting building performance metrics and supporting rapid feedback in early design stages (Candanedo et al., 2017; Wang et al., 2018; Van Quang et al., 2024). However, their application to natural ventilation in dense tropical housing remains limited

This research proposes a data-driven approach using Machine Learning (ML) to address limitations in conventional building performance analysis and interprets the key architectural factors that influence ventilation effectiveness. While ML-based simulations are an established trend, this study focuses on the Vietnamese tube house, a dense urban typology that is both widespread in Vietnam and representative of housing in tropical cities. The novelty lies not in the ML method itself (Random Forest), but in demonstrating its utility for generating rapid, actionable insights on natural ventilation in this specific context. By using simple architectural and spatial parameters, designers can quickly estimate ventilation outcomes without the technical barriers of CFD, complementing existing simulation tools and broadening access to passive design knowledge. This framework thus supports more sustainable, human-centred housing design in dense tropical cities.

2. Natural Ventilation for housing in the tropics

Natural ventilation is a key strategy in climate-responsive tropical architecture, improving thermal comfort through convective cooling and maintaining indoor air quality by diluting pollutants. The literature in this field highlights the importance of building orientation, opening size and placement, and porous elements to facilitate airflow while providing shade and privacy (Lechner, 2014). In Vietnam, several studies such as Nguyen et al. (2011) have documented and verified the effectiveness of open plans, large openings, and transitional spaces in vernacular dwellings to promote ventilation, while noting the challenge for contemporary architecture to adapt these principles to the dense urban contexts.

The recent study on modern tube houses by Nguyen-Tran et al. (2023) provides a critical foundation for this work. Through a detailed analysis of 47 modern tube houses in Ho Chi Minh city, the research established a nuanced relationship between spatial composition and airflow performance and showcased the complex role of layered openings, combining windows with attached elements like louvers, porous blocks, or façade plantation. The study revealed diverse effects: while large windows promote openness, these layered elements help reduce heat gain and ensure privacy, though they often halve the effective opening area and affect airflow. However, the relationship is not straightforward: some configurations with multiple layers still facilitated good airflow, while others restricted it. These findings revealed that airflow performance is

shaped by nonlinear, interdependent design factors. Therefore, although this knowledge is valuable for experts, it remains difficult for non-specialists, such as homeowners, to access and apply. This complexity highlights the need for a systematic, data-driven investigation capable of capturing these interactions and translating them into more accessible predictive insights.

3. Methodology

3.1. ESTABLISHING THE DATASET FOR MACHINE LEARNING

As mentioned above, the dataset for this study was derived from previous research by Nguyen-Tran et al. (2023), including 49 living rooms and 124 bedrooms as distinct data points. Each room was characterised by architectural and spatial parameters such as building position, orientation, window size, window-to-wall ratio, and a composite opening ratio that accounts for layered façade elements. Ventilation performance, the variable target of computational simulation, was represented by the airflow area ratio derived from CFD simulations, a percentage of floor area with perceptible wind speed in tropical climate ($\geq$ 0.4 m/s). This dataset provides a robust foundation, pairing validated design parameters with performance outcomes to train and test the ML model.

3.2. MACHINE LEARNING PREDICTION MODEL

A Random Forest Regression (RFR) model was employed to predict ventilation performance using existing geometric and environmental parameters (as outlined in the methodology flowchart Figure 1). We implement RFR, an ensemble method built on multiple decision trees (Breiman, 2001), using Scikit-learn (v1.7.2) in Python 3.13.7. This model is particularly suited to non-linear relationships and complex feature interactions (Biau, 2012) and is widely applied in building performance prediction (e.g., Candanedo et al., 2017; Jurado et al., 2015; Wang et al., 2018).

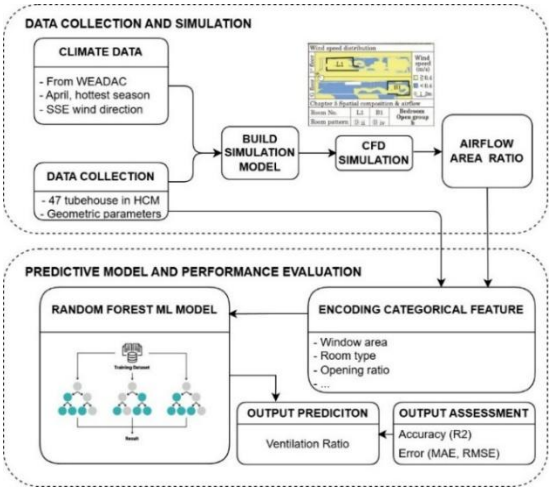


Figure 1. Methodology flowchart

We divide the dataset into 80% training and 20% testing subsets using a fixed random seed for reproducibility. Model parameters are then tuned for balance between accuracy and generalisation (n_estimators = 200, max_depth = 20, min_samples_split = 5, min_samples_leaf = 2, max_features = "sqrt") and predictions are averaged across all trees to reduce variance. Finally, we evaluate model performance using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Looking at feature importance, the actual versus predicted values, and the residual plots, we can interpret results and draw conclusion. After examining predictions for five randomly selected test samples and analysing error patterns, we assess the model's reliability for early-stage design applications.

4. Results

4.1. MODEL PERFORMANCE

For model performance, Figure 2 shows that the data points capture the overall trend, indicating a positive correlation between actual and predicted values. The RFR achieved an R^2 of 0.661, indicating that the input variables explain 66.1% of the variance in ventilation ratios, a reasonable result given the inherent complexity of natural ventilation (Van Quang et al., 2024). Results with MAE at 16.77 and RMSE at 19.36 indicate low average errors, though the higher RMSE suggests a few deviations. Manual inspection of five randomly selected test cases revealed errors ranging from 15.0 to 39.4 percentage points (mean: 24.5). This is consistent with the overall MAE of 16.77 and confirms the overprediction tendency at low actual values and underprediction at high values discussed below.

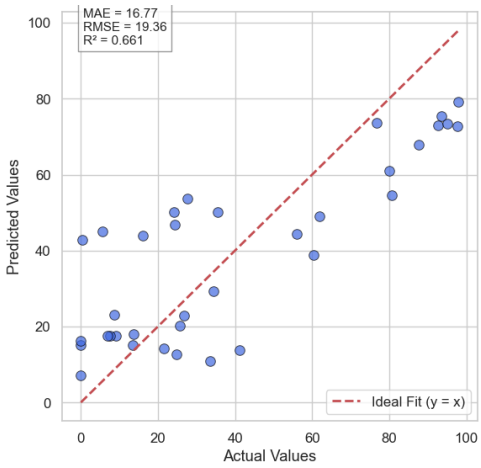


Figure 2. Actual and Predicted Values

Notably, we observe deviations at lower actual values (<40), where the model tends to overpredict ventilation performance, possibly because the training set contains predominantly weak ventilation scenarios. In contrast, in the mid to high ranges (>90),

attenuation bias occasionally occurs, reflecting the model's difficulty in capturing extreme cases. Figure 3 confirms these observations, illustrating the distribution of residuals between the actual and predicted ratios. However, a noticeable clustering pattern shows negative residuals (overpredictions) are most dominant in the lower range of expected values (approximately 15-25) and extend into the mid-range (30-50), indicating overestimated performance in these zones. In contrast, in the higher predicted range (60-80), positive residuals are frequent, reflecting the attenuation bias (underprediction) seen in Figure 2. Furthermore, the residual spread increases slightly as predicted values rise. This result suggests that the current model does not fully capture the influence of non-linear interactions or contextual parameters.

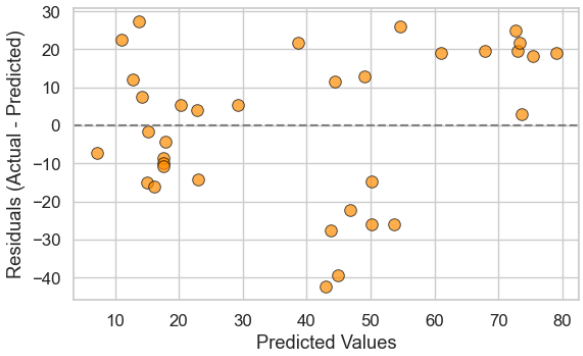


Figure 3. Predictions errors of actual ventilation ratios

4.2. CORRELATION OF DESIGN PARAMETERS AND VENTILATION

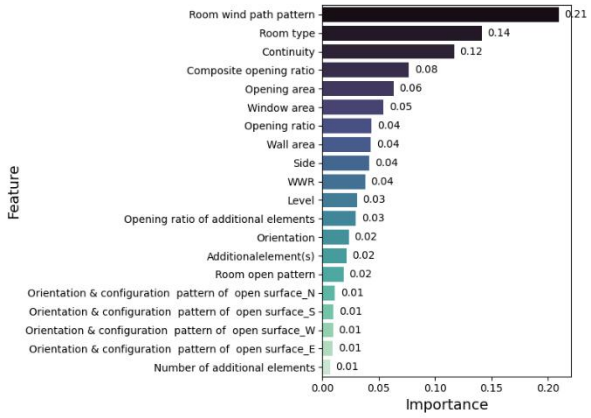


Figure 4. Important features contribute to enhance airflow

Next, the RFR model identified several key variables that significantly influence the ventilation ratio in tube houses. Figure 4 illustrates that the most dominant factor is Room wind path pattern (importance = 0.21), determining the airflow pathways and the degree of cross ventilation achievable within the room layout. The strong influence

of the wind path pattern highlights the critical role of spatial configuration and room connectivity in guiding airflow. This is validated by pairwise regression results shown in Figure 5, revealing a clear correlation between wind path factor and ventilation ratio ($R^2 = 0.485$).

The next dominant factors are Room type (importance = 0.14, $R^2 = 0.349$) and Continuity (importance = 0.12, $R^2 = 0.255$), highlighting how room function and spatial connectivity condition airflow distribution. Other important parameters in design, including Composite opening ratio (0.08), Opening area (0.06), and Window area (0.05), show a similar result, each contributing to the model's performance. These variables, while capturing the porosity of the building envelope, confirm that not all openings contribute equally to ventilation.

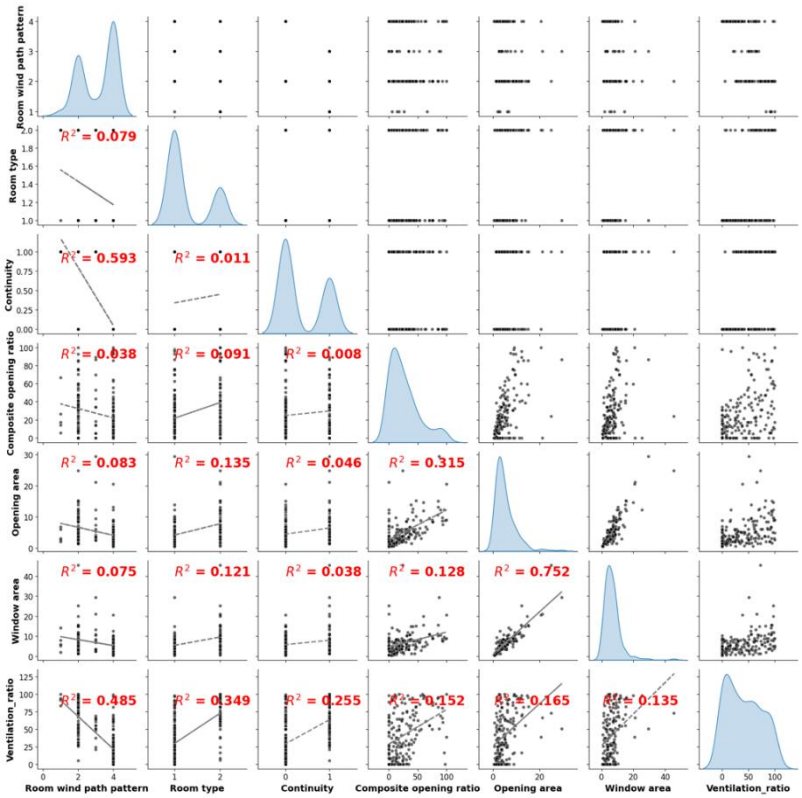


Figure 5. Pairwise Relationships of Top Features with Regression Fit and R^2

Also, other factors shown in Figure 4, such as building orientation, WWR, and additional elements, were found to have an even weaker influence. For example, Orientation (importance = 0.02) challenges the traditional assumption that cardinal direction is a dominant factor in this dense urban setting, where local spatial configuration is paramount.

Overall, the results show that ventilation performance depends primarily on spatial design and opening configuration, while factors like orientation and wall ratios play secondary roles. This underscores the need to integrate spatial morphology and façade design strategies from the early design stage to achieve effective natural ventilation.

5. Discussions

5.1. DESIGN STRATEGIES FOR ENHANCED NATURAL VENTILATION

The results provide actionable insights for design and clarify the role of predictive modelling in architectural practice, demonstrating that effective natural ventilation in Vietnamese tube houses depends less on conventional factors (e.g., orientation, window size) and more on spatial and façade design. Hence, two interrelated design principles emerge as critical: (1) spatial continuity and room function, and (2) controlled porosity in both façade and plan.

5.1.1. Spatial continuity and room function:

The model identifies wind path patterns and spatial continuity as key predictors of ventilation performance, showing that airflow is shaped not only by openings but by interior connectivity. In deep, narrow tube houses, maximising continuity through open floor plans, lightweight partitions, glass dividers, or open furniture layouts enhances air movement in back rooms more effectively than layouts with large but isolated windows. Strategic placement of doors and windows to organise cross-ventilation further facilitates air penetration into the interior, preventing stagnation near window zones, a limitation often observed when relying solely on large openings.

Room function also influences ventilation outcomes. Living rooms should be positioned to open onto a front yard or light well and maintain continuity with the kitchen, dining area, or balcony. This can be combined with a permeable façade using large windows or vegetation to admit breezes while providing shading and privacy. Bedrooms, in contrast, require privacy yet adequate airflow to avoid dampness and discomfort; placing them near light wells or ventilation openings ensures effective air paths. Bedrooms on higher floors generally benefit from stronger ventilation, while those on lower levels may require supplementary mechanical systems.

5.1.2. Porosity in façade and plan:

The findings reaffirm that the effectiveness of openings depends on their composite porosity rather than sheer size. The composite opening ratio, accounting for louvers, vegetation, and perforated blocks, proved a more reliable predictor of airflow performance than total window area. Thus, the quality of porosity, not its quantity, is decisive. A façade with high visual openness but obstructed airflow (e.g., dense plants or solid screens) results in poor ventilation, while a carefully designed porous envelope can provide shade, privacy, and efficient air movement. Examples include sun-shading screens with optimal mesh density, partial greenery on balconies, or strategically placed ventilation bricks. Cross-ventilation is further enhanced by locating windows at both ends of the plan (front and rear façades).

At the plan level, light wells, courtyards, and open stairwells act as vertical and

horizontal mechanisms facilitating stack and cross-ventilation deep into the building. Importantly, the model demonstrates that in dense urban contexts like Ho Chi Minh City, where adjacent buildings block prevailing winds, orientation plays a limited role. Architects should therefore prioritise aligning intake and exhaust openings with local micro-wind patterns and connecting them through a porous, continuous interior layout.

Solutions such as interior courtyards and light wells are particularly effective in generating a stack effect that draws air inward and brings daylight to recessed rooms. Overall, the study confirms that orientation is a minor factor compared to internal voids, corridors, and connecting doors, which enable wind to permeate the entire depth of the house, rather than being confined near windows.

5.2. IMPLICATIONS FOR PREDICTIVE MODELLING

The machine learning (ML) model, trained on 47 modern tube houses in Ho Chi Minh City, shows that data-driven methods can estimate airflow performance with reasonable accuracy, without relying on time-consuming CFD simulations. It provides a practical tool for the early design stage when designers need quick feedback to compare spatial or façade options. By entering a few basic parameters related to building geometry and configuration, architects, builders, or homeowners can obtain an initial sense of how well a design may ventilate before moving to more detailed analysis. The identified dominant factors of wind path pattern, room type, and continuity (cf. 3.2) mean that designers need to input only a limited set of spatial parameters to obtain performance estimates. This contrasts with CFD simulations that require detailed 3D geometry, mesh generation, and boundary condition specifications. The relatively lower importance of orientation in this urban context also simplifies the prediction process, as designers can focus on spatial configuration rather than cardinal direction during early-stage layout decisions.

The results also confirm the relationships discussed above, where ventilation performance depends more on spatial continuity, room arrangement, and less on the degree of openness in façades and plans, like the current design assumption. The ML model helps quantify these relationships and makes them easier to test across different layouts. Rather than replacing detailed simulations, it serves as a simpler step that can guide design adjustments and help identify layouts with better airflow potential. Furthermore, the model's ability to capture non-linear relationships is evidenced by the feature importance rankings and the pairwise regression results (cf. Figure 5). While individual features show moderate correlations (wind path pattern $R^2 = 0.485$, room type $R^2 = 0.349$), the Random Forest model achieved an overall $R^2 = 0.661$ by accounting for interactions between these parameters. This improvement demonstrates that ventilation performance depends on combinations of spatial factors rather than isolated variables, supporting the need for a data-driven approach that can handle such complexity. With MAE = 16.77 on a percentage scale, the model's predictions carry an average error of approximately 17 percentage points. This level of accuracy is sufficient for comparing design alternatives and identifying potentially problematic configurations during early design stages but should not replace detailed simulation for final design validation. Designers should be aware that predictions for low-performing rooms tend toward overprediction, while high-performing configurations may be underpredicted (cf. 3.1).

From a broader perspective, this approach can support the development of practical design and planning recommendations for dense urban housing in Vietnam. The results may contribute to ongoing discussions and updating regulations about ventilation standards and energy efficiency, such as the National Technical Regulation on Energy Efficiency Buildings QCVN 09:2017/BXD (Ministry of Construction, 2017) and the Design Standards for Ventilation and Air Conditioning TCVN 5687:2010 (Ministry of Science and Technology, 2010). The model could also be refined to reflect other climatic regions within Vietnam or similar tropical cities, using local data to adjust its predictions. In summary, the use of ML in this study demonstrates how relatively simple predictive tools can assist design decisions for natural ventilation in urban tube houses. Its main value lies in improving understanding of airflow behaviour in compact dwellings and providing a basis for responsive architectural and planning practices.

6. Conclusion

This study shows the ML model can predict ventilation performance in Vietnamese tube houses with reasonable accuracy and provide insight into the factors influencing airflow. The results confirm that spatial configuration and room connectivity play a more important role in ventilation than traditional design parameters such as orientation or opening ratio. The model offers a faster and more accessible way to assess ventilation potential during early design stages, reducing reliance on time-consuming CFD simulations.

Nevertheless, several limitations should be acknowledged. The dataset of 173 rooms (49 living rooms and 124 bedrooms) from 47 professionally designed tube houses, which provides adequate coverage of professionally designed dwellings in Ho Chi Minh City's dense urban context. However, architect-designed limits the representativeness of the training data. As a result, generalisation to informal, owner-built housing or other climatic regions remains constrained. Additionally, the ML model achieved an R^2 value of 0.661, meaning that about one-third of the variance in ventilation performance remains unexplained. This limitation reflects both the complexity of airflow behaviour in naturally ventilated building and the simplified assumption embedded in the simulation. In particular, the simulations model did not include surrounding urban morphology, street geometry, or occupant behaviour, factors known to influence real ventilation outcomes. These omissions lead to predictive uncertainty when applying the model beyond the simulation context.

Future work should expand the dataset to include a broader range of tube house typologies, including owner-built dwellings, and extend coverage to other Vietnamese climatic regions. Integrating contextual parameters and occupant behaviour variables such as window-opening habits (Zhang & Barrett, 2012) would improve prediction accuracy. Besides, testing other algorithms, such as Gradient Boosting, may also help capture the complex interactions between design and airflow.

With further refinement, the model could be integrated into simple design tools to help architects, builders, and homeowners estimate ventilation performance early in the design process. It may also inform the updates of national guidelines by providing data that reflect actual urban housing conditions. In this regard, this study has established a foundation for using predictive modelling to support more climate-responsive housing design in Vietnam and cities with similar tropical conditions.

References

Biau, G. (2012). Analysis of a random forests model. *J. Mach. Learn. Res*., 13(null), 1063-1095.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Candanedo, L. M., Feldheim, V., & Deramaix, D. (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, 140, 81-97. https://doi.org/https://doi.org/10.1016/j.enbuild.2017.01.083

Jurado, S., Nebot, À., Mugica, F., & Avellana, N. (2015). Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques. *Energy*, 86, 276-291. https://doi.org/https://doi.org/10.1016/j.energy.2015.04.039

Lechner, N. (2014). Heating, cooling, lighting: Sustainable design methods for architects. *John Wiley & sons*.

Liaw, A., & Wiener, M. C. (2007). Classification and Regression by Random Forest.

Ministry of Construction. (2017). QCVN 09:2017/BXD – National technical regulation on energy efficiency buildings. *Hanoi, Vietnam: Ministry of Construction*.

Ministry of Science and Technology. (2010). TCVN 5687:2010 – Ventilation and air conditioning – Design standards. *Hanoi, Vietnam: Ministry of Science and Technology.*

Nguyen, A.-T., Tran, Q.-B., Tran, D.-Q., & Reiter, S. (2011). An investigation on climate responsive design strategies of vernacular housing in Vietnam. *Building and Environmen*t, 46(10), 2088–2106. https://doi.org/https://doi.org/10.1016/j.buildenv.2011.04.019

Nguyen Tran, Y. K., Ueyama, M., & Murata, R. (2023). Passive Design Seen from Spatial Openness and Airflow in Contemporary Tube Houses of Ho Chi Minh City, Vietnam. *Journal of Architecture and Planning (Transactions of AIJ),* 88(806), 1249-1258. https://doi.org/10.3130/aija.88.1249

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.

Piesik, S. (2023). *Habitat: Vernacular Architecture for a Changing Climate. Thames & Hudson*

Seo, D. K., Kim, Y. H., Eo, Y. D., Park, W. Y., & Park, H. C. (2017). Generation of Radiometric, Phenological Normalized Image Based on Random Forest Regression for Change Detection. *Remote Sensing*, 9(11), 1163.

Turkbeyler, E., Yao, R., Nobile, R., Bentham, T., & Lim, D. (2012). The impact of urban wind environments on natural ventilation. *International Journal of Ventilation*, 11(1), 17–28.

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). Random Forest based hourly building energy prediction. *Energy and Buildings*, 171, 11-25. https://doi.org/https://doi.org/10.1016/j.enbuild.2018.04.008

Van Quang, T., Doan, D. T., & Yun, G. Y. (2024). Recent advances and effectiveness of machine learning models for fluid dynamics in the built environment. *International Journal of Modelling and Simulation*, 1–27. https://doi.org/10.1080/02286203.2024.2371682

Zhang, Y., & Barrett, P. (2012). Factors influencing the occupants' window opening behaviour in a naturally ventilated office building. *Building and Environment*, 50, 125–134. https://doi.org/https://doi.org/10.1016/j.buildenv.2011.10.018

From Sensing to Control

IoT for Sleep-Aware Architecture

Jing Zhong[1†], Ziyang Han[2†], Peilin Li[3], Jun Yin[4], Yuxin Zhang[5], and Shuai Lu[6*]
[1,2,4,5,6]*Tsinghua University.*
[3]*National University of Singapore.*
[1†]*zhongj24@mails.tsinghua.edu.cn, 0009-0003-5249-0152*
[2†]*han-zy25@mails.tsinghua.edu.cn, 0009-0000-6038-7559*

Abstract. The development of Human-Building Interaction (HBI) research increasingly regards residential environments as information ecosystems in which human activity and physical settings continuously interact. In recent years, adaptive design for residential spaces that emphasizes occupants' everyday physiological behaviors has advanced. However, sleep, as a primary daily activity, still largely depends on standardized control logic in smart home systems, which constrains genuinely human-centered adaptation. Accordingly, this study makes three contributions: (1) We develop a longitudinal multimodal dataset on sleep information involving 30 participants. (2) Based on this dataset, mutual information (MI) and Granger causality (GC) analyses identify time-lagged causal relationships between multimodal environmental elements and sleep. These analyses enable the construction of a personalized sleep sensitivity weighting matrix, whose robustness and adaptability are validated by random forest models, with DCW-based individual models achieving a median increase in R^2 of 0.57 and a median reduction in RMSE of 11.0 min for key sleep outcomes relative to the population baseline. (3) We propose a dual-mode closed-loop feedback framework that integrates passive regulation with active intervention. This framework can be integrated with Building Information Modeling (BIM) and Digital Twins for architectural practice scenarios centered on sleep-related information. This research broadens the scope of HBI, facilitating individualized reactions and physiologically interpretable intelligent residential architecture.

Keywords. Architectural Informatics, Human–Building Interaction, Cyber-Physical Systems (CPS), Multimodal IoT Sensing, Occupant-Centric Predictive Control

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 521-530.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

1. Introduction

Architecture is increasingly viewed as an information ecosystem where human behavior, digital perception, and the physical environment form a dynamic symbiosis through feedback and adaptation. Within this perspective, HBI has become the theoretical core driving the development of building intelligence, evolving environmental control from automation toward human-centered co-regulation systems. Within residential environments, sleep is a particularly fundamental design concern. It is both a basic physiological need and a spatial behavior shaped by the combined effects of lighting, temperature, air quality, and acoustic conditions. However, current intelligent residential building systems that regulate the sleep environment predominantly rely on population average models or static sleep comfort indices. This approach overlooks individual variations in sleep circadian rhythms, sensitivity, and adaptability, resulting in "intelligence" that remains confined to uniform control.

To bridge the gap between individual physiological perception and building control, this paper proposes an IoT-based "Sleep-Aware Building" framework, as shown in Figure 1: (1) Wearable physiological signals and environmental sensor data are integrated to construct a longitudinal dataset of 30 subjects. (2) MI and GC are employed to reveal nonlinear associations and directional time lags between environmental or physiological factors and sleep outcomes, extracting and characterizing individualized sensitivities. These sensitivities were mapped into encodable DCW to enable personalized regulation of lighting, thermal, and acoustic devices. Comparative analysis of random forest-based prediction models demonstrated that the individualized model utilizing DCW outperformed the group baseline in total sleep time, number of awakenings, and sleep efficiency. (3) The study further establishes a dual-mode feedback mechanism combining passive device regulation with active user intervention. The passive layer implements adaptive adjustment of the equipment, while the active layer collects user feedback and subjective sleep experiences via a mobile app, forming a collaborative closed loop system involving humans, devices, and the environment in real world scenarios. Overall, this study provides a replicable pathway for adaptive buildings through causally driven, interpretable personalized control, expanding the research boundaries of HBI.

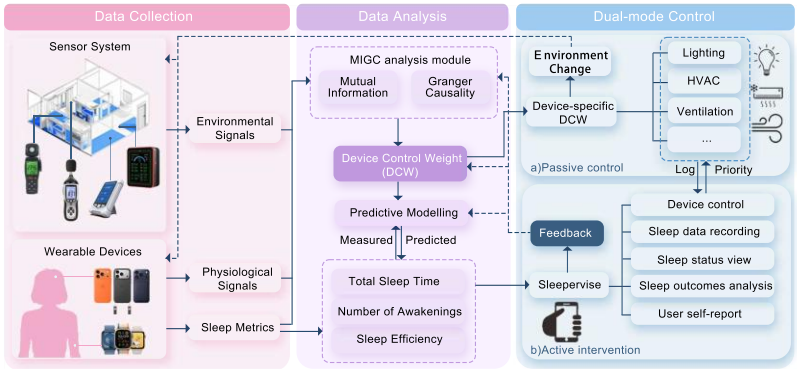


Figure 1. The IoT-based "Sleep-Aware Building" framework

2. Related Work

2.1. HUMAN–BUILDING INTERACTION

In recent years, research on HBI has shifted from a technology-dominated architecture paradigm toward a three-dimensional framework integrating people, environment, and technology, emphasizing the fusion of human experience, mechanisms, and building systems (Becerik-Gerber et al., 2022). Within this context, buildings are viewed as intelligent entities capable of perception, responsiveness, and adaptation, while users are regarded as individuals possessing unique experiences, capabilities, and needs (Economidou et al., 2023). However, most HBI studies remain focused on group level comfort, energy efficiency or interface design, with limited research exploring the relationship between individual physiological states and building response mechanisms. Current literature predominantly employs group feedback or average metrics, failing to incorporate individual sensitivity as a core variable.

2.2. IOT-BASED ENVIRONMENTAL CONTROL

With the advancement of IoT technology, building environmental control systems are transitioning from fixed settings to real time sensing and adaptive responses. The integration of wearable devices, environmental sensors, and smart controllers enables personalized environmental adjustments (Carli et al., 2020). In thermal comfort and indoor environmental quality control, numerous studies have begun utilizing IoT data and machine learning to establish individualized comfort models. Previous research has proposed generating personal thermal comfort models based on wearable device data and implementing control via IoT architectures (Alsaleem et al., 2020). Recent work further proposes IoT systems with low implementation cost as a deployment foundation for large-scale monitoring and tracking of individual thermal comfort states (Chillón Geck et al., 2024). Alongside IoT-based sensing and actuation, diffusion-based generative architectural modeling offers an interactive way to generate and refine 3D forms that can be coupled with sensor feedback for sleep-oriented optimization(Yin et al., 2025). While IoT-based smart homes have been studied for sleep assistance and environmental control, most research focuses on the relationship between smart devices and sleep behavior, lacking architectural modeling that refines individual sensitivity and closed-loop feedback chains.

3. Dataset

3.1. DATA SOURCE

This study's dataset comprises longitudinal multimodal recordings from 30 participants, integrating 43 dimensions of environmental and physiological data from wearable devices and IoT environmental sensors including an 18-in-1 air quality monitor, ozone sensor, light meter, and noise meter. After time zone alignment, the study defines “a sleep night” as the time window from 18:00 to 12:00 the following day, with each sleep night serving as an independent

observation unit. The final dataset is structured as a multimodal record with each night of sleep as the basic unit of analysis, supporting subsequent MI and GC analyses.

3.2. DATA SCHEMA AND VARIABLES

The dataset comprises two components: Sleep Outcome Metrics and Indicators describing factors influencing sleep quality. These influencing factors are categorized into Environmental Indicators and Physiological Indicators. To comprehensively evaluate the combined effects of multiple factors on sleep quality, we follow prior research frameworks and aggregate selected key measurements from the 43 factors into several composite indicators, improving interpretability and robustness while supporting subsequent causal modeling and personalized intervention design. Variables not aggregated are retained as Individual Indicators. Table 1 summarizes the definitions of all indicators used in this study.

Table 1. Summary of Variables and Measurement Indicators

Categories	Variables	Unit	Description
Sleep Outcome Metrics	SOL	min	Sleep Latency
	NWAK	count	Number of Awakenings
	TST	min	Total Sleep Time
	WASO	min	Wake After Sleep Onset
	SE	%	Sleep Efficiency
Environmental Indicators	AEI	-	Air Pollution Index (PM2.5, PM10, PM1.0, TVOC, CO_2)
	LEI	h	Light Exposure Index
	Individual Indicators		Humidity, Audio Exposure, Headphone Audio Exposure, Sound Reduction, Audio Exposure Event, Time In Day light
Physiological Indicators	CSI	-	Cardio Stability Index
	TRI	-	Thermoregulation Index
	MLI	-	Metabolic Load Index
	DAI	-	Daily Activity Intensity
	GSI	-	Gait Stability Index
	VMI	-	Vertical Mobility Index
	EAI	-	Endurance level Index
	Individual Indicators		Heart Rate, Respiratory Rate, Oxygen Saturation, High Heart Rate Event, VO2Max, Walking Heart Rate Average, Physical Effort, Running Vertical Oscillation, Ground Contact Time, Running Stride Length, Six Minute Walk Test Distance, Apple Stand Time

3.3. DATA POLICY AND CONSENT FRAMEWORK

This study treats sleep information as highly sensitive physiological data, adhering to edge computing and data minimization principles: raw data from wearable and

environmental sensors undergoes processing solely at the local level, with only derived features such as nightly statistics and MI/GC coefficients stored in the cloud. All participants provided explicit informed consent, clearly acknowledging data usage, retention periods, and withdrawal rights. Data is fully encrypted during transmission and storage, with least privilege access and auditable logging enforced.

4. Methodology

4.1. DATA PREPROCESSING

To ensure temporal and spatial comparability of multimodal signals, data from wearable devices and IoT sensors undergo time synchronization and sampling calibration. Data preprocessing comprises four steps: achieving cross-device time synchronization and sampling alignment by aggregating data within a nightly time window; handling missing and anomalous values via rule-based skipping or smoothing; performing Z-standardization on full signals and implementing local detrending.

4.2. MUTUAL INFORMATION

MI quantifies the nonlinear connection between factor X and sleep outcome Y, thereby delineating the sensitivity of each factor to sleep quality:

$$I(X;Y)=\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} \quad (1)$$

where p(x,y) denotes the joint probability distribution, and p(x), p(y) represent the marginal distributions.

4.3. GRANGER CAUSALITY

Employing VAR-based GC tests to elucidate directional and temporal causative links between sleep-related parameters X_t and sleep outcome metrics Y_t, so producing individual causal matrices with group means as the reference point:

$$Y_t=\alpha_0+\sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{i=1}^{p} \gamma_i X_{t-i} + \varepsilon_t \quad (2)$$

where p denotes the lag order, and ε_t represents the residual term.

4.4. DEVICE CONTROL WEIGHT

To characterize individual sleep sensitivity under multifactorial influences, a DCW matrix is constructed based on MI/GC results for prediction and control purposes:

$$DCW=Norm(MI)\times sgn\times \Gamma \times \Pi \quad (3)$$

where Norm(MI) represents the normalized value of MI strength; sgn determines the sign of the Pearson correlation coefficient to establish directionality; Γ denotes the composite weight of causal confidence and lag penalty; Π represents the group robustness weight.

4.5. PREDICTIVE MODELLING AND VALIDATION

This study further uses DCW as interpretable feature weights to construct sleep outcome prediction models. For each night, control scores are computed by summing the nocturnal mean of each factor $f(x)$, weighted by its corresponding DCW(x, y), which are subsequently employed to forecast TST, NWAK, and SE. The enhancement of individual models compared to group models is assessed. The model construction is as follows:

- Personal Model: Independently trains a random forest regression model for each subject (80% training-20% testing split) to capture individualized environmental sensitivity;
- Population Baseline Model: Constructed using leave-one-out (LOSO) cross-validation to generate an average prediction model for the population.

Furthermore, to evaluate robustness under real-world sensing imperfections, this study conducted simulated data missingness tests to examine performance stability under data loss.

5. Experiment

5.1. EXPERIMENTAL PROCEDURE

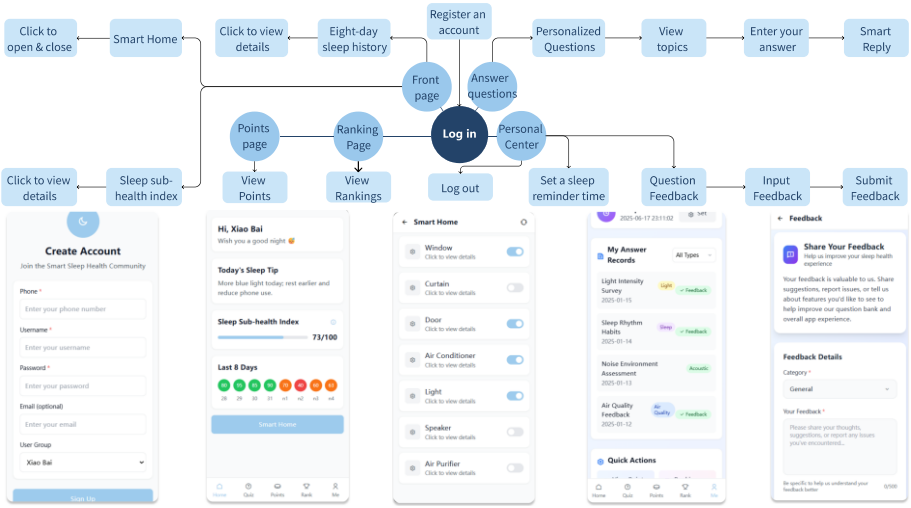


Figure 2. The companion sleep feedback and regulation app "sleepervise"

A total of 30 participants were recruited. Wearable devices, including Apple Watches, continuously recorded data for over 3 years to establish individual physiological baselines. The formal experiment lasted one month, during which IoT environmental sensors were deployed. Participants maintained their daily routines while all data was fed in real-time into a unified framework for modeling and cross-validation.

The specific experimental workflow is as follows: (1) Integrate historical physiological and sleep data, input MI/GC results, and generate personalized DCW initial values. (2) Recode according to device interfaces to connect DCW to lighting, temperature control, white noise, etc., enabling personalized passive regulation. (3) Online prediction of sleep performance, with DCW-based adaptive adjustment

strategies. (4) To support real-time interaction and feedback during the experiment, we developed the companion sleep feedback and regulation app *Sleepervise*. As shown in Figure 2, *Sleepervise* allows users to directly control devices (with priority over passive regulation), view trends and recommendations, and answer personalized questions, with feedback immediately returned to update strategies and models. (5) The system perpetually logs metrics, updating DCW and models continually. Following the trial, individuals submit subjective surveys.

5.2. RESULTS

5.2.1. DCW Matrix Results

Based on MI/GC analysis of 30 subjects, we generated individualized DCW results to reveal the underlying regulatory characteristics of sleep influenced by multidimensional physiological and environmental factors. In the DCW matrix, positive and negative weights indicate enhancing and suppressing effects, respectively, while larger absolute values reflect stronger impacts. Results indicate significant individual differences in sleep responses to acoustic, light, and activity factors. For example, as shown in Figure 3, concerning noise sensitivity, AEI demonstrated an inhibitory effect on SOL and TST for subject 15, while exerting a beneficial influence for subject 4. In contrast to outcomes averaged among groups, DCW maintains and emphasizes these distinct regulatory variations.

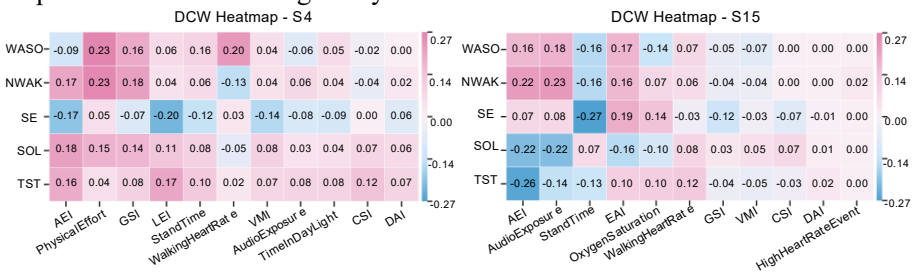


Figure 3. Individual DCW heatmaps

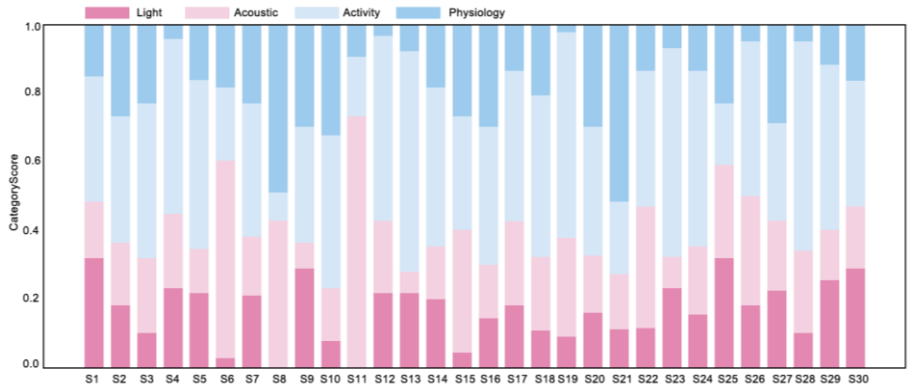


Figure 4. Distribution of multimodal influencing factors across 30 subjects

As shown in Figure 4, the primary influences for most subjects were concentrated in Physiology and Acoustic, indicating that the interaction between circadian rhythms and the acoustic environment is the core driver of sleep quality. Light and Activity exhibited stable but secondary positive effects in some individuals. This aligns with the dual-action mechanism of "physiological load and environmental disturbance" widely recognized in sleep research. Overall, DCW demonstrates strong interpretability in terms of directionality, intensity distribution, and cross-individual consistency, indicating that the model has successfully extracted key sleep-driving features from multimodal signals. This provides a data foundation for personalized regulation and environmental interventions.

5.2.2. Individual vs. Population Model Performance

This comparison is designed to quantify the benefit of personalization for sleep-aware control. The population baseline represents a deployable cold-start model learned from other occupants, while the individualized model represents an adapted model learned from the target occupant's longitudinal records. To ensure a fair comparison, both models use the same input variables and are evaluated on the same held-out nights for each occupant; the reported differences therefore reflect the performance gain attributable to personalization under a realistic deployment scenario.

We evaluated model performance using R^2 and RMSE . Higher R^2 and lower RMSE indicate better fit and accuracy. We defined:

$$\Delta R^2 = R^2_{ind} - R^2_{group}, \Delta RMSE = RMSE_{ind} - RMSE_{group}$$

Here, $\Delta R^2 > 0$ and $\Delta RMSE < 0$ indicate superior individual model performance. Significance was assessed via Wilcoxon signed-rank test ($p < 0.05$).Note that when the group baseline performs poorly, its R^2 can be negative, indicating performance worse than a mean predictor. In such cases, the resulting difference can exceed 1 after subtracting the baseline R^2.

Table 2. Main Analysis Results

Sleep Metric	Median ΔR^2	Mean ΔR^2	Wilcoxon p (ΔR^2)	Median ΔRMSE	Mean ΔRMSE	Wilcoxon p (ΔRMSE)
NWAK	**0.5687**	1.0864	**0.0005**	-0.7567	-0.6878	**0.0006**
TST	0.1244	1.1286	0.0329	**-11.0081**	**-21.8336**	0.0462
SE	0.5275	**5.5385**	0.0329	-0.0402	-0.0320	0.0251

As shown in Table 2, the comparison between individual and group models reveals that individual models demonstrate overall superiority across key sleep metrics. The mean R^2 for NWAK significantly improved ($p=0.0005$), while RMSE markedly decreased ($p=0.0006$), better capturing individual sensitivity to nocturnal disturbances. TST and SE also showed significant improvements ($p<0.05$), reflecting higher stability and fitting accuracy. Overall, individualized modeling based on DCW effectively enhances the predictive performance of sleep-environment interactions.

5.2.3. Subjective Feedback Results

To assess the framework's usability and acceptance, we collected subjective ratings

from 30 participants and 5 experts on a 1–10 scale covering sleep metrics, system features, and willingness to continue participation. Figure 5 showed consistent trends across both groups: Participants reported higher overall satisfaction and a stronger intention to continue using the system, whereas experts were more cautious but nonetheless actively evaluated the system's responsiveness and the synchronization.

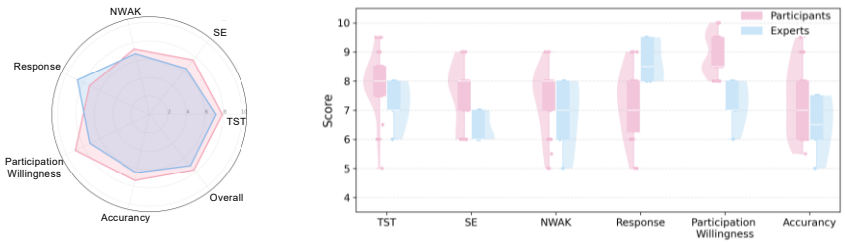


Figure 5. Subjective evaluation results from participants and experts

5.2.4. Robustness

To evaluate robustness under imperfect sensing conditions, we conducted a missing-data simulation under a missing-at-random setting driven by acoustic exposure. As reported in Table 3, the individualized model consistently maintains an advantage over the population baseline across key sleep metrics, with positive ΔR^2 and negative ΔRMSE and statistically significant differences in most cases. These results indicate that the DCW-based individualized modeling framework remains resilient to sensor data loss and exhibits strong robustness to noise and measurement uncertainty in real-world deployments.

Table 3. Simulation Results of Missing Data

Sleep Metric	Median ΔR^2	Mean ΔR^2	Wilcoxon p (ΔR^2)	Median ΔRMSE	Mean ΔRMSE	Wilcoxon p (ΔRMSE)
NWAK	0.3306	0.3848	0.0009	-0.3110	-0.3488	0.0042
TST	**2.0319**	**3.6640**	**0.0000**	**-90.8063**	**-97.6974**	**0.0000**
SE	1.2107	2.4891	0.0001	-0.0931	-0.0768	0.0001

6. Discussion

This study investigates the potential for sleep-aware intelligent adaptation in residential settings from a Human–Building Interaction perspective. DCW provides codable access to home devices to support personalized orchestration, and it can be integrated with Building Information Modeling and Digital Twins. In a representative spatial zoning workflow, environmental and physiological data streams can be mapped to Digital Twin entities such as the bedroom zone, lighting, HVAC terminals, windows, shading components, and air-cleaning devices. The individualized model then translates predicted sleep responses into actionable control setpoints, including lighting dimming schedules, ventilation or purification operation, shading adjustments, and noise-mitigation prompts. Compared with conventional rule based control with thresholding or group average modeling, our framework explicitly captures individual differences and time-lagged dependencies between exposures and sleep outcomes, and

operationalizes these insights into an interpretable closed-loop process rather than applying uniform thresholds.

The limitations of this study include an intervention period of approximately one month and a restricted timeframe. Although objective improvements did not reach strong statistical significance, positive subjective feedback suggests that the system has potential for continuous iteration and further optimization.

7. Conclusion

This study proposes an IoT framework for sleep-aware buildings from a HBI perspective. It builds a DCW-based regulation model for causal inference and closed-loop control of physiological–environmental data. DCW's encodability enables integration with building environmental control devices to adaptively regulate sleep environments within residential structures. The dual-mode feedback framework based on DCW can interface with BIM and Digital Twins, providing quantitative evidence for optimizing residential building design. Future research will further explore integrating open data with IoT building ecosystems to advance intelligent building design practices tailored to individual responses.

References

Alsaleem, F., Tesfay, M. K., Rafaie, M., Sinkar, K., & Besarla, D. (2020). An IoT framework for modeling and controlling thermal comfort in buildings. Frontiers in Built Environment, 6, 87.

Basner, M., & McGuire, S. (2018). WHO Environmental Noise Guidelines for the European Region: A systematic review on environmental noise and effects on sleep. International Journal of Environmental Research and Public Health, 15(3), 519.

Becerik-Gerber, B., Fatah gen. Schieck, A., Heydarian, A., Lucas, G., Aryal, A., Awada, M., Bergés, M., Billington, S., Boric-Lubecke, O., Ghahramani, A., Hölscher, C., Jazizadeh, F., Khan, A., Langevin, J., Liu, R., Marks, F., Mauriello, M. L., Murnane, E., Noh, H., Pritoni, M., Roll, S., Schaumann, D., Seyedrezaei, M., Taylor, J. E., Zhao, J., & Zhu, R. (2022). The field of human–building interaction for convergent research and innovation for intelligent built environments. Scientific Reports, 12, 21557.

Carli, R., Dotoli, M., Pellegrino, R., & Ranieri, L. (2020). IoT based architecture for model predictive control of HVAC in smart buildings. Sensors, 20(3), 781.

Chillón Geck, C., Alsaad, H., Voelker, C., & Smarsly, K. (2024). Personalized low-cost thermal comfort monitoring using IoT technologies. Indoor Environments, 1(4), 100048.

Daskalova, N., Yoon, J., Wang, L., Beltran, G., Araujo, C., Nugent, N., McGeary, J., Williams, J. J., & Huang, J. (2020). SleepBandits: Guided flexible self-experiments for sleep. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–13). ACM.

Daskalova, N., Metaxa-Kakavouli, D., Tran, A., Nugent, N., Boergers, J., McGeary, J., & Huang, J. (2016). SleepCoacher: A personalized automated self-experimentation system for sleep recommendations. In Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology (pp. 347–358). ACM.

Economidou, E., Gerner, N., Pichler, C., Hartl, A., & Frauenberger, C. (2023). Uncovering placemaking needs with(in) a kindergarten community: A cross-disciplinary approach to participatory design. Frontiers in Psychology, 14, 1126276.

Yin, J., Gao, W., Li, J., Xu, P., Wu, C., Lin, B., & Lu, S. (2025). ArchiDiff: Interactive design of 3D architectural forms generated from a single image. *Computers in Industry, 168*.

Varanasi Heritage Infill

Specification of a loam-based modular kit-of-parts

Abhijeet Manjunath[1] , Rutuja Rode[2] , Sonali Rane[3] and Milad Showkatbakhsh[4]
[1,2,3,4]*Architectural Association School of Architecture*
1manjunath.abhijeet@gmail.com, 0009-0002-2270-8838
[2]roderutuja@gmail.com, 0009-0000-5861-5815
[3]sonalirane.str@gmail.com, 0009-0000-2275-3212
4showkatbakhsh@aaschool.ac.uk, 0000-0001-5440-5175

Abstract. This paper specifies a system-level loam-based modular kit-of-parts for This paper presents a modular, loam-based kit-of-parts designed for heritage-sensitive infill in Varanasi, India. Addressing the city's dense urban fabric and high pilgrim density, the system prioritizes rapid, low-disruption assembly and off-site prefabrication to minimize interference with religious processions and local commerce. The methodology utilizes a test-calibrated pipeline linking soil mix design to structural performance. Field screening identified an optimal base loam (60% sand, 25% silt, 18% clay), which was bio-stabilized using lime pozzolana, barley straw, and sugarcane bagasse. Structural testing confirmed high-performing finalists with compressive strengths of and significant thermal damping properties. Full-scale panels are integrated into standardized glulam frames using a sliding top-groove detail, with rubber gaskets and horizontal battens managing hygroscopic movement. By utilizing abundant Ganges alluvium and CNC-milled molds, this system ensures material circularity and industry-grade precision. The result is a scalable, passive-cooling solution suitable for hot-humid climates.

Keywords: Kit-of-Parts, Loam Panels, Bio-stabilised Earthen Materials, Thermal Performance, Heritage Retrofit.

1. Introduction

1.1. URBAN AND CULTURAL CONTEXT

Varanasi, one of the world’s oldest continuously inhabited cities and widely regarded as the religious capital of India, hosts a continual inflow of pilgrims and tourists (Singh, & Rana, 2017). This floating population produces recurrent peaks in footfall and service demand, compounding congestion in narrow lanes, courtyards, and ghats.

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 531-540.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Local markets and service functions (hospitality, crafts, logistics, ritual services) are economically tied to this visitor flow and constitute a primary income source for residents; construction must therefore be rapid, reversible, and low disruption so livelihoods are not impaired.

1.2. MATERIAL AND ASSEMBLY RATIONALE

A principal reason to select earthen construction in this context is material proximity: the Ganges River system deposits alluvial soils along Varanasi's banks, providing abundant local loam (sand–silt–clay) historically used in building. Loam's workability, thermal mass, porosity, reusability, and low embodied carbon align with hot-humid needs and circular supply chains, and, paired with a glulam primary frame and panelised earthen inserts, enable reversible, heritage-compatible retrofits (Fathi, 1973, Gurvich, 2000). Compared with common concrete/cement-block infill, which typically has higher embodied CO_2, limited breathability, and relies on wet, disruption-prone site processes, the loam system buffers humidity, damps short-term heat gains, reduces environmental impact, and preserves fabric through dry, reversible joints. Both serve as non-loadbearing infill within a primary frame, but loam panels deliver similar speed with better indoor moisture comfort, easier panel-by-panel replacement, and greater sensitivity to historic contexts. From craft to system (Figure 1).

Traditional rammed earth is site-intensive (formwork, ramming, curing). In this research, CNC-milled moulds and a unitised panel catalogue shift production off-site, raising precision, reducing on-site labour, and compressing programme, transforming earthen construction from craft-by-parcel to a repeatable system that keeps streets and storefronts active during peak religious calendars.

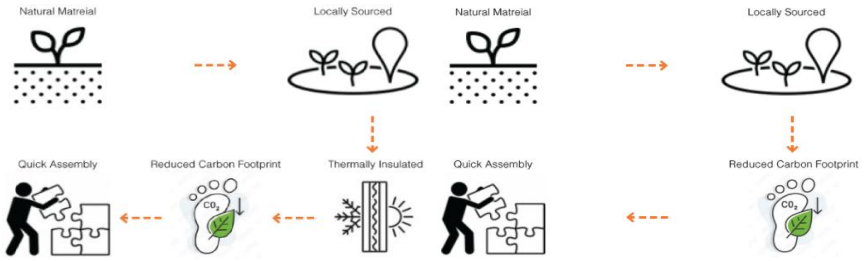


Figure 1. Rammed earth & glulam panel construction

1.3. AIM (TIME-CONSTRAINT, QUALITY & SPEED)

In a context constrained by congestion, a large floating population, limited closure windows, and the need to keep markets and services operating, this study aims to deliver a system-level loam kit-of-parts that improves panel quality, through CNC precision, consistent curing, and industry-grade finishes, while increasing manufacturing and installation speed via off-site prefabrication and dry assembly; to maintain or enhance passive-cooling performance by leveraging thermal mass, breathability, and operable shading/ventilation; and to demonstrate a reversible, tolerance-friendly assembly detail suitable for dense heritage precincts without

disrupting the local service economy.

2. Methodology

2.1. FIELD SCREENING AND BASE SELECTION

Three base mixes: Sample 1 = 70/20/10, Sample 2 = 60/25/15, Sample 3 = 65/30/5 (sand/silt/clay) (Figure 2). Drop, biscuit/shrink, and cigar/roll tests identified Sample 2 as balanced; refinement produced 2A (clay 18%) and 2B (clay 20%). Repeat drop tests selected 2A as the controlled base (Figures 3, 4, 5, 6).



Figure 2. Loam composition and mixing

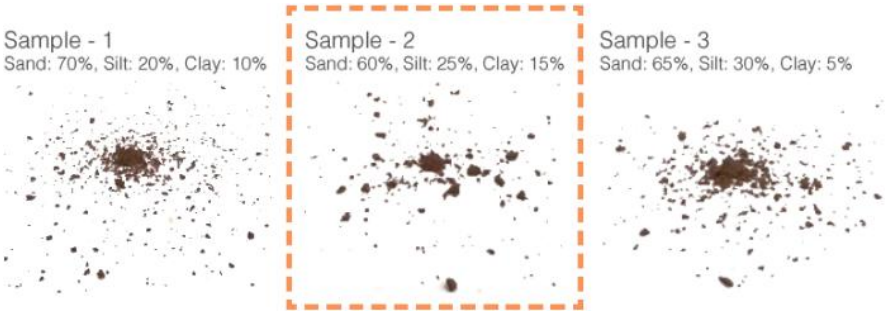


Figure 3. Drop test results



Figure 4. Cigar/roll test results



Figure 5. Biscuit test results

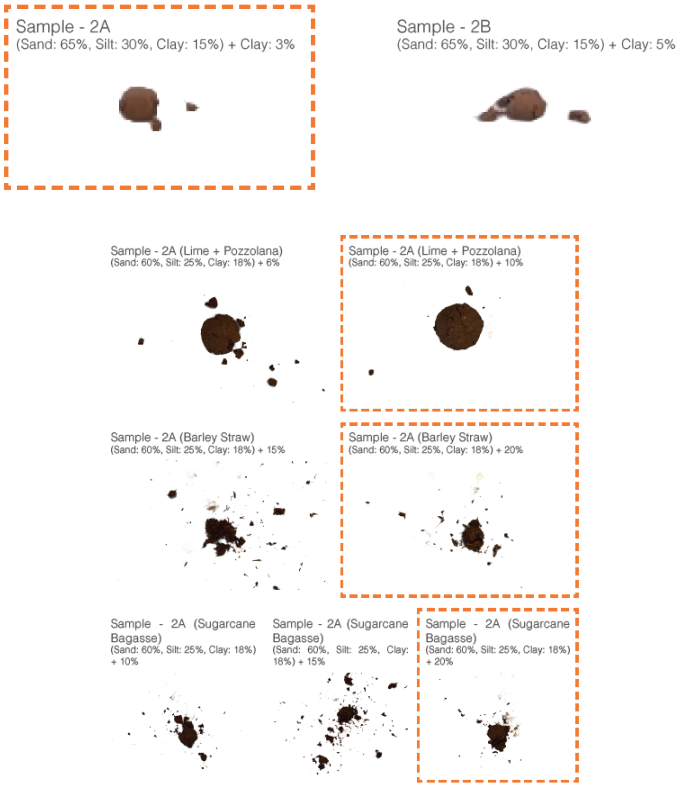


Figure 6. Drop test results

2.2. STABILISER DOWN-SELECTION

Base 2A combined with lime + pozzolana (6%, 10%), barley straw (15%, 20%), sugarcane bagasse (10%, 15%, 20%). Drop-test survivors: 2A+10% L+P, 2A+20% barley straw, 2A+20% bagasse.

2.3. MECHANICAL TESTS (CUBES)

Uniaxial compression tests were performed on 100 × 100 × 100 mm cube specimens (Figure 7) prepared from each shortlisted mix and the unstabilised 2A control, in accordance with BS EN 12390-1:2021. For each specimen, compressive strength and oven-dry density were measured and recorded (Figure 8).



Figure 7. Finalised samples for compression test

Holliday Concrete Testing Ltd.
Unit 11, Ironbridge Close, Great Central Way,London NW10 0UF
Tel: 020 8955 1707 Fax: 020 8955 1709

Concrete Test Cube Report
(BS EN 12390-1:2021, 2:2019, 3:2019, 7:2019)

UKAS TESTING 0478

Client Ref	Test Age Days	Width (mm)	Length (mm)	Height (mm)	Mass (g)	Density (kg/m³)	Failing Load (kN)	Compressive Strength (N/mm²)
TEST 1	27	97	98	100	1965	2070	21	2.2
Location of Concrete: -			Comments:	70, 20, 10%				
Mix Details:								
TEST 2	27	97	100	100	1880	1940	17	1.8
Location of Concrete: -			Comments:	60, 25, 15%				
Mix Details:								
TEST 3	27	97	99	100	1884	1960	14	1.5
Location of Concrete: -			Comments:	65, 30, 5%				
Mix Details:								
2B CLAY	15	98	98	100	2101	2190	15	1.6
Location of Concrete: -			Comments:	57, 24, 9%				
Mix Details:								
2A-LIME+POZ	15	97	96	100	2091	2250	17	1.8
Location of Concrete: -			Comments:	10%+LIME+POZZOLONA				
Mix Details:								
2A-BARLEY	15	98	96	100	**1859**	**1980**	**21**	**2.2**
Location of Concrete: -			Comments:	**20% (BARLEY STRAW)**				
Mix Details:								
2A-SUGARCANE	15	98	99	100	**1773**	**1830**	**21**	**2.2**
Location of Concrete: -			Comments:	**20% (SUGARCANE BAGASSE)**				
Mix Details:								

Figure 8. Compression test results

2.4. PANEL PROTOTYPING, FABRICATION, AND REGIONAL ASSEMBLY VALIDATION

Prototype units (≈1:3; 130 × 200 × 40 mm) were formed in milled moulds in four lifts, with a timber batten placed in the final lift to create the top groove, then air-dried under ambient conditions to stable mass before assembly mock-ups (Figure 9). Sun-drying was selected over kiln firing to minimise process energy and preserve the earthen matrix's open pore

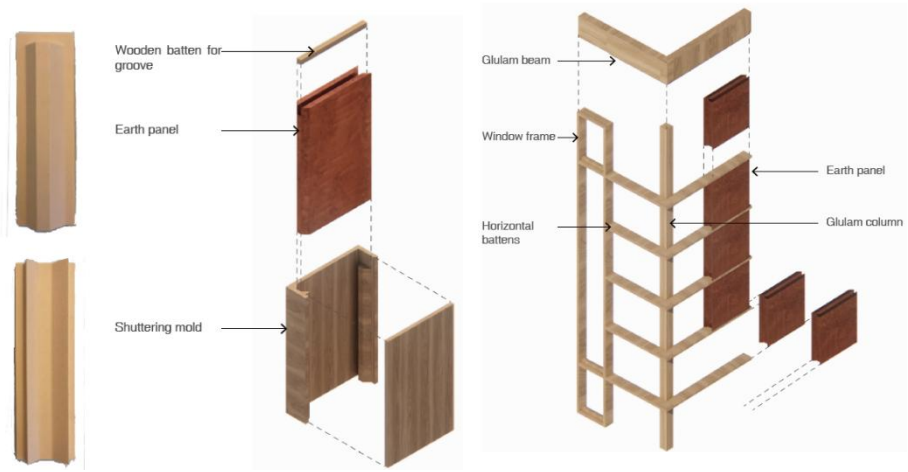


Figure 9. Foam mould & shuttering assembly for casting and regional assembly for wall panels

structure, maintaining vapour permeability and hygroscopic buffering that are critical to hot-humid comfort; kiln burning densifies the microstructure through sintering/vitrification, typically reducing moisture exchange while increasing embodied energy and emissions. For interface validation, a regional wall mock-up was assembled: panels were slid from above along thin horizontal battens that act as spacers and shock absorbers to control racking and accommodate minor dimensional tolerances from compaction and drying, while compressible rubber gaskets at vertical joints-maintained air-tightness and allowed moisture and temperature-driven movement without cracking. Together, the sun-dried fabrication route and the groove–batten–gasket assembly delivered a fast, fully reversible system that can be installed and serviced with minimal tools and site occupation, well suited to tight rights-of-way and commercially active streets where trading must continue during works, while retaining loam's key advantages of thermal mass, permeability, and reversibility in heritage contexts.

2.5. THERMAL ASSAY (TRANSIENT CONDUCTION, 1:3 PANELS)

2A+20% barley straw and 2A+20% bagasse panels (130×200×40 mm) tested over 5 minutes in a metal-plate conduction rig; initial and final surface temperatures recorded (Figures 10, 11).

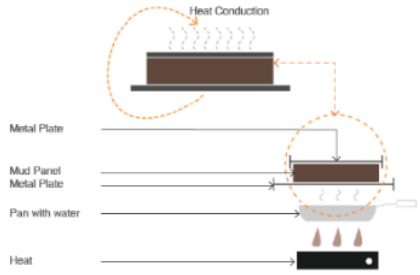


Figure 10. Foam mould & shuttering assembly for casting and regional assembly for wall panels

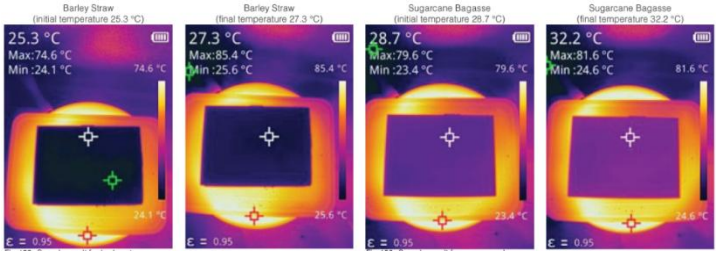


Figure 11. Sample results for barley straw & sugarcane bagasse

3. Results

3.1. MECHANICAL AND THERMAL PERFORMANCEY

Compression tests on cube specimens showed that the two bio-stabilised finalists, 2A + 20% barley straw and 2A + 20% sugarcane bagasse, each achieved ≈2.2 kN/mm² compressive strength, with corresponding dry densities of ≈1980 kg/m³ and ≈1830 kg/m³ respectively; both outperformed other candidates and were advanced to panel specification. Complementary short-cycle thermal assays indicated panel-scale damping: over a five-minute interval, barley-straw panels exhibited a temperature rise of ~2 °C, while bagasse panels rose by ~3.5 °C, with barley straw performing slightly better in the test rig.

3.2. ASSEMBLY PERFORMANCE

Earthen panels absorb and release moisture, so they swell and shrink seasonally. The groove–batten–gasket interface decouples this natural movement from the frame: grooves and battens absorb tolerance shifts while gaskets maintain air-tightness as the panel "breathes." Practically, that means panels drop in cleanly, don't bind or crack as humidity changes, and don't require wet sealants or return visits for remedial work, so installation is faster, sites can reopen sooner, and trading hours are preserved during rapid retrofits. (Timber Dev. UK, 2024, American Wood Council, 2020, IntechOpen, 2021, Meridian, 2013).

3.3. FINALISED PANEL RECIPES

Two infill options within a standardised frame (Figure 12):

- Barley-straw panel: (s:si:c + b) = 60 : 25 : 18 + 20%.
- Bagasse panel: (s:si:c + su) = 60 : 25 : 18 + 20%.



Figure 12. Final panels

4. Comparison to substitutes and environmental value

Bio-stabilised loam panels demonstrate viable service performance, achieving approximately 2.2 kN/mm² at lower densities, with short-cycle thermal assays confirming panel-scale heat-gain damping. In delivery terms, off-site fabrication (sun-dried units) and dry, reversible joints reduce on-site labour and programme risk compared with formwork-heavy, site-rammed methods, an advantage where processions and festivals limit closures and market/service incomes depend on continuous operation. Environmentally, Ganges-derived local loam cuts hauling and supports material circularity, while agro-residue stabilisers and sun-drying lower embodied impacts; the glulam primary frame further improves the footprint relative to conventional steel/concrete. Finally, the system's demountable panels and modular logic enable repair and replacement without fabric loss, avoiding long-term material lock-in and supporting adaptive, heritage-sensitive retrofit.

5. Applicability beyond the case scenario (passive cooling)

Varanasi serves as the tested case, calibrating mixes, unit dimensions, and joints in a dense, hot-humid, festival-active context with a large floating population and an economy reliant on visitors. Building on this validation, the system logic, combining thermal mass, breathability, operable shading/ventilation, and dry, reversible assembly, generalises to other hot-humid and warm climates as a passive-cooling retrofit that minimises site disruption and reduces loss of trading hours. In settings with extreme diurnal swings or prolonged wet seasons, deployment should be preceded by diurnal/seasonal hygrothermal monitoring and complemented by panel-scale flexure and in-plane shear testing to finalise thickness–span rules and connection detailing.

6. Contribution and novelty

This work delivers a closed-loop specification in which field tests, stabiliser screening,

cube mechanics, and panel-scale thermal assays culminate in a constructible, reversible joint, so that testing drives the specification end-to-end. It advances a system-level shift from site-intensive craft to off-site, industry-grade panels, enabling standardisation, distribution, time-certain delivery, and reduced disruption to markets and services. The groove–batten–gasket interface is a reversible jointing strategy tailored to earthen hygroscopy that preserves air-tightness without wet adhesives. A parametric panel catalogue is explicitly tied to micro-climatic intent shading, ventilation, and privacy, supporting staged, rapid retrofits in congested heritage fabrics. Finally, the paper offers practice-facing reflections on variability, tolerances, the assay horizon, and adoption, translating computational and material workflows into actionable specifications.

7. Practical challenges and reflections

Material variability (grading, moisture) is managed through field-test gates and a standardised 2A base before stabilisation; digital–physical tolerances from compaction/drying are absorbed by the groove–batten–gasket interface, while CNC-moulded panels improve repeatability across batches and suppliers. Given that current thermal assays are short-cycle, we plan diurnal/seasonal hygrothermal monitoring and panel-scale flexure/shear tests to close the validation loop. Supply and logistics are addressed via off-site hubs and industry-grade panels that enable distributed fabrication and just-in-time installation, protecting trading hours in busy corridors.

8. Conclusion

This study shows that a system-level loam kit-of-parts, coupled with a glulam primary frame, can satisfy both performance and programme in one of the most time-sensitive urban settings, Varanasi's congested, festival-active core where visitor-driven markets and service economies must remain open. Drawing on Ganges-derived alluvial loam, the work anchors a circular, local material chain and translates field-tested recipes and short-cycle panel assays into a constructible specification, standard unit dimensions, stabiliser bands, and a robust groove/batten/gasket interface. Together, these deliver material performance with passive cooling: bio-stabilised loam panels achieve viable compressive strengths at moderate densities and provide panel-scale thermal damping consistent with hot-humid comfort through thermal mass, breathability, and operable shading/ventilation. They also enable precision and quality at scale by moving from hand-formed, on-site ramming to repeatable, off-site fabrication with controlled quality checks and just-in-time installation. The dry, reversible assembly shortens site occupation, preserves trading hours, and reduces labour congestion, critical where processions and festivals cannot pause. Over the life cycle, demountable panels and bio-based stabilisers support repair, reuse, and recovery, avoiding material lock-in and allowing incremental upgrades without rebuilding the frame. Finally, an explicit adoption protocol, progressing through localisation, bench testing, prototyping, and deployment, together with defined parameter windows (soil-grading targets and 15–20% residue bands) makes the approach replicable beyond the case study to hot-humid and warm climates with narrow rights-of-way, seasonal surges, and limited closure windows, without compromising heritage sensitivity or environmental intent.

8.1. WHAT THIS CHANGES IN PRACTICE

Earthen construction need not be slow, site-bound, or tolerance-fragile. With CNC precision and kit-based logistics, loam panels become a standard product rather than a bespoke craft, elevating quality while shortening manufacturing and installation time. For municipalities and heritage custodians, the framework offers predictable sequencing, lower embodied impacts, and reversible interventions compatible with conservation aims. (The broader computational workflow for panel assignment and climate-responsive deployment is documented in the companion framework.)

8.2. FUTURE DISCUSSIONS

Future work will extend validation through diurnal and seasonal hygrothermal monitoring with in-situ trials to translate short-cycle panel results into whole-building comfort impacts; undertake panel-scale flexure and in-plane shear testing to refine thickness–span rules and connection detailing under wind and seismic racking; conduct cost/schedule benchmarking to compare assembly time and performance of off-site panels versus conventional infill; and establish QA/inspection protocols for moisture, compaction, and gasket integrity to institutionalise industry-grade production and maintenance. In sum, the proposed loam-based modular system, grounded in Ganges alluvium and bio-residue stabilisation, provides a fast, precise, and repairable path to passive-cooling retrofits in Varanasi and other climate-compatible regions, supporting economic continuity for market and service livelihoods while advancing a lower-carbon, heritage-respectful urban future.

Attribution Statement

ChatGPT (OpenAI, 2025) was used for language editing and grammar refinement during the preparation of this paper.

References

Fathy, H. (1973). Architecture for the poor: An experiment in rural Egypt. University of Chicago Press.

Gurvich, M. (2000). Mud, mirror and thread: Folk traditions of rural India. Mapin Publishing.

Timber Development UK. (2024). Glued laminated timber (glulam).

American Wood Council. (2020). Mass timber.

IntechOpen. (2021). Engineered wood products as a sustainable construction material: A review.

Meridian Allen Press. (2013). Cradle-to-gate life-cycle impact analysis of glulam.

OpenAI. (2024). ChatGPT Model: GPT-4 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/

Singh, R. P., & Rana, P. S. (2017). Varanasi: Sustainable development goals, smart city vision and inclusive heritage development.

Computational Urban Walkability

Integrating Perceptual and Computational Frameworks for Walkability Assessment in Campus Environments Using AI-Based Models

Sena Kaynarkaya[1], Aslı Çekmiş[2], İsmail Çetin[3], Yusuf Hüseyin Şahin[4], Gözde Ünal[5]
[1,2]*Istanbul Technical University, Faculty of Architecture, Department of Architecture.*
[3,4,5]*Istanbul Technical University, Faculty of Computer and Informatics Engineering, Department of AI & Data Engineering.*
[1]*kurkcuoglus@itu.edu.tr, 0000-0002-8434-5211*
[2] *cekmis@itu.edu.tr, 0000-0001-6558-6469*
[3] *cetini18@itu.edu.tr, 0009-0002-9251-7631*
[4] *sahinyu@itu.edu.tr, 0000-0002-4134-9423*
[5] *unalgo@itu.edu.tr, 0000-0001-5942-8966*

Abstract. This study presents an AI-driven framework for assessing campus walkability by adapting a deep learning model initially trained on a large-scale, expert-annotated urban streetscape dataset. Unlike previous campus-focused studies that relied solely on locally collected perceptual data, this approach transfers knowledge from the broader urban context to the campus scale. A convolutional neural network model designed to predict safety, accessibility, and aesthetics was fine-tuned using 2,820 Google Street View images labeled by students at Istanbul Technical University's (ITU) Ayazağa Campus. Fine-tuning increased the model's sensitivity to the campus environment while also incorporating user perception. Comparative analyses show that the fine-tuned model exhibited stronger agreement with student assessments and higher prediction accuracy than the untuned version. These results highlight the advantages of combining urban-scale expert knowledge with campus-specific perceptual inputs and demonstrate that transfer learning enhances perceptual relevance at finer spatial scales. Overall, the study highlights the potential of fine-tuning deep learning models to provide more robust, human-centered walkability assessments in campus environments.

Keywords. Walkability, Perception, Deep Learning, Campus Design, CNN

1. Introduction

Walkable cities provide substantial health, environmental, and social benefits, as residents in such areas are more active and experience better well-being (Sallis et al., 2012; WHO, 2010). Streetscapes that integrate pedestrian infrastructure, greenery, and

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 543-552.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

effective traffic management encourage walking and reduce car dependency. Policy frameworks such as the EU Urban Agenda, the European Green Deal, and the United Nations Sustainable Development Goals recognize walkability as a key component of sustainable and inclusive urban development, while rising private vehicle use continues to threaten these goals. Although numerous methods have been proposed to evaluate walkability, no standardized framework exists (Shashank & Schuurman, 2019). As walking is inherently human-oriented, recent studies highlight street-level analyses that integrate objective and perceptual perspectives. Advances in computer vision now bridge these approaches by using street imagery to link measurable spatial features with human perception (Ewing & Handy, 2009; Zavadskas et al., 2019; Li et al., 2022).

University campuses function as micro-urban environments that bring together spaces for learning, mobility, and social interaction. Campus walkability improves comfort, reduces stress, and enhances well-being while supporting sustainable design (Zhang et al., 2024). The degree of campus walkability depends on both physical conditions and user perceptions (Asod-Shekari et al., 2014; Horacek et al., 2018). Recently, AI-based studies have investigated the influence of campus environments on students across dimensions such as aesthetics, safety, depression, and vibrancy (Qin et al., 2025), as well as the restorative effects of visual qualities within campus settings (Ma, Xu & Zhang, 2023). The AI models employed in these studies were trained exclusively on campus-specific data, without incorporating the broader spatial patterns and environmental complexity present in diverse urban settings.

The principal novelty of this study lies in adapting an AI model—initially trained on a large-scale, expert-annotated streetscape image dataset—to incorporate user perceptions within campus environments, thereby enabling more context-sensitive walkability assessment at the campus scale. To achieve this, we employ a convolutional neural network (CNN)–based AI model for assessing urban walkability as a base. The model, developed by Kaynarkaya et al. (2025), is grounded in conceptual frameworks that define walkability as a perceptual construct shaped by the visual characteristics of the built environment (Ewing & Handy, 2013; Krambeck, 2006; Mouratidis, 2021; Huang et al., 2019). Expert-labeled street images, semantically segmented using Segment Anything Model 2 (SAM2) and DeepLabv3, were used to train a CNN model on 5,440 samples for predicting safety, aesthetics, and accessibility scores. Initially, 2,820 Google Street View images (GSV) were collected from the ITU Ayazağa Campus, and then the walkability of the streets in the images were rated by the students in terms of safety, accessibility, and aesthetics. This labeled data was then used to fine-tune the baseline urban walkability model, creating a campus-specific version with an aim to better reflect student perceptions. The fine-tuned models accuracy of walkability predictions exceed the previous model.

We aim to compare the predictions of both the baseline and fine-tuned models to examine how user judgment is reflected in the computational process. This approach highlights the need to integrate user perceptions into AI models on a case-by-case basis, enabling us to move beyond general inferences about urban walkability to reveal perceptual patterns in more defined spatial contexts, such as residential areas, parks, or campuses.

2. Related Works

Walking is the primary mode of movement within university campuses. As significant components of cities, campuses differ from other urban areas in terms of population composition, spatial form, land use, and travel modes (Zhang, Fisher & Wang, 2024). A well-developed built environment greatly enhances walkability and promotes students' walking comfort (Zhang et al., 2024). Therefore, numerous studies have aimed to measure the physical and psychological effects of campus environments on students, emphasizing the role of campus walking conditions in shaping students' perceptions and behaviors (Qin et al., 2025). For instance, Ding et al. (2024) highlighted the positive relationship between increased street connectivity, transportation infrastructure, and both physical activity and mental health in campus environments. Furthermore, factors such as the presence of green spaces, accessibility of facilities, proximity to open areas, and aesthetic quality are clearly associated with students' walking behaviors (Qin et al., 2025; Horacek et al., 2018). Campus greenery, in particular, has been observed to have a positive influence on students' sense of relaxation (Hipp et al., 2016).

While the physical attributes of a street environment can accurately represent its objective quality, they cannot fully capture human perception (Meng et al., 2023). Several studies have used Geographic Information Systems (GIS) and field surveys to assess how campus environments influence students' perceptions. Some of these studies have utilized online platforms such as WalkScore, PlacePulse, and CityScope (Zhang, Fisher & Wang, 2024; Qin et al., 2025; Meng et al., 2023). King et al. (2020) proposed a campus walkability assessment framework that incorporates 24 objective and perceived factors, while Zhang and Mu (2020) developed a synthetic quantitative method combining walking preference surveys with GIS-based and audit-derived objective measurements.

Recent research has also investigated the restorative effects of visual features on campuses using machine learning and street view imagery (Ma, Xu, & Zhang, 2023). Machine learning models and tools such as semantic segmentation have been applied to analyze the relationships between the physical characteristics and visual quality of campus areas (Meng et al., 2023). Generally, previous studies have used streetscape imagery and AI algorithms to assess campus walkability and spatial quality based on student perceptions. The reliance on campus-scale data and user perceptions, while insightful, narrows the analytical scope by excluding comparative insights drawn from more varied urban environments and objective spatial indicators.

In this context, the present study leverages a large-scale dataset labeled by experts who assess urban walkability. This strategy offers two key advantages: (1) employing a model trained on a broader diversity of urban environments enables more accurate and contextually grounded predictions; and (2) expert-driven labeling ensures reliable, comparable, and conceptually coherent annotations essential for rigorous walkability evaluation. Ultimately, this approach not only anchors campus-scale walkability assessment in a stronger conceptual foundation but also supports the development of a generalizable model that can be sensitively aligned with user perceptions through campus-specific recalibration.

3. The Baseline Model

The current 'Urban Walkability Model' (UWM) was trained using the Urban Walkability Dataset (UWD), which integrates both objective and subjective parameters to assess walkability (Kaynarkaya et al., 2025). The measurement criteria were derived from a synthesis of established frameworks proposed by Ewing et al. (2013), Krambeck (2006), and Franz et al. (2005). The evaluation procedure consisted of four stages: General Metric Questions (GMQ), Subjective Metric Questions (SMQ), Labeling, and Walk Hierarchy Questions (WHQ). The GMQ consisted of 34 items rated on a five-point Likert scale addressing spatial features such as "human scale" and "street greenery." The SMQ assessed the relationship between spatial features and emotional response through 10 parameters rated on a bipolar integer scale ranging from −10 to +10. The WHQ involved assessing three core dimensions of walkability—safety, accessibility, and aesthetics—adapted from Alfonzo's (2005) hierarchy of walking needs. Labeling phase, street-level objects were manually identified and categorized into predefined positive and negative classes. All parameters were evaluated perceptually, and numerical outputs generated by the SAM2 were incorporated into the analysis. SAM2, a foundation-level segmentation model trained on a broad range of object classes (e.g., buildings, streets, vegetation, grass, fences, trees, sky), enables coarse segmentation without additional fine-tuning. A total of 5,440 high-quality street images were used for labeling. From these, the Urban Walkability Dataset (UWD) was constructed, consisting of 4,352 training samples, 544 validation samples, and 544 test samples.

The evaluation was carried out by five experts holding at least a bachelor's degree in urban planning or architecture. Prior to the assessment, experts were briefed on the study's scope and participated in a structured seminar outlining the evaluation procedure. Upon seminars, participants were directed sequentially to the SMQ, GMQ, and WHQ assessment interfaces. After scoring all 5,440 images, the model training phase was successfully completed.

The model was then applied to two neighbourhoods in a large Turkish city, and its predictions showed strong agreement with survey data conducted independently by the municipality to which the neighbourhoods belonged. This correspondence between model outputs and perceptual survey responses supports the hypothesis that incorporating perceptual assessments can substantially enhance the reliability and interpretability of deep learning–based walkability analyses.

3.1. CAMPUS WALKABILITY DATASET(CWD)

The case study focuses on Istanbul Technical University's Ayazağa Campus, which spans 1,600,000 m² and accommodates 13 faculties, several institutes, the Central Library, and the Student Social Centre. Hosting more than 25,000 students, the campus also includes student dormitories, a stadium, an Olympic swimming pool, multiple cafés, and an artificial pond, making it a complex and pedestrian-intensive environment. The campus is connected to the city via a metro station offering multiple access points, thereby strengthening overall connectivity and facilitating convenient access for students, staff, and visitors. Along the campus streets, a total of 2,820 street-level images were collected—captured in both directions and from four viewpoints at

intersections (Figure 1).



Figure 1. GSV images from ITU.

The assessments were conducted by 15 senior undergraduate students enrolled in their fourth year at ITU. Images were presented to students through a customized online interface and students were asked to assign a value between -10 and +10 to the walkability of each scene in terms of safety, aesthetics, and accessibility (Figure 2). On average, each student assessed about 190 images.

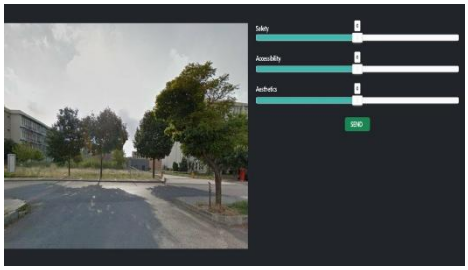


Figure 2. Online interface for CWD.

First, 2,820 images collected from the Ayazağa Campus were evaluated by students to incorporate user perception into the model. Of these labeled images, 2,225 were used to fine-tune the UWM, resulting in an adapted model. Subsequently, 595 student-rated images—never shown to either model during training—were evaluated by both the UWM and the fine-tuned model. The resulting predictions were then compared with student scores to assess model performance (Figure 3).

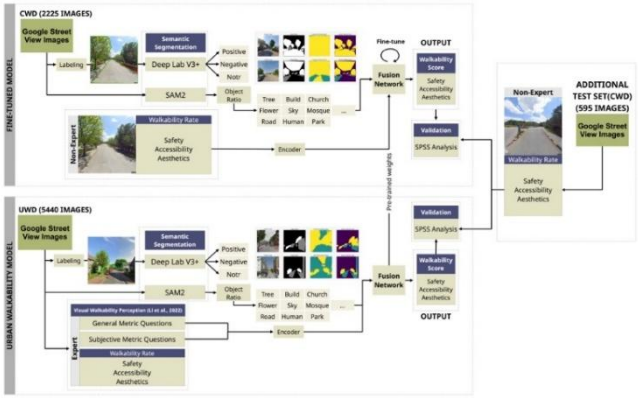


Figure 3. Our Study Framework.

By comparing the outputs of the UWM and the fine-tuned model with the scores of the held-out additionaly test set from the CWD, we investigate whether the fine-tuned model achieves higher accuracy in predicting walkability within the campus context, and how these improvements enhance its ability to represent real-world user experience.

4. Methodology

In this study, the deep learning pipeline design from (Kaynarkaya et al., 2025) is considered as a baseline. This baseline model, takes input street-level RGB images together with a predefined list of textual tags that describe positive, negative, and neutral scene elements. These inputs are processed through a segmentation backbone and a prompt-based SAM2 module (Ravi et al., 2024) to produce dense pixel-level segmentation maps. In this study, unlike the original pipeline, we did not incorporate expert-labeled walkability scores; therefore, the expert scoring branch was excluded from our architecture, as illustrated in Figure 4.

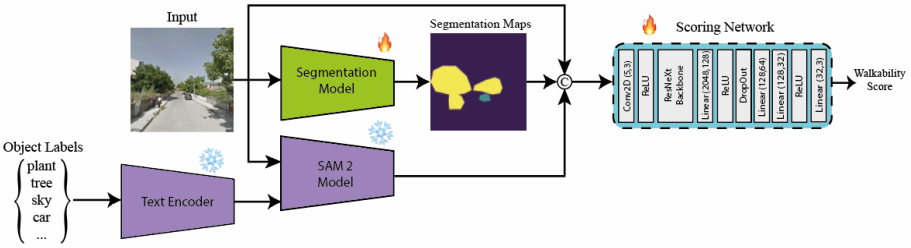


Figure 4. Modified walkability scoring pipeline adapted from Kaynarkaya et al. (2025).

During the fine-tuning process, only the Segmentation Model and the subsequent Scoring Network were updated, using a user-annotated Campus Walkability Dataset (CWD) comprising 2820 labeled street-level images collected from ITU Ayazapa campus.

While the original model relied on expert-labeled walkability scores and diverse urban imagery, we extended it by incorporating a user-labelled dataset of 2,225 campus-level images collected from ITU's Ayazağa Campus. Fine-tuning was required to adapt the baseline model to this new annotative and spatial context, while preserving the generalizable representations learned from the original urban dataset. An additional 595 randomly sampled images from the campus were reserved exclusively for testing to evaluate transferability.

Fine-tuning refers to the process of updating selected layers of a pre-trained model on a new dataset to improve performance in a target domain, while avoiding complete retraining of the network. To adapt the model, we employed a targeted fine-tuning strategy. The core of the pipeline consists of a pre-trained convolutional backbone responsible for extracting mid-level visual features such as spatial enclosure, facade continuity, and vegetation presence. During fine-tuning, early convolutional layers were frozen to preserve generic feature encodings, while deeper layers and task-specific prediction heads were re-initialized and optimized. This approach allows the model to retain fundamental visual priors—edges, gradients, surface patterns—while

adapting higher-level representations to campus-specific visual cues, such as pedestrian-dominant networks, dedicated bike lanes, landscaped walkways, signage systems, and structured green spaces.

All modules except the SAM2 segmentation backbone and the BERT-based text encoder were updated during training. Freezing these components ensured that the general semantic representations (e.g., road, building, vegetation, sky) and text embeddings remained stable, minimizing overfitting in the smaller campus dataset. The trainable branches were optimized using a learning rate of 5×10^{-4} and a batch size of 32, which follows standard transfer-learning heuristics to prevent large parameter shifts when training on reduced-scale datasets. The network was supervised to predict safety, aesthetic, and accessibility scores through a multi-objective regression head optimized via mean squared error (MSE). This objective jointly minimizes error across the three perceptual sub-dimensions, encouraging internal feature sharing and consistent behavior across correlated perceptual domains.

For computational robustness, early stopping and learning-rate scheduling strategies were adopted to prevent overfitting and improve convergence stability. Data augmentation—including random horizontal flipping, slight color jittering, and contrast adjustments—was applied to increase visual diversity and enhance generalization to varying lighting and seasonal conditions present on campus.

Model performance before and after fine-tuning was quantified using Pearson and Spearman correlation coefficients, mean absolute error (MAE), root mean square error (RMSE), and directional bias. Reductions in both MAE and RMSE, along with increased correlation scores, demonstrated that the model successfully adapted to campus-specific spatial and perceptual features. These improvements suggest that the fine-tuned model more effectively captures localized determinants of walkability—such as pavement continuity, sightline openness, and pedestrian activity—that differ in scale and expression from standard urban streetscapes.

5. Discussions

We conducted SPSS analysis comparing the outputs of the UWM and fine-tuned model with the scores from an additional test set from CWD (Table 1). The results indicate that the fine-tuned model achieves higher accuracy in predicting walkability in the campus context, and these improvements enhance its ability to represent user experience.

Student-based evaluations were compared with predictions from the UWM and fine-tuned models. Across all perceptual dimensions—safety, accessibility, and aesthetics—the fine-tuned model showed substantially stronger alignment with student judgments, with Pearson correlations rising from 0.58–0.64 (UWM) to 0.79–0.87, and similar gains in Spearman coefficients. Error measures also improved, with MAE and RMSE decreasing by roughly 1.1–1.3 and 0.8–1.3 points, respectively.

Bias analysis indicated that the UWM tended to overestimate safety and underestimate accessibility and aesthetics, whereas fine-tuning markedly reduced these systematic errors. This suggests that incorporating context-specific perceptual labels improves model calibration.

Table 1. Comparison of UWM and the fine-tuned model with additional test set from CWD.

Dimension	Model	Pearson r	Spearman ρ	MAE	RMSE	Bias
Safety	UWM	0.64	0.63	3,12	3,95	+2.10
Safety	Fine-tuned	0.87	0.85	1,98	2,68	+1.20
Accessibility	UWM	0.58	0.57	3,51	4,11	−3.20
Accessibility	Fine-tuned	0.79	0.79	2,74	3,31	−1.90
Aesthetics	UWM	0.61	0.60	3,48	4,02	−3.40
Aesthetics	Fine-tuned	0.81	0.79	2,68	3,2	−2.00

Overall, the results show that fine-tuning increases perceptual transferability, allowing predictions to more closely reflect student evaluations. This supports previous findings that safety perceptions are tied to readily identifiable spatial cues, while accessibility and aesthetics rely more on higher-order experiential factors. Incorporating user perception into refinement therefore offers a valuable strategy for enhancing walkability prediction in complex environments such as university campuses.

6. Conclusions

This study compares the output of an expert-rated model based on a large, heterogeneous urban dataset with a context-specific, fine-tuned model based on user evaluations. Both models simulate perceptual responses to built environment stimuli (safety, accessibility, and aesthetics) using expert- and additional user-rated visual data. These findings demonstrate that fine-tuning improves the model's ability to predict all three perceptual dimensions. Although further optimization is needed to better capture perceived accessibility and aesthetics, integrating user-generated perceptual data through fine-tuning marks an important step toward more human-centric computational assessment of walkability.

This study has several limitations, including the limited number of campus images and the absence of nighttime evaluations. Future work will expand the CWD to train a new model and compare it with the current fine-tuned version, while also exploring broader case-based variations such as incorporating additional campuses—including inner-city settings—and examining labeling differences across user groups (e.g., first-year vs. senior students), as well as assessing nighttime walkability.

References

Alfonzo, M. A. (2005). To walk or not to walk? The hierarchy of walking needs. Environment and Behavior, 37(6), 808–836. https://doi.org/10.1177/0013916504274016

Asadi-Shekari, Z., Moeinaddini, M., & Zaly Shah, M. (2013). Non-motorised Level of Service: Addressing Challenges in Pedestrian and Bicycle Level of Service. Transport Reviews, 33(2), 166–194. https://doi.org/10.1080/01441647.2013.775613

European Commission: Directorate-General for Communication and Leyen, U. (2019). Political guidelines for the next European commission 2019–2024: opening statement in

the European parliament plenary session 16 2019; speech in the European parliament plenary session 27 2019. Publications Office of the European Union, Available at: https://data.europa.eu/doi/10.2775/101756.

Ewing, R., Clemente, O., Neckerman, K. M., Purciel-Hill, M., Quinn, J. W., & Rundle, A. (2013). Measuring urban design: metrics for livable places: vol. 200, Washington, DC: Island Press.

Ewing, R., & Handy, S. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. Journal of Urban Design, 14(1), 65–84.

Ewing, R., Handy, S., Brownson, R. C., Clemente, O., & Winston, E. (2006). Identifying and measuring urban design qualities related to walkability. Journal of Physical Activity and Health, 3(s1), S223–S240.

Franz, G., Von Der Heyde, M., & Bülthoff, H. H. (2005). An empirical approach to the experience of architectural space in virtual reality—exploring relations between features and affective appraisals of rectangular indoor spaces. Automation in Construction, 14(2), 165–172.

Hipp, J. A., Gulwadi, G. B., Alves, S., & Sequeira, S. (2016). The relationship between perceived greenness and perceived restorativeness of university campuses and student-reported quality of life. Environment and Behavior, 48(10), 1292–1308. https://doi.org/10.1177/0013916515598200

Horacek, T. M., Dede Yildirim, E., Kattelmann, K., Brown, O., Byrd-Bredbenner, C., Colby, S., Greene, G., Hoerr, S., Kidd, T., Koenings, M. M., Morrell, J., Olfert, M. D., Phillips, B., Shelnutt, K., & White, A. (2018). Path Analysis of Campus Walkability/Bikeability and College Students' Physical Activity Attitudes, Behaviors, and Body Mass Index. American journal of health promotion: AJHP, 32(3), 578–586. https://doi.org/10.1177/0890117116666357

Huang, R., Moudon, A. V., Zhou, C., & Saelens, B. E. (2019). Higher residential and employment densities are associated with more objectively measured walking in the home neighborhood. Journal of Transport and Health, 12, 142–151.

Kaynarkaya, S., Çekmiş, A., Çetin, İ., Şahin, Y. H., & Ünal, G. (2025). Assessing walkability with deep CNNs by integrating objective and subjective urban qualities: The case of “Cittaslow” neighbourhoods. Sustainable Cities and Society, 133, 106863. https://doi.org/10.1016/j.scs.2025.106863

King, S. B., Kaczynski, A. T., Knight Wilt, J., & Stowe, E. W. (2020). Walkability 101: A multi-method assessment of the walkability at a university campus. SAGE Open, 10(2). https://doi.org/10.1177/2158244020917954

Krambeck, H. V. (2006). The global walkability index (Doctoral dissertation), Massachusetts Institute of Technology.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a deep learning approach. Urban Forestry & Urban Greening, 14(3), 675–685. https://doi.org/10.1016/j.ufug.2015.06.006

Li, Z., Liu, H., Zhang, Y., & Zhang, W. (2022). Measuring visual walkability perception using panoramic street view images, virtual reality, and deep learning. Sustainable Cities and Society, 86, 104140. https://doi.org/10.1016/j.scs.2022.104140

Ma, H., Xu, Q., & Zhang, Y. (2023). High or low? Exploring the restorative effects of visual levels on campus spaces using machine learning and street view imagery. Urban Forestry & Urban Greening, 88, 128087.

Meng, Y., Li, Q., Ji, X., Yu, Y., Yue, D., Gan, M., ... & Fukuda, H. (2023). Research on campus space features and visual quality based on street view images: A case study on the chongshan campus of liaoning university. Buildings, 13(5), 1332.

Moreno, C., Allam, Z., Chabaud, D., Gall, C., & Pratlong, F. (2021). Introducing the “15-minute city”: Sustainability, resilience and place identity in future post-pandemic cities. Smart Cities, 4(1), 93–111. https://doi.org/10.3390/smartcities4010006

Mouratidis, K. (2021). How COVID-19 reshaped quality of life in cities: A synthesis and implications for urban planning. Land use policy, 111, 105772.
Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore—Predicting the perceived safety of one million streetscapes. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 779–785).
OpenAI. (2025). ChatGPT Model: GPT-5 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/
Qin, Y., Wu, X., Yu, T., & Jiang, S. (2025). Enhancing student-centered walking environments on university campuses through street view imagery and machine learning. PloS one, 20(4), e0321028.
Sallis, J. F., McKenzie, T. L., Beets, M. W., Beighle, A., Erwin, H., & Lee, S. (2012). Physical education's role in public health: steps forward and backward over 20 years and HOPE for the future. Research Quarterly for Exercise and Sport, 83(2), 125–135.
Shashank, A., & Schuurman, N. (2019). Unpacking walkability indices and their inherent assumptions Health and Place, vol 55, 145, 154.
World Health Organization (2010). The world health report 2010—health systems financing: the path to universal coverage. Geneva: WHO.
Zavadskas, E. K., Bausys, R., & Mazonaviciute, I. (2019). Safety evaluation methodology of urban public parks by multi-criteria decision making. Landscape and Urban Planning, 189, 372–381.
Zhang, Z., Fisher, T., & Wang, H. (2024). Campus environmental quality and streetscape features related to walking activity. Journal of Asian Architecture and Building Engineering, 23(1), 405-423.
Zhang, X., & Mu, L. (2020). The perceived importance and objective measurement of walkability in the built environment rating. Environment and planning B: urban analytics and City science, 47(9), 1655-1671.
Zhang, Z., Sun, T., Fisher, T., & Wang, H. (2024). The relationships between the campus built environment and walking activity. Scientific reports, 14(1), 20330.

Prompts to Perception

A Promptable Segmentation Approach to Transparency and Safety

Zheng Fang[1] and Xuechen Li[2]
[1,2]*Harvard University.*
[1]*zheng_fang@gsd.harvard.edu, ORCID: 0009-0002-0746-7618*
[2]*xuechen_li@gsd.harvard.edu, ORCID: 0009-0002-3971-2147*

Abstract. Quantifying how façade level urban design qualities shape perceived safety from street-level imagery remains methodologically constrained. We propose an image-native framework that uses promptable segmentation models to quantify transparency directly as a window-to-wall ratio (WWR) from Google Street View and link it to crowdsourced safety judgments. Implemented for roughly 50,000 street locations in Boston, the framework recovers consistent WWR measures and learns an interpretable relationship between transparency and perceived safety, capturing variation in perceived outcomes. Relative to prior association studies based on fixed semantic parsers or coarse scene classes, this prompt-driven segmentation produces high-quality façade-opening masks and broadens the set of design-relevant elements that can be examined without bespoke training data. The resulting methodology demonstrates that WWR can be operationalized as a citywide indicator from existing imagery and used to assess where transparent façades are associated with higher perceived safety. More broadly, it offers a transferable template for using state-of-the-art vision models to support conceptually precise, scalable association studies between urban qualities and perception.

Keywords. Urban Transparency, Perceived Safety, Urban-scale Mapping, GeoAI, Image-based Perception Modelling

1. Introduction

1.1. CHALLENGES

Façade transparency—visual permeability between building interiors and the street—has long been linked to perceived safety and informal surveillance in urban design theory (Ewing and Handy, 2009). Street View Imagery (SVI) and GeoAI have made it possible to study perceived safety at scale, but prevailing approaches rely on fixed semantic parsers or coarse scene-level descriptors (such as greenness, sky share, or generic land-use classes). These pipelines treat façades as undifferentiated backdrops, rather than measuring transparency as an explicit, façade-level ratio. As a result,

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 553-562.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

transparency's theorized role remains weakly reflected: there is no image-native, scalable indicator that captures windowed frontage directly from SVI and links it to systematic variation in perceived safety.

1.2. CONTRIBUTION AND APPROACH

Given these challenges, we introduce a promptable, window-centric framework for urban association studies and use transparency–safety relations as a benchmark case. Using Grounded-SAM on SVI, we derive consistent WWR and link them to perceived safety scores obtained from pairwise comparisons. This pipeline (Figure 1) shows that WWR can be implemented as a citywide indicator directly from imagery, and that prompt-based segmentation supports more conceptually aligned, fine-grained association models than prior approaches. More generally, it establishes a transferable protocol for applying state-of-the-art vision models to test theory-driven visual hypotheses about the built environment with higher precision and scalability.

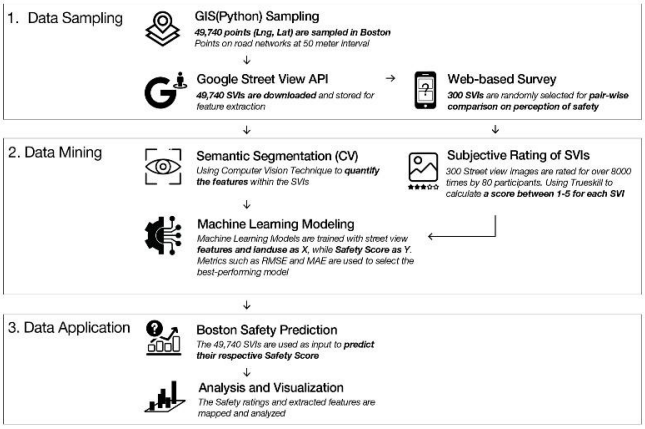


Figure 1. Framework

2. Related Works

2.1. URBAN DESIGN QUALITIES AND TRANSPARENCY

Urban design theory provides the conceptual foundation for transparency at the street edge. From Jacobs' (1961) "eyes on the street" to more formal operationalisations by Ewing and Handy (2009), visibility between interiors and the sidewalk has been linked to perceived safety and informal surveillance. Ewing and Handy (2009) identify a set of perceptual urban design qualities including imageability, enclosure, human scale, and transparency, that shape how pedestrians experience streets and relate these qualities to perceived walkability and safety. In their framework, transparency describes the degree of visual permeability between the sidewalk and adjacent buildings, conditioned by windows, doors, and other openings in the street wall. This work makes transparency explicit as an index that can be quantified and introduces field survey and rating protocol. However, these protocols rely on manual measurement, require substantial effort, and rarely achieve metropolitan coverage,

which makes them difficult to integrate with large-scale analysis of urban form. As a result, transparency remains conceptually significant but empirically underrepresented in city-scale studies.

2.2. IMAGE-BASED URBAN ANALYSIS AND PERCEPTION MODELING

Recent advances in computer vision and GeoAI have shifted urban analysis from manual measurement toward image-based methods. SVI has been widely used to model perceived safety and related perceptual attributes in urban environments. Naik et al. (2014) introduces StreetScore, a computer vision model that uses SVI and image features with support vector regression to predict the perceived safety of streetscapes. Dubey et al. (2016) extend this paradigm with Place Pulse 2.0, training deep neural network models to infer multiple perceptual attributes, including safety, across diverse cities. De Nadai et al. (2016) adopt large-scale association frameworks that connect safety scores to urban dynamics and use occlusion analysis to reveal that greenery and street-facing windows contribute to a safer appearance.

Earlier SVI-based perception models typically used either hand-crafted scene features or deep features fed into classical supervised models. When semantic segmentation was incorporated, widely used label spaces for street scenes (e.g., "building", "wall") tended to aggregate fenestration and opaque façade elements into a single class, so window openings remained visually entangled with the rest of the street wall. In practice, this meant that transparency could only be inferred indirectly rather than computed as an explicit façade-level measure. More recent work trains Place Pulse–based deep models to map multiple street-level perceptions at city scale without decomposing façade elements (Wei et al., 2022), and couple panoptic segmentation with interpretable attribution to quantify how object and façade cues shape perceived safety (Liu et al., 2023a). Yet in these settings, window openings are not treated as a dedicated category, and transparency is still not defined explicitly.

2.3. PROMPTABLE SEGMENTATION FOR TRANSPARENCY

Recent vision foundation models enable a more direct solution. Segment Anything (SAM) introduces a promptable segmentation paradigm capable of transferring across domains (Kirillov et al., 2023), while Grounding DINO extends open-set object detection to text-specified categories (Liu et al., 2023b). Grounded-SAM integrates these components into an open-vocabulary segmentation framework that can generate high-quality instance masks for arbitrary text prompts such as "window", "shopfront", or "glass door" (Ren et al., 2024). For street view-based urban research, this capability allows measurement to follow theory by extracting façade openings directly from SVI without a bespoke, tightly scoped "window" dataset or a fixed label space.

Our method follows the general SVI based perception modelling paradigm outlined above but shifts the focus from latent scene features to WWR. We apply Grounded-SAM to SVI, derive WWR for each streetscape, and relate this metric to city-specific perceived safety scores collected for our case study. Details of the pipeline are presented in the following section.

3. Methodology

Our methodology for the present study is structured as follows:

- **SVI Sampling and Retrieval.** We sample the Boston Street network at regular intervals and query the Google Street View API with fixed pedestrian-view parameters, producing a citywide set of images for analysis.
- **Open-vocabulary Segmentation.** We use Grounded-SAM to segment façade openings and PSPNet to extract sky/tree/road context, producing window-based transparency metrics and view indices for each image.
- **Perception Scoring.** We select a random subset of the SVIs for a web-based perception survey, where respondents view two images at a time and answer "Which looks safer?". We aggregate the outcomes into a continuous safety score, serving as the ground truth in later modelling stages.
- **Modelling and Validation.** We relate WWR and context features and land use to perceived safety by fitting 10 regression models and selecting the best via cross-validated held-out R^2. SHAP-based attributions are used to interpret how individual features contribute to the model's safety predictions.

3.1. SVI SAMPLING AND RETRIEVAL.

We began with Boston's street network (Figure 2) from OpenStreetMap in QGIS and placed points every 50 m along the centrelines to balance coverage and redundancy at city scale (Fan et al., 2025). Each point (lat–lng) fed a small Python script that called the Google Street View API with standardized pedestrian settings: pitch = 0°, field of view = 90°, image size = 640×400, and a heading parallel to the road bearing. This favours a walker's line of sight along the corridor even if it sacrifices some façade-frontal detail.

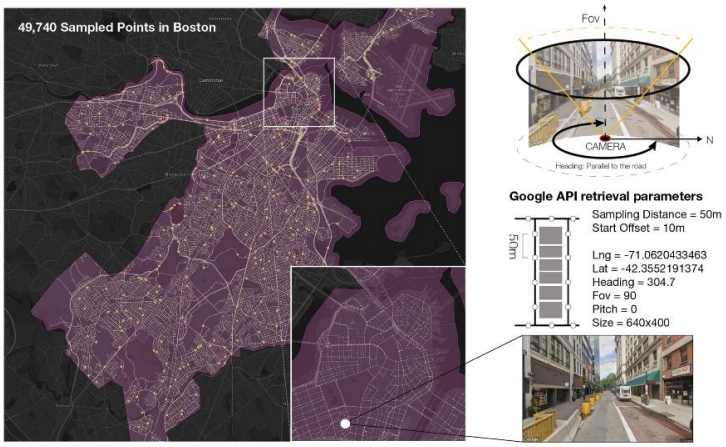


Figure 2. Sampling Method

The initial download was then lightly cleaned. We removed failed geocodes, blank frames, extreme under-exposure, and obvious occlusions, and deduplicating near-identical shots; we also kept only views oriented along the street rather than into lots or building courtyards. In parallel, we spatially joined each sample point to a parcel-based land use layer to assign a land use category (e.g., residential, commercial, industrial); these labels later enter the model as non-visual independent variables. The resulting dataset is citywide, standardized, and suitable as a consistent base for downstream segmentation and perception modelling.

3.2. SEMANTIC SEGMENTATION

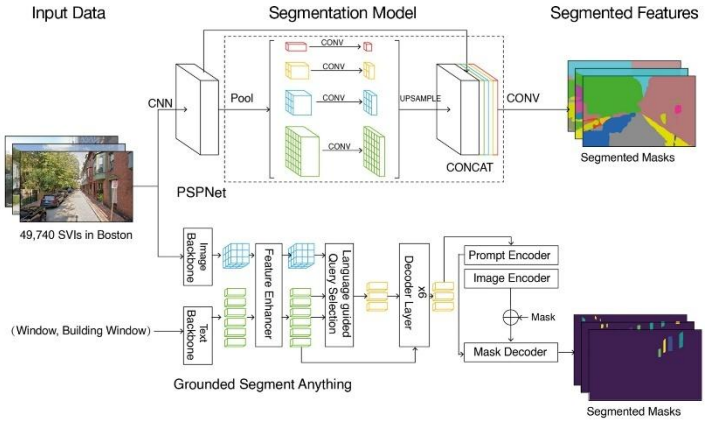


Figure 3. Segmentation Architecture

We run two segmentation branches on all 49,740 SVIs to capture both streetscape composition and façade openings (Figure 3).

3.2.1. PSPNet Semantic Segmentation

First, each image is parsed with PSPNet to obtain per-pixel masks for a list of 30 prominent features in urban scenes, including Building, sky, tree, road, grass, sidewalk, and related classes. PSPNet's pyramid pooling architecture captures global context (Zhao et al., 2017) and is widely used in streetscape "view index" work, giving stable masks for downstream tasks. We retain raw masks as well as pixel counts of all classes.

3.2.2. Open-Vocabulary Segmentation

Meanwhile, we extract façade openings with Grounded-SAM. Each image and a short prompt set (e.g., "window", "shopfront", "glass door") are encoded by separate vision and language backbones. Grounding DINO aligns prompt embeddings with image features through cross-modal attention and returns text-conditioned bounding boxes with phrase scores. We pass these boxes to SAM together with the original image; SAM uses the box as a spatial prompt over a dense image embedding to produce pixel-accurate instance masks. We merge overlapping instances and remove small artifacts,

yielding a single openings mask per image.

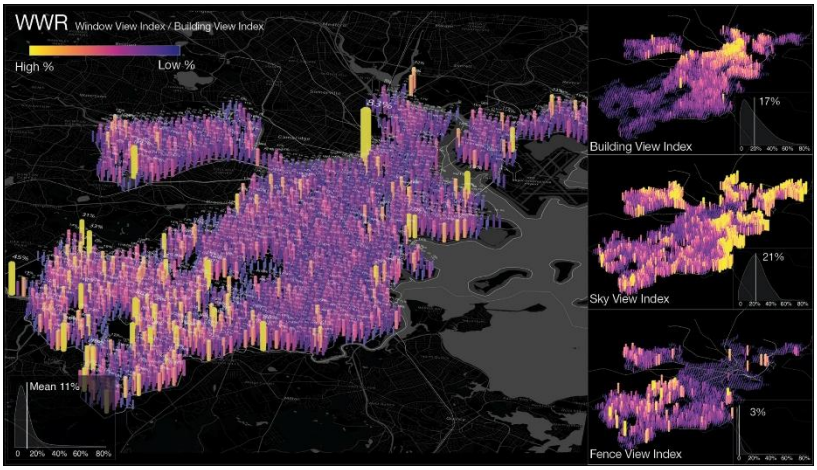


Figure 4. View Index Map

3.2.3. Masks to Features

From these masks we derive both transparency and contextual predictors. For each SVI, we compute WWR as opening pixels within the building region divided by total building pixels, adapting the standard window-to-wall convention to image space (e.g., Szcześniak et al., 2022; Suppa et al., 2025). Context view indices follow the pixel-fraction convention in SVI-based streetscape metrics (Li et al., 2015): for each semantic class, we report its pixel share over the full image. Each SVI thus yields one row in a tabular dataset where columns are theory-relevant ratios, ready for subsequent modelling.

3.3. PERCEPTION SCORING

To obtain perceptual ground truth on safety, we conduct a web-based pairwise comparison survey on a random subset of 300 SVIs, a sample size that yields roughly ten observations per predictor and follows common regression heuristics for limiting overfitting (Babyak, 2004; Harrell, 2015). Eighty respondents each completed 100 "Which looks safer?" judgements on randomized image pairs with left–right positions shuffled, producing 8,000 comparisons and roughly 50 pairwise appearances per image. We converted these outcomes into continuous safety scores using a TrueSkill-based ranking model (Herbrich et al., 2006), which infers a score per image from repeated matchups; prior work suggests that a few dozen comparisons per item are sufficient for stable rankings in this setup (Dubey et al., 2016). The resulting TrueSkill means were linearly rescaled to a 1–5 index for interpretability and used as the dependent variable in later modelling.

3.4. MODELLING AND VALIDATION

We model perceived safety scores as a function of window transparency, contextual

visual features, and land use. Using the 300 scored SVIs, we construct a feature matrix comprising WWR, pixel shares for each class, and one-hot–encoded parcel land-use categories. On this training set, we benchmark a list of supervised regression models from linear, tree-based, and ensemble methods, and select the best-performing specification based on cross-validated predictive accuracy. The selected model is then refit on the full scored sample and applied to all SVIs to generate citywide predictions of perceived safety. We then use SHAP to summarize feature attribution.

3.5. EXPERIMENTS

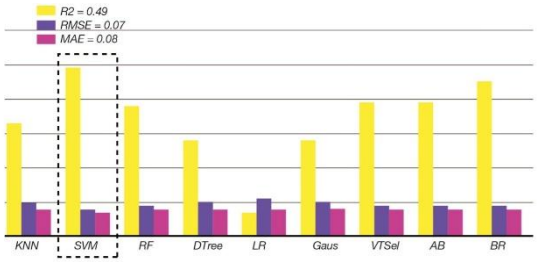


Figure 5. Model Performance

For the experimental models, we treat perceived safety as a continuous outcome derived from the TrueSkill scores. Let y_i denote the safety rating for image i on the normalized 0-1 scale. We model using $\log_{10}(y_i)$ to reduce skew and stabilize variance, but report performance on the original scale for better interpretability. The training dataset has 300 observations and 31 predictors, grouped into three categories:

- **Transparency and façade openings:** window-to-wall ratio (WWR) and an overall window-to-image ratio.
- **Semantic view indices:** PSPNet-derived pixel shares for major classes, including building, sky, tree, road, grass, sidewalk, vegetation, cars, signage, walls, fences, street furniture, and people.
- **Land use context**: parcel-based land use categories including commercial, industrial, single- and multi-family residential, mixed-use types encoded as one-hot variables.

We randomly split the 300 scored SVIs into 80% training and 20% testing sets and keep this split fixed across all experiments. On the training set, we benchmark a suite of linear and non-linear regressors (figure 5) and run light hyperparameter tuning via small grid searches, selecting settings that maximise cross-validated R^2. All Models train within seconds on a single workstation with an NVIDIA GPU. As summarised in Fig.5, a Support Vector Regressor (SVR) offers the best balance between test-set R^2, RMSE, and MAE; we therefore refit this model on all 300 scored images and use it to predict perceived safety for the full corpus of 49,740 SVIs, providing the basis for the citywide safety analysis and the SHAP-based feature attributions discussed in the following sections.

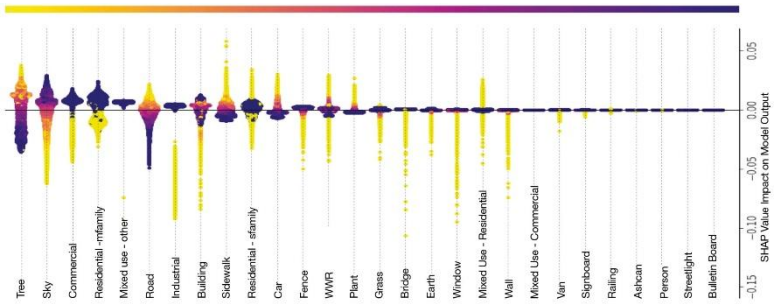


Figure 6. SHAP attribution

3.6. RESULTS

Our best specification achieves $R^2 \approx 0.50$ on held-out data, a practical signal for a single-frame perceptual outcome. SHAP attributions confirm that trees and sky are dominant cues in predicted safety, while WWR contributes a smaller but directionally consistent signal across the scored sample. Importantly, this transparency term is derived from promptable masks rather than a bespoke window dataset, demonstrating that open-vocabulary segmentation can support conceptually precise, façade-level variables in scalable perception models.

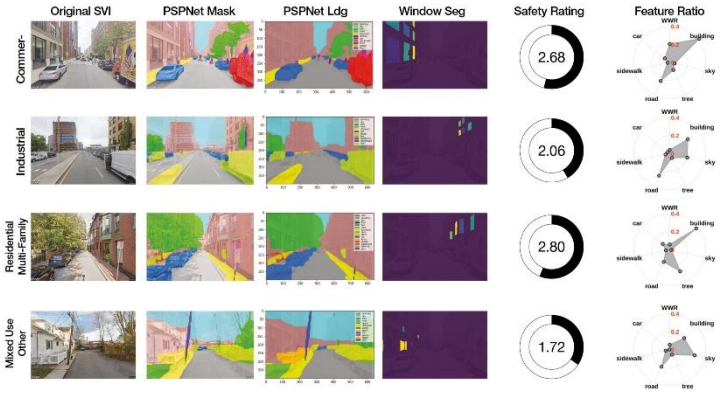


Figure 7. Sample Comparison

3.7. DISCUSSION

These results should be read as a validation of measurement rather than a claim of causal influence. WWR is estimated from automated masks, so image conditions such as glare, reflections, occlusions, and low illumination can increase measurement noise and attenuate the observed association with perceived safety.

A useful extension is to relate segmentation fidelity to perception-model behaviour: for example, auditing a stratified subset of masks and testing whether lower-quality segments correspond to larger prediction errors or higher uncertainty in safety rankings. This would clarify how transparency cues interact with image quality and survey variance in street-level perception.

More broadly, the methodological value lies in converting theory-specified façade constructs into measurable ratios without fixing the analysis to a closed label set. By swapping or expanding prompts, the same workflow can interrogate other openings and frontage elements and support higher-resolution association studies across cities using existing street imagery.

4. Conclusion

In this study, we introduce a window-centric pipeline for quantifying street-level transparency and relating it to perceived safety at city scale. Built on standardized Google Street View sampling, open-vocabulary segmentation (Grounded-SAM), PSPNet-based view indices, and an SVR model with SHAP interpretation, the workflow turns façade transparency into an image-native, reproducible metric that planners and policy makers can inspect in corridor planning, code calibration, and design review. Notably, the same framework can be adapted to a broader set of perceptual and design questions. By swapping or expanding the prompt set, the workflow can target other fine-grain elements in street views, or features related to operational programming that are difficult to capture with standard object detection models and embed perception-aware indicators into standard spatial analysis and planning workflows.

Our approach has its limitations. The perception sample is modest (300 SVIs), the explained variance is moderate, and all results are correlational rather than causal. The analysis is restricted to a single city and single-frame daytime imagery; unobserved factors such as crime, socioeconomic context, program mix, and time-of-day activity are not explicitly included. Open-vocabulary segmentation remains sensitive to reflections, occlusions, and seasonal change, and WWR inevitably captures only one facet of transparency among many, indicating that further prompt and model engineering could improve the accuracy.

Despite these constraints, the study offers a template for using state-of-the-art promptable computer-vision models to translate design qualities into quantitative variables that can be mapped, modelled, and debated in urban analysis and design.

Attribution Statement

ChatGPT (OpenAI, 2025) was used to help improve the clarity and grammar of the manuscript. Microsoft Word's built-in spelling checker was used to correct typographical errors. All conceptual content, and conclusions were developed and verified by the authors.

References

Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66(3), 411–421.

De Nadai, M., Staiano, J., Larcher, R., Sebe, N., Quercia, D., & Lepri, B. (2016). The death and life of great Italian cities: A mobile phone data perspective. *EPJ Data Science*, 5(1), 12.

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. *In Computer Vision – ECCV 2016 Workshops* (pp. 196–212). Springer.

Ewing, R., & Handy, S. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban Design*, 14(1), 65–84.

Fan, Z., Feng, C.-C., & Biljecki, F. (2025). Coverage and bias of street view imagery in mapping the urban environment. *Computers, Environment and Urban Systems*, 117, 102253.

Harrell, F. E. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2nd ed.). Springer.

Herbrich, R., Minka, T., & Graepel, T. (2006). Trueskill™: A Bayesian skill rating system. *In Advances in Neural Information Processing Systems*, 19 (pp. 569–576).

Jacobs, J. (1961). *The death and life of great American cities*. Random House.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., … Girshick, R. (2023). Segment anything. *In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W. (2015). Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, 14(3), 675–685.

Liu, Y., Chen, M., Wang, M., Huang, J., Thomas, F., Rahimi, K., & Mamouei, M. (2023a). An interpretable machine learning framework for measuring urban perceptions from panoramic street view images. iScience, 26(3), 106132.

Liu, S., Li, F., Zhang, X., Zhang, Y., Niu, Z., Zhang, J., … Zhang, L. (2023b). *Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection*. arXiv preprint.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *In Advances in Neural Information Processing Systems*, 30 (pp. 4765–4774).

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. A. (2014). Streetscore – Predicting the perceived safety of one million streetscapes. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 793–799).

Ren, T., Liu, S., Zeng, A., Lin, J., Li, K., Cao, H., Chen, J., Huang, X., Chen, Y., Yan, F., Zeng, Z., Zhang, H., Li, F., Yang, J., Li, H., Jiang, Q., & Zhang, L. (2024). *Grounded SAM: Assembling open-world models for diverse visual tasks*. arXiv preprint arXiv:2401.14159.

Suppa, A. R., Aliberti, A., Bottero, M. C., & Corrado, V. (2025). Detecting window-to-wall ratio for urban-scale building simulations using deep learning with street view imagery and an automatic classification algorithm. *Building Simulation*, 18, 2175–2199.

Szcześniak, J. T., Ang, Y. Q., Letellier-Duchesne, S., & Reinhart, C. F. (2022). A method for using street view imagery to auto-extract window-to-wall ratios and its relevance for urban-level daylighting and energy simulations. *Building and Environment*, 207, 108108.

Wei, J., Yue, W., Li, M., & Gao, J. (2022). Mapping human perception of urban landscape from street-view images: A deep-learning approach. **International Journal of Applied Earth Observation and Geoinformation, 112***, 102886.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2881–2890).

Reasoning with Generative Vision-Language Models to Enhance Low-quality Street-view Imagery for More Accurate and Interpretable Urban Building Analysis

Haoyang He[1*], Shuyang Li[2], Wanyu Pei[3], Rudi Stouffs[4], Yu Ye[5*]
[1,5]*College of Architecture and Urban Planning, Tongji University.*
[2,3,4]*National University of Singapore.*
[1]*2330091@tongji.edu.cn, 0009-0001-2524-4760*
[2]*shuyangli@u.nus.edu, 0000-0002-8780-7469*
[3]*peiwanyu@u.nus.edu, 0000-0003-2822-6418*
[4]*stouffs@nus.edu.sg, 0000-0002-4200-5833*
[5]*yye@tongji.edu.cn, 0000-0003-3402-5786*

Abstract. Street-view imagery (SVI) provides human-perspective data for urban analysis, yet low-quality images caused by occlusion, distortion, and lighting variation undermine the reliability of visual measurements. This study proposes a controllable reasoning framework that integrates generative Vision–Language Models (VLMs) for semantic inference and content restoration. Four inference strategies: direct and integrated captioning, together with direct and caption-guided regeneration, were tested using Baidu Street View Images of Shanghai. Validation results demonstrate that combining textual and visual prompts enhances both interpretability and informational accuracy in low-quality SVI analysis compared with conventional computer-vision preprocessing. By restoring façade details and improving analytical stability, the framework supports fine-grained urban analytics and design applications at low cost, offering a scalable and interpretable pathway for integrating VLMs into human-centred urban practices.

Keywords. Street-view imagery (SVI), Urban regeneration, Vision-Language Model (VLM), Image restoration, Interpretable inference

1. Introduction

Streets constitute the most ubiquitous form of urban public space, where building façades collectively shape the perceptual and physical character of the local sense of place (Chen et al., 2024). Before conducting practices on specific architectural façades, it is important for planners and stakeholders to identify representative street segments. For street building measurement, Street-view imagery (SVI) directly enables diverse urban analyses and design applications (Fan et al., 2023) related to façade colour (Yang et al., 2024), architectural style and material (Chen et al., 2024), and building function (Liang et al., 2025) These studies primarily rely on semantic segmentation, image encoding, and supervised classification techniques. However, variations in SVI

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 563-572.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

quality—stemming from differences in image providers, acquisition periods, lighting conditions, and data collection protocols—can substantially affect the reliability of visual urban measurements (Wang et al., 2025).

Existing techniques based on computer vision algorithm have been widely applied to address these issues, especially in large-scale analyses through proportional control of low-quality images (Fang et al., 2024). However, such methods primarily rely on the existing visual attributes, retaining most of the original content even if it's misleading. When confronted with perspective deformation or visual occlusion, such approaches often fail to recover the underlying scene structure. Consequently, both the semantic segmentation outcomes and the perceptual qualities of the image remain in bad condition. To mitigate these limitations, recent studies have begun employing generative Vision–Language Models (VLMs) for semantic inference and content supplementation. VLMs introduce reasoning-based visual understanding and have demonstrated promising performance in urban scene analysis, such as directly inferring semantic content from images (Liang et al., 2025), and substituting for expert evaluations (Xiao & Tang, 2025). However, as language-encoded models, VLMs inherently lack interpretability for quantitative visual interpretation, and large-scale inference tasks that depend on VLMs' API services often suffer from instability and limited reproducibility. Therefore, the principal strength of VLMs remain more reliable to make sub-symbolic reasoning rather than direct numerical assessment of visual data.

Building on this, we propose a controllable framework that integrates VLMs into the preprocessing of street-view imagery (SVI) for detailed street-level measurement and design applications. First, we evaluate VLM capacity to infer semantic content from low-quality SVIs through text-based reasoning and integrated image restoration. Second, we explore VLM's generative potential in practical urban renewal contexts, such as façade regeneration and streetscape element redesign. The results demonstrate that VLM-based framework enables more accurate and interpretable restoration of degraded SVIs, enhancing both quantitative measurement and design-oriented decision-making in urban renewal practices with fine spatial resolution.

This paper is organized as follows: Section 2 details the sample selection process, methodological design, and four VLM-based strategies developed for quantitative reasoning. Section 3 presents the validation of the proposed framework, the application of the model in street-level redesign, and potential limitations. Section 5 concludes the study and summarizes the key findings.

2. Method and experiment

2.1. OVERALL FRAMEWORK

Instead of directly generating quantitative values or assigning scores from low-quality images, this study employs existing information to refine low-quality SVIs, transforming them into higher fidelity imageries more suitable for computer vision analysis. The overall framework is illustrated in Figure 1. Specifically, we first apply semantic segmentation to identify potentially low-quality SVI points and map them to locate the poorest-quality sample regions. Next, we construct a spatial adjacency

network for location-based retrieval of SVI points, enabling the selection and assignment of information from nearby high-quality SVIs. Subsequently, we experimented with four VLM-based strategies for quantitative reasoning, including direct visual inference, integrated visual inference, direct image regeneration, and caption-guided regeneration. We then conduct expert-supervised scoring to validate these methods by comparing the efficiency of the real condition with the modified measurements. Furthermore, we explore the potential of generative VLM to support façade renewal from measurement, restoration to creative design.

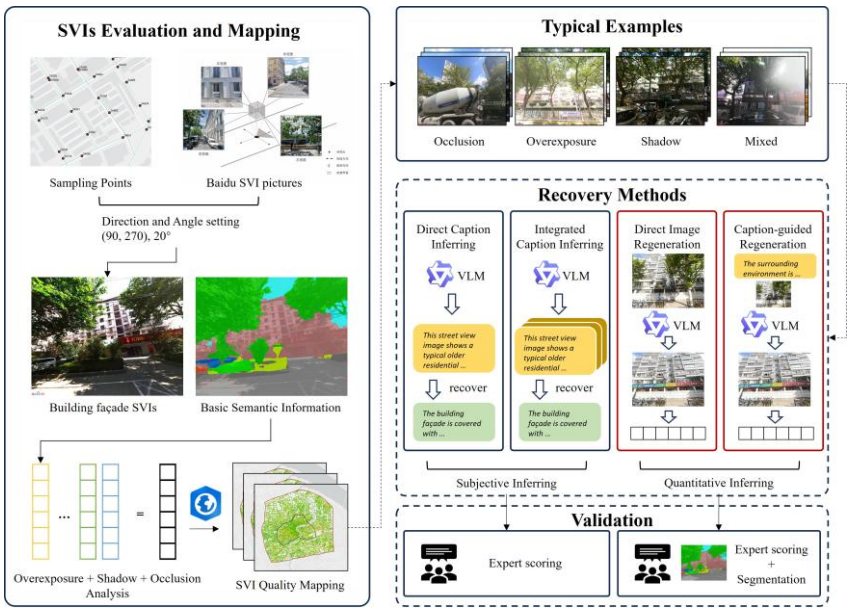


Figure 1. Method Roadmap

2.2. SAMPLES SELECTION

2.2.1. SVI collection and adjacent network building

This study employs street-view imageries from Baidu Maps as the primary dataset, as its image quality typically varies due to differences in weather conditions and capture times. To capture more comprehensive information on building façades, we only collected images from left view corresponds to 270° and the right view to 90° with an additional upward tilt of 20° on both sides of the street.

Since each sampling point is associated with a specific SVI_ID, adjacent images can be accessed through the spatial relationships of the sampling points. Specifically, for each image, we first retrieve its corresponding sampling point and identify nearby points with associated SVIs. These SVIs are then filtered via semantic analysis to support subsequent inference. Based on this process, each SVI is linked to multiple neighbouring SVIs within an adjacency network, enabling the integration of contextual semantic information from its surroundings.

In total, 55,351 sampling points were established within Shanghai's Outer Ring area, yielding 110,156 street-view images. Each image measures 1024 × 800 pixels at a resolution of 96 dpi, forming the basis for subsequent feature extraction and aggregation.

2.2.2. Low-quality SVI samples

The low-quality street-view images (SVIs) identified in this study are divided into three categories, including occlusion, overexposure, and shadow. Some of these deficiencies can be detected through computer vision–based semantic segmentation, while others can be identified via colour analysis. To ensure the representativeness of the samples, a human–machine collaborative approach was employed. First, based on existing semantic segmentation results, images with a relatively high proportion of building pixels but excessive coverage of vehicles or trees were labelled as potential cases of occlusion. Similarly, images dominated by black (dark) or white (light) tones through global colour proportion analysis were labelled for illumination-related issues. After this, we manually reviewed and validated the results, yielding a final dataset of 150 (3*50) images, comprising three primary categories including Occlusion, Overexposure and Shadow. The typical classifications and final selected samples are mapped in Figure 2.

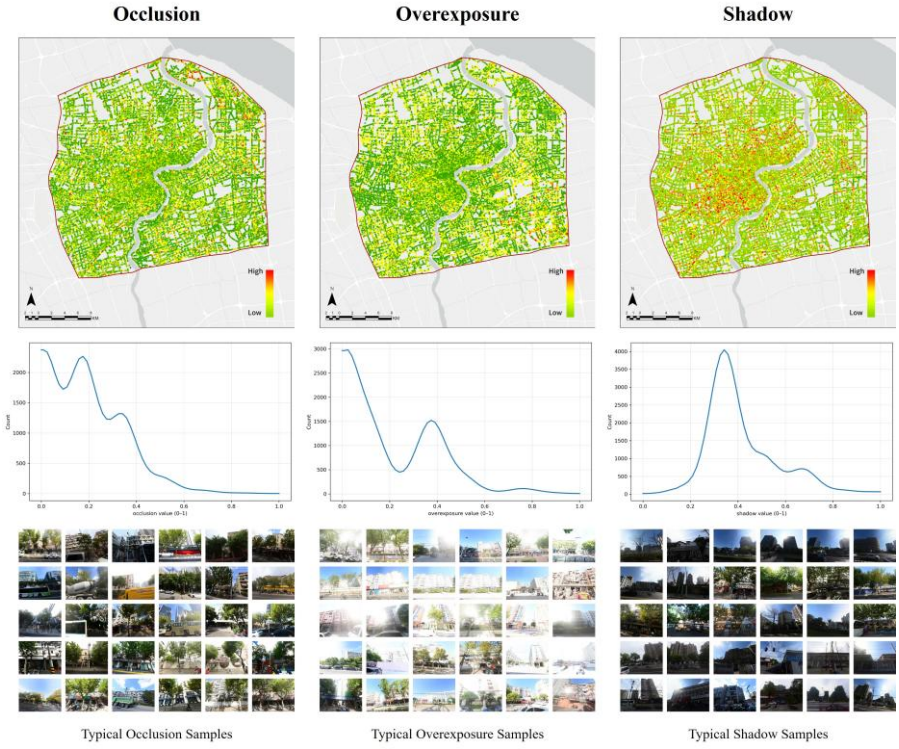


Figure 2. SVI quality mapping and typical samples selection

2.3. VLM INFERENCE

The implementation of Vision–Language Models (VLMs) in this study involves both direct interaction interfaces and API-based services. Users provide localized street-view images alongside tailored prompts to guide the VLM's visual understanding and image editing tasks (see Figure 3). Four task-specific inference strategies are developed, with prompts targeting the identification of building type, dominant colour, material composition, and quantitative estimations. All outputs are anchored to expert-validated labelled regions or predefined scoring domains for reference. Specifically, building types are classified into nine categories, building materials are grouped into eight types for ratio-based calculation, colours are expressed in hexadecimal format, and scores are assigned according to predefined rules with illustrative examples.



Figure 3. VLM caption and regeneration example

(1) **Direct Caption Inferring**. This strategy relies exclusively on the VLM's generalization capability to extract semantic information from limited visual input without any auxiliary data. The model directly processes the input image to generate structured outputs, including both categorical classifications and approximate quantitative assessments. Prompts are kept deliberately minimal to evaluate the VLM's inherent visual reasoning abilities. A prompt example is: "Please capture the key characteristics of the building and provide detailed categories of the (a) building type, (b) building material ratio, (c) main colour, (d) commercial score, and (e) façade conciseness. Please refer to the detailed samples in {file: expert classifications and score rules} and provide structured outputs in JSON format."

(2) **Integrated Caption Inferring**. This approach incorporates supplementary information from high-quality SVIs located near the low-quality image. When a clear façade visually similar to the low-quality image is available, the model first generates an image caption based on the clear façade, and this textual information is then used to enhance the interpretation of the degraded image. This strategy is particularly effective for cases involving substantial information loss, such as complete shadowing or

overexposure. A representative prompt is: "…Please refer to the detailed samples in {file: expert classifications and score rules} and provide structured outputs in JSON format. Additionally, make inferences based on the descriptive information from a nearby SVI provided in {file: Nearby SVI captions.txt}."

(3) **Direct Image Regeneration**. For tasks requiring richer semantic detail, this method leverages the VLM's generative capabilities to reconstruct refined SVI images for subsequent downstream analyses. In this process, the model directly processes the original image and applies a semantic segmentation mask to regenerate the façade. The regenerated images then support more accurate façade feature analysis. The input consists of a single street-view image accompanied by a semantic mask used for control. A typical image refinement prompt reads: "Remove occlusions in the street-view image—such as trees, trucks, and buses—while completing the building façade with plausible structural and material details. Adjust the lighting on the façade to appear consistent and natural. Do not alter any other content. Output the updated image in JPEG format." Subsequently, the same feature extraction prompt used in Strategy 1 is applied, but now using the refined SVI as the input.

(4) **Caption-guided Regeneration**. Similar to the third approach, this method introduces additional textual information derived from neighbouring high-quality SVIs for the SVI regeneration. By integrating multimodal inputs, the model achieves more context-aware image editing and produces refined reconstructions. The prompts follow a structure similar to Strategy 3 and include two components: an image refinement prompt "... Additionally, for missing information, please refer to the descriptive content from a nearby SVI provided in {file: Nearby SVI captions.txt}"—and a subsequent feature extraction prompt.

2.4. EXPERIMENT AND VALIDATION DESIGN

The four VLM-based inference strategies are applied to all 150 representative images using the qwen-image-edit model_2025. The evaluation comprises five tasks (see Table 1): two classification tasks (building type and building material), one computer vision sensing task (building colour), and two numerical inference tasks (commercial vitality and façade conciseness). For the classification results, expert validation was conducted. For the colour recognition and numerical inference tasks, both expert assessments and comparative analyses were performed to evaluate the difference between the original and generative inference results. This validation process was conducted across all categories.

Table 1. Collection of spatial and non-spatial attributes.

Experiment Task	Prediction	Validation method
Building type	27 classifications	Correct prediction (0~1)
Building material	13 classifications	Correct prediction (0~1)
Building colour	Hex conversion	CIEDE2000 (ΔE_{00}) distance
Commercial vitality	1~5 scores	Average Score Difference
Façade Conciseness	1~5 scores	Average Score Difference

3. Results

3.1. PRE-EXPERIMENT

Before scaling this approach to all the samples. The results were first subjected to pre-evaluation through single-sample comparisons, similarly followed by validation across the five defined tasks. Representative examples were selected to demonstrate the comparative performance and efficiency of the four inference methods on different kinds of low-quality SVIs (Fig. 4)

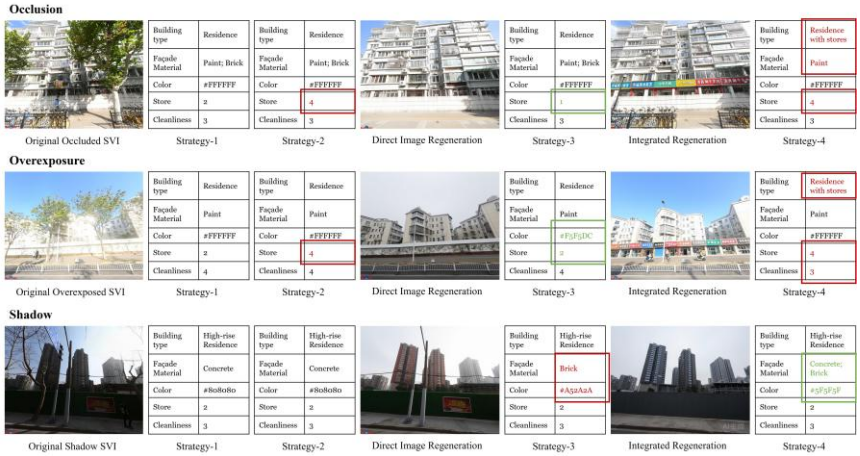


Figure 4. Typical results of four VLM-based strategies for SVI inferring (wrong inferences are labelled red and correct restorations are labelled green)

(1) For occlusion samples, Strategy-3 (direct image regeneration) produces the most reliable outputs, effectively removing obstructions while maintaining visual consistency and avoiding hallucinated elements. In contrast, Strategies-2 and Strategy-4 occasionally introduce artificial details such as additional storefronts on the ground floor, which likely over-reflect contextual features from neighbouring SVIs. (2) For overexposed samples, Strategy-2 and Strategy-4 also exhibit a higher tendency toward visual hallucination, leading to inaccurate predictions of building function and commercial vitality. Strategy-3 demonstrates greater stability, successfully reconstructing images with more balanced illumination and colour fidelity. (3) In the case of shadowed samples, performance varies considerably due to the severe information loss under low-light conditions. In this instances, supplementary information from adjacent buildings proves beneficial, particularly for recovering façade colour and material attributes. Consequently, Strategy-4 yields the most accurate and visually consistent results.

Notably, Strategies 2 and 4 exhibit general instability, primarily due to increased inference hallucination. In contrast, Strategy 3 maintains higher accuracy by relying more directly on the original semantic information.

3.2. EXPERT VALIDATION RESULTS

After refining the prompt design, we employed the VLM API service to conduct large-scale experiments using all the 150 SVI images (3 × 50 samples) across the four inference strategies. The labelling process involved two experts to ensure the accuracy of the ground-truth annotations. Validation results for all tasks and inference strategies are summarized in Table 2, with the corresponding validation methods detailed in Table 1. Classification accuracy (marked as "+", 0~1) is used for building type and material, colour is evaluated based on colour distance in RGB space, and vitality and façade conciseness are assessed via score differences from expert labelling (Distance and differences are marked as "-"). For clarity, the best-performing strategy for each task is highlighted in red.

Table 2. Expert Validation Results

Occlusion	Strategy-1	Strategy-2	Strategy-3	Strategy-4
Building type	+0.80	+0.64	+0.85	+0.75
Building material	+0.67	+0.70	+0.89	+0.78
Building colour	-71.74	-79.40	-89.21	-89.13
Commercial vitality	-1.00	-0.96	-0.92	-1.10
Façade Conciseness	-0.38	-0.38	-0.37	-0.44
Overexposure				
Building type	+0.75	+0.63	+0.88	+0.68
Building material	+0.68	+0.68	+0.91	+0.84
Building colour	-78.22	-63.08	-65.74	-55.29
Commercial vitality	-1.04	-1.34	-0.94	-1.44
Façade Conciseness	-0.48	-0.66	-0.28	-0.66
Shadow				
Building type	+0.64	+0.82	+0.84	+0.86
Building material	+0.65	+0.81	+0.86	+0.84
Building colour	-138.03	-130.38	-101.73	-130.22
Commercial vitality	-0.98	-1.50	-1.30	-1.44
Façade Conciseness	-0.65	-0.90	-0.94	-0.90
Cost (tokens/pc)	1500~2000	2500~3000	5500~6000	6000~7000

Overall, the four inference strategies demonstrate distinct performance patterns across the three categories of low-quality SVIs. Among them, Strategy-3 consistently achieves better results, due to its strong capability for visual information recovery and its reduced tendency toward hallucinated outputs.

Across the five evaluation tasks, performance varies depending on the amount and reliability of visual information available. (1) Building type and building material are the most accurately inferred attributes, with average accuracies exceeding 0.7. Strategy-3 performs best for these tasks, benefiting from its ability to reconstruct

detailed façades and support systematic reasoning of building functions. (2) Building colour shows less stability across categories, particularly in the overexposure and shadow groups, where detailed visual features are either lost or misleading. In such cases, Strategy-3 and Strategy-4, which rely on regenerated imagery, achieve higher consistency. (3) Commercial vitality and façade Conciseness exhibit task-specific sensitivities. While both metrics vary across degradation types, they generally favour direct inference over integrated approaches. In occlusion and overexposure cases, Strategy-3 yields more accurate environmental evaluations, whereas in shadowed images, Strategy-1 produces more stable outputs, indicating that shadows have minimal impact on building function and façade type recognition.

During the inference process, token consumption was also evaluated. The integrated inference strategies required a greater number of tokens due to the incorporation of additional contextual information, while the regeneration-based strategies incurred higher computational costs associated with image synthesis. Overall, Strategy-3 demonstrated the most favourable balance between accuracy and computational efficiency, producing high-quality outputs that can be directly applied to downstream tasks such as detailed façade segmentation and visual analysis.

3.3. APPLICATION EXPLORATIONS

Beyond the single-image restoration achieved through Strategy-3, we further examined the potential of VLMs for detailed street-level building regeneration based on the reconstructed SVIs. In this experiment, the regenerated images were used as a foundation for façade renewal design, enabling the model to infer and visualize possible design interventions on existing street façades (see figure 5)

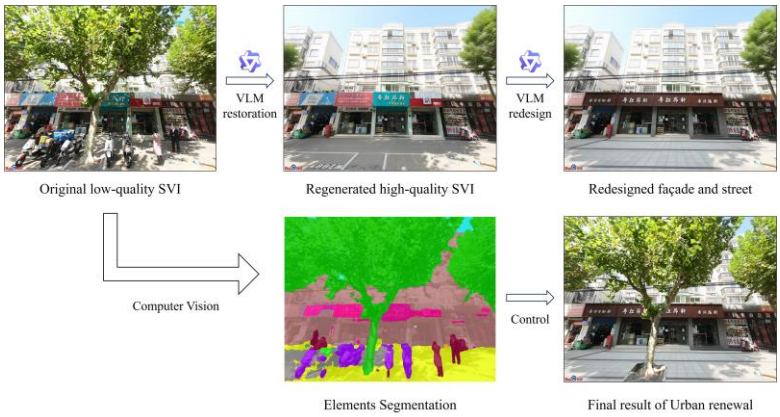


Figure 5. Single SVI regeneration for informing real urban design

We first selected a street segment exhibiting high renewal potential, characterized by façade deterioration and disordered storefront layouts. The proposed framework was then applied for the SVI restoration there, producing a clear and high-quality façade image suitable for subsequent design exploration. Subsequently, the regenerated image, together with its corresponding segmentation results, was used to guide façade element redesign through VLM-assisted inference. The outputs demonstrate that this

workflow offers an efficient and cost-effective tool for urban designers to test and visualize design concepts even under conditions of partial information loss, thereby supporting more informed and detailed urban design decision-making.

3.4. LIMITATIONS

Although this study demonstrates the effectiveness of the VLM-based framework in handling poor-quality SVIs using representative samples, broader validation across large-scale datasets and multiple cities remains limited due to the high cost of expert annotation. Furthermore, iterative prompt tuning is necessary to mitigate potential hallucinations, and extensive API usage is required to support large-scale applications.

4. Conclusion and discussion

In conclusion, this study presents an innovative framework for supplementing urban street information using generative Vision–Language Models (VLMs). Although further validation is necessary to integrate language-based encodings into robust numerical measurements, the model's generative capacity and interpretable reasoning effectively mitigate information loss related to texture and other visual deficiencies. The validated inference strategies demonstrate strong potential for large-scale urban measurements, enhancing accuracy through clearer and more descriptive building information. Moreover, by directly regenerating high-quality façade imagery, the proposed framework preserves essential architectural details that can be further leveraged for urban analytics and fine-grained design applications, offering a scalable and interpretable pathway toward data-driven urban renewal design practices.

References

Chen, X., Ding, X., & Ye, Y. (2024). Mapping sense of place as a measurable urban identity: Using street view images and machine learning to identify building façade materials. Environment and Planning B: Urban Analytics and City Science, 23998083241279992.

Fan, Z., Zhang, F., Loo, B. P. Y., & Ratti, C. (2023). Urban visual intelligence: Uncovering hidden city profiles with street view images. Proceedings of the National Academy of Sciences, 120(27), e2220417120. 10.1080/24694452.2024.2313515

Fang, Z., Jin, Y., Zheng, S., Zhao, L., & Yang, T. (2024). UrbanClassifier: A deep learning-based model for automated typology and temporal analysis of urban fabric across multiple spatial scales and viewpoints. Computers, Environment and Urban Systems, 111, 102132.

Liang, X., Xie, J., Zhao, T., Stouffs, R., & Biljecki, F. (2025). OpenFACADES: An open framework for architectural caption and attribute data enrichment via street view imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 230, 918–942.

Wang, L., Zhang, T., He, J., & Kada, M. (2025). Cross-platform complementarity: Assessing the data quality and availability of Google Street View and Baidu Street View. Transactions in Urban Data, Science, and Technology, 4(1), 22–47.

Xiao, Y., & Tang, Y. (2025). Can ChatGPT-4o assess the perceptions of streetscape change? Evidence from Shanghai, China. Sustainable Cities and Society, 130, 106674.

Yang, R., Deng, X., Shi, H., Wang, Z., He, H., Xu, J., & Xiao, Y. (2024). A novel approach for assessing color harmony of historical buildings via street view image. Frontiers of Architectural Research, S2095263524000402. https://doi.org/10.1016/j.foar.2024.02.014

Human-Informed Urban Vitality

Integrating Form, perception, and design via Heterogeneous Graph Learning

Muxin Jia[1], Kaiheng Zhang[2] and Taro Narahara[3]
[1,2,3]*New Jersey Institute of Technology.*
[1]*mj8@njit.edu, 0000-0002-2213-6027*
[2]*kz25@njit.edu, 0009-0006-9671-7269*
[3]*narahara@njit.edu, 0000-0002-2200-1253*

Abstract. Urban vitality reflects the capacity of urban environments to attract social interaction, economic activity, and cultural exchange. Measuring and predicting vitality at fine spatial scales, such as street blocks, remains challenging due to the complex interplay of spatial form, function, and human perception. This study introduces a human-informed framework that integrates dynamic street activity and static functional attributes into a composite Urban Vitality Index (UVI) and translates each urban block into a heterogeneous graph that captures building, street, and cross-layer spatial relationships. Using this representation, we evaluate multiple graph neural network models, including GCN, GAT, GIN, GraphSAGE, and the Heterogeneous Graph Transformer (HGT), to classify vitality levels across 3,930 blocks in Manhattan, NY. Among these, HGT demonstrates the strongest performance, highlighting its capacity to learn from complex, multi-type urban structures. The proposed framework provides a scalable pathway for linking urban form, function, and perception through heterogeneous graphs, offering valuable insights for data-driven urban planning and design.

Keywords. Urban vitality; heterogeneous graph; graph neural networks; urban morphology

1. Introduction

Urban vitality, first articulated by Jane Jacobs (1961), encapsulates the degree to which urban environments attract diverse social groups and foster sustained social engagement, economic activity, and cultural exchange. It reflects not only the visible intensity of human presence in space but also the underlying social and spatial mechanisms that sustain such interactions (Maas, 1984). Over the decades, two major perspectives have emerged in interpreting this multifaceted concept. From the standpoint of urban sociology, vitality is shaped by the intertwined dynamics of economic, social, and cultural life, emphasizing the collective behavior and relational networks that animate cities. In contrast, architectural and urban design approaches

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 573-582.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

interpret vitality as a manifestation of human activity grounded in the form, perception, and design of physical spaces. Together, these perspectives underscore that urban vitality is inherently human-centered and spatially embedded (Xiao et al., 2021, Chen et al., (2025).

For researchers in the urban and design domain, measuring, visualizing, and predicting urban vitality at fine spatial scales, particularly at the street-block level, are essential for understanding how spatial form and perceptual qualities shape public life. However, modeling the vitality of urban blocks remains a complex challenge. Two key obstacles persist. First, in terms of measurement, vitality cannot be directly observed but has to be inferred through proxies. Developing real-time and representative measures that reflect this complexity remains a persistent difficulty. Second, most prior predictive studies have relied on traditional statistical models or machine learning algorithms that overlook the intricate spatial and topological relationships among urban elements that fundamentally structure human experience (Xie et al., 2025).

To bridge these gaps, this study proposes a human-informed computational framework for measuring and predicting urban vitality by integrating urban form, function, perception, and design attributes within a unified geometric and visual learning system. Specifically, we develop a compound Urban Vitality Index that integrates static and dynamic proxies, combining visual activity indicators from street-level imagery with functional diversity metrics derived from POI density and entropy. Using this index, we construct heterogeneous graphs representing urban blocks, where nodes denote spatial entities (buildings and streets) and edges encode their topological relationships. Graph-level features follow the 3D framework (Density, Diversity, and Design), while node and edge attributes capture geometric and morphological characteristics (Liu et al., 2023). Finally, we apply deep graph learning models to predict urban vitality, comparing conventional architectures (GCN, GAT, GIN, GraphSAGE) with HGT, a transformer-based model optimized for heterogeneous graphs.

2. Methods

As illustrated in Figure 1, this study develops a workflow for measuring and predicting the Urban Vitality Index (UVI) at the census-block level in Manhattan, NY. The process includes three components: data development, heterogeneous graph representation, and vitality prediction. In the first stage, we prepare the target variable (UVI intensity) and corresponding graph-, node-, and edge-level features for prediction. Next, we construct 3,930 heterogeneous graphs based on each street block, incorporating building polygons and street segments to encode spatial and topological relationships. Finally, we train and evaluate graph neural network models, comparing benchmark architectures with the Heterogeneous Graph Transformer (HGT) to capture the complex heterogeneity of urban areas.

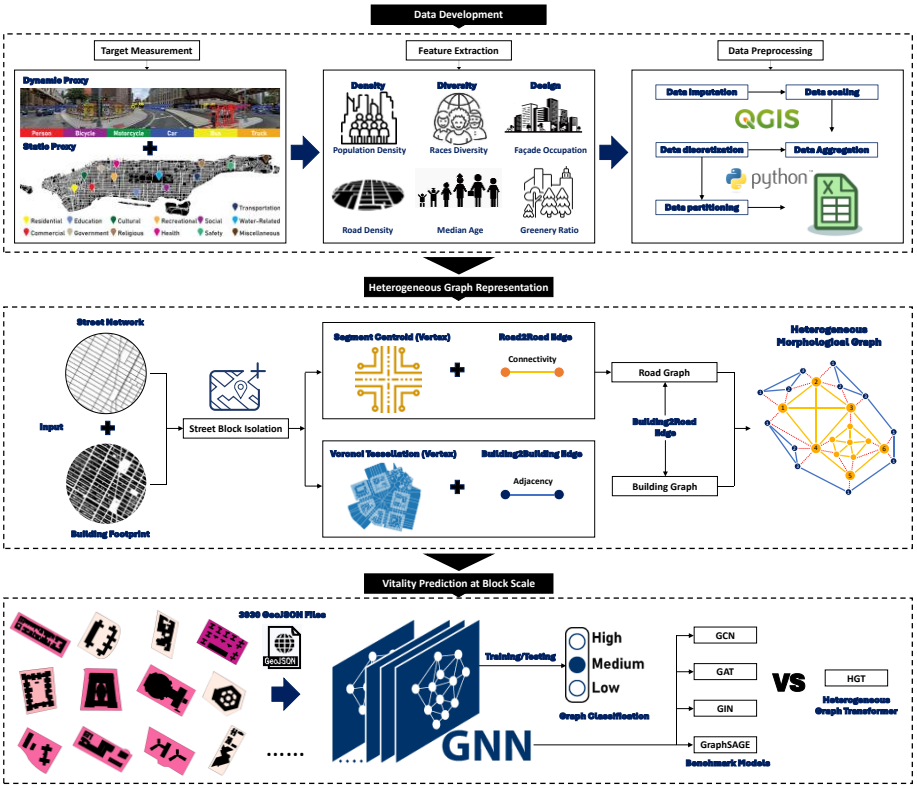


Figure 1. Research workflow.

2.1. URBAN VITALITY MODELLING

In data-driven urban analysis, the target variable represents the primary outcome that a model seeks to measure and predict. In this study, a composite Urban Vitality Index (UVI) was developed as the target variable, integrating both dynamic and static proxies of vitality. The dynamic proxy captures real-time street activity through the detection of pedestrians, bicycles, motorcycles, cars, buses, and trucks from panoramic imagery using YOLOv11s, a lightweight, high-performance object detection model pre-trained on the COCO dataset. Object counts were weighted according to their perceived contribution to vitality, as determined by the Analytic Hierarchy Process (AHP). Because sampling points were located along street centerlines, the boundaries of adjacent census blocks, a spatial allocation strategy was implemented: 50-meter buffers were generated around each point, intersected with nearby block polygons, and the weighted counts were proportionally distributed based on intersection area.

The formula for the dynamic proxy at census block i is defined as:

$$Dynamic\ Proxy_i = \sum_{j=1}^{n} \omega_j \cdot (\sum_k P_{ki} \cdot O_{ki})$$

Where:

j: Index for object categories

k: Index for sampling points

O_{ki}: Number of objects of category j at point k

ω_j: Weight assigned to object category j through AHP in Table 2

P_{ki}: Proportion of the buffer area of point k that falls within block i, calculated as:

$$P_{ki} = \frac{Area(Buffer_k \cap Block_i)}{Area(Buffer_k)}$$

The static proxy of urban vitality represents the density and functional contribution of POIs within each census block. To account for the varying significance to urban vitality, AHP-derived weights were assigned to each category, prioritizing those that promote public interaction and dynamic land use (e.g., commercial and recreational facilities).

The formula for the static proxy at census block i is:

$$Static\ Proxy_i = \sum_{l=1}^{m} \omega_l \cdot D_{il}$$

Where:

l: POI categories (1 to 13)

ω_l: Weight assigned to POI category l through AHP in Table 3

D_{il}: Density of POIs of category l in census block i

To generate a holistic assessment of urban vitality, the dynamic and static proxies were normalized and then averaged to compute the final UVI for each block:

$$UVI_i = \frac{DP_i + SP_i}{2}$$

Where:

DP_i: Normalized dynamic proxy for census block i

SP_i: Normalized static proxy for census block i

2.2. MOPHOLOGICAL GRAPH REPRESENTATION

To model the spatial configuration of urban form and its underlying relationships, we adopt a heterogeneous graph representation that encodes both the building geometry and the street network within a unified block-scale framework. The approach follows the principles of City2Graph, enabling a graph-based morphological abstraction that captures multi-type spatial entities and their contextual dependencies.

Formally, a heterogeneous graph is defined as $G = (V, E, F, T)$, where V and E denote the sets of nodes (F_V) and edges (F_G), respectively; F represents the feature vectors associated with nodes and graphs; and T indicates node and edge types. In this study, G is constructed by integrating two homogeneous subgraphs: the building graph (G_b) and the road graph (G_r), connected through inter-graph spatial relationships (E_{br}).

(1) Building Graph Construction

For each urban block B_i, all building footprints within its boundary are extracted as the base geometries. To generate a spatially continuous morphological partition, we perform morphological tessellation based on the Voronoi diagram of building centroids,

constrained by the enclosing street network. Let $\{p_1, p_2, \ldots, p_n\}$denote the centroids of n building polygons in block B_i . The corresponding Voronoi diagram $V(p_i)$subdivides the block into polygonal cells:

$$T_i = \{t_j \mid t_j = \text{Voronoi}(p_j) \cap B_i, j = 1,2,\ldots,n\}$$

Each tessellation cell $t_j$represents the spatial influence area of a building and serves as a node $v_j^b \in V_b$in the building graph. Two nodes are connected by an edge $e_{jk}^b \in E_b$if their corresponding cells share a common boundary, defined as:

$$e_{jk}^b = \begin{cases} 1, & if\ t_j \cap t_k \neq \emptyset \\ 0, & otherwise \end{cases}$$

This procedure encodes the morphological adjacency and local spatial continuity of buildings within each block.

(2) Road Graph Construction

For the road network, each street segment $s_l$is treated as an individual spatial entity. Its centroid defines a node $v_l^r \in V_r$. Two road nodes are connected by an edge $e_{lm}^r \in E_r$if their corresponding segments are topologically connected in the street network (i.e., share an intersection). This generates a graph representation that preserves the topological hierarchy and network connectivity of the street system.

(3) Cross-Graph Integration

To form the heterogeneous morphology graph, we establish cross-type edges E_{br}that link the building graph and the road graph. A cross-edge $e_{jl}^{br}$is created between a building node $v_j^b$and a road node $v_l^r$when the tessellation cell $t_j$is spatially adjacent to or intersects with the street segment s_l. Formally:

$$e_{jl}^{br} = \begin{cases} 1, & if\ t_j \cap s_l \neq \emptyset \\ 0, & otherwise \end{cases}$$

The resulting heterogeneous graph can be represented as:

$$G = \{G_b, G_r, E_{br}\} = \{(V_b, E_b), (V_r, E_r), E_{br}\}$$

This unified representation encodes both intra-layer relationships (building–building and road–road) and inter-layer linkages (building–road), enabling graph neural networks (GNNs) to learn structural and contextual features of urban morphology jointly. Multiple blocks of heterogeneous graphs are shown below for reference (Figure 2).

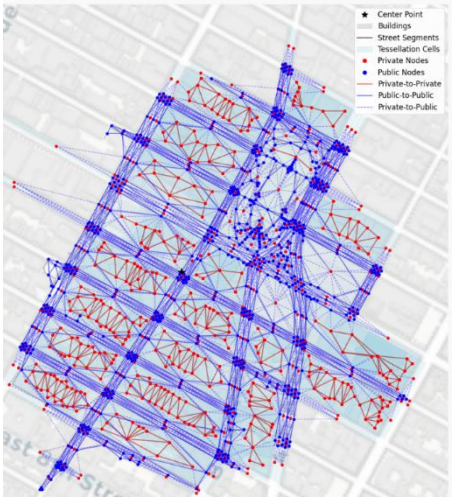


Figure 2. Morphological graphs on OSM Map

2.3. FEATURE EXTRACTION

Feature extraction is essential for effective GNN training, as it encodes morphological, functional, and perceptual attributes into node and graph representations. In this study, features are derived for both nodes and graphs within each street block of the heterogeneous morphology graph, which contains two types of nodes: building nodes and road nodes.

Both types of node features are extracted from OpenStreetMap (OSM) to represent the geometric and structural characteristics of urban fabric. Building nodes are described by area, perimeter, and compactness, reflecting the size and shape of individual structures, while road nodes are characterized by length and width, indicating the configuration and hierarchy of street segments.

Beyond node attributes, we compute graph-level features to represent the broader spatial and socio-environmental context, following the "3D" framework proposed by Cervero and Kockelman (1997): Density, Diversity, and Design. Density indicators (population and road density) and diversity indicators (racial and age diversity) are obtained from the PLUTO™ dataset, which provides detailed tax lot-level data for New York City. Design-related indicators are extracted from Google Street View (GSV) imagery to capture human-scale visual qualities. As shown in Figure 3, four panoramic GSV views are stitched and semantically segmented using the Unified Perceptual Parsing Network (UPerNet), from which the greenery ratio and façade occupation are quantified.



Figure 3. (a) Street view sampling points, (b) Street view images, (c) Object detection, (d) Semantic segmentation.

2.4. MODEL CONSTRUCTION AND TRAINING

This study performs a graph-level classification task to predict urban vitality across street blocks. The UVI is categorized into three levels, high, medium, and low, using the natural breaks (Jenks) method. A total of 3,930 heterogeneous graphs (Figure 4) are assembled and converted into a PyTorch Geometric dataset, with an 80/20 train–test split.

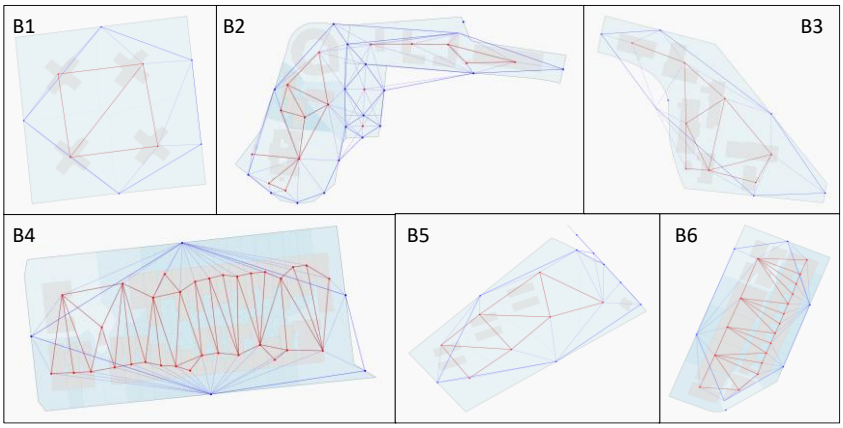


Figure 4. Morphological graph for sample blocks.

The model adopts a heterogeneous GNN architecture consisting of optional per-node-type linear projections, followed by two HeteroConv layers. Each layer applies ReLU activation and dropout for regularization. Node embeddings are aggregated through Set2Set pooling, concatenated across node types, and passed through an MLP to produce class logits.We employ GCN, GAT, GIN, and GraphSAGE as benchmark models, and compare their performance with HGT, a transformer-based architecture designed for heterogeneous graphs. Training uses the Adam optimizer and cross-entropy loss, with accuracy as the evaluation metric.

3. Experiment and results

3.1. VITALITY EVALUATION

Figure 5 shows the spatial distributions of the dynamic proxy, static proxy, and overall Urban Vitality Index (UVI) across Manhattan's census blocks. Together, these layers capture both the real-time dynamics of street activity and the structural capacity of the built environment. All indicators are normalized to a [0, 1] range using Min–Max scaling, ensuring a balanced contribution of dynamic and static components to the composite index.

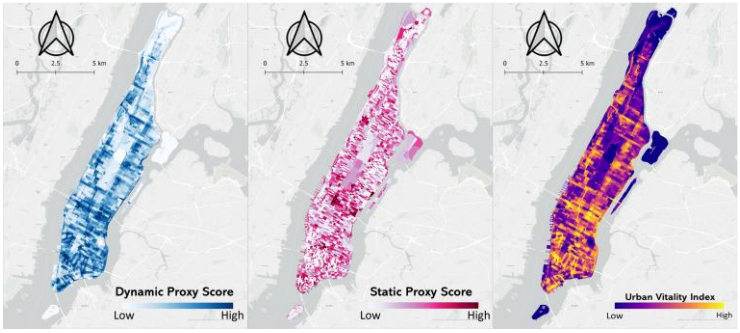


Figure 5. Spatial distribution of the dynamic proxy score, the static proxy score, and the urban vitality index.

3.2. MODEL PERFORMANCE

The evaluation results indicate that all five graph-based models demonstrate varying performance across balanced accuracy, precision, recall, and macro F1 Score. Among them, HGT (Heterogeneous Graph Transformer) clearly outperforms the others, achieving the highest mean test accuracy of 0.735, as illustrated in Table 1. The confusion matrices for 5 models are shown below as well (Figure 6).

Overall, HGT is the most robust model for this predictive task because it effectively captures the complex and heterogeneous relationships in the graph, leveraging both node and edge types through its attention-based architecture, which allows it to integrate diverse structural and functional information more effectively than the other models.

Table 1. Model Performances Comparison

Models	Mean_test_accuracy	Precision	Recall	f1_macro
sage	0.61	0.475	0.57	0.515
gcn	0.615	0.49	0.58	0.53
gat	0.62	0.515	0.595	0.55
gin	0.622	0.505	0.59	0.545
hgt	0.735	0.61	0.72	0.66

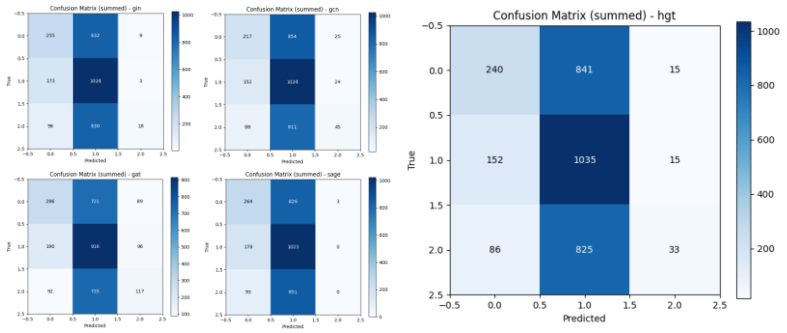


Figure 6. Confusion matrix for 5 GNN models.

3.3. ABLATION ANALYSIS

To examine the contribution of key architectural components, an ablation analysis was conducted based on the best-performing baseline model, the Heterogeneous Graph Transformer (HGT). Two ablation axes were evaluated: the inclusion of graph-level attributes and the choice of graph pooling strategy. Specifically, the baseline configuration, combining node embeddings with six graph-level features and employing Set2Set pooling, was compared against variants that removed graph-level attributes and replaced Set2Set pooling with global mean or global sum pooling. All ablation experiments were trained under identical settings (20 epochs, batch size of 16, fixed random seed, and consistent data splits) to ensure fair comparison. As shown in Table 2, model performance degradation relative to the baseline was analyzed to identify the relative importance of structural features and pooling mechanisms. The ablation result reveals that graph-level features play a dominant role in model performance. Removing these features results in a substantial decline in both accuracy and ROC-AUC, indicating that global structural information is essential and cannot be fully recovered through node-level attention alone. In contrast, replacing the Set2Set pooling mechanism with simpler global pooling strategies leads to moderate performance degradation, with sum pooling outperforming mean pooling. These findings suggest that while aggregation strategy affects discriminative capacity, learnable pooling provides additional expressive power when combined with explicit global descriptors.

Table 2. Ablation test results

Settings	Graph_features	Pooling_method	mean_test_accuracy	roc_auc_macro
Baseline	**TRUE**	**set2set**	**0.703698**	**0.76413**
Test_1	FALSE	set2set	0.500689	0.644266
Test_2	TRUE	mean	0.614787	0.670007
Test_3	TRUE	sum	0.652427	0.718919

4. Conclusion

This study presents a human-informed and graph-based framework for assessing and predicting urban vitality at the street-block scale, demonstrating how heterogeneous graph representations can effectively integrate urban form, function, and perceptual qualities (Liu, Guo & Zheng, 2025). By combining dynamic street activity measures with static functional indicators into a composite Urban Vitality Index, and by modeling each block as a multi-type morphological graph, the framework captures the spatial and contextual dependencies that shape vitality. The comparative evaluation of graph neural network models shows that transformer-based architectures, particularly the Heterogeneous Graph Transformer (HGT), offer clear advantages in learning from complex urban structures. This suggests that urban environments benefit from models capable of handling multi-type relationships and cross-scale interactions, rather than relying solely on homogeneous or topology-only representations.

While the results underscore the promise of heterogeneous graph learning for urban analysis, this study has several limitations. The vitality index relies on available imagery and POI datasets, which may not fully capture temporal variations in activity or diverse forms of public life. In addition, the graph construction focuses primarily on buildings and streets; incorporating land-use parcels, open spaces, or pedestrian pathways could further enrich the representation. Future work could extend the framework across different cities, explore temporal graph models, or integrate human-centered perception data more directly (e.g., surveys, social media, or sensory analyses). By advancing scalable and interpretable methods, this research contributes toward more nuanced, data-informed planning and design practices that better reflect the lived experience of urban environments.

References

Chen, D., Feng, Y., Li, X., Qu, M., Luo, P., & Meng, L. (2025). Interpreting core forms of urban morphology linked to urban functions with explainable graph neural network. arXiv preprint arXiv:2502.16210.

Liu, H., Guo, Q., Zhu, H., Fu, Y., Zhuang, F., Ma, X., & Xiong, H. (2023). Characterizing and forecasting urban vibrancy evolution: A multi-view graph mining perspective. ACM Transactions on Knowledge Discovery from Data, 17(5), 1-24.

Liu, Y., Guo, Q., & Zheng, C. (2025). Urban street network morphology classification through street-block based graph neural networks and multi-model fusion. International Journal of Digital Earth, 18(1), 2497490.

Maas, P. R. (1984). Towards a theory of urban vitality (Doctoral dissertation, University of British Columbia).

Xiao, L., Lo, S., Zhou, J., Liu, J., & Yang, L. (2021). Predicting vibrancy of metro station areas considering spatial relationships through graph convolutional neural networks: The case of Shenzhen, China. Environment and Planning B: Urban Analytics and City Science, 48(8), 2363-2384.

Xie, X., Mao, W., Yang, Z., Hu, S., Xu, Y., Wu, L., & Xie, Z. (2025). A hybrid spatial learning method integrating graph neural networks and explainable AI to understand urban vitality. Sustainable Cities and Society, 106920.

Research on Urban Waste Classification Optimization Integrating Spatial Analysis and Behavioral Intervention

A case study of Kowloon City, Hong Kong

Yuanxi Zhang [1] , Tiantian Lo [2] and Yi Jian [3]
[1,2] *School of Design， Hong Kong Polytechnic University.*
[3] *Department of Social Sciences and Policy Studies, The Education University of Hong Kong*
[1] *yua2zhang@polyu.edu.hk, 0009-0009-7492-2570*
[2] *tt2lo@polyu.edu.hk, 0000-0002-1992-0777*
[3]*izzyjian@eduhk.hk, 0000-0001-7957-3546*

Abstract. This paper investigates how to optimize waste sorting management in high-density cities by combining spatial analysis and behavioral intervention, using Kowloon City in Hong Kong as an example for empirical analysis. The study aims to address the mismatch between facility layout and resident behavior in current waste management, thereby improving waste recycling efficiency. Methodologically, a Geographic Information System (GIS) is used to identify waste generation hotspots and blind spots in recycling facility coverage, and a Fogg behavioral model is used to analyze the main obstacles to resident participation in waste sorting. The study finds that inconvenient facilities, insufficient sorting knowledge, high time costs, and lack of immediate feedback are key factors affecting residents' waste sorting participation. Based on this, a system optimization scheme is proposed, including intelligent sorting container design, facility layout optimization, and a multi-level behavioral guidance mechanism. The conclusion shows that the strategy of co-optimizing spatial and behavioral aspects can effectively improve waste sorting participation and resource recycling efficiency, providing a feasible path for waste management in Hong Kong and other high-density cities.

Keywords. Urban Waste Management, Waste Classification Optimization, Spatial Analysis, Behavioral Intervention, High-Density Cities

1. Introduction

As a densely populated international metropolis, Hong Kong's waste management system is facing increasingly severe pressure. Currently, the amount of municipal solid waste generated far exceeds the overall processing capacity of local facilities, creating ongoing environmental governance challenges. The latest statistics from the Hong

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 583-592.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Kong Environmental Protection Department (2023) show that Hong Kong generates approximately 15,000 tons of municipal solid waste daily, totaling over 5 million tons annually; the average daily waste disposal per capita is 1.51 kg, ranking among the highest among major international cities, highlighting the urgency of source reduction efforts.

Structurally, Hong Kong's waste mainly consists of three parts: municipal solid waste, construction and demolition waste, and special waste, with municipal solid waste being the dominant component. Further analysis reveals that municipal solid waste primarily originates from residential, industrial, and commercial activities, as well as construction and demolition projects; it also includes special waste categories such as medical waste and chemical waste. Despite the massive total amount of waste, the overall resource utilization rate remains low. Taking municipal solid waste as an example, its recycling rate is only about 31%, with 113.57 million tons ending up in landfills. Even for hazardous or difficult-to-treat special waste categories, the recycling rate remains only around 30%, reflecting significant shortcomings in the structure and efficiency of the recycling system.

Hong Kong's current municipal solid waste management strategy suffers from systemic limitations, primarily manifested in an overemphasis on the expansion and technological upgrading of end-of-pipe treatment facilities, while failing to fully incorporate the interaction mechanism between "people, space, and behavior" into the overall governance framework. Existing policies largely focus on expanding landfill capacity and constructing waste-to-energy facilities. While technologically advanced, these policies largely neglect the crucial behavioral logic of "people" as the primary agents of waste generation and sorting, leading to a structural disconnect in the implementation of the management strategy.

Furthermore, the mismatch between behavioral patterns and spatial resource allocation has become a core obstacle to residents' participation in waste sorting and recycling. Specifically, the layout of recycling points fails to adequately respond to community population density and mobility characteristics; sorting guidelines suffer from inconsistent information and poor visibility; and the design of recycling containers often fails to align with residents' daily routines and usage habits.

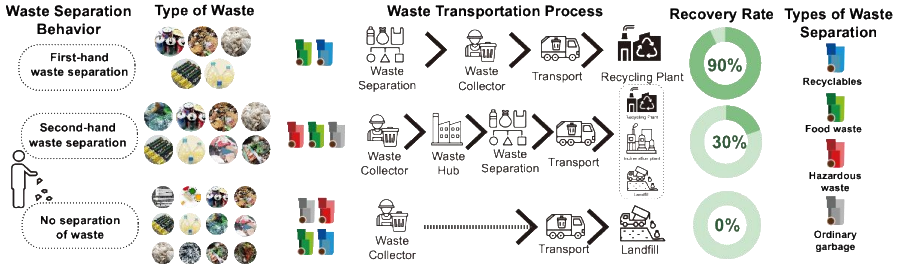


Figure1. Relationship between different behaviors and waste disposal outcomes

These factors collectively weaken the convenience and sustainability of public participation, creating a gap between willingness to sort and actual behavior, hindering the full realization of the potential for source reduction and resource recycling.

This study aims to construct a waste sorting management system framework that integrates spatial optimization and behavioral intervention to systematically improve the recycling efficiency of urban solid waste in Hong Kong. This framework is based on the Fogg Behavior Model and combines spatial analysis techniques to establish a collaborative mechanism among "people, facilities, and environment," thereby overcoming the current disconnect between behavioral guidance and spatial allocation in waste management.

2. Theoretical Framework and Methodological Analysis

This study focuses on constructing a collaborative "human-space-behavior" framework for waste management, systematically integrating policy practice insights, behavioral science theories, and spatial analysis techniques. The aim is to propose an optimized path for high-density urban waste recycling systems. The following discussion covers five aspects: policy foundation, regional characteristics, theoretical relationships, data system, and analytical framework.

2.1. POLICY PRACTICE FOUNDATION: HONG KONG PILOT EXPERIENCE AND THE KEY ROLE OF SOURCE SEPARATION

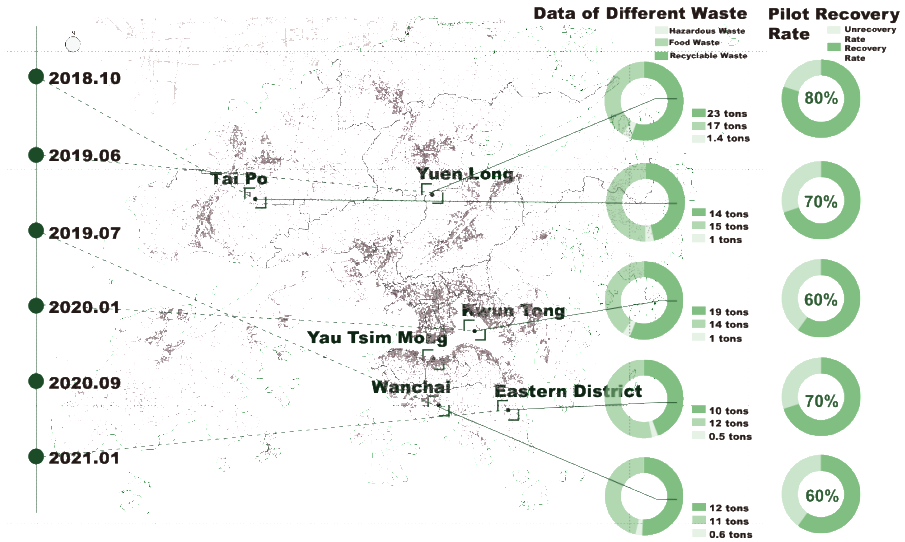


Figure 2: Analysis of Hong Kong's Community Pilot Policy

The community waste sorting pilot projects implemented in Hong Kong in recent years provide an important policy practice basis for this study. According to the Waste Disposal Ordinance passed by the Legislative Council (2021), multiple pilot results show that there is a significant positive correlation between recycling rate and the degree of implementation of source separation. This experience confirms that the key to improving recycling efficiency lies in encouraging residents to effectively sort waste at the generation stage, providing a practical basis for this study's focus on the "people-

behavior-space" interaction mechanism.

2.2. TYPICALITY OF THE STUDY AREA: THE MULTIDIMENSIONAL REPRESENTATIVENESS OF KOWLOON CITY DISTRICT

This study selects Kowloon City as the empirical research object, and its typicality is reflected in the following three aspects: First, as a high-density built-up area, Kowloon City is highly representative in terms of population density, building form and public space constraints; second, its mixed urban functions result in a complex waste structure, covering multiple types of waste, including household waste, commercial waste and recyclable resources; third, the existing waste management system in this area combines a public recycling network, three-color recycling bins and private collection services, providing a complete research object for analyzing system bottlenecks and optimization potential.



Figure 3: Left: Land use; Right: Building type

2.3. THEORETICAL INTEGRATION: THE CORRELATION BETWEEN THE FOGG BEHAVIORAL MODEL AND THE 9R PRINCIPLES

To overcome the limitations of traditional behavioral interventions, this study theoretically combines the Fogg behavioral model with the 9R principle of sustainable environmental management. The Fogg model explains three elements of individual behavior at the micro level—motivation, ability, and triggering cues—while the 9R principle constructs a systematic cycle from awareness to action.

2.4. DATA SYSTEM CONSTRUCTION: COLLECTION AND PROCESSING OF SPATIAL AND BEHAVIORAL MULTIDIMENSIONAL DATA

Spatial data: This includes the location distribution of waste collection facilities, land use attributes, building classifications, and the spatial coverage of waste bins. GIS is used for geocoding and spatial measurement to establish a matching relationship between facility service areas and buildings/land uses.

Behavioral data: Through structured questionnaires and on-site behavioral observations, data on residents' willingness to sort waste, their knowledge level, barriers to participation, and actual waste disposal behaviors are collected. The correlation between behavioral motivations and facility use is analyzed.

2.5. SYSTEM ANALYSIS FRAMEWORK: A THREE-TIER PATH FROM DIAGNOSIS TO OPTIMIZATION

Based on the aforementioned theories and data, this study constructs a three-tiered analytical framework of "diagnosis-analysis-optimization": Spatial Diagnostic Analysis: Utilizing GIS visualization technology and waste heat maps, spatial mismatch areas characterized by "high waste generation - low recycling facility coverage" are identified, clarifying facility blind spots and service overlap areas. Behavioral Barrier Analysis: Using the Fogg model, the root causes of residents' waste sorting failures are analyzed from three dimensions: motivation, ability, and triggering mechanism deficiencies. System Optimization Design: A collaborative solution integrating intelligent infrastructure and behavioral guidance mechanisms is proposed, including the rational deployment of intelligent waste bins and a layout optimization prompt system based on spatial simulation.

3. The Optimized Scheme

This chapter systematically presents the key findings of this study, unfolding from three levels: spatial diagnosis, behavioral barrier analysis, and derived system optimization solutions. The aim is to comprehensively reveal the causes and solutions to the problem of solid waste management in high-density cities.

3.1. SPATIAL DIAGNOSTIC RESULTS

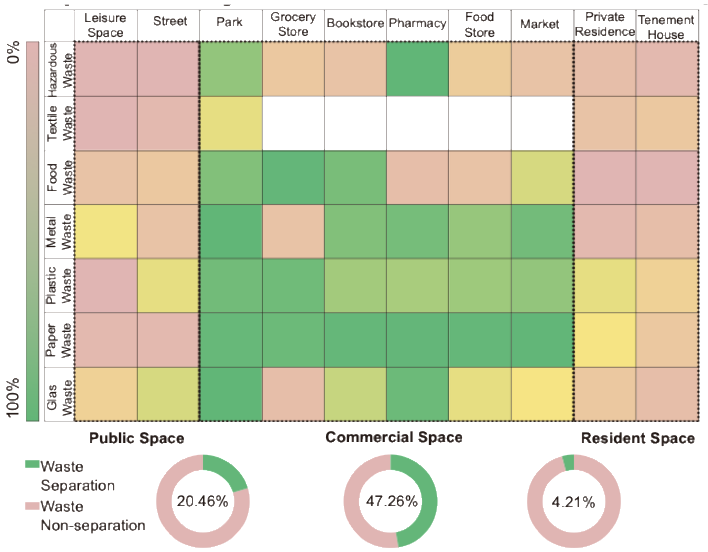


Figure 4. Heatmap analysis and Scenarios classification

Through GIS spatial analysis, this study accurately identified several waste generation hotspots within Kowloon City. These areas are highly concentrated in commercial districts and some ultra-high-density residential areas, with waste generation intensity

significantly higher than the urban average. However, spatial overlay analysis reveals a key contradiction: these high-load areas precisely face the problem of insufficient recycling facility coverage.

Precise calculations of the facility service range show that approximately 23% of residents in the district are outside the standard service radius, forming a typical "facility coverage blind spot." Further analysis shows a spatial mismatch between the distribution of existing facilities and population activity density. Facilities in some high-demand areas are operating beyond capacity due to insufficient quantity or unreasonable layout, while facilities in surrounding areas have low utilization rates. This spatial mismatch directly constitutes an objective physical barrier to residents' participation in waste sorting.

3.2. RESULTS OF BEHAVIORAL BARRIER ANALYSIS

Based on the Fogg behavioral model, the survey analysis clearly revealed a significant "willingness-to-action" gap among residents regarding waste sorting. Data showed that although 68% of respondents supported and were willing to participate in waste sorting, only 35% were able to translate this into consistent daily practice.

Attribution analysis of behavioral barriers identified four key barriers and their respective percentages: Inconvenient facilities, such as remote locations and inflexible disposal times (42%); Insufficient knowledge of waste sorting, particularly unclear definitions of complex categories (28%); Perceived high time cost, with the sorting process perceived as cumbersome and time-consuming (18%); and Lack of immediate feedback and positive incentives, making it difficult to maintain motivation (12%). These results systematically elucidate the core targets of behavioral intervention from three dimensions: motivation, ability, and triggering conditions.

3.3. SYSTEM OPTIMIZATION SCHEME

Based on the comprehensive spatial and behavioral analysis results, this study proposes a system optimization scheme integrating "space-technology-behavior," specifically including the following three core components:

Intelligent Sorting Container Design: Addressing the issues of insufficient knowledge and ease of disposal, this study proposes introducing intelligent sorting containers with automatic identification, weighing, and compression functions. The design emphasizes a user-friendly interface, providing immediate sorting guidance to lower the cognitive and operational barriers for residents from the source.

Facility Layout Optimization Recommendations: Based on the spatial diagnostic results, a layout optimization strategy based on the principles of "comprehensive service coverage" and "scenario adaptation" is proposed. This includes adding collection points in identified coverage blind spots and high-density waste concentration areas; and adopting differentiated facility configuration and management models according to the different waste generation characteristics and time patterns of commercial and residential areas.

Behavioral Guidance Mechanism: A multi-layered behavioral guidance system was designed. Effective triggering mechanisms were provided by optimizing the container identification and community notification systems. An intelligent system-

based prompting mechanism was constructed to provide immediate feedback on correct categorized behaviors, thereby enhancing behavioral motivation. Through optimization of the entire system, the time and effort required by residents were minimized, fundamentally empowering them and promoting the long-term sustainability of their behavior.

4. Discussion

This study systematically explores the optimization path of waste recycling systems in high-density urban environments by integrating spatial analysis and behavioral insights. Based on an empirical study in Kowloon City, we delve into key strategies for improving waste sorting efficiency from four dimensions: technological intervention, spatial restructuring, theoretical innovation, and practical application.

4.1. SPACE DIAGNOSTIC RESULTS AND FACILITY OPTIMIZATION PLANS

This study, based on a comprehensive assessment using GIS spatial analysis and heat map overlay, identified three representative waste-generating hotspots in Kowloon City. Through observational data on waste sorting behavior, it revealed the intrinsic correlation between different spatial forms and resident participation.

In public space scenarios (Scenario A), the observed waste sorting participation rate was approximately 20%. These areas mainly consist of street-side rest areas and small public squares, characterized by high mobility and short dwell times. Due to the lack of dedicated waste sorting facilities and the fact that most users are short-term transients, waste sorting behavior relies primarily on individual environmental awareness, resulting in low overall participation.

The waste sorting participation rate in commercial areas (Scenario B) was relatively high, reaching 47%. This area features diverse businesses, including food sort es, retail markets, bookstores, and grocery stores, generating waste primarily consisting of packaging materials, paper products, and kitchen waste. Sound commercial management regulations, centralized waste collection points, and frequent cleaning and maintenance constitute the basic conditions for waste sorting. Simultaneously, the relatively standardized spatial layout of commercial spaces provides favorable conditions for the rational placement of sorting containers.

The situation in residential areas (Scenario C) is particularly noteworthy. Despite the stable and substantial amount of waste generated in residential areas, the participation rate in waste sorting is only 4%. The coverage of waste sorting facilities in private residences and multi-story buildings is low, and the existing recycling point layout fails to effectively align with residents' daily habits. Furthermore, spatial constraints within residential buildings and differences in property management models further hinder the effective implementation of waste sorting systems. This result indicates that relying solely on residents' self-discipline, without corresponding spatial support and systemic guidance, is insufficient to establish an efficient waste sorting mechanism in residential areas.

The comparative analysis of the three scenarios clearly shows that waste sorting participation depends not only on residents' environmental awareness but is also

significantly influenced by spatial function, facility configuration, and management models. Commercial areas, due to their standardized operation and management, perform better in waste sorting; while residential areas and public spaces face bottlenecks in waste sorting due to their respective spatial limitations. This finding provides important empirical evidence for developing differentiated optimization strategies tailored to the characteristics of different scenarios.

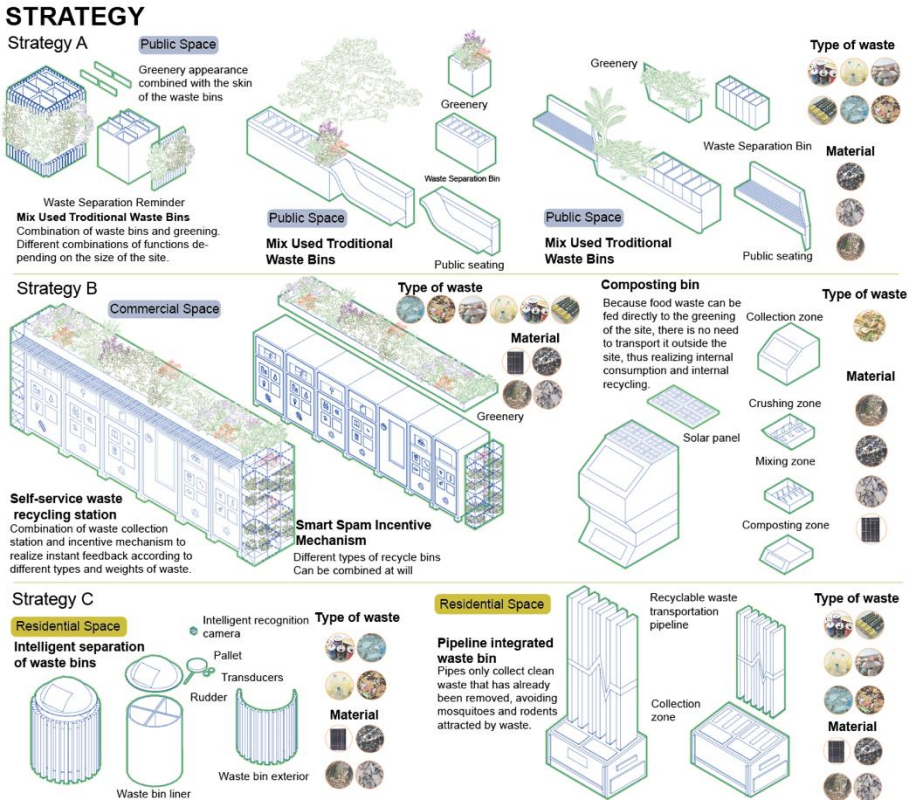


Figure5. Different facilities in different scenarios

4.2. BEHAVIORAL BARRIER ANALYSIS RESULTS

Behavioral barrier analysis further reveals the underlying mechanisms of residents' "high willingness, low action" phenomenon. Insufficient facilities raise the behavioral threshold, vague understanding weakens intrinsic motivation, and lack of feedback weakens the motivation for continued participation. The systematic superposition of these three barriers leads to the dilemma of "knowing what to do, but not doing it" in waste sorting practice. This study empirically demonstrates that only through coordinated intervention of spatial optimization and behavioral guidance, constructing a "convenience-cognition-incentive" three-in-one support system, can the path from willingness to sort waste to sustained action be effectively opened up, and a practical solution for waste management in high-density urban environments be provided. This

chapter aims to explore the significance of the research results, analyze the effectiveness and innovation of the proposed "space-behavior coordinated optimization strategy," and discuss its theoretical contributions, practical value, and future improvement directions.

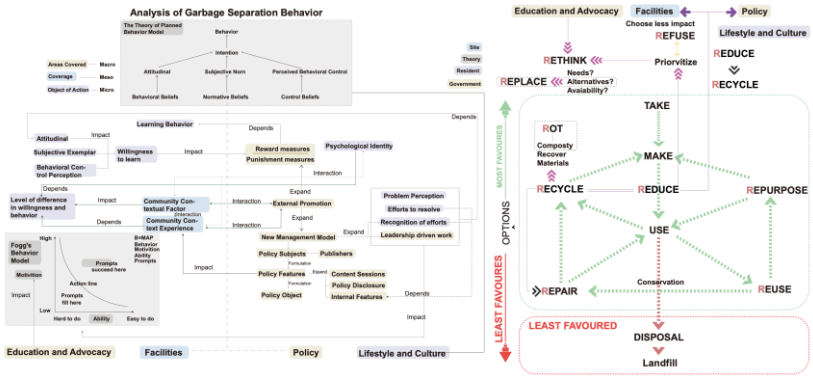


Figure 6. Fogg model and 9R analysis

4.3. SPATIAL LAYOUT OPTIMIZATION SCHEME



Figure 7. Spatial distribution of different facilities

At the spatial strategy level, differentiated facility layout schemes are key to addressing complex urban environments. This study proposes that facility layout should follow the "space-behavior" matching principle and formulate corresponding strategies based on the characteristics of different functional areas: In public spaces, compact recycling facilities should be set up at key rest nodes, taking into account both accessibility and basic functionality; in commercial areas, medium-sized recycling points should be set up at intersections of main roads and internal streets, taking into account shop density and pedestrian flow paths to match the high waste generation; in residential areas, a hybrid "centralized + decentralized" model should be adopted, with basic recycling containers at building entrances and fully functional recycling stations in community centers, forming a multi-layered recycling network. This precise spatial response

strategy can effectively overcome specific challenges in different scenarios.

5. Conclusion

This study proposes a "space-behavior co-optimization strategy" suitable for high-density cities through systematic spatial diagnosis and behavioral analysis, and has completed preliminary verification at both the theoretical framework and empirical analysis levels. The results show that combining precise spatial planning with intelligent behavioral guidance can effectively improve the sorting efficiency and resource recycling rate of urban solid waste.

It should be noted that this study still has certain limitations, mainly that the optimization scheme has not yet undergone large-scale field verification. Although spatial simulation and behavioral analysis provide direction for system optimization, its actual effectiveness still needs to be verified through subsequent pilot deployments and long-term tracking. Future research will conduct controlled experiments in selected communities to evaluate the sustained effectiveness of the intelligent system and spatial reconstruction strategies. Meanwhile, current research mainly focuses on everyday life scenarios, but in the future, it can be further extended to different types of urban functional areas to verify the universality and scalability of the framework and gradually improve the high-density urban waste management system.

Attribution Statement

ChatGPT (OpenAI, 2024) is used to (1) suggest the initial structure of a manuscript and (2) improve the fluency of the text; Grammarly (Grammarly In., 2024) is used to correct spelling and grammar errors.

References

Armayones, M., Gómez-Zúñiga, B., Robles, N., Pousada, M., & Hernández, E. (2020). Relationship between Behavior Fogg Model and Behavior Change Technique Taxonomy 93: Implications for Behavior Designers. .

Chung, S., & Poon, C. (1994). Waste Recycling Policy in Hong Kong. Waste Management & Research, 12, 21 - 32. https://doi.org/10.1177/0734242X9401200104.

Wu, Z., Yu, A., & Poon, C. (2020). Promoting effective construction and demolition waste management towards sustainable development: A case study of Hong Kong. Sustainable Development, 28, 1713-1724. https://doi.org/10.1002/sd.2119.

Wang, Y., Shi, Y., Zhou, J., Zhao, J., Maraseni, T., & Qian, G. (2021). Implementation effect of municipal solid waste mandatory sorting policy in Shanghai.. Journal of environmental management, 298, 113512 . https://doi.org/10.1016/j.jenvman.2021.113512.

Govindan, K., & Zhuang, Y. (2022). Analysis of factors influencing residents' waste sorting behavior: A case study of Shanghai. Journal of Cleaner Production. https://doi.org/10.1016/j.jclepro.2022.131126.

Grammarly inc. (2024) Grammarly: A wrting and Grammar checker App (Version 14.1198.0)[Large Language Modell htps: /app.grammarly.com

Lo, A., & Liu, S. (2018). Towards sustainable consumption: A socio-economic analysis of household waste recycling outcomes in Hong Kong.. Journal of environmental management, 214, 416-425 . https://doi.org/10.1016/j.jenvman.2018.03.029.

OpenAI. (2024). ChatGPT Mode!: GPT-4 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt

From Evaluation to Design

Vector-Similarity Search for Analogue-Based Land-Use Design Exploration

Kiarash Kiany[1] and Jinmo Rhee[2]
[1,2]*School of Architecture, Planning, And Landscape, University of Calgary, Calgary, Canada.*
[1]*kiarash.kiany@ucalgary.ca, 0009-0006-3309-3903*
[2]*jinmo.rhee@ucalgary.ca, 0000-0003-4710-7385*

Abstract. This paper introduces a prototype for decision-support tool that uses embedding-based vector similarity search to help planners move from land-use performance analysis to actionable guidance. Unlike conventional evaluation methods that Identify underperforming areas, the prototype returns context-aware candidate alternatives (analogue parcels) to support planner judgement. The system builds on the Land-Use Fitness Index (LUFI), a fiscal–spatial indicator that highlights parcels requiring attention. It serves as the diagnostic layer: once an underperforming parcel is identified on the map, the prototype allows planners to query it and receive cases of similar parcels that perform better, maintaining contextual similarity while highlighting fiscally stronger analogues. Applied to the Canadian city of Airdrie, the prototype shows how performance evaluation and design exploration can operate within a single workflow, positioning data-driven tools as collaborators in municipal decision-making.

Keywords. Land-Use Planning, Embedding-Based Vector Search, Fiscal Analysis, Spatial Analysis

1. Introduction

Urban land-use evaluation conventionally emphasizes *diagnosis*—identifying whether development aligns with plans or where fiscal and spatial inefficiencies occur. One of the most common approaches to the evaluation is the performance-based approach, which focuses on analytically describing problems. Planners can gain understanding of how a land-use zone performs fiscally and spatially, but lack mechanisms to translate those insights into actionable planning directions.

In earlier phases of this research, we developed the LUFI, a fiscal–spatial diagnostic index that integrates spatial analysis with financial return on investment (ROI) to assess land-use performance. LUFI is used for identifying parcels that underperform within their spatial contexts. The index then guides the planner to parcels where fiscal inefficiencies coincide with spatial mismatches. Still, the index alone does not indicate *how* those areas might improve—reflecting a broader limitation shared by most

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 593-602.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

evaluation systems.

Building on this limitation, this paper addresses the analysis–guidance gap by introducing a prototype decision-support tool—a vector-similarity search system that uses LUFI and spatial attributes to retrieve higher-performing analogue parcels with comparable spatial contexts. The novelty is operational: it turns parcel embeddings into actionable prompts for early-stage zoning review, supporting both within-use benchmarking and context-aware alternatives, while keeping decisions with planners. The prototype is applied to the LUFI dataset for the city of Airdrie, Alberta, Canada, as a test case. The remainder of the paper reviews related work (Section 2), details the method (Section 3), reports the case-study evaluation (Section 4), and concludes with implications and future directions (Section 5).

2. Background

2.1. LAND-USE EVALUATION AND FISCAL ANALYSIS

Land-use evaluation has long emphasised conformance—the extent to which development aligns with adopted plans or zoning (Alexander & Faludi, 1989). While useful for monitoring compliance, conformance captures only part of actual performance. Performance-based evaluation addresses this gap by focusing on measurable outcomes such as density, accessibility, environmental efficiency, and economic productivity (Talen, 1996), often assessed through multi-criteria methods (Chen, 2014). However, these approaches typically depend on complex indicators and expert-assigned weights, raising challenges for replication and for interpreting fiscal implications.

Fiscal analysis offers a more tangible basis for evaluating land-use patterns by comparing tax revenue with the cost of infrastructure and public services (Peterson, 2009). ROI expresses this balance as a benefit-to-cost ratio, enabling comparisons across land-use categories and highlighting where municipal expenditures exceed returns (Burns, 2018). Yet fiscal metrics operate independently of spatial context, limiting their ability to explain how parcel configuration or adjacency influence performance.

2.2. EMBEDDING-BASED VECTOR SIMILARITY SEARCH

Recent advances in urban analytics show that machine-learning models can encode places as compact vector representations—urban embeddings— capturing co-variation in mobility, facilities, land-use structure, and other spatial attributes. Fan et al. (2021) further show that embeddings trained in one metropolitan region explain a substantial share of spatial structure in another, indicating cross-city transferability.

Planning research has begun to apply embedding methods to design-oriented tasks. Case studies show how unsupervised and supervised learning can cluster urban form and evaluate redevelopment scenarios (Podrasa et al., 2021). Multi-view graph learning approaches further integrate human mobility, land-use attributes, and road networks into unified embeddings, enabling more holistic spatial representation (Wang et al., 2020).

Building on these developments, this research encodes each parcel and its

surrounding context as a multidimensional vector and retrieves higher-performing analogues using cosine similarity. Although approximate nearest-neighbour methods such as FAISS are widely used for large-scale vector search (Johnson, Douze, & Jégou, 2017), the prototype uses direct vectorized cosine comparison across the full municipal dataset, which remains efficient at this scale. This adaptation of embedding principles yields a vector-similarity framework that turns diagnostics into exploratory, context-aware land-use suggestions by making latent similarities explicit and enabling comparison of how analogous configurations perform under different land-use decisions.

3. Methods

3.1. LAND-USE FITNESS INDEX

LUFI links fiscal performance with spatial coherence by clustering adjacent parcels of the same land-use category and assessing the consistency of their fiscal outcomes. LUFI was developed and validated previously; its formulation and limitations are detailed in Kiany (2024).

First, parcel-level ROI (*ROIp*) was calculated using municipal data provided by the city of Airdrie. The dataset included parcel geometries, land-use categories, and parcel tax revenue (T_p). A uniform expenditure rate (EX_r) of 2.45\$/m^2 was calculated by dividing the city's annual service budget by its total municipal area. Parcel expenditure (EX_p) was computed by multiplying its area by EX_r. ROI_p was then calculated as T_p minus EX_p, yielding a consistent fiscal measure across all parcels.

Second, ROI_p values were first classified into ten Natural Breaks (Jenks) classes (ROI_c). Adjacent parcels with the same land-use category were dissolved into land-use clusters (C_p). For each C_p, parcel areas were grouped by ROI_c and summed within each C_p. The ROI_c with the largest total area in that cluster was then identified as the dominant ROI_c area. LUFI was then calculated as the ratio of the dominant ROI_c area to the C_p area. LUFI ranges from 0 to 1, where 1.0 indicates complete fiscal consistency and lower values indicate clusters with inconsistent fiscal performance.

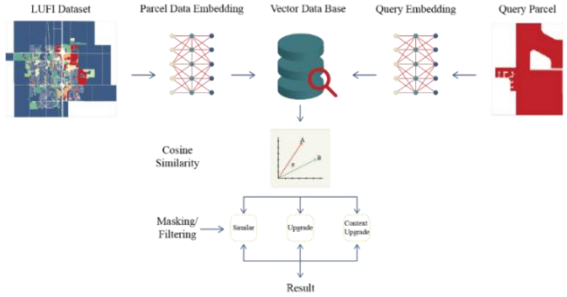
3.2. VECTOR SIMILARITY SEARCH PROTOTYPE

To connect LUFI with spatial attributes and identify which land uses may be most appropriate for each parcel, we develop a decision-support system based on vector similarity search. Within the system, each parcel (P) is represented as a structured fiscal–spatial embedding vector (V) that integrates both intrinsic attributes and contextual attributes.

The intrinsic attributes is a vector (IV), describing parcel-level performance and land-use identity, consisting of three terms: each parcel's standardized ROI ($_{S}ROI$), normalized LUFI values ($_{n}LUFI$), and encoded land use categories (LU) by a one-hot vector—for example, if the land-use categories consist of Commercial, Industrial, Residential, and Public, a parcel classified as Industrial is represented with a one-hot encoding of [0,1,0,0]. In our dataset, dimension of the vector |IV| is 11, corresponding to one value for $_{S}$ROI, one value for $_{n}$LUFI, and a 9-dimensional one-hot land-use vector (9 different categories of land use (Airport, Commercial, Future-Development,

Hybrid, Industrial, Mixed-use, Public, Residential, Rural)).

Figure 1. Overview of the vector-similarity workflow: parcel attributes and neighbourhood context are transformed into fiscal–spatial embeddings, which are compared using cosine similarity to retrieve the closest analogue parcels.



$$IV = [sROI, nLUFI, LU]$$

The contextual attributes is another vector (CV), describing the spatial setting of P and its three nearest parcels (P_i, $i \in \{0,1,2\}$), where P_i 's land-use category (LU_i) and a normalized distance (d_i) between the centroids of P and itself P_i. Thus, each neighbour contributes a $(9+1)$-dimensional block, yielding a total contextual vector of size |CV|=30.

$$CV = [LU_i, d_i], \quad i \in \{0,1,2\}$$

We then apply feature weighting to IV and CV so that different dimensions contribute at different strengths. In particular, assigning a higher weight to the contextual attributes (Wc) increases the influence of local spatial configuration in similarity comparisons relative to the intrinsic attributes (Wi). This ensures that parcels with comparable surrounding land uses—and not only similar fiscal values—are prioritized in the similarity ranking. In our implementation, IV is weighted by Wi=1, and CV by Wc=1.5.

Therefore, by concatenating two weighted vectors,

$$\mathbf{V} = [\underbrace{sROI, nLUFI, LU}_{\text{Intrinsic Attributes}}, \underbrace{LU_0, d_0, LU_1, d_1, LU_2, d_2}_{\text{Contextual Attributes}}]$$

where the dimension of V is 41.

Applying L2-normalization to all parcel vectors, the resulting representations serve as parcel embeddings.

When a query parcel is input into the search system, it is likewise converted into a vector V. Cosine similarity is then computed between the query vector and all parcel embeddings to retrieve parcels with similar feature profiles. The system supports two retrieval modes for different planning tasks. Same-Used Benchmarking and Context-Aware Alternatives.

Same-Use Benchmarking begins by restricting the search space to parcels with the same land-use category (LU) as the query. Within this subset, candidates must initially meet two conditions; (i) their ROI must be at least as high as the query parcel's and (ii) their LUFI must also be equal or higher. If no candidates satisfy both conditions, the system relaxes them in a controlled order—first dropping the LUFI requirement, then

the ROI requirement—so it only returns a fiscally weaker parcel when no better analogue exists. This mode supports incremental, within-zoning benchmarking by identifying stronger-performing parcels located in comparable spatial contexts.

Context-Aware Alternatives scans all embeddings by excluding the query parcel, all parcels with unnecessary land-use types (e.g., Future-development here), outliers—embeddings with drastically lower either ROI or LUFI values. The remaining parcels are then processed as follows. Cosine similarity is computed between the query vector and all remaining parcel embeddings, and the ten most similar parcels are selected as initial candidates. For each candidate, ROI gains are calculated as the difference between its ROI and that of the query parcel, and LUFI gains are computed in the same manner. The candidates are then ranked according to their ROI and LUFI gains, with priority given to ROI, and the highest-ranking parcel is returned. This mode enables planners to select contextually plausible yet fiscally superior alternatives that fall outside the query parcel's current land-use class.

Once the query search is completed, the land-use category of the returned parcel serves as an alternative for the query parcel. This alternative may suggest an adjustment in land-use intensity or a change in land-use category, depending on planning objectives and subsequent expert interpretation.

4. Evaluation

To evaluate the prototype, we conducted a case study using data from the city of Airdrie. The municipal dataset included 27,000 parcel geometries, land-use categories, and parcel-level tax revenues. These data were processed through the ROI and LUFI methods described in Section 3.1, and the resulting maps are shown in Figure 2. After constructing the parcel embeddings, we carried out two evaluations. The evaluation is intended to assess internal coherence and qualitative plausibility of retrieved alternatives, rather than to establish predictive accuracy or normative planning outcomes.

First, we conducted an embedding analysis using dimensionality reduction to assess the quality of the fiscal-spatial vector construction. This analysis examined whether parcels with similar intrinsic and contextual attributes formed coherent structures in the reduced space.

Second, we conducted a spot-sample evaluation using an interface built in Streamlit running the prototype in the back-end. In this evaluation, we qualitatively assessed the returned alternatives by comparing each query parcel with its alternative suggestion. Confirming whether intrinsic and contextual attributes were preserved in the retrieved result.

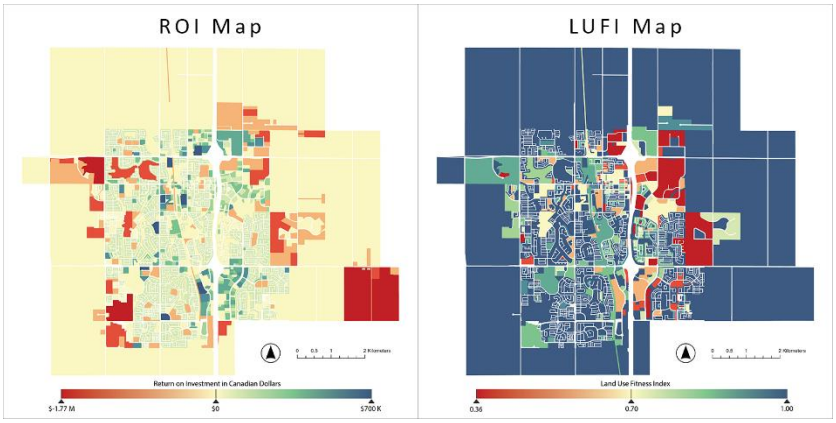


Figure 2. Return on investment(left), and the land-use fitness index maps(right) which demonstrate the parcel level performance of land-uses within the city

4.1. EMBEDDING ANALYSIS

By projecting the 41-dimensional parcel embeddings using UMAP—a dimensionality-reduction technique for visualizing high-dimensional vectors (McInnes, Healy, & Melville, 2018)—we assessed whether the constructed fiscal–spatial embeddings preserved meaningful structure. If the embeddings are well aligned with the intrinsic attributes—sROI, nLUFI, LU—and contextual attributes—LUi, di—parcels with similar profiles should appear in coherent regions of the reduced space.

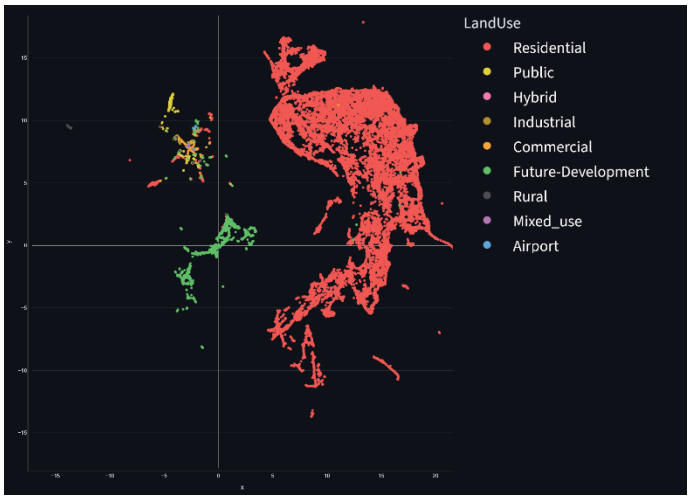


Figure 3. UMAP projection of parcel embeddings showing cluster structure by land-use category.

Figure 3 shows the 2D UMAP projection. Residential parcels form a large, continuous cluster (shown in red), reflecting the spatial regularity and consistent contextual mix typical of suburban residential areas. A distinct cluster of future-development parcels appears below and to the left (green), representing underperforming parcels that share similar contextual surroundings. The upper-left region contains a mixed grouping of commercial, industrial, and mixed-use parcels; this reflects their shared contextual patterns, especially in areas where industrial and commercial parcels are interspersed with public uses. Within this mixed cluster, centrally located points in this cluster correspond to higher-performing commercial and industrial parcels, which aligns with their more favourable fiscal values and contextually consistent environments.

Overall, the UMAP projection reveals coherent clusters aligned with major land-use categories and spatial contexts, with mismatched parcels appearing as outliers. This indicates that the embedding captures structure suitable for similarity and alternative retrieval. This analysis is used as a diagnostic check on embedding structure and does not constitute a performance evaluation of the retrieval system itself.

4.2. SPOT-SAMPLE ANALYSIS

The prototype was implemented in a Streamlit interface that allows users to select a parcel, compute its embedding, and retrieve its closest analogue. In the first example (Figure 4), an industrial parcel was queried, and the system returned a nearby industrial parcel with slightly higher ROI_p and LUFI values and a nearly identical contextual composition. This illustrates how the similarity mode identifies structurally comparable parcels within the same land-use category (Both the query and the alternative parcel share the same three immediate neighbours (Industrial–Industrial–Industrial)).

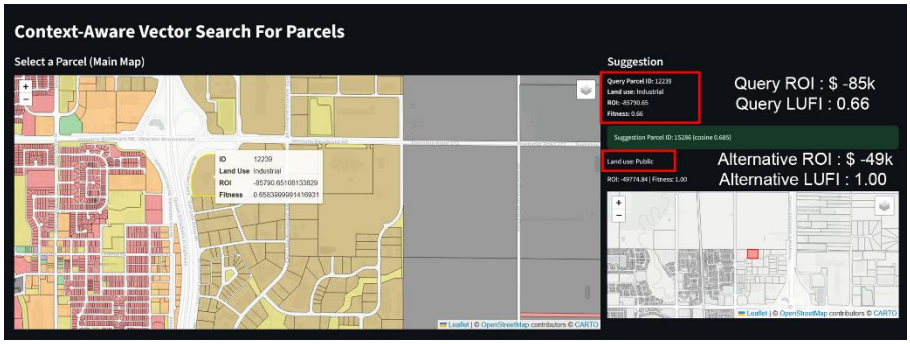


Figure 4. Closest analogue returned for an industrial query parcel, showing a structurally similar industrial site.

Figure 5 shows the Same-Use Benchmarking mode. Selecting the industrial query parcel returns the closest higher-performing industrial analogue whose ROIp and LUFI meet or exceed the query, providing an alternative within the same land-use category under identical zoning.

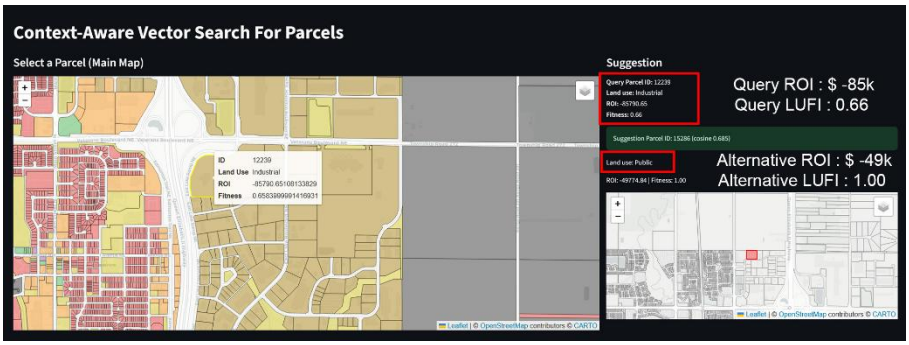


Figure 5. Same-use upgrade result for the industrial parcel.

The context-aware mode widens the search to all land-use categories while preserving contextual similarity (Both the query and the alternative parcel share the same three immediate neighbours (Industrial–Industrial–Industrial)). In Figure 6, querying the same industrial parcel returns a public parcel as the alternative suggestion—identified as the closest higher-performing analogue within a comparable contextual configuration.

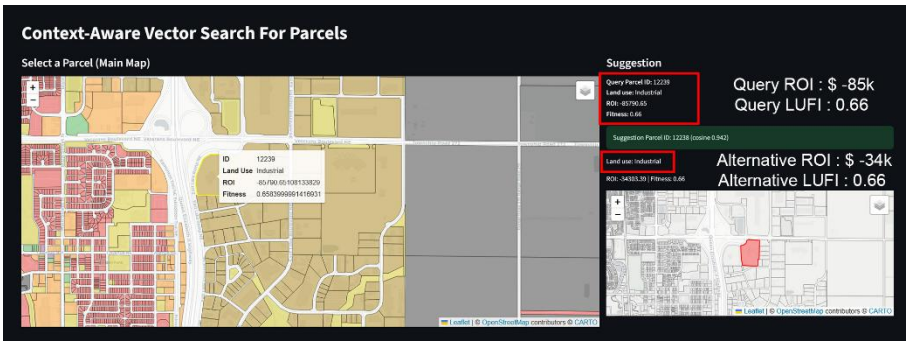


Figure 6. Context-aware alternative for the industrial parcel, returning a higher-performing public parcel in a comparable spatial configuration.

In the final example (Figure 7), a future-development parcel is queried. The prototype compares its contextual attributes (The query parcel's three nearest neighbours include two Industrial parcels and one Commercial parcel, whereas the retrieved alternative parcel is surrounded by three Industrial parcels, indicating a more homogeneous local context in the retrieved case.) to the rest of the dataset and returns a commercial parcel as the alternative suggestion—indicating the existing development type that most closely aligns with its underlying spatial context.

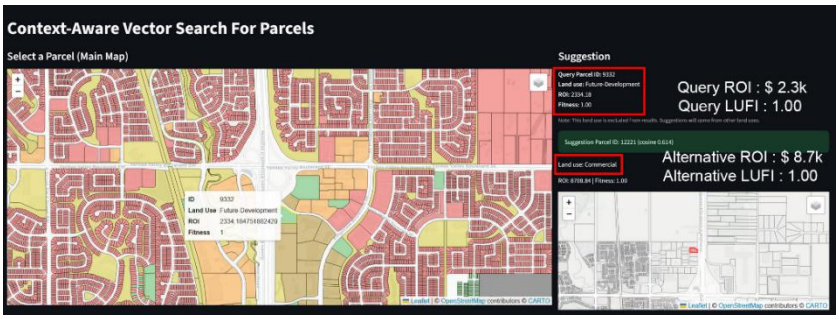


Figure 7. Similarity result for a future development parcel, identifying a commercial configuration as the closest contextual match.

The spot-sample evaluation examined the qualitative plausibility of the alternative land-use suggestions generated by the prototype. Using the Streamlit interface, sampled parcels across different land-use categories were queried, and each retrieved alternative suggestion was compared directly to its query. For each pair, we assessed whether the returned parcel functioned as a credible alternative land use by checking for higher ROI_p and LUFI values and verifying that its contextual attributes aligned with those of the query parcel. Across the examined queries, the prototype returned alternatives that were fiscally stronger than the query parcel and broadly aligned in contextual attributes, indicating feasibility for early-stage decision-support prompting.

5. Conclusion and Future Directions

This paper introduced a prototype that retrieves parcels with similar fiscal–spatial characteristics and returns analogue parcels as candidate land-use alternatives for a query parcel. By integrating both parcel-level intrinsic attributes and the spatial context of immediate neighbours, the system enables planners to compare underperforming parcels with higher-performing analogues elsewhere in the city and use these comparisons to support more context-aware zoning review and amendment.

The returned land-use category functions as an analogue-based prompt for early-stage optioning, to be interpreted through local planning mechanisms rather than treated as an automated prescription. Feasibility depends on constraints (zoning, servicing capacity, transition costs), especially when alternatives cross land-use classes; integrating a policy/servicing constraint filter would improve practical applicability while keeping planners in control.

In practice, the prototype is most relevant for early-stage planning tasks where multiple pathways are plausible, such as targeted rezoning review, identifying candidates for intensification, or prioritizing redevelopment areas for further study. The system does not determine the "best" land use; it surfaces fiscally stronger analogues in comparable contexts so planners can frame and test scenarios with clearer precedent.

The prototype uses heuristic weighting and centroid-based contextual representation, and was assessed on a single municipality with an exploratory evaluation focused on embedding coherence and qualitative plausibility rather than robustness. Future work will expand evaluation using more systematic protocols—such as retrieval stability tests, baseline comparisons, and practitioner review—and

extend testing across additional municipalities to better establish reliability and real-world utility.

References

Alexander, E. R., & Faludi, A. (1989). Planning and plan implementation: Notes on evaluation criteria. *Environment and Planning B: Planning and Design, 16*(2), 127–140.

Batty, M. (2018). *Inventing future cities*. MIT Press.

Burns, L. (2018). How to calculate and visualize value per acre in your city. *Urban Prosperity*.

Chen, J. (2014). Multi-criteria land-use suitability analysis for residential development. *Journal of Urban Planning and Development, 140*(1), 04013002.

Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvàsy, G., Mazaré, P.-É., Lomeli, M., Hosseini, L., & Jégou, H. (2024). The Faiss library. *ScienceDirect*.

Fan, Z., Zhang, Y., Li, X., Zhang, X., & Wang, S. (2021). Learning urban region embeddings from multiple data sources. *arXiv*. arXiv:2110.12371

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 855–864).

Johnson, J., Douze, M., & Jégou, H. (2017). Billion-scale similarity search with GPUs. Facebook AI Research.

Kiany, K. (2024). *Integrating fiscal and spatial analysis in data-driven land-use evaluation* (Master's thesis, University of Calgary). University of Calgary Digital Repository.

Li, Y., Zhang, T., & Chen, Q. (2022). Vector representation of urban morphology using deep graph embeddings. *Computers, Environment and Urban Systems, 95*, 101808.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems, 30* (pp. 4765–4774).

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.* arXiv:1802.03426.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*. arXiv:1301.3781

OpenAI. (2025). ChatGPT (GPT-5 Thinking) [Large language model]. Used for language refinement during manuscript preparation. https://openai.com/chatgpt

Peterson, G. E. (2009). *Unlocking land values to finance urban infrastructure*. World Bank.

Podrasa, R., Vukadinović, S., Wössner, U., & Schrenk, M. (2021). A data-driven planning tool using artificial intelligence to inform urban design scenarios. In *REAL CORP 2021: Proceedings of the International Conference on Urban Planning, Regional Development and Information Society*.

Wang, D., Liu, C., Wang, Y., & Lin, X. (2020). Multi-view joint graph representation learning for urban region embedding. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 4373–4379).

Wang, Y., Wu, F., & others. (2022). Spatial planning implementation effectiveness: Review and research prospects. *Land, 11*(8), 1279. https://doi.org/10.3390/land11081279

Xu, H., Zhao, P., & Shen, Q. (2021). Discovering urban functional zones via POI embeddings and spatial clustering. *Environment and Planning B: Urban Analytics and City Science, 48*(6), 1564–1579.

Zhang, J., Li, P., & Liu, Y. (2023). Urban knowledge graphs and embedding models for city intelligence. *Computers, Environment and Urban Systems, 103*, 101906.

Predicting Pedestrian Flow Spatio-temporal Heterogeneity via Conditional GANs:

Focusing on Urban Street Market Dynamics

Yujing Wu[1], Yuhan Yao[2], Waishan Qiu[3], Yuankai Wang[4] and Jie Shen[5]
[1,5]*Tongji University*
[2] *Politecnico di Milano,*
[3,4] *The University of Hong Kong*
[1]*2510009@tongji.edu.cn, 0009-0001-7426-4434*
[2]*yuhan.yao@mail.polimi.it*
[3]*waishanq@hku.hk*
[4]*u3011546@connect.hku.hk*
[5]*Jieshen@tongji.edu.cn*

Abstract. Street markets significantly influence pedestrian dynamics in high-density cities. However, their irregular spatio-temporal patterns present a challenge for urban prediction and planning. This study asks: How can a data-driven generative model accurately capture and predict the complex, transient pedestrian distributions induced by street market activities? We propose a novel computational framework to manage the high volatility of informal urban activities. A multi-channel Pix2Pix-based conditional generative adversarial network (cGAN) is leveraged to predict heterogeneous pedestrian distributions influenced by street markets in Shanghai. The methodology integrates multimodal urban data, including urban morphology, market influence zones, and real-time crowd data. To enable context-specific, detailed analysis, four temporal sub-models (Full-Day, Morning, Afternoon, Evening) were constructed. The results demonstrate that the Full-Day model achieves high accuracy, performing robustly in market core zones. Performance varies temporally: Morning/Evening models effectively capture structured commuting patterns, while the Afternoon model reflects increased diversity and randomness in midday behavior. The generative approach achieves superior spatial resolution compared to traditional methods. This robust, data-driven tool enables scenario-based simulations for adaptive market siting and supports rapid response strategies critical for resilient planning against informal street dynamics.

Keywords. Urban street markets, pedestrian flow prediction, conditional GAN, spatiotemporal heterogeneity, multimodal urban data

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 603-612.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

1. Introduction

Street markets are a global urban phenomenon essential for enhancing urban vibrancy, fostering social interactions, and driving economic resilience. Global policies, such as Shanghai's 2023 guidelines, now increasingly support standardized flexible work models and street vending activities.

In high-density environments, however, the periodic operation of markets and their embedding within the irregular urban fabric amplify the spatio-temporal heterogeneity of crowd dynamics, leading to unpredictable surges in pedestrian flow that can cause localized congestion and traffic interruptions (Sun et al., 2016). Traditional planning approaches, relying on static models and conventional spatial data (e.g., POIs), are ill-equipped to manage these highly nonlinear systems. Existing vitality prediction methods struggle to analyze the nonlinear interactions between complex spatial configurations and informal economic activities, resulting in planning decisions that lack quantitative assessment of instantaneous spatio-temporal variations in pedestrian flow.

Addressing the need for better planning, this study analyzes how street markets influence pedestrian flow in Shanghai's high-density districts to optimize site selection. We propose an innovative multi-channel Conditional GAN framework (cGAN). The novelty of this framework lies in its methodology: it first establishes a robust data foundation by successfully integrating multimodal urban data (morphology, market influence, and dynamic flow), which are then rigorously encoded at the pixel level. This approach allows the cGAN to overcome the limitations of traditional linear models by capturing the complex, nonlinear interactions between the built environment and transient pedestrian activity. Ultimately, the framework offers enhanced practical value through visualized, real-time planning feedback, aiding urban planners in dynamically adjusting market configuration and capacity.

2. Related Work

2.1. URBAN STREET MARKET (USM)

Urban Street Markets (USMs) are a vital part of the informal economy (Bromley, 2000). Historically, they were often viewed merely as characteristic of less developed regions, but they've gained renewed recognition for their cultural and economic significance (Carpenter, 2018). The central conflict in governing USMs lies in the tension between informal vending and urban governance. Their use of public space (Samuelson, 1954) frequently results in critical issues such as congestion, safety risks, and sanitation (Lata et al., 2019).

Despite the gravity of these urban conflicts and their acknowledged significance, academic research is predominantly qualitative (Carpenter, 2018), lacking systematic empirical, quantitative, or spatial analysis (Bromley, 2000). Studies often focus on conceptual discussions (e.g., regulation) rather than actionable planning insights. Although recent efforts have begun to examine the spatial logic of USMs (Peimani and Kamalipour, 2022) using geo-statistics and GIS (Lan et al., 2021), these remain case-specific and insufficiently integrated with broader urban factors (e.g., land use and transport). Pedestrian flows are key to understanding urban vitality (Wang et al., 2023).

USMs expand the spatial and temporal dimensions of service access (Bromley, 2000). In conclusion, a critical gap remains in urban-scale, data-driven spatial and temporal analysis adequately linking USMs, pedestrian flow, and broader urban dynamics.

2.2. GENERATIVE ADVERSARIAL NETWORK (GAN)

Deep Learning (DL) offers crucial data-driven support for urban planning by extracting spatio-temporal dependencies (Cai et al., 2021; Zhang et al., 2022). GANs (Goodfellow et al., 2014) are a game-theory-based network utilizing adversarial processes to synthesize realistic data, making them powerful for automated design (Boim et al., 2022; Liu et al., 2021). GANs have proven versatile in urban contexts, including morphological simulation (Wang et al., 2025; Zhou et al., 2025), layout generation (Wu and Biljecki, 2023), and dynamic flow prediction (Zhang et al., 2021). However, most studies focus on physical features (buildings, roads), neglecting the influence of social activities on urban vitality (Ye et al., 2018), which manifests in "activity spaces" like street markets (Carmona, 2003). To bridge this gap, this research trains a GAN model to simulate pedestrian flow using both physical and social variables to assess the impact of street vending.

3. Methodology

3.1. RESEARCH FRAMEWORK

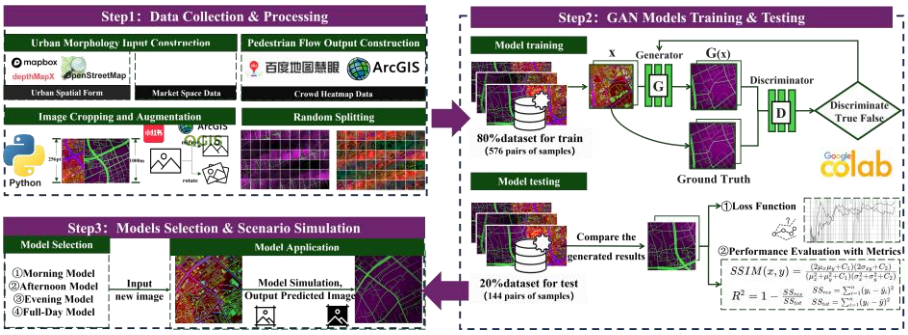


Figure 1. Workflow diagram of the Multi-channel cGAN

The study follows three main steps (Figure 1):

- Data collection and processing. We collected foundational urban space and pedestrian flow samples, focusing on market activity zones to capture necessary spatiotemporal context. The dataset was augmented using Python for robust training.
- GAN models training and testing. The Pix2Pix neural network was employed to train models using paired urban and flow images. Performance was evaluated via quantitative metrics to assess accuracy and generalization.
- Model selection and scenario simulation. New spatial configurations were input into the trained model to automatically generate pedestrian flow predictions for various planning scenarios.

3.2. DATA PROCESSING AND DATASET CONSTRUCTION

The research utilized three primary data sources focused on Shanghai's high-density inner ring. (1) Urban morphology data, detailing roads and building height, was acquired from Mapbox and OSM Buildings. (2) Street market space data, defining 29 distinct market events, was generated through manual digitization of market polygons. Pedestrian flow data was incorporated using the Baidu Huiyan API, providing grid resolution at 200 × 200 meters.

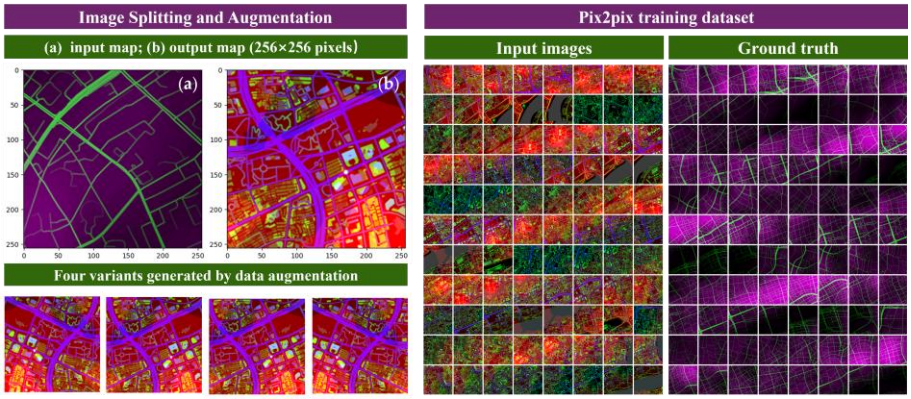


Figure 2. Standardized Dataset Construction

Input map overlaid urban elements assigned unique 8-bit RGB colors for GAN processing (Li and Zheng, 2023). This composite map rigorously encoded building attributes (function/height), modeled road structure using Space Syntax integration indices, and quantified street market influence via isochrone service area buffers ranging from 300m–2500m.

The pedestrian flow data was pre-processed using Bayesian Kriging interpolation to create the output map, where the pixel RGB values reflect crowd density (people/km^2). To capture daily spatio-temporal heterogeneity, the day was divided into four representative time periods by averaging data from specific hourly data, generating four corresponding output maps (Figure 2).

The final Pix2Pix training dataset consists of 2880 paired image samples, generated by segmenting the input and output maps into standardized 256 × 256 pixel patches, each representing a 1000 × 1000m urban area. After initial filtering, 720 high-quality pairs were retained and augmented in real-time through flipping and normalization to generate the final samples (80% for training, 20% for testing).

3.3. EXPERIMENTAL MODEL FRAMEWORK

3.3.1. The GAN Architecture

Given the spatio-temporal heterogeneity of street market activities, this study trains four separate models: the Morning Model (MM), Afternoon Model (AM), Evening Model (EM), and Full-Day Model (FDM), corresponding to distinct pedestrian flow

heatmaps. The trained models are validated on separate test data against real values to confirm accuracy.

The framework employs a CGAN to enhance the model's controllability by incorporating conditional inputs. The Pix2Pix (Isola et al., 2017), a CGAN-based algorithm, is selected for its efficiency in image-to-image translation, enabling the synthesis of targeted pedestrian flow maps (y) directly from urban space images (x) (Li and Zheng, 2023). The training process proceeds until a Nash equilibrium is established, where the discriminator can no longer reliably distinguish between real and generated samples (Goodfellow et al., 2014; Aithal et al., 2020).

3.3.2. Structure and Loss Functions

The Generator (G) adopts a U-net architecture with skip connections (Chen et al., 2024) to produce the output: $y = G(x, z)$. The Discriminator (D) utilizes a PatchGAN structure that evaluates the realism of local image patches, outputting a probability that the input pair is real (Isola et al., 2017). The training objective combines an adversarial loss ($\mathcal{L}_{cGAN}$) with pixel-wise reconstruction losses. While the $L1$ loss (Mean Absolute Error, MAE) is employed to promote visual sharpness and suppress blurring, the $L2$ loss (Mean Squared Error, MSE) is simultaneously incorporated to enhance structural consistency and mitigate overfitting. The final objective for the generator (G) is formulated as (1):

$$G^* = \arg\min_G \left[\max_D \mathcal{L}_{CGAN}(G, D)\right] + \lambda \mathcal{L}_{L1}(G) + \beta \mathcal{L}_{L2}(G) \quad (1)$$

where λ and β are hyperparameters balancing the reconstruction terms. Following established benchmarks in urban morphology generation (Zhou et al., 2023), we set λ = 100 and β = 150. These values were empirically selected to align generated flow maps with the input spatial features while ensuring pixel-level accuracy.

3.3.3. Accuracy Assessment

Model generalization and accuracy are evaluated using a unified test set through both qualitative inspection of visual realism and four quantitative metrics: Structural Similarity Index (SSIM), Coefficient of Determination (R^2), Fréchet Inception Distance (FID), and Kernel Inception Distance (KID) (Wang et al., 2025). SSIM measures the structural and visual similarity between the generated and reference images, while R^2 reflects the correlation and goodness-of-fit between predicted and actual values. The FID and KID metrics are employed to evaluate the statistical distance between the deep feature distributions of the real and generated images. Lower values for FID and KID indicate higher fidelity and realism in the generated outputs.

4. Results

4.1. MODEL PERFORMANCE AND TRAINING BEHAVIOR

Four temporal Pix2Pix models were trained to capture the dynamic heterogeneity of pedestrian flows across daily cycles. Analysis of the training loss curves (Figure 3) revealed the models' dynamics. FDM achieved the strongest overall performance and stability, recording the lowest total generator loss and $L1$ loss. This demonstrates an optimal balance between adversarial effectiveness and image reconstruction fidelity for full-day prediction. Conversely, the high generator loss of MM suggested limitations in training effectiveness, while AM exhibited the weakest structural recovery based on its highest $L1$ loss, likely due to greater spatial heterogeneity during the afternoon period.

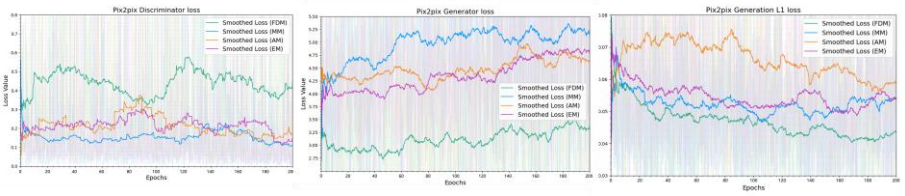
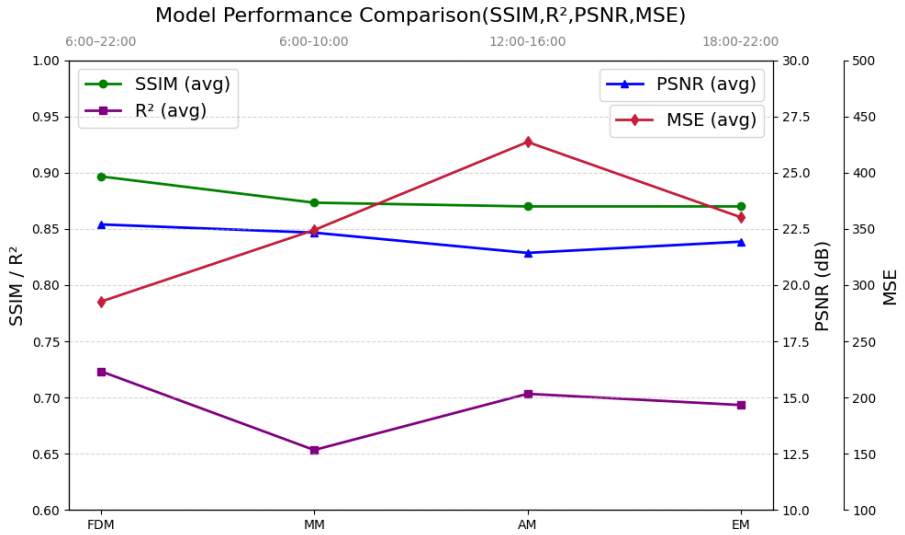


Figure 3. Training process represented by loss curves of Pix2Pix

4.2. QUANTITATIVE EVALUATION OF FOUR MODELS



The generative performance was evaluated across time and four hierarchical spatial zones defined by market service radius (Core, Adjacent, Transitional, and Peripheral).

Figure 4. Model performance comparison (SSIM, R^2, PSNR, MSE)

FDM consistently outperformed time-specific models in the quantitative assessment (Figure 4), achieving the highest average SSIM (0.887), R^2 (0.723), and PSNR (22.740), along with the lowest MSE (286.365). This indicates that aggregating pedestrian dynamics over an entire day yields a more stable and learnable signal. Temporally, all models showed the weakest predictive capacity during the MM (R^2 of 0.653), reflecting the stochasticity of early-day commuting. Spatially, the Core service zone (300–1000m) showed the highest prediction accuracy (R^2: 0.717 to 0.906), confirming that areas near market cores exhibit more regular and learnable flow patterns, while the Adjacent zone (500–1500m) presented the greatest modeling challenge.

4.3. VISUAL-BASED QUALITATIVE EVALUATION

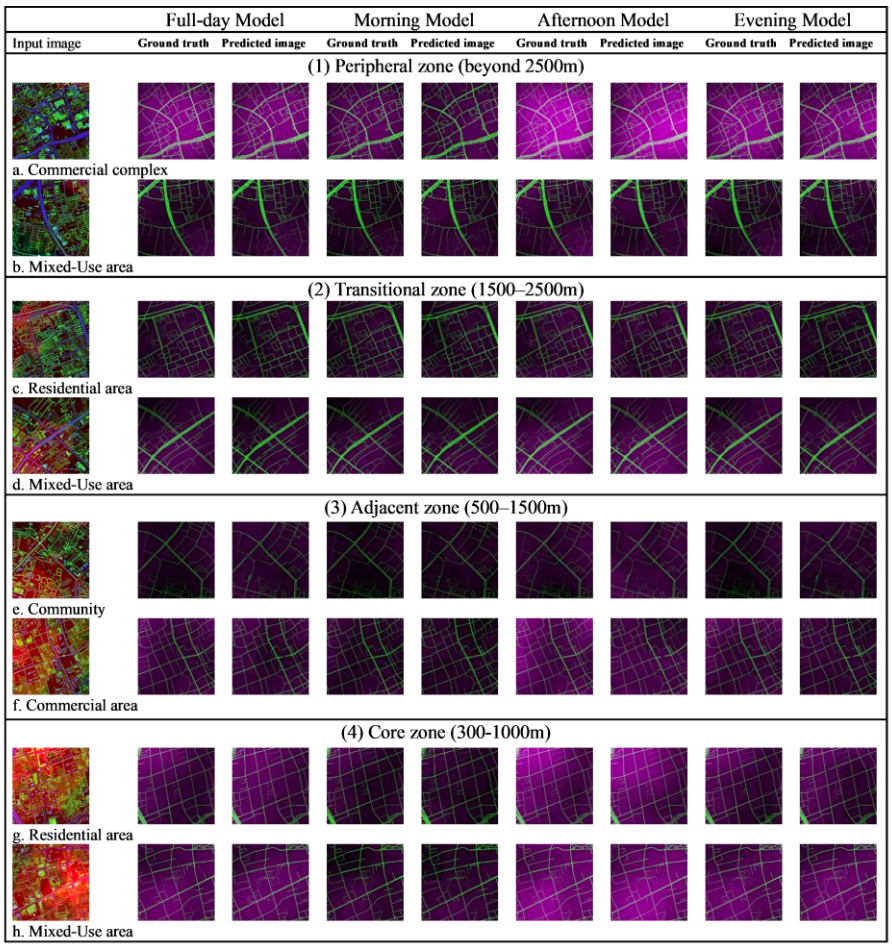


Figure 5. Re-generated images of the test set using four models

Visual analysis (Figure 5) confirmed that all models successfully replicated road

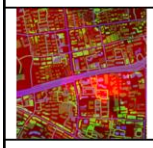
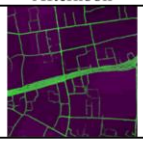
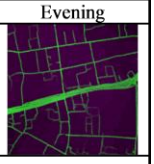
networks and captured the expected concentric clustering of pedestrian density near street markets, with FDM showing superior continuity. However, models tended to underestimate pedestrian activity in residential zones with low functional mix, suggesting they overly prioritize market-centric attractions. Model performance varied temporally; the AM showed the most noticeable decline, attributed to the complexity and increased randomness of midday activities (such as diverse leisure and dining movements) which challenged the model's predictive stability. Regarding practical application, FDM is best suited for dynamic, full-day monitoring, while MM and EM are ideal for short-term forecasting and managing crowds during peak hours.

5. Discussion and Conclusions

This study successfully utilized a multi-channel Pix2Pix-based cGAN to predict pedestrian flow distribution under street market influence, demonstrating its ability to integrate multimodal urban data and capture spatio-temporal heterogeneity.

The model's performance highlights a crucial spatial logic: accuracy strongly correlates with market influence, with the FDM achieving superior overall accuracy, particularly in core service zones. This success reframes street markets as dynamic, integrable elements and spontaneous spatial mechanisms that activate urban edges. This aligns with global trends advocating for the strategic integration of informal economies (Ray, 2020), supporting the view that markets are bottom-up catalysts of spatial activation and essential components of resilient urban systems (Wang et al., 2023; Ye et al., 2018).

Pedestrian flows simulation with street markets (1 km2)

Input Image	Predicted Results				
	Time	FullDay	Morning	Afternoon	Evening
	**Images**				
	RGB range	(61,0,61)-(69,0,69)	(36,0,36)-(45,0,45)	(61,0,61)-(79,0,79)	(50,0,50)-(64,0,64)
Predict people/km²		1984	1500	2266	1984

Pedestrain flows simulation without street markets (1km²)

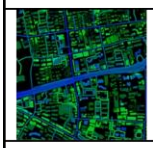
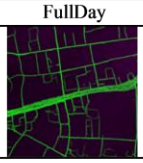
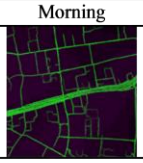
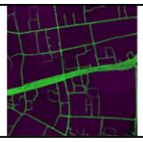
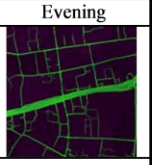
Input Image	Predicted Results				
	Time	FullDay	Morning	Afternoon-	Evening
	**Images**				
	RGB range	(33,0,33)-(51,0,51)	(33,0,33)-(43,0,43)	(43,0,43)-(61,0,61)	(33,0,33)-(43,0,43)
Predict people/km²		1500	1278	1732	1500

Figure 6. Simulated Pedestrian Flows with and without Street Markets (1 km²)

Scenario simulation confirmed that the presence of street markets consistently raises pedestrian density (predicted increase of 17% to 32%), particularly during peak

afternoon periods (Figure 6). These findings inform adaptive zoning strategies: high-density surges require dynamic capacity adjustments (e.g., widening pedestrian buffers), while lower-impact areas support flexible siting to activate underused spaces.

Future research will focus on enhancing the model's robustness and generalizability. This includes exploring transfer learning techniques to validate its effectiveness across diverse urban settings beyond Shanghai. To improve temporal realism, dynamic contextual factors such as weather and special events must be integrated. Furthermore, evaluation metrics will shift from purely image quality assessment toward measuring the model's functional relevance for actionable urban planning decisions.

Attribution Statement

ChatGPT (OpenAI, 2025) was used to (1) improve flow, (2) correct errors in spelling and grammar, and (3) enhance the visual presentation and clarity of Figure 4. The authors verified all AI-suggested edits to ensure scientific accuracy and original intent.

References

Aithal, B. H., Das, S. K., and Subrahmanya, P. P. (2020). Urban 3D structure reconstruction through a generative adversarial network model. Arabian Journal for Science and Engineering, 45(12), 10731–10741. https://doi.org/10.1007/s13369-020-04850-7

Boim, A., Dortheimer, J., and Sprecher, A. (2022). A machine-learning approach to urban design interventions in non-planned settlements. CAADRIA Proceedings,Proceedings of the 27th Conference on Computer Aided Architectural Design Research in Asia (CAADRIA) [Volume 1], 1, 223–232. https://doi.org/10.52842/conf.caadria.2022.1.223

Bromley, R. (2000). Street vending and public policy: A global review. International Journal of Sociology and Social Policy, 20(1/2), 1–28. https://doi.org/10.1108/01443330010789052

Cai, C., Guo, Z., Zhang, B., Wang, X., Li, B., and Tang, P. (2021). Urban morphological feature extraction and multi-dimensional similarity analysis based on deep learning approaches. Sustainability, 13(12), 6859. https://doi.org/10.3390/su13126859

Carmona, M. (2003). Public places - urban spaces: The dimensions of urban design, second edition. Retrieved from https://www.semanticscholar.org/paper/Public-Places-Urban-Spaces-%3A-the-dimensions-of-Carmona/d51cd2b1ee3c7ea84bef63bdaac05f382a2cf084

Carpenter, D. M. (2018). Street vending in the united states: A unique dataset from a survey of street vendors in America's largest cities. Cityscape, 20(3), 245–256.

Chen, R., Zhao, J., Yao, X., He, Y., Li, Y., Lian, Z., … Li, H. (2024). Enhancing urban landscape design: A GAN-based approach for rapid color rendering of park sketches. https://doi.org/10.3390/land13020254

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014, June 10). Generative Adversarial Nets. Presented at the Neural Information Processing Systems. Retrieved from https://www.semanticscholar.org/paper/Generative-Adversarial-Nets-Goodfellow-Pouget-Abadie/86ee1835a56722b76564119437070782fc90eb19?p2df

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5967–5976. https://doi.org/10.1109/cvpr.2017.632

Lan W., Hao C., Jiang Y., and Duan J. (2021). Research on Spatial Characteristics of Morning Street Vendors in the Main Urban Area of Nanjing. Modern Urban Research, (2), 14–19.
Lata, L., Walters, P., and Roitman, S. (2019). A marriage of convenience: Street vendors' everyday accommodation of power in dhaka, bangladesh. Cities, 84, 143–150. https://doi.org/10.1016/j.cities.2018.08.002
Li, Q., and Zheng, H. (2023). Prediction of summer daytime land surface temperature in urban environments based on machine learning. Sustainable Cities and Society, 97, 104732. https://doi.org/10.1016/j.scs.2023.104732
Liu, Y., Luo, Y., Deng, Q., and Zhou, X. (2021). Exploration of campus layout based on generative adversarial network. Proceedings of the 2020 DigitalFUTURES, 169-178
OpenAI. (2025). ChatGPT Model: GPT-4 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/
Peimani, N., and Kamalipour, H. (2022). Mapping the spatiality of informal street vending. Journal of Urbanism: International Research on Placemaking and Urban Sustainability, 0(0), 1–20. https://doi.org/10.1080/17549175.2022.2150267
Ray, K. (2020). SOAS address: Rethinking street vending. Gastronomica, 20(1), 1–15. https://doi.org/10.1525/gfc.2020.20.1.1
Samuelson, P. A. (1954). The pure theory of public expenditure. The Review of Economics and Statistics, 36(4), 387. https://doi.org/10.2307/1925895
Sun, Z., Bell, S., and Scott, I. (2016). How does street vending contribute to walkability? A report on a study in yuncheng, China. Environment-Behaviour Proceedings Journal, 1(4), 203. https://doi.org/10.21834/e-bpj.v1i4.166
Wang, M., Xiong, Z., Zhao, J., Zhou, S., Wang, Y., Ikram, R. M. A., … Tan, S. K. (2025). Generative adversarial networks for climate-sensitive urban morphology: An integration of Pix2Pix and the cycle generative adversarial network. Land, 14(3), 578. https://doi.org/10.3390/land14030578
Wang, Y., Qiu, W., Jiang, Q., Li, W., Ji, T., and Dong, L. (2023). Drivers or pedestrians, whose dynamic perceptions are more effective to explain street vitality? A case study in guangzhou. Remote Sensing, 15(3), 568. https://doi.org/10.3390/rs15030568
Wu, A. N., and Biljecki, F. (2023). InstantCITY: Synthesising morphologically accurate geospatial data for urban form analysis, transfer, and quality control. ISPRS Journal of Photogrammetry and Remote Sensing, 195, 90–104. https://doi.org/10.1016/j.isprsjprs.2022.11.005
Ye, Y., Li, D., and Liu, X. (2018). How block density and typology affect urban vitality: An exploratory analysis in shenzhen, China. Urban Geography, 39(4), 631–652. https://doi.org/10.1080/02723638.2017.1381536
Zhang, W., Ma, Y., Zhu, D., Dong, L., and Liu, Y. (2022). MetroGAN: Simulating urban morphology with generative adversarial network. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. https://doi.org/10.1145/3534678.3539239
Zhang, Y., Wang, S., Chen, B., Cao, J., and Huang, Z. (2021). TrafficGAN: Network-scale deep traffic prediction with generative adversarial nets. IEEE Transactions on Intelligent Transportation Systems, 22(1), 219–230. https://doi.org/10.1109/tits.2019.2955794
Zhou, S., Wang, Y., Jia, W., Wang, M., Wu, Y., Qiao, R., and Wu, Z. (2023). Automatic responsive-generation of 3D urban morphology coupled with local climate zones using generative adversarial network. Building and Environment, 245, 110855. https://doi.org/10.1016/j.buildenv.2023.110855
Zhou, S., Xu, X., Xu, H., Zhao, Z., Yuan, H., Wang, Y., … Wu, Z. (2025). From heat resilience to sustainable co-benefits: Adaptive urban morphology generation based on multimodal data fusion and a novel generative framework. Sustainable Cities and Society, 127, 106452. https://doi.org/10.1016/j.scs.2025.106452

Who's Strolling Under the Sun

Study on Adaptive Behaviors in Urban Public Spaces Under High-Temperature Exposure

Yujing Wu[1], Shiqi Dai[2], Zanwen Ma[3], Menghan Wang[4], Xintong Lyu[5], Waishan Qiu[6] and Chongyang Ren[7]
[1]*Tongji University*
[2] *Tsinghua University*
[3,5,6] *The University of Hong Kong*
[4] *Shenzhen University*
[7]*University College London*
[1]*2510009@tongji.edu.cn, 0009-0001-7426-4434*
[2]*daisq220@gmail.com, 0009-0009-0602-0217*
[3]*u3643068@connect.hku.hk*
[4]*2450111003@mails.szu.edu.cn*
[5]*lyuxt24@connect.hku.hk, 0009-0002-3236-8291*
[6]*waishanq@hku.hk, 0000-0001-6461-7243*
[7]*chongyang.ren.25@ucl.ac.uk, 0009-0008-3330-834X*

Abstract. Urban heat stress increasingly impedes outdoor public life, yet the micro-scale dynamics between thermal discomfort and spatial attraction remain underexplored. This study investigates the "heat–behavior" coupling mechanism through a comparative analysis of Piazza Navona (Rome) and Senado Square (Macau). Integrating ENVI-met UTCI simulations with a computer vision-based tracking framework (YOLOv8/DeepSORT), we quantified pedestrian spatiotemporal responses under high-temperature exposure (UTCI > 32°C). Individual behavioral attributes, including age, gender, moving state, solo activity, and adaptive strategy, were extracted from six one-hour segments, then spatialized and aggregated into a 2m grid for cross-site and cross-temperature comparison. While findings acknowledge a general trend of activity contraction under heat stress, the study reveals a critical divergence in adaptive logic: Piazza Navona exhibited a "Climate-Dominant" avoidance pattern where users retreated to shaded edges, whereas Senado Square demonstrated "Thermal Resilience", where strong landscape attractors counterbalanced thermal discomfort, sustaining high activity intensity despite severe heat. The research further identifies demographic-specific adaptation thresholds, with youth populations showing higher thermal tolerance. The proposed framework provides actionable evidence for climate-adaptive design, demonstrating that enhancing landscape attractiveness can effectively elevate the thermal tolerance threshold of public spaces.

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 613-622.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

Keywords. Urban public space, high-temperature exposure, adaptive behavior, computer vision

1. Introduction

Since the beginning of the 21st century, global warming and intensified heatwaves have led to a continuous rise in surface temperature, significantly increasing urban heat risks (Meehl & Tebaldi, 2004). Combined with ongoing urbanization—projected to add 2–3 billion new urban residents by 2050 (Stone, Jr, 2012), the scope and intensity of extreme heat stress are expected to expand further (Mora et al., 2017).Simultaneously, the societal demand for safe, inclusive, and high-quality public spaces continues to grow, as emphasized by the United Nations Sustainable Development Goal (SDG 11.7).

Classical space–behavior theories (Gehl, 1987; Lynch et al., 1960; Whyte, 1980) have long established that spatial quality is the primary driver of public life. However, these foundational theories were largely developed under "neutral" climatic assumptions. In the context of climate change, environmental comfort—specifically thermal stress—has emerged as a critical disruptor that competes with spatial attractiveness (Yang et al., 2024). This convergence presents a key theoretical tension: Under high-temperature stress, how do users mediate the trade-off between thermal discomfort and social interaction?

Recent studies highlight that microclimatic variations directly affect pedestrians' route choices, willingness to stay, and subjective thermal comfort (Silva et al., 2024). The microclimate is strongly influenced by urban morphology, exhibiting significant spatiotemporal heterogeneity (Kim & Brown, 2023; Rahman et al., 2021). Under intense radiation, shaded zones become strong attractors for social activity, while exposed areas are quickly avoided (Levenson et al., 2025). However, while traditional thermal assessments based on weather stations or remote sensing often miss fine-scale variations (Wu et al., 2025), pedestrians primarily experience heat at the micro level. Furthermore, thermal stress is heterogeneous: older adults exhibit lower heat tolerance (Kenny et al., 2010), and high-intensity activities increase physiological thermal load. Consequently, adaptive behaviors (such as seeking shade or avoiding peak hours) are key to reducing individual heat stress (Zhao et al., 2025).

Despite progress, three key gaps persist: (1) Conceptual: current studies often overlook "thermal resilience" driven by spatial attraction, treating heat merely as a deterministic repellent; (2) Demographic: heterogeneity in heat tolerance remains underexplored; and (3) Methodological: traditional observation fails to quantify the fine-grained coupling between microclimate and behavioral dynamics.

The emergence of AI and computer vision, utilizing tools like YOLO and DeepSORT, shifts the paradigm toward data-driven analysis, enabling the automated quantification of these complex dynamics (He et al., 2025). To address these gaps, this study develops an AI-assisted environment–behavior coupling framework. Integrating ENVI-met UTCI simulations with computer vision, we analyze Piazza Navona (Rome) and Senado Square (Macau) to investigate the non-linear trade-offs between thermal stress and public life. This study aims to move beyond simple validation of heat avoidance to reveal the underlying mechanisms of adaptive behavior, providing

evidence for climate-adaptive urban design that balances physical comfort with spatial vitality.

2. Methodology

2.1. RESEARCH FRAMEWORK

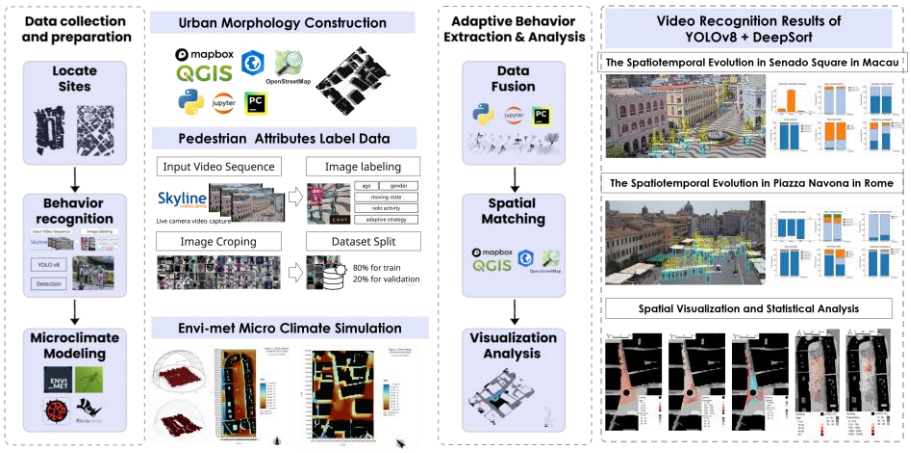


Figure 1. Workflow diagram

To bridge the gap between macro-scale urban heat island effects and micro-scale behavioral adaptation, this study develops a spatiotemporal coupling framework (Figure 1), which consists of three core modules:

- Microclimate Simulation Module. The ENVI-met model is employed to simulate the thermal environment, utilizing the Universal Thermal Climate Index (UTCI) to quantify physical heat exposure conditions.
- Behavior Recognition and Classification Module. Using YOLOv8 and DeepSORT on public video feeds, pedestrian detection and tracking are executed to enable individual-level trajectory extraction and behavior classification. This module yields attributes such as demographic composition, activity intensity, and spatial aggregation patterns.
- Spatiotemporal Coupling Analysis Module. Data from the previous modules are integrated and visualized (via thermal maps, behavioral statistics, and coupling diagrams) to characterize the correlations between microclimatic conditions and pedestrian adaptive behavioral patterns.

2.2. DATASET CONSTRUCTION

2.2.1. Study Area and ENVI-met Modelling

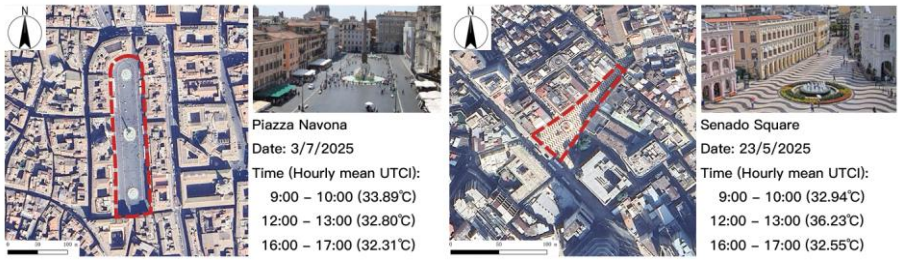


Figure 2. Study area

This study selected Senado Square (Macau, China) and Piazza Navona (Rome, Italy) as comparative case sites based on three criteria (Figure 2): Thermal comparability (average daily temperature ≈ 32°C); morphological consistency (similar elongated forms and widths of ≈ 50m); and cultural contrast (Asian vs. European contexts), offering a comparative framework for thermal adaptation strategies. To examine behavioral patterns associated with varying heat loads, three typical daytime periods were analysed: 09:00–10:00, 12:00–13:00, and 16:00–17:00. Although all periods may exceed 32°C, variations in solar altitude and shading distribution lead to distinct spatial patterns of sun–shade, allowing for an examination of the relationship between dynamic thermal conditions and behavioral adaptation.

Thermal simulations were conducted using ENVI-met version 5.6.1. Site plans of Senado Square and Piazza Navona were imported as BMP base maps for model construction. The modelling process sequentially integrated buildings, vegetation, and ground surfaces. The simulation domains were set to 70 m × 100 m × 40 m for Senado Square and 70 m × 180 m × 40 m for Piazza Navona, with spatial resolutions of dx = 2 m, dy = 2 m, dz = 1 m.

Among various indicators of heat exposure, such as meteorological observations , remote-sensing-derived land surface temperature , and composite thermal comfort indices—the Universal Thermal Climate Index (UTCI) provides the most comprehensive assessment. UTCI integrates multiple meteorological parameters (air temperature, wind speed, relative humidity, mean radiant temperature) and human physiological variables (metabolic rate, clothing insulation), thereby reflecting actual thermal perception more accurately. The simplified formula is:

$$UTCI = T_a + f(U, RH, M, R) \quad (1)$$

T_a is the air temperature (°C), U is the wind speed (m/s), RH is the relative humidity (%), M is the metabolic rate (W/m²), and R is the radiation index.

Regarding the evaluation standard, UTCI > 32°C (defined as "Strong Heat Stress") was adopted as the high heat exposure threshold for this study (Binarti et al., 2020). While acknowledging that subjective tolerance varies with acclimatization (e.g., Macau residents vs. Rome tourists), this metric serves as a consistent physical baseline to quantify environmental heat potential across distinct climatic contexts. The BIO-met

module within ENVI-met was used for the pixel-wise computation of UTCI.

2.2.2. Computer Vision Data Collection and Processing

Publicly accessible SkylineWebcams were used to capture pedestrian activities in Senado Square and Piazza Navona during three representative daytime periods. Each one-hour video stream was converted into sequential image frames at 15 fps, and minor cropping was applied to remove occluded regions.

To extract detailed behavioral information, a combined YOLOv8–DeepSORT framework was implemented for pedestrian detection and multi-object tracking (MOT). YOLOv8 provided frame-level pedestrian bounding boxes (Hussain, 2024), while DeepSORT maintained identity consistency, enabling continuous trajectory extraction (Chen et al., 2024). Five key pedestrian attributes—gender, age, moving state, solo activity, and adaptive strategy—were manually defined and annotated using the CVAT platform. Approximately 1,000 labeled images were prepared for each attribute category, sampled via FFmpeg and manually annotated. These datasets were split into training and validation subsets at a 3:1 ratio to train five independent YOLOv8n-cls classifiers.

Validation performance varied across categories (Table 1). Given the inherent complexity in classifying 'Age' and 'Strategy', these attributes were interpreted as aggregated statistical distributions rather than individual diagnostics. This macroscopic approach mitigates the impact of instance-level classification noise on the overall behavioral patterns. For spatial analysis, homography transformation was applied using ground control points to convert image-space coordinates into real-world metric coordinates. These georeferenced trajectories were then used for spatial aggregation and statistical analysis (e.g., calculating visit frequency and assigning attribute modes) essential for subsequent spatiotemporal analysis.

Table 1. Validation Performance Metrics of the Five Pedestrian Behavioral Attribute

Attribute	Class labels	Val. accuracy (%)	Weighted F1 (%)
Gender	male; female	85.5	85.5
Age	0 – 14; 15 – 30; 30 – 45; 45 – 60; 60+	76.8	76.7
Moving state	sitting; standing; squat; walking	96.4	96+
Solo activity	enjoying scene; playing phone; taking photo; with stroller; with suitcase	95.3	95+
Adaptive strategy	none; holding umbrella; wearing hat; wearing sunglasses	69.0	68.2

2.3. HEAT–BEHAVIOR COUPLING ANALYSIS

A 2 m × 2 m fishnet grid was generated in ArcGIS Pro to standardize the spatial resolution of all data. Pedestrian attributes (visit frequency and five behavioral categories) were aggregated per grid using the mode, and UTCI mean values were joined through spatial overlay. All thermal and behavioral variables were merged into

six consolidated spatiotemporal layers covering 10:00, 12:00, and 16:00 for both plazas. UTCI and behavioral modes were visualized through grid-based thematic maps to examine spatial correspondence. The attribute tables were exported as CSV for cross-site and cross-period statistical comparisons under different thermal conditions. Results were interpreted alongside video observations and local micro-environment features to characterize the heat–behavior coupling dynamics.

3. Results

3.1. THERMAL ENVIRONMENT SPATIOTEMPORAL PATTERNS

This section compares the UTCI distribution of Piazza Navona and Senado Square during summer high-temperature periods (Figure 3). UTCI levels range from Slight heat (26–32°C) to Moderate (32–38°C) and High heat (38–46°C) (Binarti et al., 2020).

Piazza Navona shows rapid heating but strong spatial contrast modulated by alternating building shadows. Most areas fall within Moderate heat, with High heat occurring locally. Morning shading mitigates heat on the east side; by noon, Moderate heat becomes widespread, and by 16:00, only the eastern sun-exposed zone reaches High heat. In contrast, Senado Square is characterized by consistently higher heat stress. Its urban canyon morphology facilitates persistent heat accumulation, with Moderate–High heat dominating all periods. At noon, nearly the entire square reaches high thermal stress levels, while only narrow shaded margins remain cooler. Afternoon shading reduces but does not eliminate high-heat zones.

Overall, Senado Square presents stronger and more uniform heat exposure across the day, whereas Navona's high-heat zones are localized and largely shaped by shifting shadows.

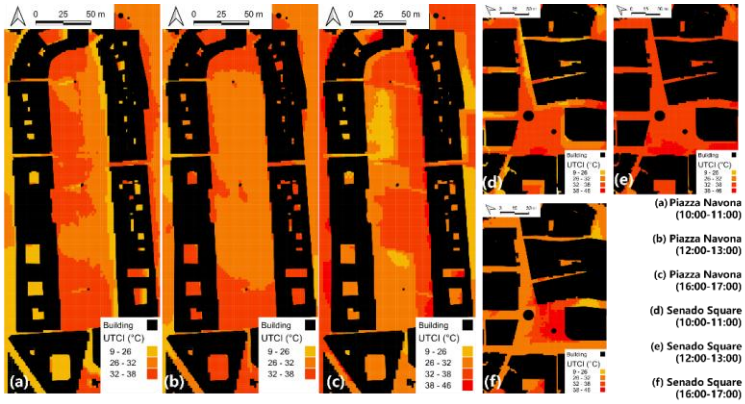


Figure 3. Comparative UTCI Distribution Maps (Piazza Navona vs. Senado Square)

3.2. BEHAVIORAL DIFFERENTIATION ACROSS THERMAL THRESHOLDS

This section contrasts behavioral patterns using UTCI = 32°C as the threshold separating the Low Thermal Sensations Zone (UTCI < 32°C , i.e., non-high heat) from

the High Thermal Sensations Zone (UTCI > 32°C, i.e., high heat). The analysis focuses on Piazza Navona (16:00) and Senado Square (10:00).

3.2.1. Heat-Behavior Coupling Analysis in Piazza Navona (16:00)

The 16:00 period at Piazza Navona represents a typical high-heat scenario, shaped by strong solar exposure and accumulated surface heat. Low-thermal-load grids occupied slightly more than half of the area, while high-heat grids formed a substantial portion of the square. Heat stress coincided with a clear spatial constraint: pedestrian activity in hot zones dropped markedly, with visit frequency and activity intensity showing substantial declines. This strong negative association between elevated thermal load and spatial usability is clearly quantified by comparing activity metrics (Figure 4).

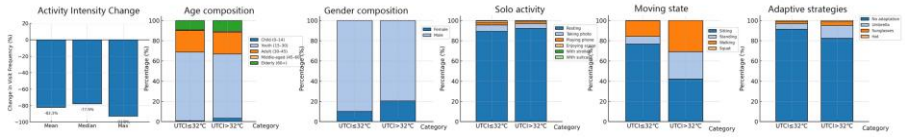


Figure 4. Comparison of Pedestrian Attribute Modes Across Thermal Zones (Navona, 16:00)

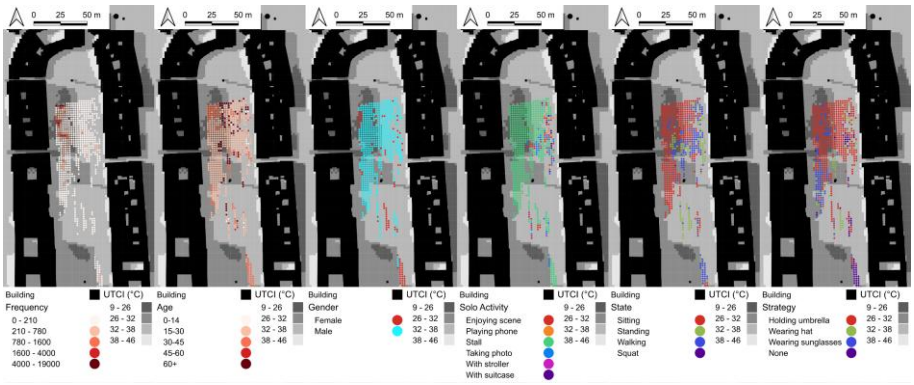


Figure 5. Spatial Distribution of Pedestrian Attributes (Navona, 16:00)

Behavioral patterns exhibited a clear shift from lingering to short-duration, functional movement (Figure 5):

- Age and gender: The pedestrian population remained youth-dominated. However, the share of female pedestrians increased under high heat, suggesting greater use of sun-protection strategies and higher behavioral adaptability.
- Moving State: Sitting—dominant in cooler areas, declined sharply in hotter zones, giving way to more standing and walking behaviors.
- Adaptive Strategy: Active heat-coping behaviors became more pronounced, with sun-shading tools such as umbrellas becoming notably more prevalent.
- Solo activity: High-exposure or stationary activities decreased, replaced by more transient and heat-avoidance-oriented behaviors.

3.2.2. Heat-Behavior Coupling Analysis in Senado Square (10:00)

At 10:00, Senado Square entered an early stage of rising solar radiation and accumulating surface heat. High-thermal-load grids already dominated the square, indicating an overall hotter microclimatic condition at this hour. Figures 6 and 7 illustrate the differences in pedestrian attribute modes and spatial distribution across thermal zones. A notable pattern emerged: the high-heat zone was more active than the cooler areas, anchored largely by the presence of key landscape features such as the fountain and flower beds. These attractions compensated for thermal discomfort and sustained high visitation. The relatively uniform distribution of activity further suggests stable user expectations and effective self-adjustment despite increasing heat exposure.

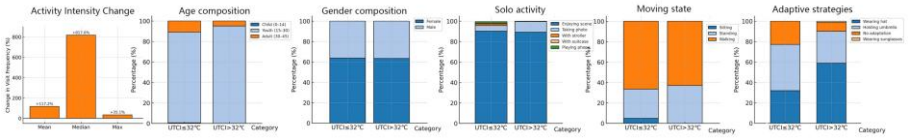


Figure 6. Comparison of Pedestrian Attribute Modes Across Thermal Zones (Senado, 10:00)

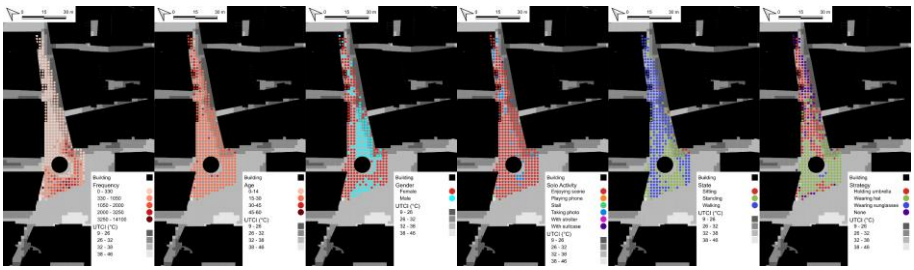


Figure 7. Spatial Distribution of Pedestrian Attributes (Senado, 10:00)

Behavioral adaptation was dynamic and co-existent:

- Age and gender: coincided with a distinct shift toward youth-centric usage: the youth proportion reached 95.1% as middle-aged users decreased by 55.6%, suggesting youth's higher thermal resilience. Conversely, the gender distribution remained stable, indicating no significant heat-related filtering by sex during this phase.
- Moving state: Sitting dropped entirely to 0.0%, reflecting a behavioral logic focused on avoiding prolonged contact with accumulated surface heat; movement prioritized walking and short-term standing.
- Adaptive strategy: Strategy selection heavily favored portability and manoeuvrability. Wearing a hat became the most prevalent strategy, reflecting a shift toward low-burden personal protection suitable for the walking-dominant activity pattern.
- Solo activities: The user base became sharply youth-centric under high heat, consistent with this demographic's stronger mobility rigidity. While core "Enjoying scene" activity remained stable, highly sensitive groups (e.g., those "with stroller")

exited completely, signifying a functional adaptation to maintain spatial vitality despite the thermal stress.

3.3. HIGH-HEAT ADAPTATION: A CROSS-SITE COMPARISON

The comparative analysis of high-heat environments highlights a "same-heat, different-structure" phenomenon, where adaptation is mediated by the interplay of landscape function and cultural context. Piazza Navona exhibited a "Climate-Dominant" pattern characterized by spatial contraction and retreat. High heat was associated with a marked reduction in activity intensity, where static behavior gave way to dynamic movement. Adaptation relied on individual strategies to maintain core social functions, prioritizing thermal avoidance. In contrast, Senado Square demonstrated "Thermal Resilience" and functional co-existence. Activity intensity was sustained or even increased in high-heat zones, facilitated by the compensatory effect of landscape attractiveness. The behavioral logic favoured mobility, characterized by the absence of sitting and a sharply youth-centric user base. Adaptation shifted to portable strategies that seamlessly integrated with the dominant tourism and walking function, thereby maintaining overall spatial vitality despite severe thermal stress.

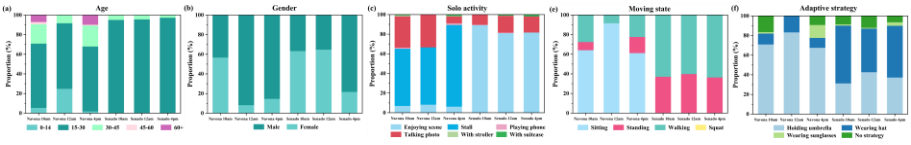


Figure 8. Comparative Distribution of Behavioral Attributes in High-Heat Zones

4. Findings and Discussion

This study challenges the assumption that heat simply suppresses activity, revealing a dynamic "Attraction-Comfort Trade-off." While high thermal stress (UTCI > 32°C) was generally associated with activity contraction, the response magnitude was mediated by spatial attraction. A critical divergence emerged: Piazza Navona exhibited a "Climate-Dominant" mode characterized by passive avoidance, whereas Senado Square demonstrated "Thermal Resilience," where landscape affordances counterbalanced physiological discomfort. Despite the focused scope, distinct adaptive logics were evident: users in Senado favored high-mobility protection (e.g., hats), while those in Navona prioritized energy conservation. Consequently, design strategies must differentiate: "Climate-Dominant" typologies require dynamic shading synchronized with peak heat, whereas "Attraction-Dominant" hubs demand localized thermal refuges (e.g., misting) to mitigate health risks for users prioritizing engagement over comfort. Methodologically, this AI-microclimate framework offers a non-intrusive approach to quantify these dynamics. Future work should expand the dataset and incorporate physiological monitoring to further validate these coupling mechanisms across diverse contexts.

Attribution Statement

ChatGPT (OpenAI, 2025) was used to (1) improve flow, (2) correct errors in

spelling and grammar, and (3) enhance the visual presentation and clarity of Figures 4 and 6.

References

Binarti, F., Koerniawan, M. D., Triyadi, S., Utami, S. S., & Matzarakis, A. (2020). A review of outdoor thermal comfort indices and neutral ranges for hot-humid regions. Urban Climate, 31, 100531. https://doi.org/10.1016/j.uclim.2019.100531

Chen, J., Shi, H., Pan, W., & Sun, D. (2024). Characterizing the supportive environment of informal spaces on cold region university campuses to enhance social interaction behavior. Buildings, 14(6), 1529. https://doi.org/10.3390/buildings14061529

Gehl, J. (1987). Life between buildings: Using public space. Island Press.

He, K., Li, H., Zhang, H., Hu, Q., Yu, Y., & Qiu, W. (2025). Revisiting gehl's urban design principles with computer vision and webcam data: Associations between public space and public life. Environment and Planning B: Urban Analytics and City Science.

Hussain, M. (2024). YOLOv1 to v8: Unveiling each variant–a comprehensive review of YOLO. IEEE Access, 12, 42816–42833. https://doi.org/10.1109/ACCESS.2024.3378568

Kenny, G. P., Yardley, J., Brown, C., Sigal, R. J., & Jay, O. (2010). Heat stress in older individuals and patients with common chronic diseases. Canadian Medical Association Journal, 182(10), 1053–1060. https://doi.org/10.1503/cmaj.081050

Kim, S. W., & Brown, R. D. (2023). Development of a micro-scale heat island (MHI) model to assess the thermal environment in urban street canyons. Renewable and Sustainable Energy Reviews, 184, 113598. https://doi.org/10.1016/j.rser.2023.113598

Levenson, M., Pearlmutter, D., & Aleksandrowicz, O. (2025). An observational analysis of shade-related pedestrian activity. Buildings and Cities, 6(1), 398–414.

Lynch, K., Lynch, K., Lynch, K. M., & Lynch, K. (1960). The image of the city. MIT Press.

Meehl, G. A., & Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. Science, 305(5686), 994–997.

Mora, C., Dousset, B., Caldwell, I. R., Powell, F. E., Geronimo, R. C., Bielecki, C. R., Counsell, C. W. W., Dietrich, B. S., Johnston, E. T., Louis, L. V., Lucas, M. P., McKenzie, M. M., Shea, A. G., Tseng, H., Giambelluca, T. W., Leon, L. R., Hawkins, E., & Trauernicht, C. (2017). Global risk of deadly heat. Nature Climate Change, 7(7), 501–506. https://doi.org/10.1038/nclimate3322

OpenAI. (2025). ChatGPT Model: GPT-4 | Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/

Rahman, M. A., Dervishi, V., Moser-Reischl, A., Ludwig, F., Pretzsch, H., Rötzer, T., & Pauleit, S. (2021). Comparative analysis of shade and underlying surfaces on cooling effect. Urban Forestry & Urban Greening, 63, 127223.

Silva, T., Lopes, A., & Vasconcelos, J. (2024). A micro-scale look into pedestrian thermophysiological comfort in an urban environment. Bulletin of Atmospheric Science and Technology, 5(1), 18. https://doi.org/10.1007/s42865-024-00082-x

Stone, Jr, B. (2012). The City and the Coming Climate: Climate Change in the Places We Live (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139061353

Whyte, W. H. (1980). The social life of small urban spaces. Conservation Foundation.

Wu, H., Zhang, G., Liu, Y., & Ming, Y. (2025). Addressing neighborhood effect averaging problem: Distinguishing mobility-based and residence-based heat exposure. Building and Environment, 280, 113139. https://doi.org/10.1016/j.buildenv.2025.113139

Yang, S., Dane, G., & Arentze, T. A. (2024). Modeling pedestrians' activity time-use choices in a virtual urban public space: The influences of the environment and affective experience. Cities, 154, 105399. https://doi.org/10.1016/j.cities.2024.105399

Zhao, H., Xu, G., Xu, Y., Zhao, L., & Meng, Q. (2025, May 21). Path-based Urban Shade Metrics: Integrating Pedestrian Thermal Comfort into Urban Design Guidelines.

Agent-Based Urban Simulation

From Pixels to Predicates

Structuring urban perception with scene graphs

Yunlong Liu[1] and Shuyang Li[2] and Pengyuan Liu[3] and Yu Zhang[4] and Rudi Stouffs[5]
[1,4]*Southeast University*
[2]*Future Cities Laboratory, Singapore-ETH Centre.*
[3]*University of Glasgow*
[2, 5]*National University of Singapore*
[1]*lyl_arch@seu.edu.cn, 0009-0001-8392-7420*
[2]*shuyangli@u.nus.edu, 0000-0002-8780-7469*
[3]*pengyuan.liu@glasgow.ac.uk, 0000-0002-5443-5910*
[4]*zhangyuseu@seu.edu.cn*
[5]*stouffs@nus.edu.sg, 0000-0002-4200-5833*

Abstract. Perception research is increasingly modelled using streetscapes, yet many approaches still rely on pixel features or object co-occurrence statistics, overlooking the explicit relations that shape human perception. This study proposes a three-stage pipeline that transforms street view imagery (SVI) into structured representations for predicting six perceptual indicators. In the first stage, each image is parsed using an open-set Panoptic Scene Graph model (OpenPSG) to extract object–predicate–object triplets. In the second stage, compact scene-level embeddings are learned through a heterogeneous graph autoencoder (GraphMAE). In the third stage, a neural network predicts perception scores from these embeddings. We evaluate the proposed approach against image-only baselines in terms of accuracy, precision, and cross-city generalization. Results indicate that (i) our approach improves perception prediction accuracy by an average of 26% over baseline models, and (ii) maintains strong generalization performance in cross-city prediction tasks. Additionally, the structured representation clarifies which relational patterns contribute to lower perception scores in urban scenes, such as (graffiti) – [on] – (wall) and (car) – [parked on] – (sidewalk). Overall, this study demonstrates that graph-based structure provides expressive, generalizable, and interpretable signals for modelling urban perception, advancing human-centric and context-aware urban analytics.

Keywords. Urban Perception, Street View Imagery, Graph Neural Networks, Scene Graph, Spatial Relationship

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 625-634.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

1. Introduction

Street view imagery (SVI) has become a mainstream data source for urban analytics because it captures façades, signage, vegetation, and other micro-scale cues at eye-level details that are underrepresented in overhead remote sensing and traditional statistics (Biljecki & Ito, 2021). Recent reviews highlight a surge in SVI studies on urban perception, gaining increasing importance in planning, transportation, and public health (Ito et al., 2024).

Early analytical pipelines heavily relied on hand-crafted or pixel-level features (e.g., GIST descriptors, colour, texture) and later adopted deep convolutional neural network (CNN) features to detect or segment visual elements within images. These models typically identify what objects are present and roughly where they are located. Subsequent studies incorporated object counts or semantic segmentation to move beyond raw pixel data (Naik et al., 2014; Dubey et al., 2016; Liu et al., 2023). However, such representations still describe presence rather than relationships. For example, "a car parked on a sidewalk" versus "a car moving along the roadway" conveys different meanings because human judgments depend on interactions among objects and their spatial context, not merely on which objects exist. Reducing both cases to the same categorical set {car, road} omits relational cues that influence perception, prompting a shift from simple co-occurrence statistics toward structured representations that explicitly encode inter-object relationships.

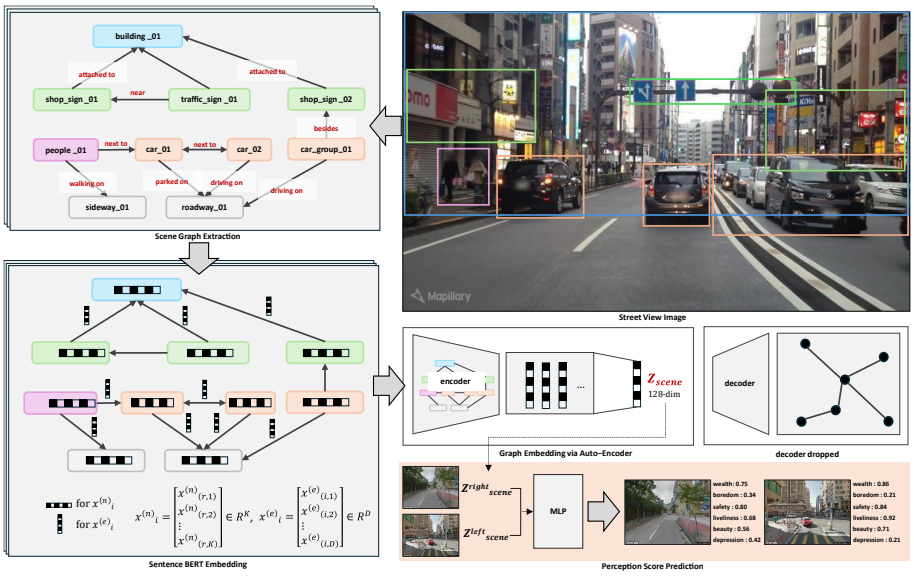


Figure 1. Research framework overview

Scene graphs (SGs) provide structure by representing images as object–predicate–object triplets (Johnson et al., 2015). Recent advances extend this paradigm to Panoptic Scene Graphs (PSG), which ground nodes on panoptic segments and jointly predict relations, thereby enhancing object–context coherence in complex scenes (Yang et al.,

2022). However, urban environments are exceptionally diverse, and region-specific street furniture, signage, or cleanliness cues often do not appear in predefined label sets of conventional models. Open-vocabulary PSG/SG methods address this limitation by leveraging vision-language learning to expand label sets, allowing the model to recognize new objects and relations beyond the original dataset. This flexibility improves the model's ability to describe and reason about real-world urban scenes (Zhou et al., 2024; Zhang et al., 2023) (see Figure 1). Overall, these developments indicate that open-vocabulary and panoptic scene graphs encode richer semantic and contextual information for urban perception than traditional pixel-based or object-count approaches.

To evaluate whether relation-aware representations improve urban perception prediction, we employ datasets that provide human perception labels and street-level imagery for model training and validation. Specifically, we combine Place Pulse 2.0 (Dubey et al., 2016) and Mapillary Street-Level Sequences (Warburg et al., 2020) to enable both supervised learning on perceptual judgments and cross-city generalization testing under real-world visual conditions.

Place Pulse 2.0 provides a large, multi-city benchmark of crowdsourced pairwise comparisons across six dimensions: safe, lively, boring, wealthy, depressing, and beautiful (Dubey et al., 2016), allowing the model to learn from human evaluations of urban scenes. Mapillary Street-Level Sequences provide extensive, geo-tagged street-level imagery collected under diverse conditions (Warburg et al., 2020). We selected subsets from Tokyo and Amsterdam as independent test domains to assess cross-city generalization, ensuring that the trained model can transfer effectively beyond the cities included in Place Pulse.

The study design and main findings are summarized as follows:

- Proposed pipeline: We develop a pipeline (see Figure 1) that transforms street view imagery into scene graph representations and predicts six perceptual dimensions.
- Performance improvement**:** Encoding SVI as structured scene graphs yields higher prediction accuracy than image-only baselines (ResNet (He et al., 2016), ViT (Dosovitskiy et al., 2021), CLIP (Radford et al., 2021)) across all perception dimensions, demonstrating that object–relation structures benefit prediction.
- Cross-city generalization**:** Models trained on Place Pulse 2.0 maintain stable performance on Mapillary subsets, indicating that the graph-structured representations improve cross-city transferability.

2. Methodology

2.1. RELATIONAL PARSING

The Place Pulse 2.0 dataset is used to train the model, providing supervision across six perceptual dimensions. The dataset is randomly divided into training/validation/test partitions in a 6:3:1 ratio. Each image is then converted into a scene graph following the pipeline illustrated in Figure 2. In this representation, nodes correspond to urban entities, and edges denote semantic relations between them. Object and predicate texts are encoded with Sentence-BERT (Reimers & Gurevych, 2019), a text embedding

model, which maps textual descriptions into high-dimensional semantic vectors. This process can be formalized as:

$$x_n = f_{SBERT}(attr_x)$$

where $attr_x$ denotes the textual description of an object or predicate (e.g., "building", "on", "sidewalk"), and $f_{\text{SBERT}}(\cdot)$ represents the Sentence-BERT encoder that maps each text into a semantic vector$x_n \in R^{d_t}$.

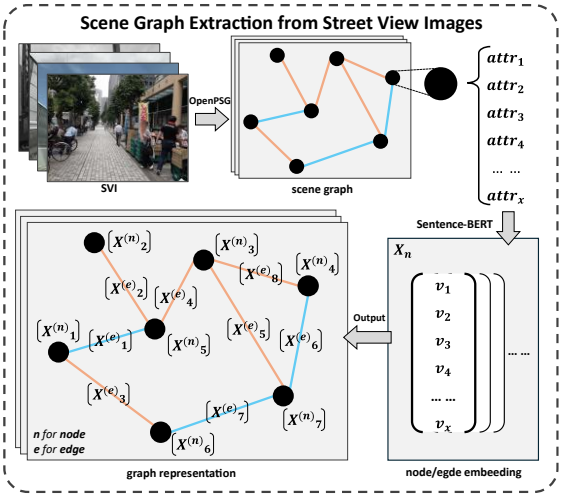


Figure 2. Scene graph construction and feature encoding

2.2. GRAPH EMBEDDINGS

To learn meaningful representations from scene graphs, we employ GraphMAE, a self-supervised model framework for learning structured graph representations without manually labelled data (Hou et al., 2022). GraphMAE masks portions of the input graph and trains the model to reconstruct the missing information, thereby forcing it to understand the underlying spatial and semantic structure of the urban scene (see Figure 3). After training, the decoder of GraphMAE is discarded and the encoder produces scene-level embeddings (128-dim in our implementation) for downstream urban perception prediction.

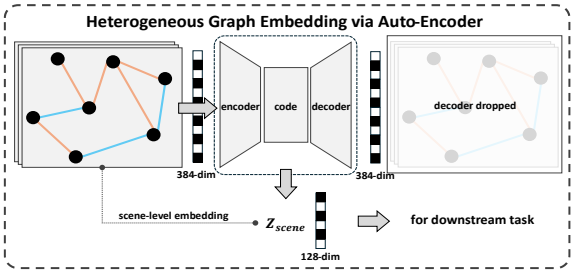


Figure 3. Masked graph autoencoding for scene-level representation learning

2.3. PERCEPTION INFERENCE

To enable the model to learn from human perceptual judgments, we adopt a pairwise-comparison learning approach as the training mechanism. This approach trains a model to determine, for each pair of images, which scene is perceived more positively on a given perceptual dimension. Classical formulations such as Bradley–Terry and RankNet can translate binary comparisons into continuous perceptual scores (Bradley & Terry, 1952; Burges et al., 2005). In our implementation, each pair of scene embeddings is passed through a weight-shared multilayer perceptron (MLP) to yield scores, as shown in Figure 4. The probability of one scene being preferred over another is computed as $P(left > right) = \sigma(s_{left} - s_{right})$, and the model is trained via cross-entropy using the true pairwise labels.

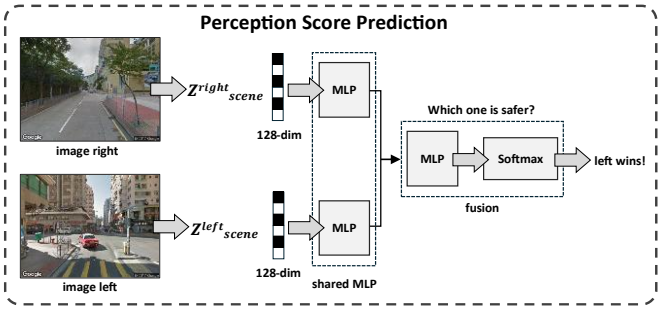


Figure 4. Pairwise perception prediction framework

Different backbones are compared for the Bradley–Terry head, including ResNet-50, ViT-B/16, and CLIP ViT-B/32. We evaluate model performance on the Place Pulse 2.0 dataset with accuracy per dimension and mean accuracy. To further assess cross-city generalization, the trained models are tested on Mapillary street-level imagery from Tokyo and Amsterdam.

3. Results

The first experiment examines how structured representations improve perception prediction relative to image-based models (ResNet50, ViT-B/16, and CLIP ViT-B/32). All approaches were trained and evaluated within the same dataset, allowing a controlled comparison of representational capacity.

The baseline models shared the same pairwise comparison formulation, implemented with identical data splits and loss functions as the graph-based approach. Each image-only baseline was pretrained on large-scale datasets but trained here with frozen backbones to ensure fair evaluation.

Results show that the graph-based model achieves higher accuracy, precision, and area-under-curve (AUC) scores across all perceptual dimensions (Table 1). The largest gains appear in wealth and liveliness predictions (Table 2), where human perception depends strongly on compositional relationships rather than raw texture or colour.

Table 1. Performance comparison across baseline and graph-based models

Model	AUC	accuracy	recall	f1	precision
CLIP	0.61	0.63	0.60	0.55	0.60
CNN	0.82	0.73	0.78	0.68	0.64
ViT	0.79	0.71	0.74	0.72	0.67
GraphMAE	**0.84**	**0.87**	**0.89**	**0.85**	**0.83**

Table 2. Perception prediction performance across six perceptual dimensions

Model	beautiful	boring	depressing	lively	safety	wealthy	average
CLIP	0.59	0.64	0.59	0.64	0.64	0.66	0.63
CNN	0.74	0.69	0.67	0.77	0.76	0.77	0.73
ViT	0.72	0.67	0.63	0.73	0.74	0.74	0.71
GraphMAE	**0.87**	**0.84**	**0.83**	**0.88**	**0.87**	**0.90**	**0.87**

3.1. CROSS-CITY GENERALIZATION TEST

The second experiment investigates whether relational representations generalize across distinct urban contexts. To minimize cultural bias, we selected Tokyo and Amsterdam from the Mapillary dataset, two cities that differ in spatial layout, building typology, and streetscape. The pre-trained graph-based model is directly applied to predict on the manually annotated subsets from Tokyo and Amsterdam. These subsets are annotated using the same pairwise comparison method as the Place Pulse dataset, where annotators select from two street-view images the one that better corresponds to a specific perceptual dimension (e.g., safety, liveliness).

Results reveal a moderate decline in performance when evaluated on the Tokyo–Amsterdam subset compared to the full Place Pulse dataset. Test accuracy decreases by approximately 5.6%, while the AUC shows a 3.5% reduction (Table 3), indicating that the relational embeddings can capture higher-level spatial patterns that remain interpretable and predictive across different cities.

The structured graph-based model exhibits strong cross-city generalization, indicating that relational semantics provide a transferable and robust foundation for urban perception prediction.

Table 3. Cross-city generalization performance on Mapillary subsets (Tokyo and Amsterdam)

Metric	Place Pulse 2.0	Tokyo–Amsterdam	Change
Accuracy	0.84	0.79	−5.6%
AUC	0.87	0.84	−3.5%

3.2. QUALITATIVE ANALYSIS

We further inspect selected examples from the Mapillary subset used in the cross-city

generalization experiment. Figure 5 arranges representative scenes from Tokyo and Amsterdam along six perceptual dimensions: safety, liveliness, boredom, wealth, depression, and beauty, ordered from lower to higher predicted scores.

A close look at Figure 5 reveals consistent relational motifs across both Tokyo and Amsterdam. In the higher-scoring columns, scenes tend to exhibit spatial enclosure: pedestrians appear near façades, trees or hedges align with sidewalks, and street elements (lamps, benches) create a human-scale rhythm. These object-relation patterns correspond to perceptions of safety, liveliness, and beauty. In contrast, the lower-scoring columns often display road segments with few adjacent structures or human presence, producing a sense of emptiness or exposure. These scenes typically lack relational connectivity, such as people–sitting on–benches or bicycles–parked beside–building, or plants–growing along–walls, which the patterns associate with more favourable perceptual scores. While these patterns are broadly consistent across cities, sampling imbalances remain: Tokyo's subset includes more façade-rich streetscapes, whereas Amsterdam's subset contains more open-road or green-dominant views, which may partially influence the perceptual scores.

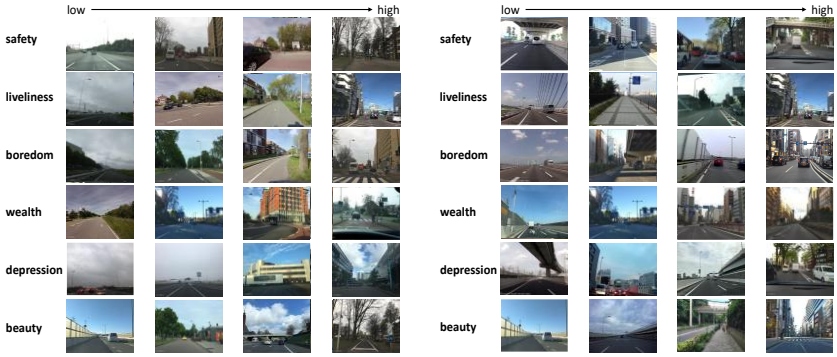


Figure 5. SVI examples by perception score, Tokyo(right) and Amsterdam(left)

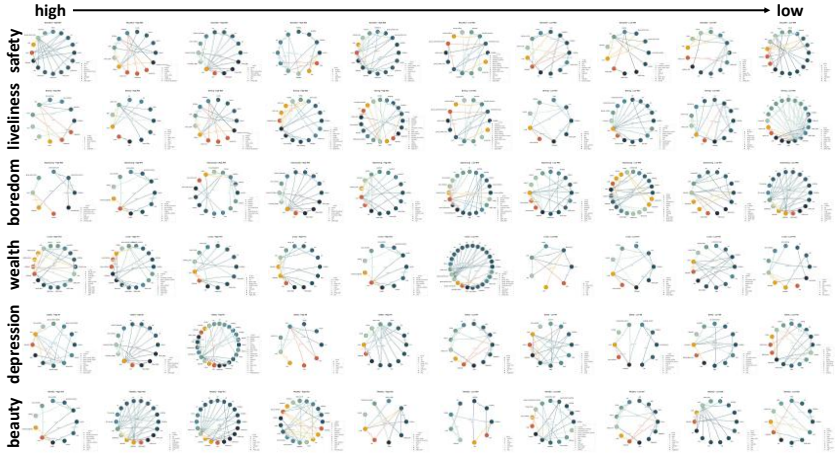


Figure 6. Scene graph thumbnail overview

To complement the image grid, Figure 6 visualizes the predicted scene graphs for six perceptual dimensions. Higher-scoring scenes generally display richer relational topology (characterized by more edges) and greater subject diversity (reflected in more node types), suggesting a denser web of interactions among urban elements (see examples in Figure 7). In contrast, some low-scoring cases also contain many nodes and edges but concentrate them within a limited set of repetitive relations—for example, repetitive triplets like car–near–car or road–beside–road. Despite their apparent complexity, these graphs lack the heterogeneity and cross-category links that correlate with positive perceptions. This contrast underscores that variety and cross-type relations, rather than graph density, are more strongly aligned with favourable human perception.

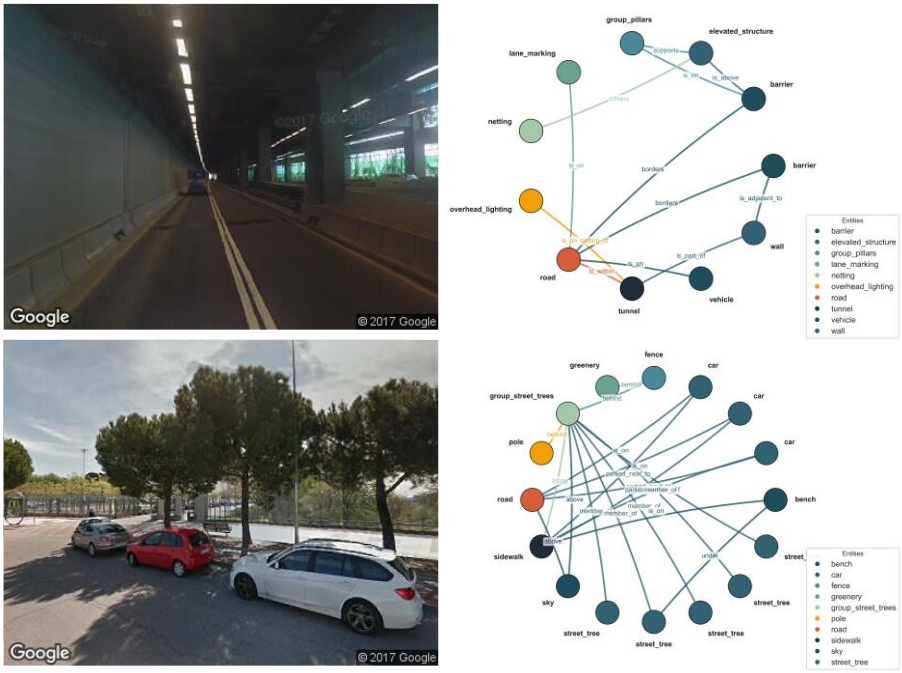


Figure 7. Street views and scene graph examples (top: low beauty score; bottom: high beauty score)

The results suggest that relational structure is a key factor for predicting human perception. Graph-based representations effectively capture spatial and functional interactions, such as proximity between people and buildings, or alignment of vegetation with walkable space, which remain stable across contexts. By embedding these semantics into a deep learning framework, the model bridges the gap between visual appearance and human interpretation, enabling both scalable and interpretable perception analysis.

4. Conclusion

This study demonstrates that representing street view imagery through structured scene

graphs substantially enhances the modelling of urban perception. Compared with image-only baseline models, the proposed approach consistently achieves higher predictive accuracy across all perceptual dimensions. Moreover, the proposed approach also exhibits cross-city generalization in independent tests conducted in Tokyo and Amsterdam, demonstrating that relational structures captured by scene graphs are transferable across diverse urban contexts.

This approach also holds practical values for evidence-based urban design. The relational patterns identified by the model can serve as diagnostic cues, revealing where perceived discomfort or disorder may arise from spatial misalignment or missing human-environment interactions.

Nonetheless, several limitations point to directions for future research. The cross-city evaluation was based on a limited subset, which may introduce potential cultural or seasonal biases, and expanding it with more representative samples would improve statistical reliability. In addition, the OpenPSG-based scene parsing could be further improved through domain-specific fine-tuning to better capture relational pattern nuances unique to different urban environments. Finally, future work should consider integrating socio-demographic data to address the inherent subjectivity of perception, as well as exploring dynamic scene graphs for video-based analysis.

Attribution Statament

ChatGPT (OpenAI, 2025) was used to (1) improve writing and (2) assist with coding.

References

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika*, 39(3–4), 324–345. https://doi.org/10.1093/biomet/39.3-4.324

Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., & Hullender, G. (2005). Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning* (pp. 89-96). https://doi.org/10.1145/1102351.1102363

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. In *European Conference on Computer Vision (ECCV)* (LNCS 9905, pp. 196–212). https://doi.org/10.1007/978-3-319-46448-0_12

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR 2021)*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2016.90

Hou, Z., Liu, X., Cen, Y., Dong, Y., Yang, H., Wang, C., & Tang, J. (2022). GraphMAE: Self-supervised masked graph autoencoders. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)* (pp. 594–604). https://doi.org/10.1145/3534678.3539321

Ito, K., Kang, Y., Zhang, Y., Zhang, F., & Biljecki, F. (2024). Understanding urban perception with visual data: A systematic review. *Cities*, 152, 105169. https://doi.org/10.1016/j.cities.2024.105169

Johnson, J., Krishna, R., Stark, M., Li, L.-J., Shamma, D., Bernstein, M., & Fei-Fei, L. (2015, June). Image retrieval using scene graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3668-3678).

Liu, Y., Chen, M., Wang, M., Huang, J., Thomas, F., Rahimi, K., & Mamouei, M. (2023). An interpretable machine learning framework for measuring urban perceptions from panoramic street view images. *Iscience*, 26(3).

Naik, N., Philipoom, J., Raskar, R., & Hidalgo, C. (2014). Streetscore-predicting the perceived safety of one million streetscapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 779-785).

OpenAI. (2025). GPT-5 [Large Language Model]. https://openai.com/chatgpt/

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *Proceedings of International Conference on Machine Learning* (pp. 8748-8763). PMLR.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3980–3990). https://doi.org/10.18653/v1/D19-1410

Warburg, F. R., Hauberg, S., López-Antequera, M., Gargallo, P., Kuang, Y., & Civera, J. (2020). Mapillary street-level sequences: A dataset for lifelong place recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020)* (pp. 2626–2635). IEEE. https://doi.org/10.1109/CVPR42600.2020.00270

Yang, J., Ang, Y. Z., Guo, Z., Zhou, K., Zhang, W., & Liu, Z. (2022). Panoptic scene graph generation. In *European Conference on Computer Vision (ECCV)* (pp. 178-196). https://doi.org/10.48550/arXiv.2207.11247

Zhang, Y., Pan, Y., Yao, T., Huang, R., Mei, T., & Chen, C. W. (2023). Learning to generate language-supervised and open-vocabulary scene graph using pre-trained visual-semantic space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2915-2924). https://doi.org/10.1109/CVPR52729.2023.00285

Zhou, Z., Zhu, Z., Caesar, H., & Shi, M. (2024, September). Openpsg: Open-set panoptic scene graph generation via large multimodal models. In *European Conference on Computer Vision* (pp. 199-215). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72684-2_12

Gene-up: A Somatic Neural Model for Rebalancing Urban Entropy

Vania Arlene[1], Tasos Varoudis[2] and Provides Ng[3]
[1,2,3]*Bartlett School of Architecture, University College London.*
[1]*vania.arlene@gmail.com, 0009-0008-3401-7464*
[2]*t.varoudis@ucl.ac.uk*
[3]*provides.ism@gmail.com, 0000-0001-6975-4642*

Abstract. Informational mismanagement often deteriorates urban systems through a disruptive dichotomy of hyper-regulation and uncontrolled informality. Using East London as a testbed, this research addresses the limitations of top-down control and advocates a shift from exogenous to endogenous framework that decodes latent information in the emergent patterns of collective inhabitation. To what extent can body movement be understood as an informational code that recalibrates urban entropy? The goal is to examine how motion data can enable spatial homeostasis through a 'gene-up' approach: a methodological framework treating movement as a micro-informational 'gene', capable of reprogramming urban systems via self-regulation. The objectives are to: 1) expose the persistence of structural entropy; 2) capture interactions of 32 body-point segmentation tasks; 3) discern emergent collective axial patterns; and 4) simulate optimizations on pattern-genes within generative urban models. Our methodology combines data mapping, deep-learning neural networks trained on site-specific kinematic datasets, clustering (k-means) and evolutionary algorithms. Findings revealed that conventional planning models tend to overlook body movement as somatic data that correlate with nodes of informal practices and governance lapses. By encoding body movement patterns as fundamental matrixes of urban code within generative workflows, the 'gene-up' model produces adaptive morphologies responsive to lived motion. Design populations are evaluated through reward-penalty scores that balance optimized visibility fields with motion parameters to activate spatial zones, hence redefining designer's roles as choreographers of the urban fabric.

Keywords. Urban Entropy, Gene-up, Pathosformel, Deep Neural Network, Evolutionary Algorithm, Urban Homeostasis

1. Introduction

Cities are no longer mere agglomerations of physical infrastructure; they are multi-layered information systems: memories as depth, dialogues as exchange, gestures as genes, and behaviours as codes. The way this information is interpreted directly shapes the city's form and function; yet, in the 21st century, urban systems are breaking down

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 635-644.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

under conditions of informational mismanagement (Franke & Gailhofer, 2021). Contemporary cities sit precariously between two extremes: top-down technocracy and bottom-up complexity. On one side, lie hyper-regulated, algorithmically curated environments that suppress spontaneity in the pursuit of efficiency and security (Kitchin et al., 2019). On the other, informal settlements produce their own dense, illegible codes, generating frictions and disorder (Kamalipour & Iranmanesh, 2021). For instance, East London intertwines the master planned Olympic Park, showcasing order and legacy with areas of high population turnover and poverty such as Newham (Santos et al., 2025). This dichotomy reveals a deep inadequacy in current paradigms: traditional planning and computational determinism demonstrated a shared neglect of human scale and needs, which led to social displacement, exclusion, and marginalisation.

The recalibration of order and disorder in urban developments can be understood as urban entropy. Derived from thermodynamics and information theory, the concept offers a critical lens for rethinking the dynamics of planning (Netto et al., 2025). Entropy is not viewed as deterioration, but as a fluctuating condition that demands progressive readjustment (Schrödinger, 2012). This refers to spatial homeostasis, the state of equilibrium where informational flow achieves coherence without suppressing heterogeneity and competing forces continuously interact within the urban form to maintain systemic balance (Netto et al., 2025). In East London, however, the persistent tensions between rapid redevelopment and longstanding informal economies disrupt spatial homeostasis, representing how urban entropy manifests as imbalance (Evans & Jones, 2008). This highlights the need for recalibrating planning approaches to restore systemic equilibrium without erasing the area's inherent heterogeneity (Netto et al., 2025).

This study introduces the concept of 'gene-up', a methodological framework that treats body gestures as micro-informational 'genes', the smallest programmable units of urban code. Analogous to biological genes that encode life, these motion patterns are seen as encoding the order/disorder within cities. The research develops and tests the gene-up framework as a mixed-methodology combining data analytics from urban mapping in East London with deep-learning pose detection neural network, clustering and evolutionary algorithms to capture, classify, and simulate optimized iterations of generative models and somatic data as urban code (Varoudis & Penn, 2021). The contribution of this study lies in demonstrating that pose-detected movement clusters can be translated into generative design asset-constraint that measurably alter simulated movement and visibility fields in East London test sites.

2. Decoding Urban Anatomy

A seminal expansion of living organisms as systems counteracted entropy by generating internal order through molecular "codes", laying the groundwork for this research (Schrödinger, 2012). In the mid-20th century informational theory, entropy was reconceptualized as a measure of uncertainty or surprise in message transmission, providing a statistical framework that shifted focus from physical disorder to informational unpredictability (Shannon, 1948; Coutrot et al., 2022). Cybernetics refined this by framing information as negentropy; order and predictability, that enables control and feedback in systems (Wiener, 2019). This distinction is more than

semantics or syntax: it operates at a level of system behaviour and regulation. Is urban information (people's movements, behaviours, interactions, decisions) adding unpredictability (disorder/entropy) or creating pattern and predictability (order/negentropy)?

This dialectic between negentropy and entropy is mirrored in urban design discourses debating the respective merits of planned versus emergent city forms (Ng, 2022). However, sustaining urban homeostasis remains inherently challenging, as systems must negotiate the tension between order and disorder to remain resilient and responsive over time. The theory of dissipative structures offers a compelling framework, positing that complex systems persist through perpetual flux between organized states and disorder, with entropy serving not solely as decline but as a crucial driver for renewal and transformation (Prigogine & Stengers, 1984; Gallopin, 2020). Entropy here is neither a simple force of decay nor disorder alone; rather, it is deeply intertwined with the processes of regeneration, self-organization, and transformation. Entropy becomes a catalyst for renewal, allowing urban systems to reorganize, adapt, and evolve in response to internal dynamics and external shocks.

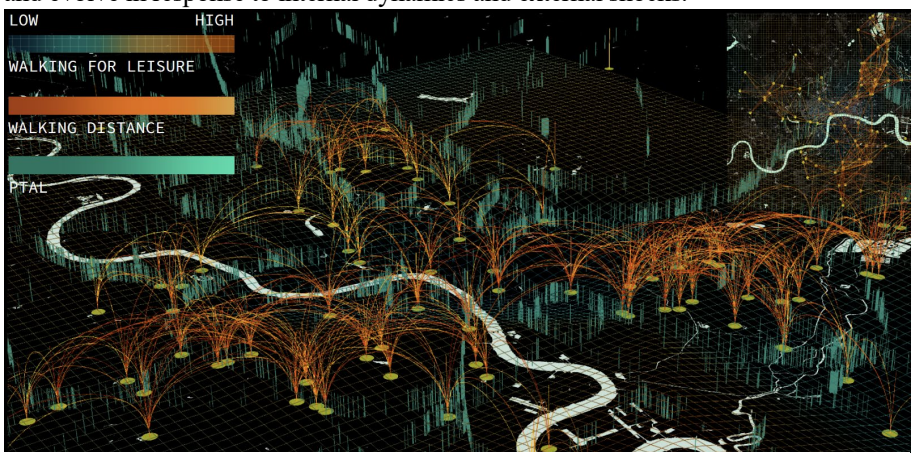


Figure 1. Interconnectivity between nodes of third places in London

Neither top-down or bottom-up interventions have yet proven fully capable of addressing the complexities of urban life in the digital age, highlighting a critical gap in current planning paradigms. The imperative shift from exogenous control to an endogenous framework decodes the emergent patterns of collective inhabitation as latent information (Wade, 2009). Ultimately, this theoretical synthesis underscores why urban design and governance must move beyond classical binaries and embrace innovative design paradigms capable of mediating between these extremes. The gene-up framework proposed in this study decodes and encodes bodily information, where inhabitants as agents in urban systems aim to achieve a responsive spatial homeostasis that is neither rigidly deterministic nor anomie-driven, but generative and human-centred. Within this, informational code is defined as the correlation between embodied human activity and its spatial translation, expressed through metrics such as spatial density, motion clustering, and entropy-based distribution indices.

3. Data Mapping: East London

The dialectic of urban entropy/negentropy is evident within East London's post-industrial landscape, where redevelopment dynamics expose layered conditions of informational friction. Spatial data correlations reveal zones marked by extreme juxtaposition, where planned regeneration schemes operate adjacent to areas of deprivation. By integrating stacks of datasets on the LSOA level, we gathered scores of Index of Multiple Deprivation (IMD) for wellbeing, scores of Public Transport Accessibility Transport (PTAL) for accessibility, areas of Greenspace Information for Greater London (GiGL) for distribution and typology of green infrastructure, among others (see Figure 2), are cross-referenced against community flows between home, work, and third places, hence defining nodes of collective engagement. For instance, street segments on the R1000 metric with high angular integration coincide with corridors of dense walking population and high public transport access, whereas areas with high IMD scores (worst at 1st to 2nd decile) and fragmented GiGL distribution correspond to lower choice segments, indicating localised pockets of spatial entropy.

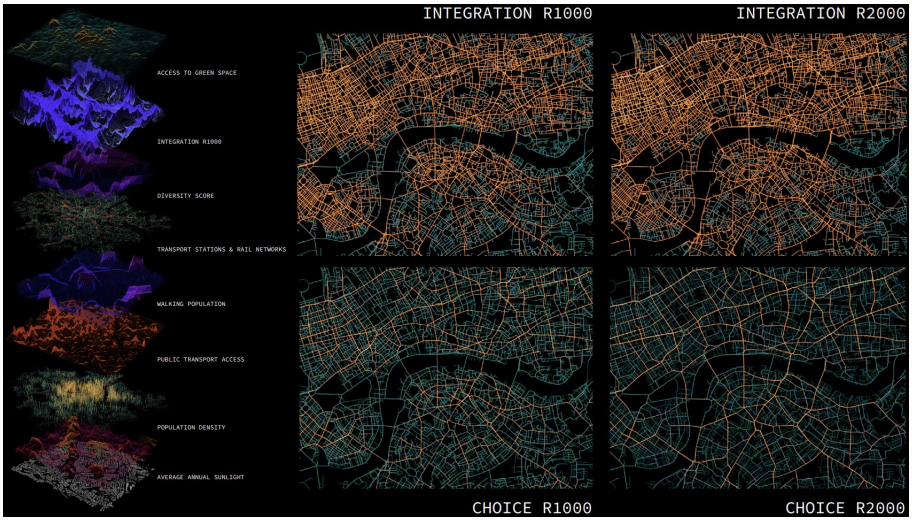


Figure 2. Stack of datasets (left) and space syntax (right) of London

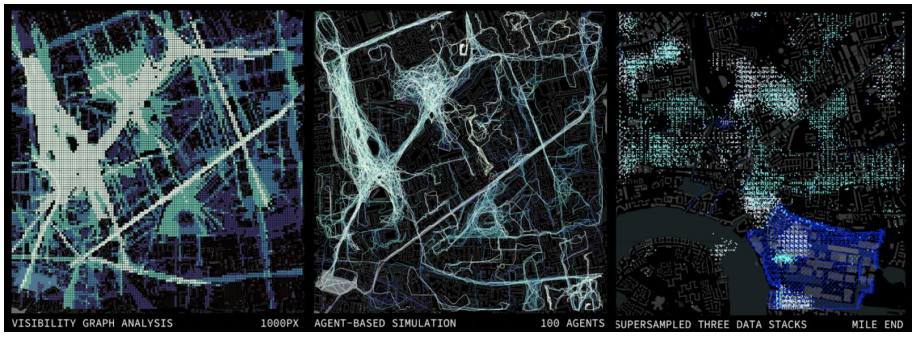


Figure 3. Visibility Graph Analysis (left), Agent simulation of 100 agents released from any point (middle), GAN pix2pix of three data stacks (left) in Mile End, East London

The objective is to interpret these variables not as discrete measures, but as interacting signals within a complex system of spatial information exchange. K-Means clustering and PCA (see Figure 4) correlations reveal fragmentations of physical and social architectures; economic disparities and informational barriers that disrupt the homeostatic potential of the urban system. High-deprivation zones coincide with reduced walking percentages, accessibility indices, and modal diversity, delineating what can be termed 'mobility deserts'. Within these, residents face constrained route options and extended travel times. This structural disorder culminates mobility and flow imbalance between well-connected and marginal zones, measurable as patterns of uneven informational distribution; a state of high entropy, characterized by uneven spatial accessibility and unpredictability in movement patterns.

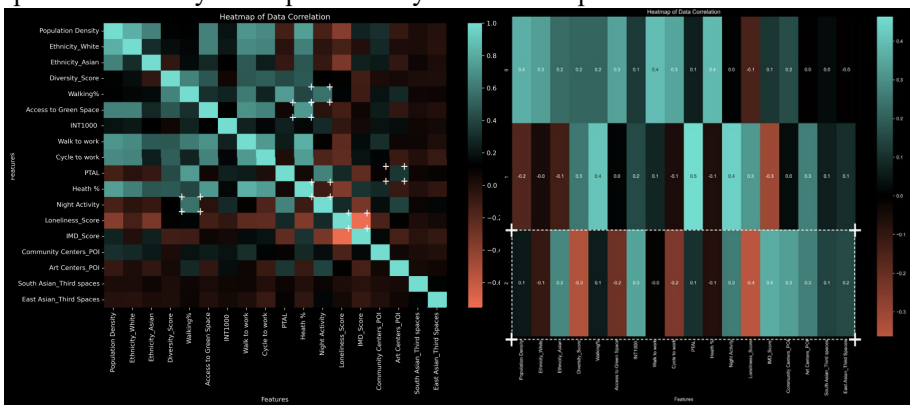


Figure 4. K-Means clustering (left) and PCA (right) of spatial data analytics

Accordingly, traces of pedestrian movement are expressed in micro-scale, human behaviours. In accessible and well-integrated zones, movement trajectories illustrate higher path conformity relative to formal spatial order. However, in entropic zones, movement data captures irregularities, such as path deviations from official layouts, seen from informal paths etched across undeveloped plots that contradict the planned infrastructure. These paths are not evidence of chaos but are a form of endogenous re-ordering. Each choice is a physical manifestation of latent information, a data point indicating an attempt to introduce negentropy (see Figure 3). Thus, East London serves as a diagnostic context in which urban entropy emerges from mismanaged spatial generation, distribution, and interpretation. These emergent patterns of mobility are the 'genes' that the gene-up framework proposes to decode by recalibrating the system, not through further top-down imposition, but by fine-tuning the information generated by the city's anatomy. The stack of datasets, K-means clustering, and PCA results ground movement as a somatic analysis in relation to well-being, accessibility, and ecological resilience, in observed spatial conditions.

4. Locomotive Morphologies

To gather idiosyncratic movement data over time, the pipeline utilizes a deep learning pose detection neural network (via MediaPipe) that is a permutation and transformation invariant, which allows it to handle the unstructured nature of the point cloud data. This enabled the classification segmentation tasks despite rotations and translations to

capture interactions and relative positions between the 32 body points by frame. CNN localizes anatomical points (respectively left and right legs: ankles, knees, hips, shoulders, elbows), filtering 3D coordinate points through confidence score for accuracy ($s \geq 0.80$) (see Figure 5). Outputs are 2D trajectory traces (with depth estimates where camera geometry allows) sampled at high temporal resolution. Each person-time sequence is a high-dimensional multivariate time series of joint coordinates and velocities.

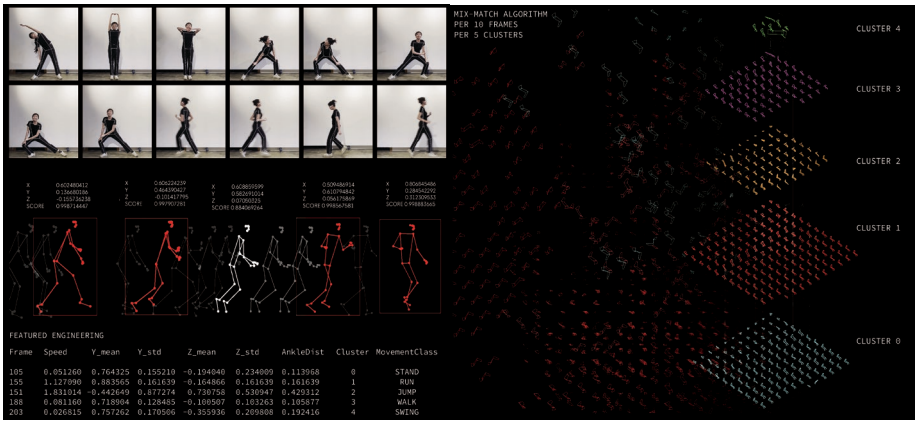


Figure 5. Pose detection NN and clustering algorithm (k-means) of movement typologies

To manage dimensionality and reveal latent structure, they are projected into low-dimensional embeddings via t-SNE after standardization and temporal windowing. These embeddings represent candidate "genes": recurrent kinesthetic signatures that carry informational content about spatial affordances (see Figure 6). The optimal *k* was

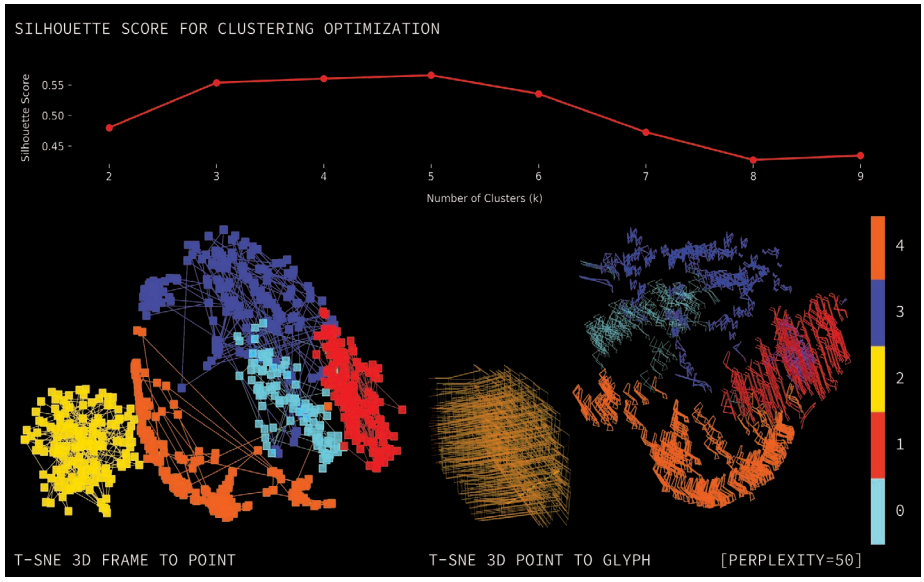


Figure 6. t-SNE of high-dimensional movement/trajectory data

chosen based on the silhouette score across *n* cluster counts to inspect cluster stability, where feature-engineering incorporates axial variations (speed, vertical movement, ankle to ankle distance, standard deviation) to determine its movement clusters (K-means into movement typologies). Each cluster becomes a genotypic code linked to its spatial context, enabling downstream design objectives to be grounded in empirically observed behaviour rath er than assumed archetypes (see Figure 7).

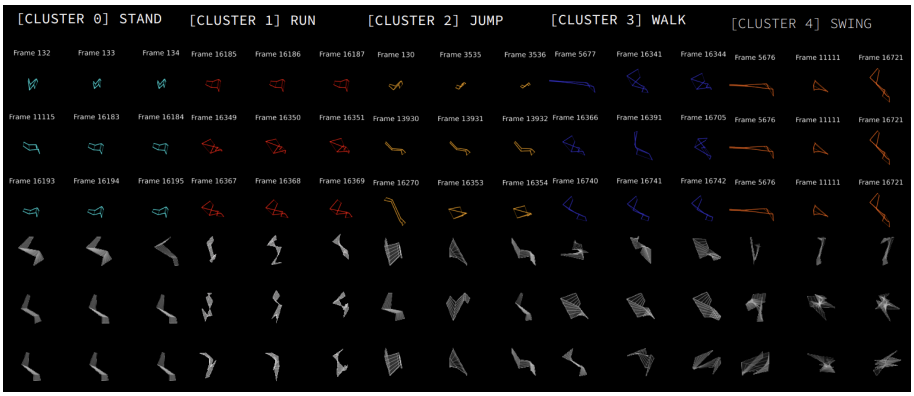


Figure 7. Snippet of shape catalogue of movement clusters (k-means optimized of 5 clusters), from segmented task of trajectory lines (top) and generated "skin" geometries (bottom) by frame

5. Evolutionary Algorithm

The multi-objective optimization algorithm bridges somatic data that condition the generation of urban layouts, so that design search is explicitly guided by somatic information. While traditional computational models enforce order through predefined logics, the algorithmically mediated approach generates complexity through iteration, selection, and adaptation to mirror how social and spatial systems continuously self-organize. Using the clustered movement typologies as parameters in the fitness evaluation of candidate layouts, solutions are rewarded when simulated trajectories reproduce, diversify, or strategically redirect these observed motion patterns.

For each generation, design variants are assessed on three main metrics: distance between ankles (movement impact), length of isovist (visibility) between 10-500m, and overlapping areas between skin geometries and isovist fields (efficiency). High-performing solutions (display widely distributed movement paths with minimal overlap, where the algorithm internalises the somatic 'genes' as drivers of spatial self-organization. Conversely, low-performing solutions indicates clustered movements with concentrated overlap, which pushes the algorithm to identify unfavourable conditions. These parameters form a reward-penalty system that aim to shift high movement typologies to activate dead zones (and vice versa), transforming underutilized areas into new centres of spatial engagement or interactions. This data-driven optimization process directly links urban morphologies to measurable patterns of collective movement, enabling spatial designs that respond to human behavioural dynamics. Each design iteration negotiates spatial intelligence encoded by human motion. The impact of employing evolutionary algorithms (see Figure 8) transform embodied movement data into legible and generative spatial patterns. What emerges

from the evolutionary process is not a definitive plan but an adaptive ecology of spatial configurations that can evolve with the dynamics of local life.

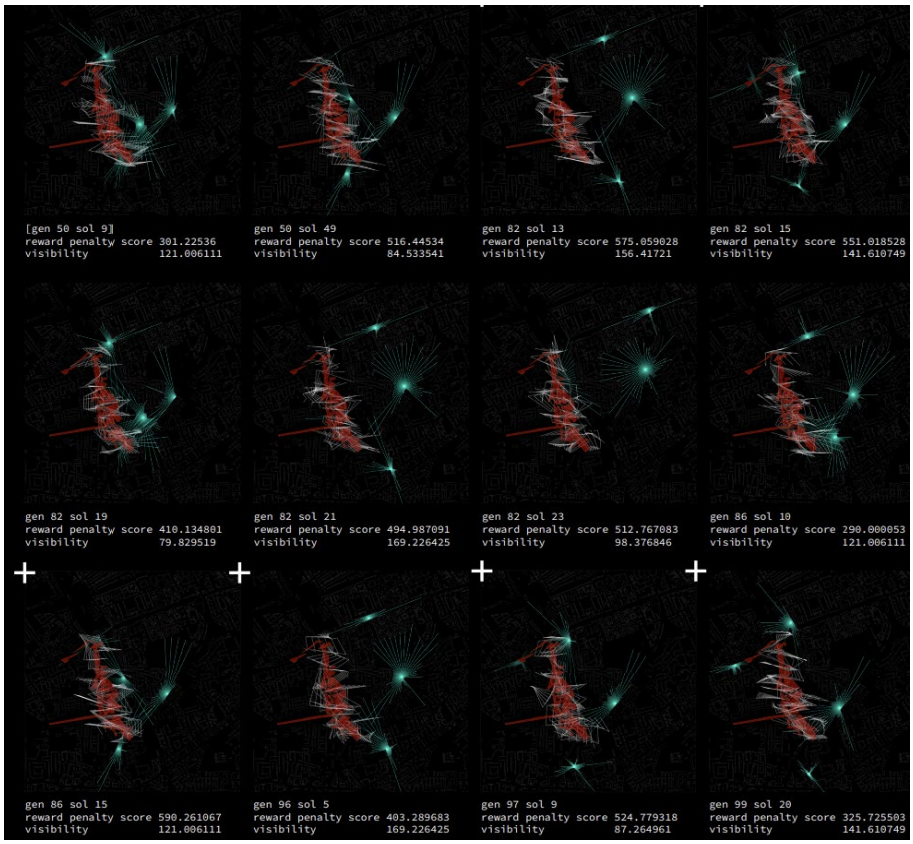


Figure 8. Evolutionary algorithm with reward-penalty scores to optimize

6. Discussion

The study has proposed a new model for understanding and designing cities as a dynamic field of entropic balance, mediated through the informational currency of bodily movement through a tripartite 'gene-up' framework: (1) By recasting embodied motion as an active urban 'gene', this research demonstrates that the micro-patterns of everyday movement are not mere by-products of urban form but vital agents in scripting the city's informational and spatial evolution, (2) the integration of pose detection, motion clustering, and evolutionary generative algorithms provided an operational chain linking the abstract and the embodied, translating lived kinetic behavior into adaptive spatial morphology, and (3) the 'entropic way of design' proposed here advocates for cities that thrive on fluctuation and feedback, resisting rigid dichotomies of order versus chaos.

In this sense, balanced urban systems emerge from continuous negotiation rather than fixed equilibrium, where designers function less as master planners and more as

curators of urban rhythm, facilitating conditions in which informational diversity and legibility coexist. Nevertheless, there exists several limitations. First, the data collection, being both spatially and temporally bound, potentially captured only a partial segment of the broader kinesthetic landscape, risking an incomplete representation of movement typologies. Second, the reductive translation of rich, affective motion data into simplified computational forms inevitably entailed an interpretive loss, as the subtle textures of lived experience resisted straightforward codification. Finally, the operationalization of the evolutionary algorithm, while generative, was itself guided by subjective value-laden choices in its parametric reward-penalty system, meaning the resulting optimizations were ultimately determined by a predefined and contestable conception of spatial performance.

Looking forward, future research should expand the gene-up methodology across heterogeneous urban contexts. 'Gene-up' as a rigorous methodology embodies movement as a quantifiable yet generative form of urban intelligence and as a disciplinary rethinking of how urban design guides emergent order. Integrating kinesthetic data with multimodal environmental and sensory datasets could yield richer models of urban information exchange and deepen understanding of embodied-environmental correlations.

7. Conclusion

This research set out to confront the critical dysfunction of contemporary urban systems; an informational imbalance that manifests as a persistent tension between hyper-ordered environments and uncontrolled informality. It proposes a redirection of focus toward the most fundamental unit of urban life: the body in motion. Through posture, gesture, and spatial relation, bodies continually generate subtle yet powerful information that expresses urban life itself.

Through this convergence of theory, computation, and embodiment, the gene-up framework reframes the city as a self-regulating, adaptive entity that evolves through the informational interplay of order and disorder, structure and movement. The study's generative, entropic approach demonstrates that the vitality of urban space lies not in eliminating complexity but in cultivating its balance. By translating embodied motion into design intelligence, urban systems gain the capacity to recalibrate themselves continuously, expressing both the freedom of human behaviours and the coherence of shared spatial logic. In essence, this analytical pipeline translates the ephemeral, high-dimensional complexity of observed human movement into a structured and machine-readable vocabulary. The resulting clusters serve as the 'pattern-genes': a finite set of empirically grounded kinesthetic typologies that represent the fundamental units of embodied information within the urban environment. With these behavioural codes identified and quantified, they are now ready to be deployed as the genetic input for the evolutionary design process, closing the loop from analysis to synthesis. This research advances a humanistic and critical trajectory for computational urbanism by approaching bodily movement as performative inscription rather than ephemeral social expression. Body language becomes not merely behaviour but a dynamic matrix of urban code: a living script through which entropy can be rebalanced.

Acknowledgements

This research is part of the Bartlett School of Architecture Research Cluster 14 and originated as a collaborative team project. We would like to thank the Greater London Authority for providing open-access datasets that enriched our spatial analytics, Space Syntax Lab for developing the DepthmapX toolkit, and Google for the open-source MediaPipe framework.

References

Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., Dalton, C., Wiener, J. M., Hölscher, C., Hornberger, M., & Spiers, H. J. (2022). Entropy of city street networks linked to future spatial navigation ability. *Entropy of city street networks linked to future spatial navigation ability*, 104-110.

Evans, J., & Jones, P. (2008). *Urban Regeneration in the UK: theory and practice.* Sage Publications.

Franke, J., & Gailhofer, P. (2021). Data Governance and Regulation for Sustainable Smart Cities. *Frontiers in Sustainable Cities*.

Gallopin, G. C. (2020). Cities, Sustainability, and Complex Dissipative Systems. A Perspective. *Frontiers in Sustainable Cities*.

Kamalipour, H., & Iranmanesh, A. (2021). Morphogenesis of Emerging Settlements: Mapping Incremental Urbanism. *Land*, 10.

Kitchin, R., Coletta, C., Evans, L., Heaphy, L., & Donncha, D. M. (2019). Smart cities, algorithmic technocracy and new urban technocrats. In M. Raco, & F. Savini, *Planning and Knowledge: How New Forms of Technocracy Are Shaping Contemporary Cities* (pp. 199-212). Bristol: Policy Press Scholarship Online.

Netto, V. M., Peres, O. M., & Cacholas, C. (2025). Entropy and the City: Origins, Trajectories and Explorations of the Concept in Urban Science. In D. Rybski, *Compendium of Urban Complexity* (pp. 233-271). Springer.

Ng, P. (2022). Preemptive Futures: Entropic and Negentropic Information in Speculative Design. *The 10th Conference on Computation, Communication, Aesthetics*, (pp. 85-101). Coimbra, Portugal.

Prigogine, I., & Stengers, I. (1984). *Order Out of Chaos: Man's New Dialogue with Nature.* Heinemann.

Santos, G. L., Sa, A. M., & Condessa, B. (2025). Olympic regeneration vs. social (in)justice: Value capture as a referee. *Cities*.

Schrödinger, E. (2012). *What is Life?* Cambridge University Press.

Shannon, C. E. (1948). *A Mathematical Theory of Communication.* Harvard University.

Smith, G. J. (2020). The politics of algorithmic governance in the black box city. *Big Data & Society*, 2.

Varoudis, T., & Penn, A. (2021). Spectral Clustering and Integration: The Inner Dynamics of Computational Geometry and Spatial Morphology. *Formal Methods in Architecture*, 243-250.

Wade, A. (2009). *Urbanism & the Production of Knowledge: The Case of Informality.* UCL Development Planning Unit.

Wiener, N. (2019). *Cybernetics or Control and Communication in the Animal and the Machine.* The MIT Press. https://doi.org/https://doi.org/10.7551/mitpress/11810.001.0001

A Case for Socially Driven Pedestrian Simulations in Urban Environments

Puja Bhagat[1], Jonathan Wong[2], and Milad Showkatbakhsh[3]
[1,2,3]*Architectural Association School of Architecture*
[1] *puja-h.bhagat@aalumni.org, 0009-0002-6286-822X*
[2] *jonathan.wong2@aalumni.org, 0009-0007-7014-0673*
[3] *Showkatbakhsh@aaschool.ac.uk, 0000-0001-5440-5175*

Abstract. Understanding pedestrian movement is a tenet of urban design, crucially linked to the viability of urban spaces. As such, a subset of this discipline focuses on the frameworks for modelling and simulating such behaviours. Designers traditionally employ pedestrian simulations to achieve this. However, these tools, such as Mass Motion and Legion, miss a key consideration: humans are social beings. Pedestrian movement is strongly influenced by the social characteristics of space and the social nature of others. This research investigated a novel, socially driven pedestrian simulation tool developed by the authors, called Kova PedSim (formerly H.I.V.E.), a Grasshopper3D plug-in which integrates social behaviours as a key algorithm to simulate human movement. The authors conducted a series of experiments using Kova PedSim which compared the performance of socially driven pedestrian simulations and traditional ones to quantify the benefits of such an approach. The presented work showcased the importance of embedding social characteristics into digital workflows. Using Kova PedSim, designers gain access to previously undiscovered metrics, allowing them to leverage new data-driven insights. This enables the opportunity to evolve beyond using pedestrian simulations as simply analytical tools and rather as generative workflows to meaningfully embed humanistic dimensions within the design process.

Keywords. pedestrian simulations, agent-based modelling, urban analytics, human-centred design, generative design

1. Introduction

Understanding pedestrian movement is crucially linked to the viability of urban spaces. As such, a subset of this discipline focuses on the frameworks of human movement, and more importantly, how to model and simulate such a behaviour. To achieve this, designers typically employ pedestrian simulations, which offer the possibility to digitally encode human-like movement and thereby potentially predict how users may navigate the space. However, simulating human-like behaviour is not simple. The social context of an urban environment is a critical element in simulating human movement (Wijermans, 2011, p. 6). Thus, selecting a simulation's algorithm is vital in predicting such complex behaviours. While several pedestrian simulation tools exist,

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 645-654.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

they miss a key consideration: humans are social beings. Human movement is influenced by the social characteristics of space and the social nature of others. Without them, these tools detach from the realities of urban environments.

This research investigates a novel, socially driven pedestrian simulation tool developed by the authors, called Kova PedSim (formerly H.I.V.E.), a Grasshopper3D plug-in which integrates social behaviours as a key engine within its algorithm. By leveraging the ability for pedestrians to engage with one another and their environment, Kova PedSim more reasonably simulates human movement. The authors utilized Kova PedSim to conduct a series of experiments to compare the performance of a socially driven pedestrian simulation with a traditional one. The authors then extracted the key findings, which showcased the benefits and shortcomings of such an approach.

2. Literature Review

2.1. THEORETICAL APPLICATIONS OF URBAN HUMAN MOVEMENT

The relationship between spatial configuration and movement has long been studied through the concept of space syntax. Space syntax is a set of theoretical frameworks which describe architecture's influence on movement and behaviour, particularly when applied within urban environments. Hillier and Hanson, who first developed this concept, explain the built environment "provides the material preconditions for the patterns of movement, encounter and avoidance which are the material realisation - as well as sometimes the generator - of social relations… On this basis, it establishes a descriptive theory of how spatial pattern can, and does, in itself carry social information and content" (Hillier and Hanson, 1984, p. ix - xi). Space syntax also quantifies these interactions, detailing metrics related to spatial connectivity, visibility, and choice, among others (van Nes and Yamu, 2021, p.35-86). As such, this concept defines key criteria that showcase how spaces can actively shape movement and social interaction.

While space syntax provides a ruleset for how these behaviours take place, this is primarily a top-down approach whereby creating arrangements of space presume behaviours. Pedestrian simulations, conversely, codify behaviours into a simulation engine, following a bottom-up system. Based on a ruleset of movements, behaviours emerge from the urban environment rather than be prescribed to certain locations. This can be used to develop novel insights and enable a more generative design process.

2.2. EXISTING TOOLS TO SIMULATE HUMAN MOVEMENT

The creation of such a simulation engine relies heavily on the design of the algorithms used to codify human behaviour. Several pedestrian simulation tools exist which aim to simulate such movement. Mass Motion, for example, utilizes an agent-based system where agents obtain individualized movement parameters to represent the variable distribution within a population. Additionally, these agents are capable of decision making to achieve a determined goal. An agent can understand which paths are the shortest or quickest and which are counter-productive to its goal (Challenger et. al., 2011, p. 36-37). Mass Motion goes further to introduce situational behaviours, where agents may act differently in an evacuation scenario as opposed to a commuter scenario (Challenger et. al., 2011, p. 39). Other tools, such as Legion, expand upon these

behaviours. Their agents' individualized parameters include profiles for luggage capacity and familiarity with their environment (Alexandersson and Johansson, 2013, p. 14). Agent movements follow the least effort principle, where the selected path is determined by the following metrics: minimum time, costs, and congestion, and maximum speed (Challenger et. al., 2011, p. 16). Beyond that, users can define various nodes which affect the locations of decision-making, pedestrian crossing, and local destination choices (Alexandersson and Johansson, 2013, p. 15-16).

Through these algorithms, these pedestrian simulations attempt to capture core human movements in an environment. However, these engines focus on "necessary" activities, such as running errands or waiting for the bus. In doing so, they miss out on a subset of activities which are "social" in nature and strongly influenced by the social characteristics of a space (Gehl, 2006, p. 9-12). This induces new behaviours such as interacting with other agents and diverting from local or global journeys towards interest points, greatly impacting spatial navigation. Without these other behaviours, pedestrian simulations detach themselves from the realities of urban environments. While integrating social behaviours within simulated human movement is an established research domain (Kaminka and Fridman, 2018, p. 1-27) (Koh and Zhou, 2011, p.1-23), Kova PedSim aims to implement such social mechanisms into a comprehensive simulation tool which can be leveraged within a design workflow.

3. Kova PedSim: A Novel Approach to Pedestrian Simulations

Kova PedSim, a Grasshopper 3D plugin developed by the authors, builds upon the frameworks of existing pedestrian simulations but leverages a novel engine which integrates social behaviours as a key mechanism of agent movement. In doing so, the tool more reasonably simulates how humans navigate. While Kova PedSim offers three engines of social pedestrian simulations, this research focuses on the physics-based engine which is most appropriate for the scale and context of the experiments.

To develop the basis of human movement, Kova PedSim's physics-based engine uses the social-forces model, where forces "represent the effect of the environment… on the behaviour of the described pedestrian... [and] evokes the physical production of an acceleration or deceleration force as reaction to the perceived information that he/she obtains about his/her environment" (Helbing and Molnár, 1998, p. 3). In Kova PedSim, these forces include traditional reactions to environmental stimuli, such as collision avoidance and optimal routing but also introduce new social stimuli. Firstly, agents may interact with others. This is based on agent social levels, a global social level, and a small random factor, among others. Additionally, agents can move based on social levels of surrounding elements. Locally, agents may temporarily divert their journeys to investigate highly social interests along their path. Globally, agents use environmental and programmatic social weightings to influence their routing process. As such, agents may prioritize routes with a more socially active journey, rather than simply the shortest route. Socially driven movement is also implemented at the agent level. Individualized agent parameters affect an agent's movement and social behaviours, such as social level, age, personal space, fitness, and more.

Through the integration of these behavioural and social mechanisms, the tool enables a more semi-deterministic algorithm of human movement, modelling pedestrian movements which go beyond traditional simulations to capture human

interactions and chance encounters which emerge from the nuances of the surrounding context. Thus, Kova PedSim enables new emergent patterns and behaviours to develop, providing a more holistic understanding of the design environment.

4. Methodology

Three sets of experiments were conducted using Kova PedSim to understand the performance of a social pedestrian simulation as compared to a traditional one. Each set tested a common urban condition around Trastevere, Rome. This location was selected due to the highly social nature of Italian culture, along with the variety of urban configurations in Trastevere. Within one experiment set, one socially driven and one traditional pedestrian simulation was conducted. Both experiments were evaluated by the same criteria to understand the insights provided by each simulation framework.

All simulations were conducted using Kova PedSim to maintain consistency. The traditional experiment used a 0.0 global social value to disable all social mechanisms. All simulations specified a population of 1000 agents generated from four unique agent types. Within each, agents were randomly assigned parameters from defined ranges to enable a diverse population (Figure 1). Each agent had a randomly selected start and end destination from the list of eligible gates. All gates had a flow rate of 10 agents per iteration. Additional programs were located following the geography of each site. Each experiment was conducted for 3000 iterations to ensure a valid data set.

Teenagers	**Young Adult**	**Adult with Children**	**Elderly**
Social Level: 0.4 - 0.6	Social Level: 0.7 - 0.9	Social Level: 0.2 - 0.4	Social Level: 0.5 - 0.7
Journey Length: 1	Journey Length: 1	Journey Length: 1	Journey Length: 1
Base Velocity: 0.8 - 1.0m	Base Velocity: 0.8 - 1.0m	Base Velocity: 0.5 - 0.8m	Base Velocity: 0.5 - 0.8m
Age: 13 - 19 Years	Age: 20 - 29 Years	Age: 25 - 40 Years	Age: 60 - 80 Years
Fitness: 0.6 - 1.0	Fitness: 0.6 - 1.0	Fitness: 0.3 - 0.8	Fitness: 0.3 - 0.6
Count: 250 Agents	Count: 250 Agents	Count: 250 Agents	Count: 250 Agents

Figure 1. A description of agent types used within all simulations

All simulations were evaluated using the same four criteria. The first criteria calculated the total foot traffic per mesh cell, which was measured by counting the number of agents who entered a cell's bounds during the simulation. The second criteria evaluated the density of agents in each cell at any given time. Both maximums and averages were recorded every 50 iterations to understand the temporal fluctuations of this data. The third criteria measured the foot traffic near interest points, which was calculated by averaging the number of agents who passed through each program type's cell location. Finally, the fourth criteria calculated the number of iterations required for an agent to complete its journey. The goal of defining such criteria was to understand the differences in agent behaviour and subsequently, the design insights related to foot traffic distribution, emerging congestion, and efficiency of travel, among others.

5. Experiment Results

5.1. EXPERIMENT 01: CONVERGENCE URBAN CONDITION

The first experiment tested a convergence condition in the north of Trastevere, where multiple streets intersected in one zone and programs were consolidated in small pockets within the urban fabric. The set-up defined each street entrance as both a gate

and target. All other points were located and categorized geographically (Figure 2).



Figure 2. Experiment 01 Environment Set-up

As evident in the results, a consideration for social behaviours produced different movement patterns within each simulation. While both simulations exhibited similar amounts of footfall (Figure 4, Graph 5), key variations emerged when analysing the travel paths and timing of agent movement.

Agents within the traditional simulation solely followed the shortest path route towards their destinations without considering the surrounding context. However, agents following socially driven behaviours diverted their paths, trading a slightly

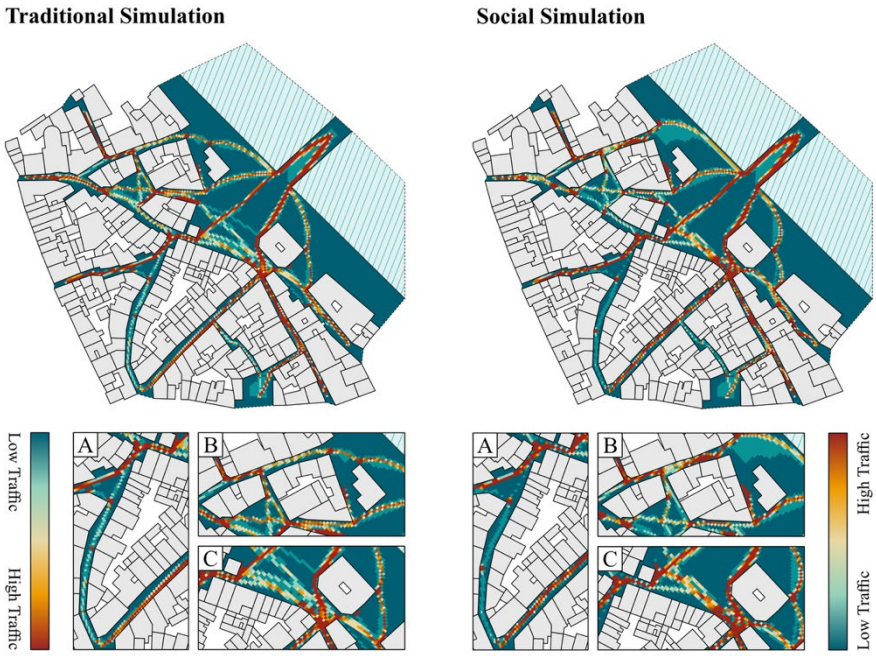


Figure 3. Experiment 01 Footfall Heatmap Results

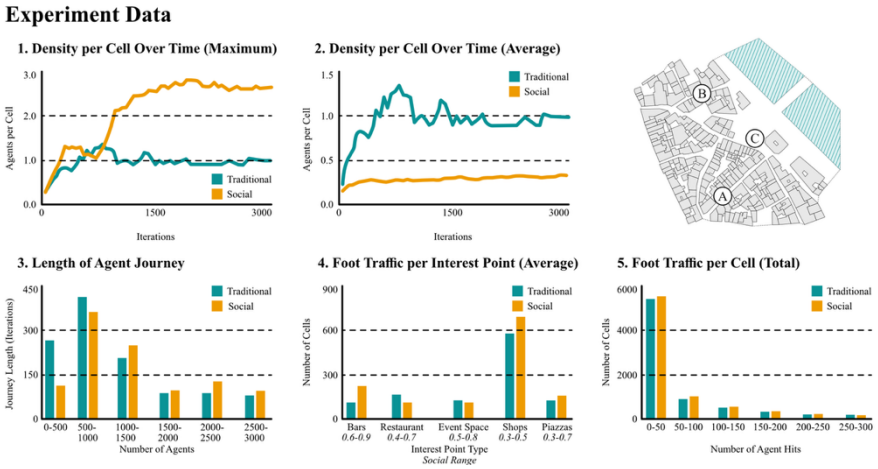


Figure 4. Experiment 01 Experiment Data Results

longer journey for a more socially engaging experience. This can be seen in Figure 3, Callouts A and B, where socially driven agents altered their global and local routes towards streets with a higher concentration of socially valuable points, showcasing a 19% increase in foot traffic near interest points on average (Figure 4, Graph 4). This type of altered route behaviour also affected the overall densities within the urban fabric. While the traditional simulation showcased a generally low maximum density with agents being evenly dispersed, the socially driven simulation revealed a series of highly dense pockets of agent movement, increasing maximum densities by around 220% on average (Figure 4, Graph 1). This is seen in Figure 3, Callout C, where the central piazza became more congested. This was due to the diverted foot traffic and the larger quantity of agents stopping to socialize here, more closely reflecting reality.

Based on these findings, designers can leverage the knowledge gained from the socially driven pedestrian simulation to unlock new layers of information which were not previously visible. New design insights showcased emerging areas of social hotspots towards high value programs and locations of increased congestion throughout the urban fabric. Within a generative mindset, designers can leverage this new social data within their design decisions to strategically place key urban entities to control movement and create social pockets within the environment.

5.2. EXPERIMENT 02: IRREGULAR URBAN CONDITION

The second experiment tested an irregular urban condition in the north-east, where a variety of street conditions wove throughout a residential neighbourhood. As opposed to Experiment 01, programs were dispersed evenly within the space, without a particular set of main streets. As before, the set-up defined each street entrance as gates and targets for all agents. All interest points were geographically placed and defined (Figure 5).

The results revealed how the socially driven pedestrian simulation showcased redistributed human movements within the urban fabric which were not present in the traditional simulation. While the average and maximum densities remained similar

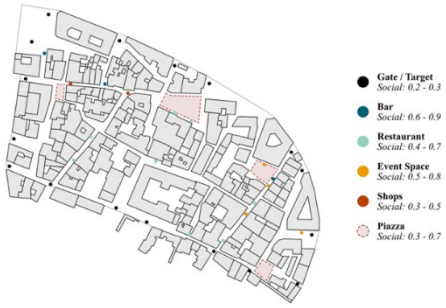


Figure 5. Experiment 02 Environment Setup

between both experiments (Figure 6, Graph 1), the specific locations for such density were relocated (Figure 6, Heatmap). Without social behaviours, agents tended to heavily utilize a few key streets within the urban fabric, leaving others to have reduced foot traffic. This concentrated the urban densities within those key areas. On the other hand, the socially driven agents showcased a higher usage throughout the entire urban fabric, more evenly distributing the pockets of high congestion (Figure 6, Callouts A, B, and C). This is due to the different arrangement of urban space, where interest points were evenly dispersed throughout. This encouraged a diversity in agent movements, where their route was influenced by the social level of the programs they encountered. This can be seen in Figure 6, Graph 3 and 4, where agents embarked on journeys which

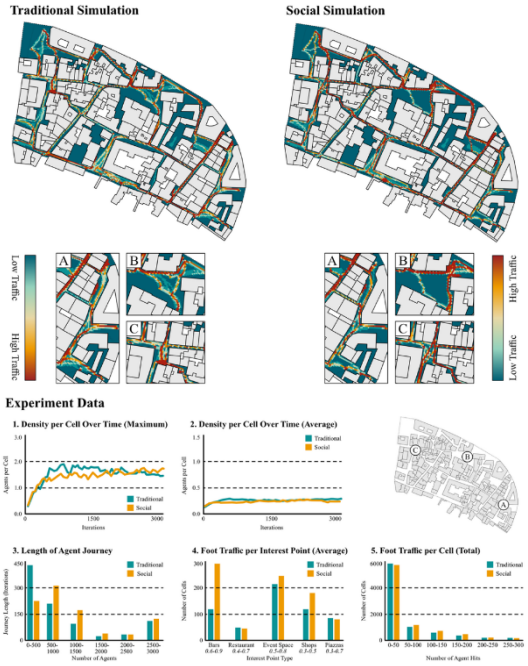


Figure 6. Experiment 02 Footfall Heatmap Results

were on average 8% longer as to re-route towards highly social programs.

By leveraging the social data in this experiment, designers can uncover new insights about human movement based on the urban form and its spatial properties. Without a social layer, it was perceived that congestion was concentrated on certain streets. However, when considering a social influence, the more likely human movement in this space revealed a lack of truly central zones, where movement is more evenly distributed. Using this layer of knowledge, designers can understand how strategically placing social programs can intentionally divert agent movement. In turn, they can disperse areas of high congestion and create new interest points within the urban fabric.

5.3. EXPERIMENT 03: GRID-LIKE URBAN CONDITION

The final experiment tested a grid-like urban condition in central Trastevere. While this site condition was also embedded within a residential neighbourhood, the program



Figure 7. Experiment 03 Environment Setup

distribution here was condensed to a small group of streets, leaving the remaining streets devoid of interest points. As before, the experiment defined site entrances as gates and targets. All remaining points were geographically located and programmed (Figure 7).

In this experiment, the socially driven simulation revealed new urban flows within



Figure 8. Experiment 03 Footfall Results

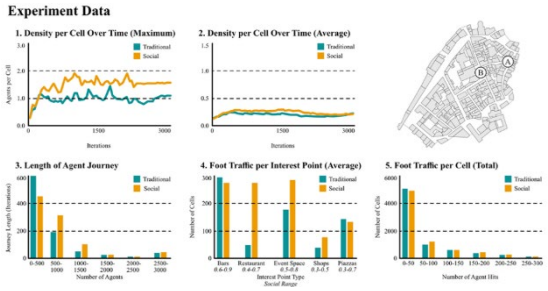


Figure 9. Experiment 03 Experiment Data Results

the environment. Paths which were previously highly populated in the traditional simulation became completely abandoned when agents gained a social understanding of the urban fabric. This is because paths which were technically the shortest route were residential, devoid of any interest points. When agents considered social factors, they re-directed their paths towards the active main street, as seen in Figure 8, Callouts A and B. As a result, agent journeys increased in length by about 20% on average and maximum urban densities increased by about 50% (Figure 9, Graphs 1 and 3).

Using this knowledge, designers can understand how the shortest path is not necessarily the most optimal route when considering social factors. Agents tended to walk towards more active and engaged zones, even if those journeys were longer or more congested. This was particularly evident in areas which are highly residential, where surrounding streets which are devoid of programs are less utilized. As such, when embedding this data in a generative framework, designers could leverage socially active elements to control urban flow and create emerging urban main streets.

6. Discussion

In each experiment, a comparison between a traditional pedestrian simulation and a socially driven one demonstrated how social movements of pedestrians express unique emergent behaviours in response to specific urban conditions. While agents in a traditional simulation widely ignored their context and strictly followed a shortest path journey, socially driven agents tended to divert their journeys towards locations of higher social value. This shifted human flow patterns and related densities within the urban fabric. Such trends were most clearly revealed when looking at the micro-scale zones and related data sets within the larger urban fabric. Through this, the research emphasized that pedestrian movement is highly influenced by the social programming and organization of the urban fabric, as explained by space syntax principles, and in turn, showcased the importance of considering social behaviours in urban design.

While the research established a strong foundation, it should be noted that the conducted simulations are not validated against real-world conditions and are primarily intended as a comparative study. Future research should be conducted to more accurately tune the simulation towards real-world behaviours, particularly to match local movement patterns and specific demographic distributions. This would empower architects and designers to make more informed and sensitive design decisions.

7. Conclusion

Through an exploration of Kova PedSim, the research highlighted how a consideration for the social nature of humans in pedestrian simulations reveals a new layer of information, which architects and designers can leverage in generative design processes. Socially driven movements exhibited different types of emergent behaviours based on the urban condition, much in line with the principles of space syntax. Each experiment demonstrated a shift in movement from a straight-line, shortest walk approach to seemingly less intentional and more diverse travel paths. This emphasized that a consideration of social behaviours could provide a more reasonable model for pedestrian flow through an urban space, as individuals may take a more socially active route when available. With this knowledge, a designer would be better equipped and more well-informed on the challenges or opportunities which exist in the space. They may leverage such information to mitigate congestion, redirect movement to under-utilized zones, or identify locations and design opportunities for informal socialization. Using Kova PedSim, designers gain access to previously undiscovered metrics, allowing them to leverage new data-driven insights and move beyond understanding pedestrian simulations as an analytical tool but rather a generative one.

References

Alexandersson, S., & Johansson, E. (2013). Pedestrians in microscopic traffic simulation: Comparison between software Viswalk and Legion for Aimsun [Master of Science Thesis i]. Chalmers University of Technology.

Challenger, R., Clegg, C. W., & Robinson, M. A. (2011). Simulation Tools (Understanding Crowd Behaviours). University of Leeds.

Gehl, J. (2006). Life between buildings: Using public space (6. ed., 2. issue). The Danish Architectural Press.

Kaminka, G. A., & Fridman, N. (2018). Simulating urban pedestrian crowds of different cultures. ACM Transactions on Intelligent Systems and Technology (TIST), 9(3), 1-27.

Koh, W. L., & Zhou, S. (2011). Modeling and simulation of pedestrian behaviors in crowded places. ACM Transactions on Modeling and Computer Simulation (TOMACS), 21(3), 1-23.

Helbing, D., & Molnár, P. (1998). Social force model for pedestrian dynamics. Physical Review E, 51(5). https://doi.org/10.1103/PhysRevE.51.4282

Hillier, B., & Hanson, J. (1984). The Social Logic of Space (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511597237

van Nes, A., & Yamu, C. (2021). Analysing Linear Spatial Relationships: The Measures of Connectivity, Integration, and Choice. In A. van Nes & C. Yamu (Eds.), Introduction to Space Syntax in Urban Studies (pp. 35–86). Springer International Publishing. https://doi.org/10.1007/978-3-030-59140-3_2

Wijermans, F. E. H. (2011). The CROSS Model—A Computational Crowd Model. In Understanding crowd behaviour: Simulating situated individuals. University of Groningen, SOM research school.

A Multi-Agent Grasshopper Component for Behaviour-Driven Furniture Relocation under Dynamic Solar–Shade Conditions

Moonflower

Ayşe Ebrar Kavak[1], Beyza Gül Topçu[2], Sema Alaçam[3], and Orkan Zeynel Güzelci[4]
[1,2,3,4]*Istanbul Technical University, Faculty of Architecture*
[1]*kavak21@itu.edu.tr, 0009-0001-3822-2405*
[2]*topcube20@itu.edu.tr, 20009-0005-3385-498x*
[3]*alacams@itu.edu.tr, 30000-0002-5979-3282*
[4]*guzelci@itu.edu.tr, 40000-0002-5771-4069*

Abstract. This study introduces Moonflower, a multi-agent simulation component developed in Grasshopper to model how humans and movable outdoor furniture adapt to dynamic solar–shade conditions. While architectural simulations commonly treat shade as an environmental performance parameter, this study positions shade as a behavioural attractor that influences spatial occupation and micro-scale reconfiguration. The research addresses a gap between human-centred agent-based models and environmental analysis workflows by asking how the integration of human and non-human agents can represent adaptive outdoor behaviour in a computationally explicit manner. The methodology combines Ladybug-driven solar simulations (a climate and environmental analysis plugin for Grasshopper) with a custom C#-based agent system, enabling people, tables, chairs, and temporally shifting shade zones to interact within a unified computational environment. A field study conducted at Istanbul Technical University, Faculty of Architecture (Taşkışla Campus) provides empirical observations of daily shade-driven relocation patterns, which are used to inform and contextualise model calibration. Results demonstrate that furniture movements are strongly correlated with the temporal progression of shaded areas, leading to distinct patterns of spatial use and comfort over time. By explicitly linking environmental conditions with behavioural adaptation, Moonflower provides a framework for designing responsive and climate-aware public spaces.

Keywords. Multi-agent simulation, Behaviour-driven shade modelling, Movable outdoor furniture, Human-environment interaction, Grasshopper

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 655-664.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

1. Introduction

In outdoor public spaces, especially in warm climates, shade strongly shapes how people occupy and adapt to their surroundings. Shade is produced by both natural and built elements (e.g., trees and canopies), providing microclimatic comfort and acting as spatial attractors that influence where people sit, rest, or temporarily settle. Early observational studies, including Whyte (1980) on urban squares and Perec (1975) through detailed accounts of everyday urban life, show that sunlight, shade, and subtle temporal changes play a decisive role in patterns of public space use. Although computational environmental tools have advanced significantly, most digital simulations still treat shade primarily as a climatic or performance metric rather than as a factor that influences behavioural adaptation and spatial decision-making.

Agent-based models have become essential tools for analysing urban and architectural environments (Chen, 2012). They enable the representation of collective pedestrian dynamics, including movement, spacing, and local interactions, alongside spatial occupancy patterns (Firas et al., 2023). However, in most of these studies, the agents are exclusively human, while non-human elements such as movable furniture remain passive or static. In many campus courtyards and public plazas, tables and chairs are frequently repositioned by users throughout the day in response to changing sunlight and shade conditions. Recent behavioural mapping research emphasises the importance of flexible seating and shaded comfort in shaping spatial use (Loo and Fan, 2023), yet this dimension of human–environment interaction has received limited attention in computational design research, particularly in models that couple behavioural logic with dynamic environmental data.

This study introduces Moonflower, a multi-agent component developed in Grasshopper using C#. It couples a solar–shade analysis workflow (Ladybug Tools, a climate and environmental analysis plugin for Grasshopper) with an agent system to simulate how humans and movable elements treated as agents (e.g., tables and chairs) respond to shifting shade patterns over time. By combining environmental inputs with decision rules, Moonflower models shade as a behavioural attractor and connects agent decision-making to solar–shade dynamics.

2. Background

2.1. AGENT BASED MODELLING IN DESIGN RESEARCH

Agent-based modelling is a computational approach used to represent human behaviour, spatial dynamics, and decision-making in urban and architectural environments. By simulating individual agents and their interactions, studies have examined collective pedestrian dynamics, evacuation patterns, and pedestrian comfort under varying spatial conditions (Chen, 2012; Karoji et al., 2019). In these applications, agent behaviours are typically defined through local interaction rules, and simulation outputs are used to describe emergent patterns at the level of groups or space use. Recent studies have also developed hybrid or data-driven frameworks that connect agent-based modelling with environmental or social variables, including applications in design evaluation and optimisation contexts (Sun et al., 2022; Firas et al., 2023).

2.2. ENVIRONMENTAL AND SOLAR–SHADE ANALYSIS IN OUTDOOR SPACES

Environmental analysis tools and digital workflows integrating solar radiation, daylight, and energy modelling are widely used to assess outdoor and microclimatic conditions with high temporal resolution (Kastner and Doğan, 2021; Lin and Brown, 2021). Research applying solar–shade analysis has addressed topics such as façade performance, energy optimisation, and the effectiveness of shading for pedestrian areas (Del Campo et al., 2020; Tomasi et al., 2024). Empirical studies in outdoor settings have also examined the relationship between shade, comfort, and social interaction (Whyte, 1980; Elnabawi and Hamza, 2020). In parallel, studies on public-space use have discussed the role of movable furniture, seating edges, and microclimatic relief in shaping patterns of outdoor activity (Loo and Fan, 2023; Loo and Zhang, 2024). Computational design research has further explored how environmental and spatial performance measures can inform design decision-making, including canopy and shading strategies evaluated alongside structural and social considerations (Van Ameijde et al., 2022).

2.3. RESEARCH GAP AND MOONFLOWER FRAMEWORK

Across these bodies of work, behavioural simulation and solar–shade analysis are commonly developed as separate workflows, and computational models that simultaneously consider human movement, movable outdoor furniture, and time-varying shade remain limited. In addition, agent-based studies often treat furniture and shading as static context rather than as elements that can influence spatial use over time.

Moonflower responds to this gap by proposing a unified simulation framework in which environmental conditions and agent-based processes are addressed together. Implemented in Grasshopper, the framework establishes a conceptual link between solar–shade conditions and adaptive spatial behaviour, forming the basis for examining how human use and furniture arrangements evolve in response to changing environmental conditions. The specific structure, parameters, and operational logic of this framework are detailed in the methodology section.

3. Methodology

The methodology combines environmental simulation, empirical observation, and multi-agent behavioural modelling within a unified Rhinoceros–Grasshopper workflow. The overall process is organised into four stages: (1) solar–shade input generation, (2) spatial model preparation, (3) agent-based simulation, and (4) output extraction and interpretation (Figure 1).

The workflow begins by generating solar–shade inputs in Ladybug Tools. Geographic information and analysis periods are processed through components such as Construct Location, Analysis Period, and Sunpath to represent solar positions across different times of the day and year. A Direct Sun Hours analysis then produces maps that distinguish areas of higher and lower solar exposure. To make these results usable within the simulation, the ground surface is divided into a regular sampling grid, and each sampling point is assigned a numerical value corresponding to direct sunshine duration. Spatial geometries of the study area—including ground surfaces, shading

elements (e.g., trees and canopies), building masses, obstacles, and initial furniture locations—are modelled in Rhinoceros and referenced directly within Grasshopper (Figure 1). These geometries are evaluated against the solar–shade data to establish the environmental and geometric context required for agent movement and collision evaluation.

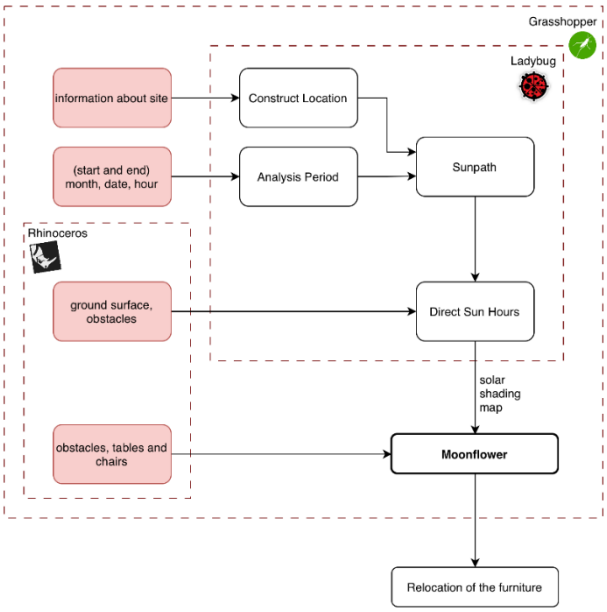


Figure 1. Workflow overview

The Moonflower simulation component is implemented in Grasshopper using C# and represents both human agents and movable-object agents (i.e., furniture such as tables and chairs) within the same computational environment. Human agents define a perception field by sampling solar values at nearby grid points within a specified radius and comparing candidate positions to their current location. Movement follows a local decision rule: agents move towards candidate points with lower solar exposure; if values are equal, a random deviation is added. Furniture agents operate through complementary logic that considers local demand (e.g., nearby people) and shade status, and reposition when predefined thresholds are exceeded. Furniture relocation is constrained by geometric collision checks, including bounding-box tests and Brep–Brep intersection checks.

The simulation progresses in iterative cycles in which environmental conditions are updated, agents reassess their perception fields, and movement vectors are recalculated. Outputs include trajectories of human movement, furniture relocation patterns, and temporally varying maps of spatial use. The internal operational logic of Moonflower is summarised in the flow diagram (Figure 3): analysis points are first verified against environmental values; if inputs are missing, the process terminates safely. Valid points are then randomised according to environmental values and sorted to identify candidate locations, which are evaluated to generate as many collision-free targets as there are

seating elements. At each candidate location, a seating template is translated in the XY plane and tested for collisions; accepted points are added to a temporary obstacle set to prevent overlap in subsequent placements. Once a sufficient number of valid targets is found, seating geometries are moved to the accepted XY targets while preserving original Z elevations, and the resulting geometries are produced as output.

4. Case Study

This section presents on-site observations conducted at the study location to document how shading conditions change throughout the day and how users respond to these variations. The observations describe recurring patterns of spatial use under shifting sunlight and shade and provide contextual grounding for interpreting the behaviours generated by the Moonflower Grasshopper component.

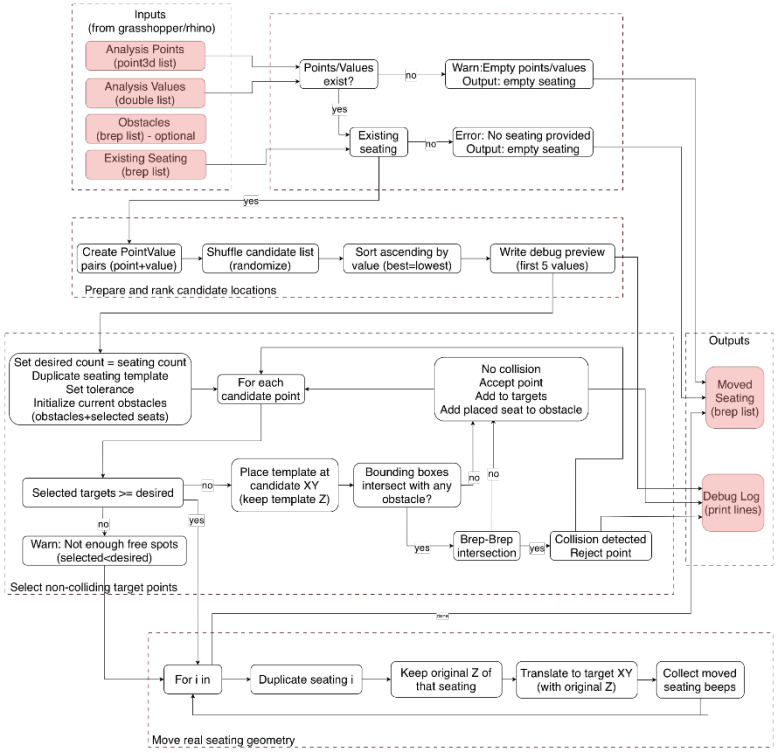


Figure 2. Grasshopper definition of Moonflower

4.1. CONTEXT AND CONSTRAINTS

The study area is located in Istanbul Technical University, Faculty of Architecture (Taşkışla Campus), Istanbul, Türkiye. The courtyard setting is subject to summer conditions in Istanbul, where high solar exposure strongly influences outdoor use. Summer temperatures often reach around 30°C, and midday periods can involve intense solar radiation. The site's south–southwest orientation produces variable

shaded zones across the day and year, particularly between the summer solstice and the equinoxes. Long daylight hours and limited wind-driven cooling increase the value of shaded areas for outdoor comfort. These climatic and spatial conditions form the context in which users make shade-seeking choices and, in some instances, reposition movable furniture.

The case-study setup includes the courtyard boundary, obstacles such as trees and building masses, and initial furniture locations recorded at the start of the observation period (Figure 4). The furniture movement zone defines the area within which tables and chairs were observed to be repositioned during the day. These spatial inputs provide the geometric context for interpreting both the on-site observations and the configuration changes produced by the Moonflower Grasshopper component.

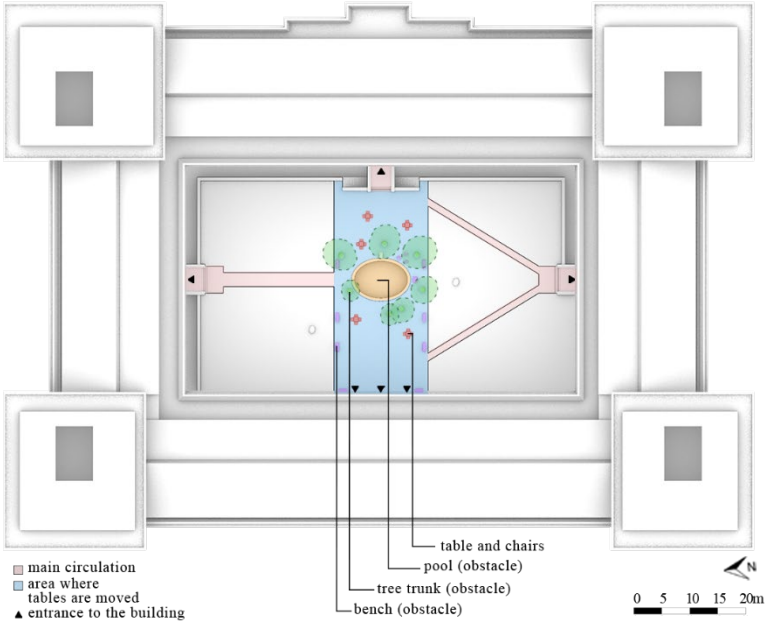


Figure 3. Spatial layout of the study area and furniture movement zone

Environmental conditions were characterised using Ladybug Tools to compute solar positions and direct sunlight values for the selected dates and time periods. These values were mapped onto the site geometry to describe the changing shade availability across the courtyard and to support the interpretation of observed user locations and furniture movements. Spatial inputs included the courtyard boundary and obstacles such as trees and building masses, together with the initial positions of tables and chairs. At the case-study level, these inputs define the environmental context (shade availability) and spatial constraints (accessible and collision-free areas) against which relocation behaviour is read, while implementation details of the agent system are presented in the methodology section.

4.2. SHADE DISTRIBUTION AS ENVIRONMENTAL INPUT

Figure 5 presents the hourly distribution of shade across the courtyard for the selected summer dates. The maps distinguish areas that remain consistently sun-exposed, zones that receive temporary shade during specific hours, and regions that retain longer periods of shade beneath tree canopies. This time-dependent shade information provides the environmental reference used to interpret when and where shade-seeking occupancy and furniture relocation were observed on site.

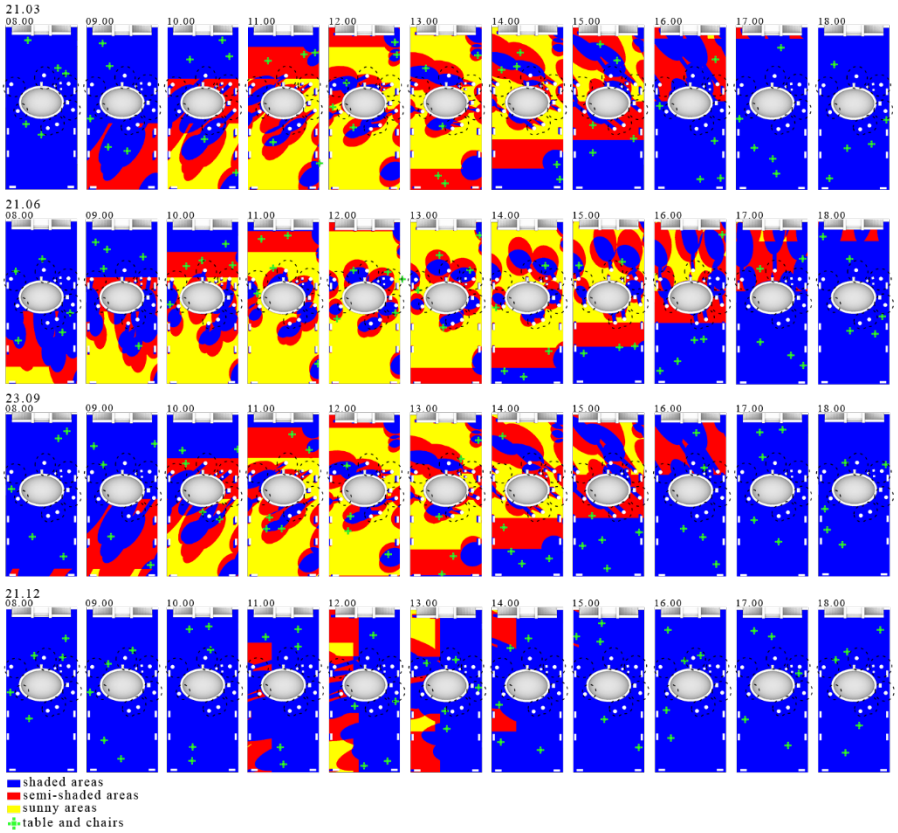


Figure 4. Hourly shade distribution across the study area for selected summer dates.

The hourly patterns also make the temporal movement of shadows legible, showing how shaded edges expand, shift, and contract throughout the day. Reading these transitions alongside the observed locations of users and seating provides a consistent basis for describing shade-related spatial use without relying on exact positional matching. The maps therefore support case-study interpretation by clarifying which areas offer persistent shade and which areas provide short-lived relief at specific times.

4.3. OBSERVED BEHAVIOUR AND COMPARISON WITH SIMULATION OUTPUTS

On-site observations conducted on 7 July show a clear alignment between user behaviour and the progression of shade across the day. As solar exposure increased towards midday, activity became increasingly concentrated beneath tree canopies and along shaded edges, while fully exposed areas showed a marked decline in use. Several instances of user-driven furniture relocation were recorded, with tables and chairs repositioned to follow newly formed shaded zones (Figure 6). These observations indicate that users can actively adjust the courtyard's seating configuration in response to changing microclimatic comfort conditions.



Figure 5. Time-lapse photographs showing shade movement and furniture relocation

The observations were quantified by recording the number of users in shaded and sun-exposed areas at specific time intervals between 09:30 and 16:30. Human-agent

density outputs generated for the same time windows were then compared with the field records in terms of temporal and spatial distribution. Rather than evaluating exact point-by-point agreement, the comparison focuses on whether both datasets exhibit consistent time-dependent preferences for shaded areas and similar shifts in concentration zones as shade moves.

The simulation reproduced these tendencies by generating movement trajectories that concentrated human agents in lower-exposure areas and by triggering furniture agents to reposition when predefined comfort thresholds were exceeded. Across both datasets, shade-seeking behaviour resulted in comparable time windows of increased use and similar areas of concentration. These observations support the use of shade as a behavioural attractor for describing micro-scale adaptation in outdoor settings, while also indicating how movable seating can contribute to comfort-driven reconfiguration.

5. Discussion and Conclusion

This study demonstrates that modelling humans, movable furniture, and shade within a unified multi-agent simulation can provide a richer account of outdoor spatial behaviour than approaches that focus on pedestrians alone. By coupling time-varying solar–shade information with agent decision-making, Moonflower shows how microclimatic variation—particularly the temporal progression of shade—can influence where people choose to stay and when furniture is repositioned during the day. Compared with common workflows in which environmental analysis and behavioural simulation are handled separately, the contribution of this study lies in operationalising shade as a time-dependent input within the same simulation loop that also represents movable seating as an active element of spatial change.

A key contribution is the treatment of shade as a behavioural attractor rather than a static environmental output. This framing supports the modelling of spatial arrangements that emerge from both environmental conditions and user-driven reconfiguration. In contrast to many agent-based models where objects remain passive context, Moonflower explicitly represents movable furniture as agents whose relocation is evaluated alongside human movement under changing shade availability. The framework also offers practical value for design exploration: by visualising movement trajectories and shifts in spatial use over time, it can inform early-stage considerations related to seating distribution, canopy placement, and microclimatic interventions.

The findings should be interpreted in light of the study's scope and constraints. The empirical example is site-specific, small in scale, and based on a controlled observation setting, which limits the extent to which bottom-up complexity can be demonstrated in the current paper. Accordingly, the results are presented as evidence of plausibility for the coupled behaviour–environment framework rather than as a comprehensive validation across diverse conditions.

Future work will extend Moonflower to incorporate thermal comfort indices, group behaviour dynamics, and real-time sensor data. Testing the framework across multiple sites with varying spatial configurations and social contexts will be essential for evaluating robustness and generalisability. Integrating predictive models or machine learning may also support automated calibration and scenario comparison,

strengthening the framework's potential as a decision-support tool for outdoor environments.

References

Chen, L. (2012). Agent-based modeling in urban and architectural research: A brief literature review. *Collection of Frontiers of Architectural Research,* 1, 166-177. https://doi.org/10.1016/j.foar.2012.03.003

Firas, A. D., Wei, Y., & Edin, J. (2023). Campusim: An Integrated Parametric BIM for Campus Design Simulation and Optimization*. In 28th International Conference on Computer-Aided Architectural Design Research in Asia: Human-Centric, CAADRIA 2023 (pp.471-480).* The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).

Karoji, G., Hotta, K., Hotta, A., & Ikeda, Y. (2019). Pedestrian dynamic behaviour modeling. *In Proceedings of the 24th international conference on computer-aided architectural design research in Asia: intelligent and informed, CAADRIA* (pp. 281-290). The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).

Kastner, P., & Dogan, T. (2022). Eddy3D: A toolkit for decoupled outdoor thermal comfort simulations in urban areas. *Building and Environment*, 212, 108639. https://doi.org/10.1016/j.buildenv.2021.108639

Lin, J., & Brown, R. D. (2021). Integrating microclimate into landscape architecture for outdoor thermal comfort: a systematic review. *Land*, 10(2), 196.

Loo, B. P., & Fan, Z. (2023). Social interaction in public space: Spatial edges, moveable furniture, and visual landmarks. *Environment and Planning B: Urban Analytics and City Science*, 50(9), 2510-2526.

Loo, B. P., & Zhang, F. (2024). Design of public open space: Site features, playing, and physical activity. *Health & Place*, 85, 103149. https://doi.org/10.1016/j.healthplace.2023.103149

OpenAI. (2025). *ChatGPT Model:* GPT-4 (It was used to improve flow)| Temp: 0.7 [Large Language Model]. https://openai.com/chatgpt/

Perec, G. (1975/2010). *An attempt at exhausting a place in Paris* (M. Lowenthal, Trans.). Wakefield Press.

Sun, S., Sun, C., Duives, D. C., & Hoogendoorn, S. P. (2023). Neural network model for predicting variation in walking dynamics of pedestrians in social groups. *Transportation*, 50(3), 837-868. https://doi.org/10.1007/s11116-021-10263-8

Tomasi, M., Nikolopoulou, M., Giridharan, R., & Löve, M. (2024). A design workflow for effective solar shading of pedestrian paths*. Building and Environment*, 261. https://doi.org/10.1016/j.buildenv.2024.111718.

Van Ameijde, J., Ma, C. Y., Goepel, G., Kirsten, C., & Wong, J. (2022). Data-driven placemaking: Public space canopy design through multi-objective optimisation considering shading, structural and social performance. *Frontiers of architectural research,* 11(2), 308-323. https://doi.org/10.1016/j.foar.2021.10.007

Whyte, W. H. (1980/2001). *The Social Life of Small Urban Spaces*. New York: Project for Public Spaces.

Root-Inspired Generative Network

An Agent-Based Model for the Design of Context-Aware Architectural Structures

Kwong Lau Hung[1] and Toni Kotnik[2]
[1]*School of Architecture, The Chinese University of Hong Kong*
[2]*Department of Architecture, Aalto University*
[1]*gordonhung@cuhk.edu.hk, 0009-0000-8943-7484*
[2]*toni.kotnik@aalto.fi, 0000-0001-7640-0009*

Abstract. Natural root systems exemplify self-organizing networks that grow adaptively in response to complex environmental stimuli, revealing a bottom-up logic where efficient structures emerge from local interactions rather than centralized control. Building on this biological paradigm, this research investigates how the self-organizing principles of growth in natural root systems can be abstracted into an agent-based computational model to generate adaptive architectural forms, augmenting the architect's ability to create highly efficient, context-aware design. The model is developed by translating key biological mechanisms, including tropism, branching, and grafting, into geometrical operations that govern the iterative growth of agents in an algorithm written using Python within Rhino/Grasshopper. The model is subsequently applied in two conceptual design proposals in Hong Kong, demonstrating its capacity to generate proto-architectural systems for slope support and pedestrian circulation by adapting to specific site constraints. The outcome bridges the gap between biological principles and architectural synthesis by offering a generative conceptual framework for creating architectural forms with enhanced structural efficiency and ecological integration within complex environments at an early design stage.

Keywords. root systems, agent-based modelling, bio-inspired computation, generative design, context-aware structures

1. Introduction

The architectural discipline is confronted by escalating environmental issues, challenging conventional top-down design methods. These traditional approaches are inherently limited in their capacity to handle complex site conditions, often producing architectural forms asynchronous with their dynamic contexts. Already since the 1950s and the emergence of system theory and cybernetics, scientists have argued for a systemic approach, framing design as a "network of interrelated requirements" (Alexander, 1979), thereby establishing a foundation for bottom-up design thinking.

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 665-674.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

This systemic logic finds a powerful analogue in biological systems that exhibit robust bottom-up intelligence developed over billions of years of evolution. D'Arcy Thompson described biological form as a "diagram of forces" (Thompson, 1992), a morphology emerging from context-aware growth. In particular, the root system of Banyan trees presents a highly efficient, decentralized structural network developed through local interactions with environmental stimuli like gravity and moisture (Fig. 1). This presents a potent model for an exploration of architectural methodology where form emerges as a system in direct dialogue with its context, rather than manually imposed.



Figure 1. Root system of Banyan trees adapting to its context for nutrient and support

Inspired by the context-awareness of root growth, this research develops an agent-based algorithm that translates the self-organizing principles in root systems into a generative tool for proto-architectural form-finding. Scripted in Python within the Rhino/Grasshopper environment, the algorithm does not aim for exact mimicry of roots but offers a conceptual interpretation that abstracts their growth logic to generate structural and circulatory organizational networks for architectural design.

The goal is to augment designers' capability to create complex, context-aware structures for circulation and support. This research bridges a critical gap between biological intelligence and architectural synthesis, demonstrating a bio-inspired systemic approach of architectural form-finding where structure emerges from direct dialogue with contextual constraints and opportunities.

2. Literature Review

This research is grounded in the shift towards a biological paradigm of architecture, which moves from object-making into the emergence of performative, adaptive systems. Hensel's "performance-oriented" approach frames performance as a dynamic outcome of interactions between subject, environment, and material organization (Hensel, 2013). Similarly, Weinstock's studies of morphological processes in nature provide a framework for understanding emergence and self-organization as generative principles (Weinstock, 2010) echoing Frei Otto's earlier research (Otto, 2011). This discourse establishes a crucial intellectual foundation but often remains conceptual, lacking practical methodologies for early-stage design.

Computational methods including L-Systems (Prusinkiewicz, 1991), Boids (Reynolds, 1987), and Conway's Game of Life (Berlekamp, 2003), demonstrate how complex order emerges from simple, local rules. These tools are powerful for simulation but often remain in representation or analysis, lacking a direct pathway to

synthesize emergent patterns into architectural proposals that integrate spatial, programmatic, and tectonic requirements.

A dominant response is the material and fabrication-oriented approach pioneered at the ICD/ITKE, University of Stuttgart. Their research integrates computational form-finding and physical realization by making material properties and fabrication constraints primary design drivers (Menges, 2012). While crucial for materialization, its strength in articulating physical systems reveals a need for a complementary conceptual methodology to generate the underlying adaptive organizational logic.

Positioned within these gaps, this research develops a root-inspired generative methodology. It aims to provide a framework for the conceptual design stage, translating the logic of biological growth into organizational networks. By focusing on the abstract generation of context-aware topologies, this work seeks to complement simulation-based analysis and material-oriented approaches, offering a missing link in a comprehensive bio-inspired design workflow.

3. Methodology

To create a computational algorithm by translating natural growth patterns in root systems into organizational networks for design applications, the methodology unfolds in three main sections: Geometrical Translations, Parametric Variations, and External Adaptations. It translates biological mechanisms of root growth into geometrical operators, building up a feedback loop that demonstrates emergent properties resulting from the interaction of its agents and their adaptation to external stimuli (Fig. 2a). The system as a network of flow is then applied to real-world design scenarios in the next stage to generate proto-architectural solutions for complex design conditions.

The algorithm operates under an iterative agent-based model where root growth is represented by a series of discrete movements of a point over time. In every iteration, a compound-vector with predefined magnitude directs the point to the next location. As this process repeats, a trajectory of movement in 3D is generated, simulating the growth of a continuous root in nature (Fig. 2b).

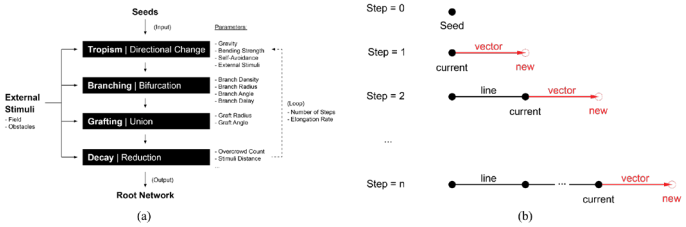


Figure 2. (a) Feedback loop composed of geometrical operators; (b) Iterative growth of roots

3.1. GEOMETRICAL TRANSLATIONS

The first stage focuses on translating three essential biological mechanisms: Tropism, Branching, and Grafting, into abstract geometrical operations. These mechanisms are responsible for directional growth, bifurcation, and union of root paths respectively, together they regulate the distribution of the system. Subsequently, the geometry-based rules translated are composed into an algorithm using Python in Grasshopper.

3.1.1. Tropism

The concept of tropism describes the directional growth response in plants to an external stimulus. The movement in response to environmental stimuli, such as gravitropism on gravity, hydrotropism on water, and thigmotropism on obstacles, demonstrates the crucial ability of root system to optimize resource uptake, avoiding collision with external objects, and offer structural support (Gilroy & Masson, 2008).

To translate the continual growth of natural roots into discrete movements over time in digital space, a compound-vector is calculated for each root tip in every iteration to determine its current growth direction. The resultant direction is the vector sum of forces exerted by multiple contextual elements and therefore subjected to constant deviation throughout its growth process (Fig. 3a).

While natural roots detect and avoid obstacles through direct physical contact, the algorithm implements an obstacle detection mechanism to respond to collisions with obstacles in digital space. To detect obstacle collisions, the algorithm checks whether the next growing point (root tip) lies inside any obstacle in each iteration. Upon collision, the growing point is moved to the closest point on its surface. Repeated operation allows the root system to walk along the surface of the obstacle when collision occurs and navigate across an artificial matrix (Fig. 3b).

Although the behaviour of self-avoidance between roots remains debated in plant physiology, this system emulates the self-avoidance between roots with each point within the system by exerting a repulsive force on all its neighbours within a certain radius, with a magnitude calculated using an inverse distance weighting method where the magnitude of repulsion is inversely proportional to the distance between points. This causes neighbouring roots to grow apart from each other and distribute themselves evenly in space for optimal flow of resource and forces (Fig. 3c).

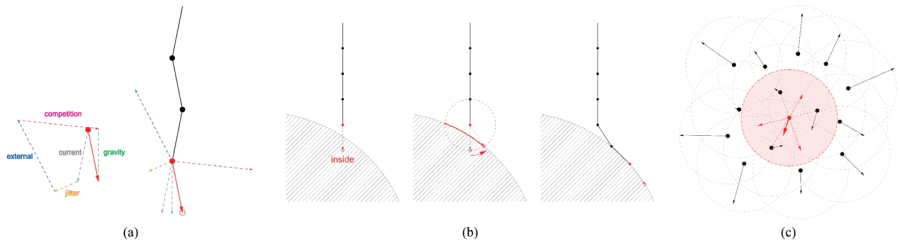


Figure 3. (a) Compound-vector as the sum of environmental forces; (b) Geometrical operation of obstacle detection and avoidance; (c) Self-avoidance mechanism between growing roots

3.1.2. Branching

As a fundamental process in root development secondary to elongation, branching involves the formation of lateral roots from meristems located in the pericycle. This mechanism allows root systems to maximize its area of contact in the most nutrient-rich region of the soil matrix for efficient resource capture (Atkinson et al., 2014).

While the branching frequency of natural roots is regulated by nutrient and hormonal signals in response to contextual conditions (Smit et al., 2000), the translated mechanism uses a local sparsity evaluation that counts the points in its neighbourhood

with fixed frequency, branching off only when no point is close by (Fig. 4a). This mechanism emulates the branching behaviour that avoids overcrowding in certain area of the soil matrix to ensure optimal spatial distribution and resource capture.

Although the branch angle is often fixed in natural roots, the branching azimuth around the axial direction of the root is often determined at random (Smit et al., 2000). The system instead takes the compound-vector of influence on its location to determine the branch azimuth and branch point towards a desirable area in the soil matrix (Fig. 4b). This adds further context-awareness to the system, not only directional changes, but also branch pattern influenced by contextual conditions.

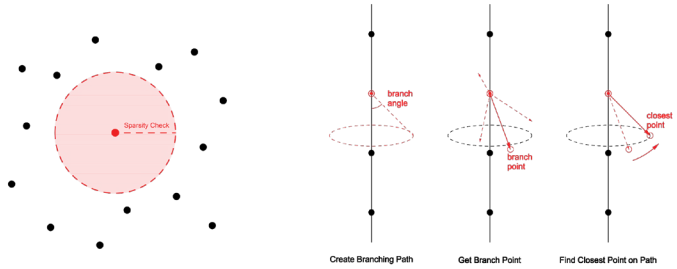


Figure 4. (a) Local sparsity evaluation; (b) Geometrical operation of branching

3.1.3. Grafting

In addition to the directional growth of tropism and bifurcation of branching, grafting is the morphological union resulting from two or more distinct roots in pressured contact with each other during the growth process (Graham & Bormann, 1966). It contributes to added mechanical support and efficiency to nutrient flow within the plant by interconnecting different distinct flow paths in the system (Loehle & Jones, 1990).

While natural grafting happens under physical pressure of roots colliding with each other (Graham & Bormann, 1966), the union of roots in the system is triggered by a proximity evaluation between root tips, followed by the redirection of their growing paths. If the distance between two growing points falls below the distance threshold and the angle between their direction of growth is smaller than a predefined angle, the two growing paths will be redirected into an average point, continuing their growth as a united path in the coming iterations (Fig. 5). This process allows the root system to self-regulate its proliferation by reducing the number of growing branches when the population density of branches exceeds the limit of its environment.

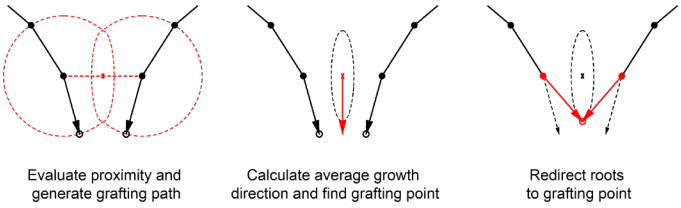


Figure 5. Geometrical operation of grafting

3.2. PARAMETRIC VARIATIONS

This part investigates the emergent properties, such as spatial dispersion, structural hierarchy, and interconnection, derived from the interaction of agents under geometrical rules by varying the growth parameters in the algorithm. The first variation modulates the spatial dispersion of the system by adjusting the parameters of self-avoidance (Fig. 6a). Increasing either the avoidance strength or the detection radius causes individual root points to experience stronger repulsive forces from their neighbours. This forces the system to grow and branch away from densely populated zones, resulting in higher degree of spatial dispersion. The next variation controls the structural hierarchy by changing the parameters of branching (Fig. 6b). The branch radius defines the sparsity required for a point branch, while branch density sets the frequency of this sparsity evaluation. Reducing either parameter causes more frequent lateral root initiation and a higher density of branches. This produces a more densely hierarchical system with higher number of branching orders. In contrast to the expansive tendencies of the previous variations, the last variation enhances the interconnection of the system through grafting (Fig. 6c). The key parameters are the grafting radius (the maximum distance between points for grafting) and the allowable angular difference between their growth directions. By increasing these thresholds, more roots are allowed to fuse, resulting in a more interconnected spatial network.

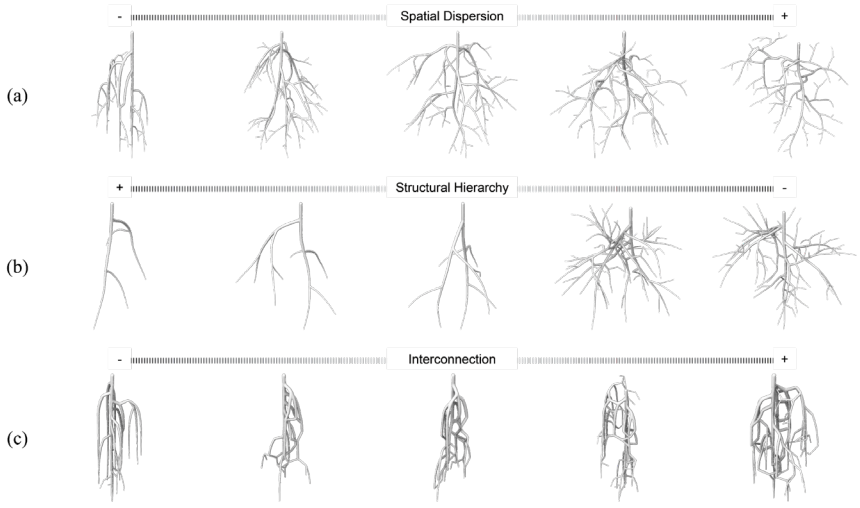


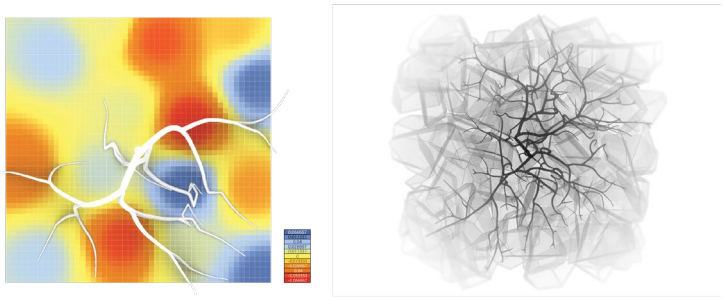
Figure 6. (a) Spatial dispersion; (b) Structural hierarchy; (c) Interconnection

3.3. EXTERNAL ADAPTATIONS

In a natural setting, roots grow adaptively in response to gradients of environmental stimuli and navigates through the obstacles it encounters when seeking nutrient and anchorage. To simulate these adaptations, two types of external stimuli, force fields and obstacles, are introduced to test how the system adapts to its context.

The first adaptation introduces a computational force field constructed from attraction and repulsion points (Fig. 7a). As a root grows, each point in the force field exerts an influence vector upon the growing tip with magnitude calculated via inverse distance weighting. These vectors influence the root's growth direction, allowing it to navigate towards favourable conditions and away from unfavourable ones, demonstrating a direct translation of environmental gradients into growth paths.

The second test introduces external obstacles such as surfaces, solids, and matrices into the system, analogous to ground, rocks, and soil in nature (Fig. 7b). Guided by the object avoidance mechanism defined in the geometrical translation, the growth direction and branching pattern of the root system is redirected according to the shape of obstacles it encountered and are distributed evenly across the surface.



(a) (b)

Figure 7. (a) Adaptation to force field; (b) Adaptation to obstacles

4. Design Applications

Drawing from the conceptual parallel between root systems and architecture, the following two proto-architectural proposals apply the context-awareness of root systems into structural and circulatory organization strategies in design.

4.1. FORCE FLOW

The first design application is a supporting system for a building on a natural slope in Hong Kong. It uses the root system to dynamically distribute supporting structures, interfacing between the irregular landscape and the building mass above. Instead of imposing a rigid grid system onto the natural landscape, the goal is to design an alternative structural system that distribute loads adaptively to minimize on-site alteration, therefore preserving and coexisting with the natural ecosystem.

To adapt to the site, key contextual elements such as transportation networks, pedestrian paths, and hydrological features, are abstracted into a computational repulsion field. These elements are translated as line-based repellents, each assigned a strength value based on its significance (Fig. 8a). Using inverse distance weighting, this field influences root growth, directing the structural network away from sensitive areas. Combined with the 3D model of the terrain, it creates a comprehensive boundary condition that guides the generative process toward a structure that adapts to its context.

The generative process starts with defining a building mass and "seeds" between

the building and the landscape. The seeds grow under the contextual field and terrain geometry, dynamically redirect, branch, and graft, resulting in a spatial structure emerged from the continuous interaction between its internal growth principles and the external site constraints (Fig. 8b). A subsequent Finite Element Method (FEM) analysis using Karamba3D revealed that the root system distributes structural stresses more evenly, avoiding the stress concentrations typical of column-and-beam grids.

The outcome is a context-aware structural support system that distributes its members spatially to efficiently transfer forces to the ground, while intelligently avoiding intersection with critical site features. The resulting morphology demonstrates a rich blend of structure and habitat, creating a covered space that is both functionally robust and ecologically integrated (Fig. 8c, 8d).

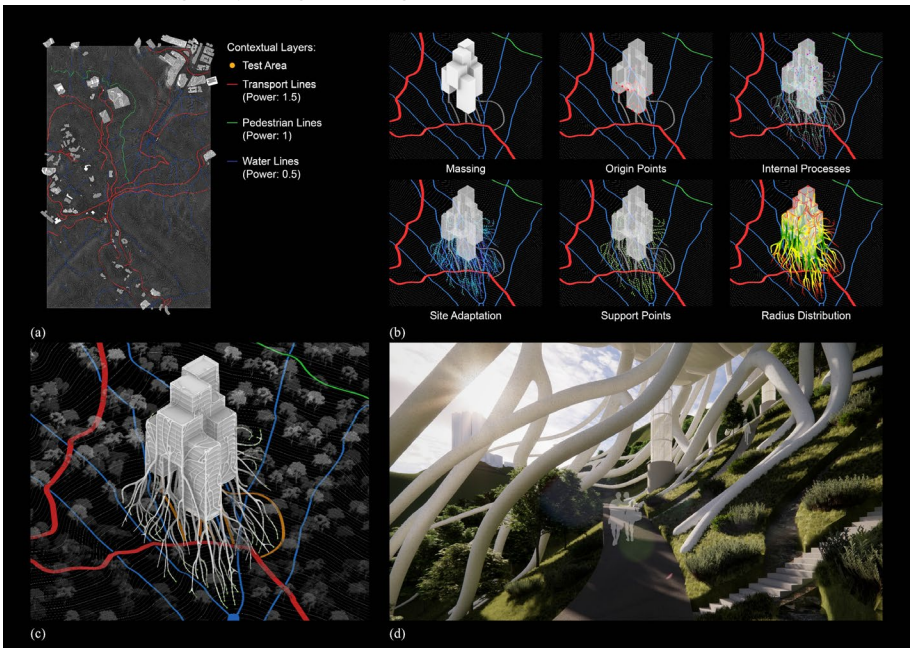


Figure 8. (a) Contextualization of site elements; (b) Generative process of support system; (c) Isometric view of support system; (d) Blend of structure and natural habitat in the covered space

4.2. HUMAN FLOW

The second design application is an elevated pedestrian network in Wan Chai, a densely populated district in Hong Kong. Here, the root system is interpreted as a circulatory network for directing human flow in an urban context. Compared to conventional top-down planning, this application treats the urban fabric as a soil matrix, under which a network of elevated circulation paths grows adaptively to distribute human flow across the area and reconnect different parts of the city.

To computationally contextualize the design, key urban elements are abstracted into a data field. Obstacles such as vehicle roads are defined as repulsion elements, while key destinations, including the harbourfront, historic alleys, and public

transportation nodes, are mapped as attraction points (Fig. 9a). The goal is to generate a network that connects the harbourfront and historic alleys via the transport nodes, creating an elevated pathway through the city while avoiding ground-level traffic.

The generative process initiates from “seeds” placed along an existing central footbridge. These points navigate the volumetric massing of surrounding towers, branching and grafting in response to spatial conditions while extending tendrils into the urban fabric. Global attraction forces pull the growth towards the harbour and old alleys, while local attraction points pull it towards transportation nodes to create access points (Fig. 89b). Vertical level changes are introduced only at branching or grafting points, otherwise kept planar to ensure walkability. The thickness of the path is determined by its simulated "age," transitioning from the densely populated central area into the remote peripheral connections.

The outcome is a speculative design for an elevated, root-inspired pedestrian network that appears to grow through the city (Fig. 9c, 9d). It reinterprets the urban fabric as a living soil matrix to be navigated, not a blank slate to be drawn upon. This model proposes an alternative to top-down master planning, suggesting a methodology for human flow to emerge from the organic, adaptive interactions of a bio-inspired computational system with its context.

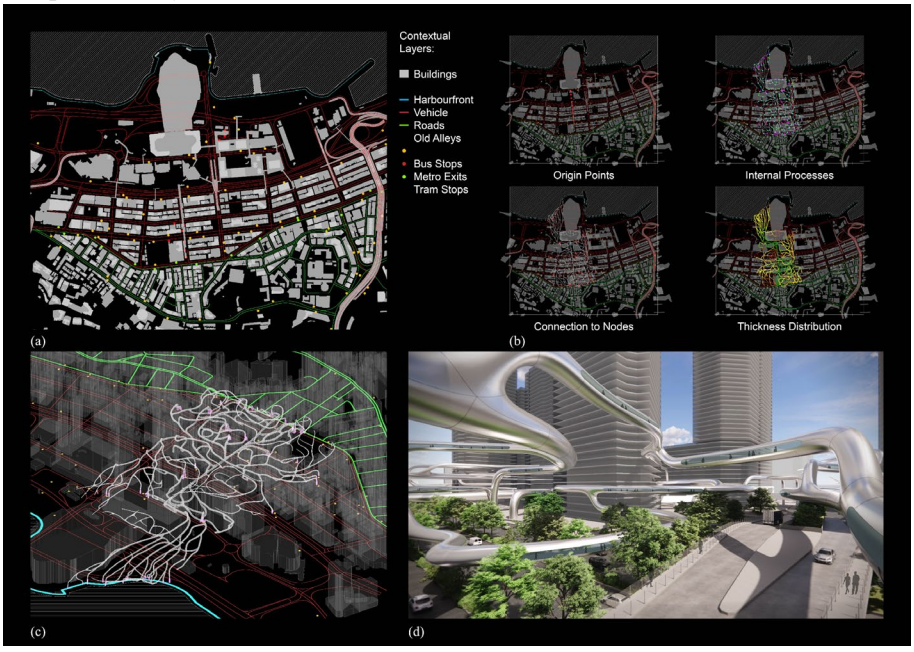


Figure 9. (a) Contextual elements mapped in colour; (b) Generative process of path system; (c) Isometric view of path system; (d) Render of elevated walkway over ground level traffic

5. Conclusion

This research demonstrates the translation of abstracted growth mechanisms of root systems into geometrical operations, enabling the generation of viable proto-

architectural solutions for context-aware structural and circulatory organization networks. In response to the top-down design approach, this research provides an alternative methodology for augmenting the designer's capability to handle complex design problems, exploring a solution space where form is defined by system-level parameters and boundary conditions that adapt to the site condition. The results address the research gaps by providing an operational method for translating biological principles into an architectural context, advancing beyond pure simulation into architectural design proposals, and acting as a conceptual framework for material-oriented approaches.

The proposed proto-architectural methodology has several limitations. The algorithm abstracts the geometrical aspect of biological root growth and focuses primarily on speculative and conceptual potential rather than on detailed tectonic resolution. Therefore, the materialization of these organizational networks as buildable structures remains a critical area for future investigation. Future work will focus on integrating material and structural feedback directly into the growth algorithm, developing real-time interactive models for design exploration, and physically testing the structural strength of the generated networks. This research serves as a conceptual starting point for further development towards materialization, learning from the bottom-up intelligence of nature to develop more responsive, efficient, and ecologically integrated architectural designs.

References

Alexander, C. (1979). Notes on the synthesis of form (10. print). Harvard University Press.

Berlekamp, E. R. (with Conway, J. H., & Guy, R. K.). (2003). Winning Ways for Your Mathematical Plays (2nd ed). CRC Press.

Gilroy, S., & Masson, P. H. (Eds). (2008). Plant tropisms (1st ed). Blackwell Pub.

Graham, B. F., & Bormann, F. H. (1966). Natural root grafts. The Botanical Review, 32(3), 255–292. https://doi.org/10.1007/BF02858662

Hensel, M. (2013). Performance-Oriented Architecture: Rethinking Architectural Design and the Built Environment (1st edn). Wiley. https://doi.org/10.1002/9781118640630

Loehle, C., & Jones, R. H. (1990). Adaptive Significance of Root Grafting in Trees. Functional Ecology, 4(2), 268–271.

Menges, A. (Ed.). (2012). Material computation: Higher integration in morphogenetic design. Wiley.

Otto, F. (2011). Occupying and connecting: Thoughts on territories and spheres of influence with particular reference to human settlement (B. Burkhardt, Ed.; 2., unrev. ed). Menges.

Prusinkiewicz, P. (with Hanan, J. S., Fracchia, F. D., Lindenmayer, A., Fowler, D. R., Boer, M. J. M., & Mercer, L.). (1991). The Algorithmic Beauty of Plants (1st ed). Springer New York.

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, 25–34. https://doi.org/10.1145/37401.37406

Smit, A. L., Bengough, A. G., Engels, C., Van Noordwijk, M., Pellerin, S., & Van De Geijn, S. C. (Eds). (2000). Root Methods. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-04188-8

Thompson, D. W. (1992). On growth and form (The complete revised edition). Dover Publications.

Weinstock, M. (2010). The architecture of emergence: The evolution of form in nature and civilisation (1. publ). Wiley.

Beyond Traditional Metrics: A High-Fidelity Latent-Representation Framework for Urban Morphological Typology Refinement and Quantification

Yu Meng[1], Xinru Wang[2] and Chengyu Sun[3]
[1,2,3] *College of Architecture and Urban Planning, Tongji University; Shanghai Key Laboratory of Urban Regeneration and Spatial Optimization Technology*
[1]*mengyu@tongji.edu.com, 0009-0007-0905-2885*
[2]*wangx_ru@126.com, 0000-0002-5686-5957*
[3]*ibund@126.com, 0009-0004-1316-5581*

Abstract. Traditional urban morphological analysis faces dual limitations: manually constructed metrics suffer from information loss and irreversibility, while existing deep learning features lack structured constraints and morphological semantic reliability. To overcome these challenges, this study proposes a VAE-based framework for urban morphological feature representation, efficiently constructing a high-fidelity latent representation. Validated using Shanghai's 2.5D morphological data, the framework achieves three key outcomes: First, the VAE successfully achieves high-fidelity reconstruction, proving its capture of all structural features. Second, validation via a CNN prediction model confirms the latent features are fully compatible with traditional morphological semantics. Most critically, in unsupervised clustering, the latent representation successfully identified 16 refined morphological types, achieving a Silhouette Coefficient (0.44) more than double that of traditional metrics (0.23). This performance breakthrough proves the latent representation captures urban features that surpass the discriminative power of traditional metrics, effectively quantifying statistically significant gradient morphological types. This study provides a structured, reliable foundational feature representation for large-scale urban morphological analysis, promising applications in accurate form-benefits prediction and controllable morphological generation.

Keywords. Urban Morphology, Variational Autoencoder, Latent Representation, Data-Driven Analysis, Cluster Analysis

1. Introduction

The study of urban morphology is rapidly evolving toward a large-scale, data-driven paradigm. However, in handling vast amounts of high-resolution 2.5D urban data, we face fundamental paradigm challenges related to high-dimensional data processing and morphological semantic quantification.

On one hand, traditional morphological metrics (such as FAR, etc.) are manually

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 675-684.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

constructed to compress urban form into limited-dimensional metrics. These metrics generally offer good interpretability in existing analyses. Yet, this extreme compression leads to significant information loss, manifesting as: irreversibility in morphological reconstruction (the difficulty in inverting the process to reconstruct the urban fabric from these metrics) and restricted discovery of morphological typologies (insufficient information limits the discovery of finer-grained form types). In downstream urban analysis tasks, the information loss and feature redundancy (e.g., the multicollinearity problem among metrics (Sun et al., 2024)) collectively create a performance bottleneck.

On the other hand, deep learning-based image feature extraction methods (such as CNN features (Cai et al., 2021), perceptual hashing (Yao et al., 2019), CAE (Dong et al., 2019), MRCAE (Cai and Chen, 2022), DAE (Ran et al., 2023), etc.) have overcome the limitations of manually constructed metrics. However, they exhibit some limitations in terms of scientific credibility and structural stability for morphological analysis. CNN features or perceptual hashing lack reversibility, which prevents internal validation through reconstruction loss, while autoencoders like CAE and DAE cannot guarantee the structured and continuous nature of the latent space, restricting their reliability in analysis. Furthermore, these methods find it difficult to provide a general morphological feature representation that is high-fidelity, structured, and possesses morphological semantic compatibility.

Given these limitations, this study aims to propose a complete framework that automatically constructs a latent representation capable of efficiently encoding and high-fidelity decoding urban morphology via unsupervised learning, thereby achieving a performance breakthrough in the quantification task of morphological typology refinement.

2. Methods

2.1. A HIGH-FIDELITY LATENT REPRESENTATION AND VALIDATION FRAMEWORK

The latent representation is a semantically rich, low-dimensional vector extracted by a deep learning encoder (Rhee and Krishnamurti, 2023), offering the potential to reveal finer-grained urban typologies. While standard autoencoders (e.g., CAE) achieve compression, they often yield unstructured, discontinuous latent spaces, compromising morphological semantic stability.

This study constructs a complete framework for feature representation and validation, realized by a Variational Autoencoder (VAE) (Kingma and Welling, 2014) to encode urban morphological features. The VAE is employed to enforce structured and continuous properties in the latent space, providing a more stable semantic manifold compared to traditional CAE/DAE. Previous research, such as that which applied VAE to extract geometric prototypes of building footprints for building type evaluation and vulnerability analysis (de-Miguel-Rodriguez et al., 2025), has proven the feasibility of VAE in this domain. However, confining the application only to 2D geometry and low-dimensional latent spaces results in insufficient information and may not be compatible with morphological semantics.

Therefore, this research intends to achieve effective optimization in terms of original information dimension and validation methodology. We defined three specific research objectives, along with corresponding validation experiments and evaluation criteria:

Exp1-High Fidelity and Reversibility: To overcome information loss and irreversibility. The VAE model is used to achieve reversibility, assessed by quantitatively evaluating the reconstruction difference (e.g., MAE), ensuring full capture of the urban fabric structure while highly compressing the information.

Exp2-Morphological Semantic Compatibility/Embeddedness: Aligned with the "Probing Task" paradigm established in representation learning, a CNN predictive model is used to validate morphological semantic compatibility by assessing the prediction accuracy of traditional metrics using the latent vector as input. This cross-validation effectively bridges the gap between latent coding and human-understandable urban logic. The rationale is that the accurate prediction of traditional metrics serves to demonstrate that the latent space is not a stochastic geometric encoding but has successfully internalized fundamental urban morphological rules—such as FAR—that are well-established in urban discourse. This confirms that the model is processing "morphological data" with inherent physical significance rather than performing arbitrary classification, thereby establishing the latent representation as a high-level "digital proxy" for traditional morphology in the digital dimension. This assessment is an essential prerequisite for the reliability of subsequent typology refinement and quantification.

Exp3-Performance Breakthrough in Typology Refinement: To significantly outperform the traditional metric system with statistical validity. Unsupervised clustering algorithms (using K-means++ and Hierarchical Clustering for dual verification) are employed, and performance is quantified using the Silhouette Coefficient (Rousseeuw, 1987) and verified via ANOVA tests to confirm the statistical validity of the refined typologies (rather than being purely black-box results).

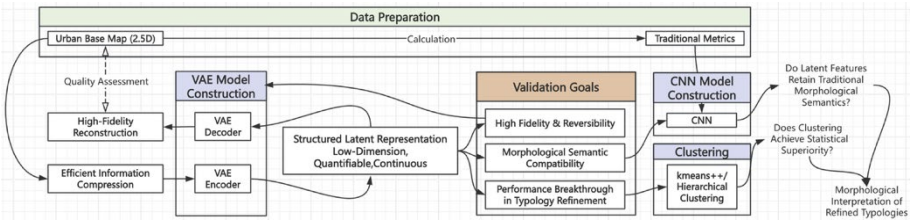


Figure 1. A High-Fidelity VAE-Based Framework for Urban Morphological Typology Refinement and Validation

2.2. DATA PREPARATION

2.2.1. Image Data

An automated procedure extracted 26,897 valid samples from the 2.5D urban map of Shanghai. Each sample represents a 300×300m area, converted into a 512×512 dual-channel TIFF image. The first channel records building footprints, and the second records building stories. The dataset was processed using perceptual hashing for

deduplication and min-max normalization to ensure quality and feature diversity.

2.2.2. Metric Data

This study employs 10 commonly used urban morphological metrics (Table 1) as the benchmark system for validation. These metrics were independently calculated using automated methods based on the GIS data. It must be noted that the artificial metric system suffers from an inherent methodological dilemma: a limited set risks significant information loss, while adopting a vast set introduces feature redundancy. This redundancy necessitates dimensionality reduction techniques (e.g., PCA), which in turn compromises the original interpretability of the metric system by adding a new black-box layer. Balancing these limitations, we chose a limited set of metrics with relatively high independence for this research.

Table 1. urban morphology metrics used in research

Metrics	**Formula**	**Explanation**
Floor Area Ratio, FAR	$FAR = \sum A_f / A_g$	The land development intensity.
Building Cover Ratio, BCR	$BCR = \sum A_c / A_g$	The land occupation degree by buildings.
Building Numbers, BN	$BN = COUNT(A_c)$	The number of buildings.
Volume Density, VD	$VD = \sum V_i / A_g$	The intensity of space development.
Mean Floors, MF	$MF = AVERAGE(F_i)$	The average number of floors of all buildings in the area.
Mean Building Coverage, MBC	$MBC = AVERAGE(A_c)$	The average coverage area of all buildings in the area.
Mean Floor Area, MFA	$MFA = AVERAGE(A_f)$	The average floor area of all buildings in the area.
Nearest Neighbour Index, NNI	$\mathrm{NNI} = \frac{\bar{d}}{0.5/\sqrt{BN/A_a}}$	The aggregation or dispersion of building distribution
Hoover Index of Coverage, CHI	$HI = \frac{1}{2}\sum\left\|A_{c,i} - \overline{A_c}\right\| / \sum A_c$	The non-uniformity in building coverage area distribution
Hoover Index of Volume, VHI	$HI = \frac{1}{2}\sum\|V_i - \overline{V_i}\| / \sum V_i$	The non-uniformity in volume distribution

A_g: The total area of the analysis grid; $\bar{d}$: The mean distance from each building to its nearest neighbour; F_i, $A_{c,i}$, $A_{f,i}$, V_i: The floor number, coverage area, floor area, and volume of the i-th individual building, respectively.

3. VAE Latent Representation Construction and High-Fidelity Validation

3.1. VAE CONSTRUCTION AND TRAINING

A symmetrical VAE model was constructed, in which encoder compresses the 2×512×512 image into a 32×32×32 latent space, while the decoder is responsible for high-fidelity reconstruction of the original image from this latent representation (Figure 2a). Training employed a composite weighted loss function. Structure and hyperparameters were adjusted to balance fidelity and efficiency; convergence was

achieved within 40 epochs (Figure 2b). Later fine-tuning steps incorporated ground-truth difference evaluation to refine accuracy.

3.2. VAE TRAINING RESULTS

The model achieved high-fidelity reconstruction with a low MAE. This low MAE, combined with the visual indistinguishability of the original and reconstructed images (Figure 2c), demonstrates high information fidelity and reversibility. This establishes the necessary foundation for the subsequent Morphological Semantic Compatibility (Exp 2) and Typology Refinement (Exp 3) validations.

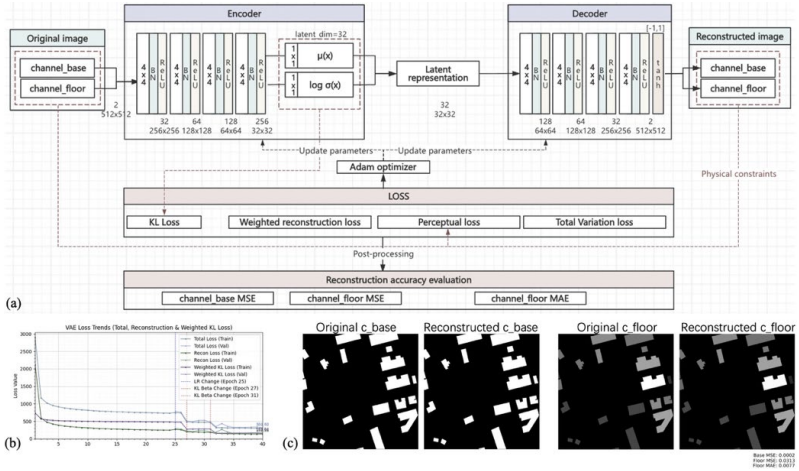
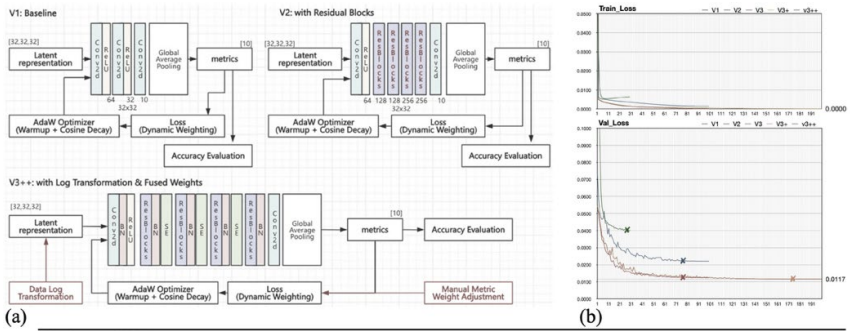


Figure 2. Schematic of VAE Structure, Training Process and Encoding-Reconstruction Effect of the Optimal Model



Metric	MAE						
	V1[baseline]	V2	V3	V4(CBAM)	V5(ResNeXt)	V3+	V3++
FAR	0.0348	0.0175	0.0146	0.0225	0.0441	0.0131	**0.0131**
BCR	0.0023	0.0019	0.0015	0.0032	0.0042	0.0015	**0.0015**
BN	2.7125	1.4961	1.0466	1.5414	2.7072	0.8881	**0.2440**
VD	0.1389	0.0533	0.0427	0.0752	0.1352	0.0400	**0.0399**
MF	0.7666	0.3737	0.1686	0.2493	0.4052	0.1580	**0.1580**
MBC	257.0247	158.8654	79.2137	94.7279	334.4862	53.4054	**0.7888**
MFA	2293.5329	1199.5233	571.9949	742.3950	2490.8229	421.4699	**3.6992**
NNI	0.0849	0.0478	0.0373	0.0491	0.0739	0.0364	**0.0365**
CHI	0.0393	0.0241	0.0123	0.0186	0.0297	0.0114	**0.0115**
VHI	0.0466	0.0255	0.0114	0.0184	0.0290	0.0107	**0.0107**

(c)

Figure 3. Progressive Optimization of the CNN Prediction Model and Training Performance

4. Validation of Morphological Semantic Compatibility

This section confirms the morphological semantic compatibility of the latent representation by assessing its predictive capability over the 10 traditional morphological metrics (Table 1). A CNN-based prediction model was constructed that takes the 32×32×32 latent vector as input.

The prediction model architecture was optimized through multiple iterations (Figure 3a, b), integrating strategies like Residual Blocks and Attention Mechanisms. The Optimal Model (V3++) achieved the lowest prediction error (MAE≈0.0117) by incorporating a logarithmic transformation pre-processing step to address the skewed distribution of certain metrics.

As depicted in Figure 3c, the optimal model achieved high-accuracy prediction for all traditional metrics. This result confirms that the latent representation fully embeds the information provided by these traditional metrics, and, following Exp 1, overcomes the poor semantic interpretability limitation found in existing image feature extraction methods (e.g., perceptual hashing). Crucially, this establishes the foundation for semantically analyzing the refined typologies in Exp 3.

5. Validation of Performance Breakthrough in Morphological Typology Refinement

Based on the high information fidelity and morphological structure compatibility already demonstrated by the latent representation (Sections 3 and 4), this section aims to quantitatively validate its superior performance in morphological typology refinement compared to traditional metrics through unsupervised clustering analysis.

5.1. OPTIMAL CLUSTERING SETUP AND QUANTITATIVE PERFORMANCE SUPERIORITY

Clustering analysis was conducted with rigor and multiple validations:

- **Data Dimensionality Reduction:** UMAP (McInnes et al., 2018) was applied to latent vectors to preserve topological structure. To establish a rigorous baseline, the latent space was compressed to n=5 — half the dimensionality of the traditional metric system (n=10). Despite this dimensional disadvantage, the latent representation's superior discriminative power rendered PCA-reduction on traditional metrics unnecessary, as its performance significantly exceeds even the full-dimensional metric baseline.
- **Optimal K Determination:** Both the Elbow Method and the Silhouette Coefficient were utilized. Dual validation was performed using K-Means++ and Hierarchical Clustering.

The Elbow Method curve (Figure 4) for both features showed a sustained gentle decrease, lacking a clear inflection point. This suggests that urban typologies follow a smooth and continuous gradient transition pattern, rather than clear-cut boundaries.

The quantitative superiority is evident in the Silhouette Coefficient comparison (K-Means++). Traditional metrics peaked low (≈ 0.23 at K=5), reflecting inherent information loss and poor cluster cohesion. In contrast, the latent representation

achieved its optimal clustering number at K=16, with a significantly superior clustering quality (Silhouette Coefficient ≈ 0.44). This superiority (approximately double the traditional metrics' coefficient) was cross-algorithmic, demonstrating the latent feature is more discriminative.

This high clustering quality is sustained at K>16 (consistently≈ 0.4 in Hierarchical Clustering), suggesting the latent representation enables effective typology refinement across various scales beyond the optimal K.

One-way Analysis of Variance (ANOVA) was performed using traditional metrics (based on the K-Means++ K=16 results) to quantitatively validate the statistical effectiveness of the clustering. The results showed that the F-statistics for all metrics were high, with p-values approaching 0 (Figure 6 c), indicating high statistical significance.

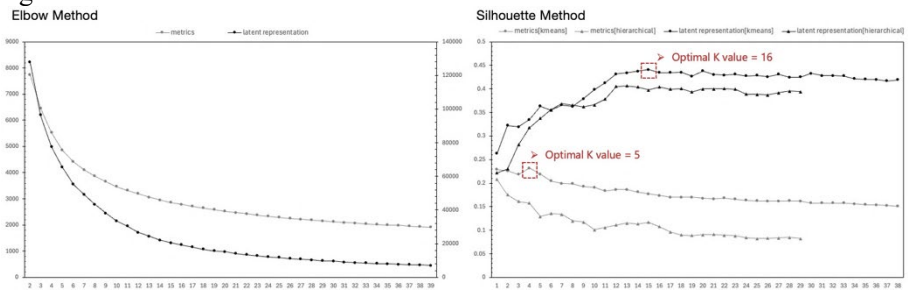


Figure 4. Optimal Clustering Number Search Based on Latent Representation and Metrics

5.2. MORPHOLOGICAL CONTRAST: INSIGHTS FROM METRICS AND LATENT REPRESENTATIONS

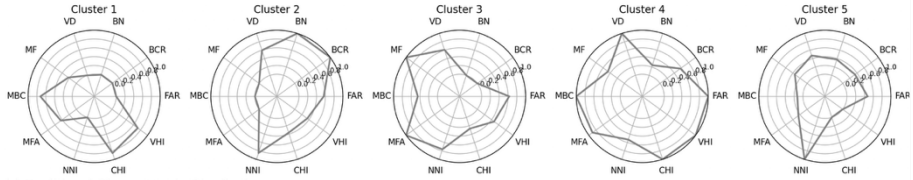
Morphological contrastive analysis was conducted on cluster centroid samples and radar charts (solid line representing metric centroid value) to highlight the latent representation's value in discovering fine-grained typologies.

5.2.1. Traditional Metric-Based Typologies (K=5)

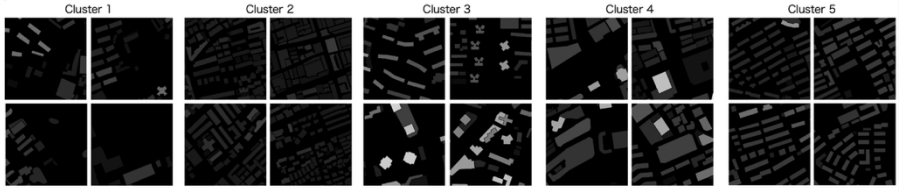
The 5 typologies identified based on traditional metrics (Figure 5) visually describe macroscopic morphological differences, but their refinement capacity is limited. They generally distinguish between extreme types such as small volume dense fabric (C2) and high development intensity fabric (C4), along with more mixed and uniform typologies (C3, 5), and block edge type (C1).

5.2.2. Latent Representation Typologies (K=16)

To confirm the semantic validity of the latent-discovered typologies, morphological contrastive analysis was conducted on cluster centroid samples and their metric structures (Figure 6). Besides intuitively distinguishable extreme types, such as Cluster 7 (Irregular Dense Old City Fabric), we focus on comparing morphological differences among other similar gradient types:



(a) Traditional Cluster Metric Distribution

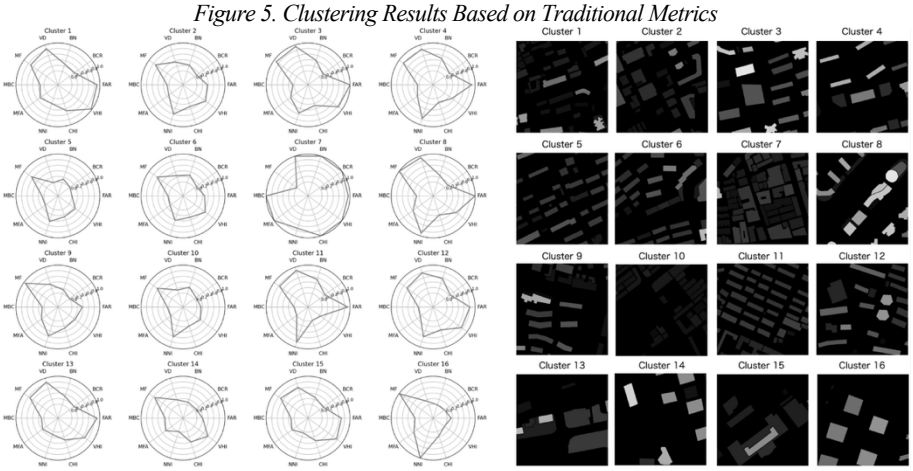


(b) Representative Cluster Centroid Samples

	FAR	BCR	BN	VD	MF	MBC	MFA	NNI	CHI	VHI
F_statistic	8044.9957	12584.8415	11827.2723	8248.4972	8235.7799	321.5549	1159.9587	1974.4154	3429.3649	4759.4171
p_value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

(c) ANOVA Statistical Validation of K=5 Clustering

Figure 5. Clustering Results Based on Traditional Metrics



(a) Traditional Cluster Metric Distribution

(b) Representative Cluster Centroid Samples

	FAR	BCR	BN	VD	MF	MBC	MFA	NNI	CHI	VHI
F_statistic	92.0061	786.3051	208.7857	98.9201	199.0079	171.4976	44.2434	23.0199	105.3576	70.5183
p_value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

(c) ANOVA Statistical Validation of K=16 Clustering

Figure 6. Clustering Results Based on Latent Representation

- **Group A (C3, 4, 8, 11, 12):** These groups share similar overall metric structures, characterized by high FAR and mean stories. The differentiations mainly involve building size and regularity: C8 has higher mean building area and volume density than C4, confirmed by sample images showing a large-volume dominant fabric; C11 possesses the lowest coverage and volume heterogeneity, corresponding to the most uniform and regulated fabric; C12 has a greater number of buildings and lower FAR, showing a mid-rise dense transitional morphology; C3 shows high volume heterogeneity despite relatively uniform footprint area, characterized by vertically heterogeneous forms.

- **Group B (C1, 13, 15):** The main difference from group A is a more dispersed distribution. C13 displays lower volume heterogeneity compared to C1 while C15 features lower building density, FAR, and volume uniformity than C1.
- **Group C (C2, 9):** Volume density is lower than the preceding two groups. C9 has higher mean building stories and lower volume heterogeneity than C2, indicating a more unified and taller mid-rise cluster fabric.
- **Group D (C14, 16):** These are close to Group C but exhibit more unbalanced metric values. C14 shows higher volume heterogeneity and higher mean stories than C16, and the distribution is more scattered.
- **Group E (C5, 6, 10):** Both FAR and building coverage are at medium-to-low levels. Compared to C5, C10 has higher volume density and a more clustered distribution, and C6 has higher building density.

These comparisons conclusively prove that the latent-discovered sub-clusters are not black box results, but possess interpretable morphological semantics. It successfully identifies statistically significant gradient morphological types at a scale where traditional metrics cannot effectively differentiate.

6. Conclusion

This study proposes a complete VAE-based framework for urban morphological feature representation and validation, which efficiently constructs a high-fidelity latent representation. This framework successfully overcomes the key limitations of traditional metric systems regarding information fidelity and performance, as well as the constraints of existing deep learning methods concerning semantic compatibility. Through this paradigm, we achieved a substantial performance breakthrough in morphological typology refinement.

Beyond technical validation, this latent representation offers a more holistic quantification of urban fabric. Unlike traditional metrics that often reduce morphology to a collection of independent scalar values (e.g., FAR, BCR), the VAE-derived latent space functions as a structural manifold that preserves the implicit topological relationships and textural complexity of urban forms. This non-anthropocentric feature discovery enables the identification of fine-grained statistical gradients, providing a reliable digital proxy for complex morphological analysis that transcends human-predefined categories.

Consequently, this framework provides a robust foundation for modelling urban performance and exploring morphological design spaces. Utilizing this high-fidelity proxy is expected to enhance the precision of predicting complex urban phenomena by providing more comprehensive spatial inputs. Furthermore, the continuous structure of the latent space establishes a potential pathway for data-driven morphological exploration, offering new possibilities for future research into more structured urban simulation and generative analysis.

Acknowledgements

This study is supported by a project of National Natural Science Foundation of China

(Grant No.52478030). The authors would also like to thank the Shanghai Key Laboratory of Urban Regeneration and Spatial Optimization Technology (Tongji University) for their support in providing the data used in this study.

Attribution Statement

Google Gemini (Google, 2025) was used to assist with manuscript translation, language refinement and conciseness.

References

Cai, C., Guo, Z., Zhang, B., Wang, X., Li, B., & Tang, P. (2021). Urban morphological feature extraction and multi-dimensional similarity analysis based on deep learning approaches. *Sustainability*, 13(12), 6859.

Cai, J., & Chen, Y. (2022). A novel unsupervised deep learning method for the generalization of urban form. *Geo-Spatial Information Science*, 25(4), 568–587.

de-Miguel-Rodriguez, J., Requena-Garcia-Cruz, M. V., Romero-Sánchez, E., & Morales-Esteban, A. (2025). Automated building typology clustering and identification using a variational autoencoder on digital land cadastres. *Results in Engineering*, 26, 105232.

Dong, J., Li, L., & Han, D. (2019). New quantitative approach for the morphological similarity analysis of urban fabrics based on a convolutional autoencoder. *IEEE Access*, 7, 138162–138174.

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. Paper presented at the 2nd International Conference on Learning Representations, *ICLR 2014*, Banff, Alberta, Canada.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426.*

Ran, W., Yu, J., & Yin, L. (2023). Machine-Learning-Driven Comparative Study Morphological Taxonomy in Screen-Based Building Clusters. In *Habits of the Anthropocene: Scarcity and Abundance in a Post-Material Economy-Proceedings of the 43rd Annual Conference of the Association for Computer Aided Design in Architecture, ACADIA 2023*, 2, 596-605.

Rhee, J., & Krishnamurti, R. (2023). Latent Space to Support Virtual 3D Models. In *Cultural Space on Metaverse* (pp. 41-52). Singapore: Springer Nature Singapore.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

Sun, C., Meng, Y., Wang, X. (2024). Scale Mismatching: Prevalent Compactness Indexes of Urban Form DO NOT Work in Computational Urban Design. *ACCELERATED DESIGN-Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia, CAADRIA 2024*, 2, 455–464.

Yao, J., Huang, C., & Liu, P. (2019). Research on urban fabric identification and evaluation based on artificial intelligence. *Housing Science*, 39(11), 9–14.

Exploring Pedestrian-Tactile Paving Interactions in High-Density Urban Spaces

A Computer Vision and Spatial Analytics Approach

Jiqian Huang[1], Yean Yang[2], Xiaoxi Guo[3], Shiliang Wang[4], Waishan Qiu[5], Xintong Lyu[6] and Chongyang Ren[7]
[1]*South China University of Technology.*
[2,5,6]*The University of Hong Kong.*
[3]*Tongji University.*
[4] *University of Jinan.*
[7]*University College London.*
[1]*huangjiqian@hnu.edu.cn , 0009-0009-9579-5072*
[2]*yeanyang794@connect.hku.hk, 0009-0003-7786-9994*
[3]*2010242@tongji.edu.cn, 0009-0001-5072-618X*
[4]*shiliangwang@xauat.edu.cn, 0009-0002-8075-0911*
[5]*waishanq@hku.hk, 0000-0001-6461-7243*
[6]*lyuxt24@connect.hku.hk, 0009-0002-3236-8291*
[7]*chongyang.ren.25@ucl.ac.uk, 0009-0008-3330-834X*

Abstract. Tactile paving is a key element of inclusive cities, yet how it functions within dense everyday flows remains unexplored. This study proposed a computer vision framework to quantify pedestrian–tactile paving interactions at scale. We analyzed over 500 hours of fixed-camera video from eight sidewalks and intersections in four Japanese cities. Using YOLOv8, DeepSORT, and homography-based projection, we reconstructed pedestrian trajectories and related them to tactile paving. Trajectories were classified into four interaction types, including avoiding, approaching, acrossing, and overlapping, under everyday flows and in the presence of persons with disabilities. Results showed that avoiding was the dominant behavior, accounting for 47.9%. Tactile paving was safer in the morning and became increasingly obstructed toward the evening peak. Sidewalks showed higher yet more volatile avoidance than intersections. When persons with disabilities appeared, avoidance rose to 52.0% but was higher during the evening than morning, especially at intersections. The framework revealed micro-scale accessibility dynamics and can support time and site-specific strategies for inclusive space design and management.

Keywords. Tactile Paving, Pedestrian Trajectory, Computer Vision, Inclusive City, Accessibility Evaluation

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 685-694.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

1. Introduction

Inclusivity has become a central principle of contemporary urban governance, emphasizing the right of all residents to move safely, independently, and with dignity in everyday life (Zhao et al., 2025). Accessible facilities such as tactile paving, ramps, and audible signals constitute an essential material dimension of urban inclusivity. Their spatial layout, continuity, and daily use shape not only mobility opportunities but also the extent to which cities genuinely accommodate diverse bodies and abilities.

Tactile paving, originating in Japan and globally diffused since the 1990s, is now the most widespread guidance system for visually impaired pedestrians. However, the mere presence of tactile paving does not guarantee smooth mobility. A French disability survey shows that 90% of visually impaired adults still experience mobility difficulties, and 29% never go out without assistance (Pigeon et al., 2026). Perceived safety, route continuity, and the need to avoid contested spaces all influence travel choice and independence (Rosa et al., 2021).

Existing research on tactile paving has focused on standardization and specifications, road-crossing assistance, and tactile pattern design (Watanabe et al., 2014). User experience surveys and laboratory-based experiments have deepened our understanding of perception thresholds and navigation strategies (Yang & Chen, 2025), but they typically involve self-reported behavior or controlled settings. What remains insufficiently addressed is a fundamental question: how is tactile paving actually used and negotiated within the complex, dynamic flows of everyday urban life?

Recent advances in computer vision and spatial analytics provide new tools to observe micro-scale interactions in public space. Deep learning-based pedestrian detection and tracking have already been deployed for traffic safety, crowd management, and micro-social interaction analysis (van Ameijde et al., 2025). These methods are particularly suitable for high-density urban contexts, enabling continuous observation of who, when, and how pedestrians respect or occupy tactile paving.

Therefore, we developed a computational framework to automatically extract pedestrian–tactile paving interactions from large-scale videos and to uncover their spatio-temporal patterns in Japan, the birthplace of tactile paving. By integrating detection, classification, trajectory reconstruction, and ground-plane spatial analysis, we systematically quantified the extent to which tactile paving is actually respected. This framework provides empirical evidence on the dynamics of accessibility in dense space, moving beyond static design standards to behaviorally grounded evaluation.

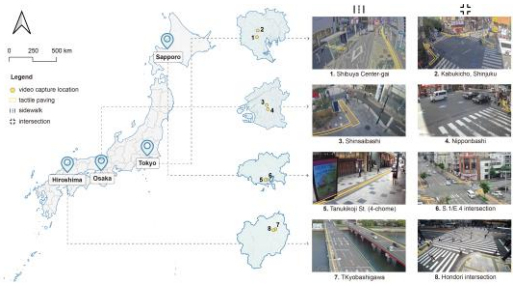


Figure 1. Research area

2. Methods

2.1. RESEARCH AREA AND FRAMEWORK

Japan, as one of the most urbanized countries worldwide, has developed unified accessibility standards for tactile paving under the "Act on Promotion of Barrier-Free Transportation." Considering data availability, spatial distribution, and urban population size, we selected Tokyo, Osaka, Sapporo, and Hiroshima as case studies. In each city's central district, we chose one sidewalk and one signalized intersection, representing the typical scenarios of free pedestrian flows and traffic-controlled flows, respectively (Figure 1).

The overall research framework comprises three stages (Figure 2).

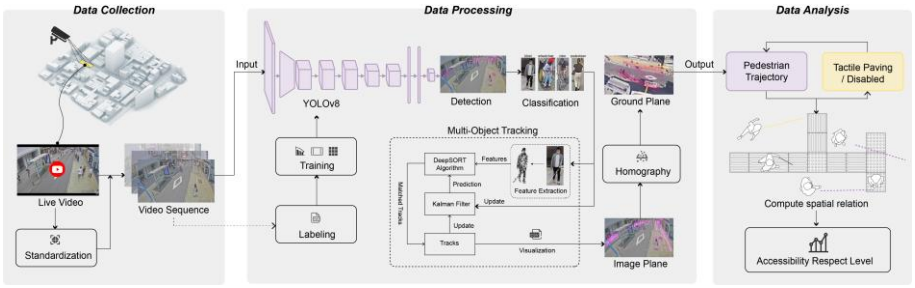


Figure 2. Research framework

2.2. DATA COLLECTION AND PREPROCESSING

We collected video data from publicly accessible, fixed-position YouTube live streams. Continuous recording was conducted from 2–8 September 2025, covering 09:00–18:00 daily at all eight sites. In total, more than 500 hours of raw footage were obtained. To standardize subsequent processing, all videos were resampled to a width of 1280 pixels with proportional scaling, converted to 10 fps, stripped of audio, and compressed with H.265 for storage.

2.3. PEDESTRIAN DETECTION, CLASSIFICATION, AND TRACKING

The YOLOv8 series, a mature high-performance object detector, was used as the core detection engine. On the COCO dataset, it achieves an average precision of 87% at an IoU of 0.5, which is sufficient for person detection in surveillance scenes (Nimma et al., 2025).

For category recognition, we began with a publicly available surveillance dataset focusing on persons with disabilities (9,868 images) that included labels such as normal, blind, stroller, and wheelchair. To adapt the model to Japanese street scenes, we further fine-tuned it using our own data. Using OpenCV, we sampled five frames per hour from one weekday (6 September) and one weekend day (8 September), manually annotated them in Roboflow, and identified blind pedestrians based primarily on white canes or guide dogs. Distant, barely discernible pedestrians were removed to reduce noise. The final dataset included 3,259 labeled pedestrian instances, including 809 blind pedestrians, 870 wheelchair users, 941 normal pedestrians, and 639 riders.

Then we trained an image classification model on this dataset, with 70% training and 30% validation set. Inputs were resized to 128×128 pixels, trained with a batch size of 16 for 150 epochs, and the weights with the best validation performance were retained.

In each video frame, a pre-trained YOLOv8x model detected pedestrians and output bounding boxes. All detections labeled as blind or wheelchair were manually verified. Each bounding box was represented by four parameters (u, v, r, h), where (u, v) is the center of the box, r is the scale factor, and h is the aspect ratio. Cropped pedestrian images were passed to the classifier to obtain the final category label.

To link detections into individual movement paths, we used DeepSORT. The tracker predicts box motion with a Kalman filter and matches detections across frames with the Hungarian algorithm, assigning a persistent track ID to the same pedestrian. We used the bottom center of each bounding box as an approximate footpoint. By connecting the footpoints associated with the same ID over time, we reconstructed continuous pedestrian trajectories.

For spatial analysis of interactions, image coordinates were mapped to a ground-plane coordinate system via homography. For each camera, we selected four control points in image space $\{(u_i, v_i)\}$ and corresponding points in map space $\{(X_i, Y_i)\}$ to solve the homography matrix H, which maps image plane to ground plane positions. Tactile paving was manually digitized as polygons, and both trajectory points and tactile paving geometry were projected into the same ground system.

2.4. INTERACTION TYPE CONSTRUCTION AND COMPUTATION

Based on the geometric overlap and the directional relationship between the pedestrian trajectory and the tactile paving, we defined four pedestrian-tactile paving interactions, including avoiding, approaching, acrossing, and overlapping (Figure 3). To delimit the spatial range within which pedestrians are likely to be influenced by tactile paving and nearby pedestrians, we referenced proxemics for unfamiliar individuals (1.2–3.6 m) and adopted a 2.4 m filtering radius (Hall, 1966). Within this radius, trajectory segments were retained for interaction classification.

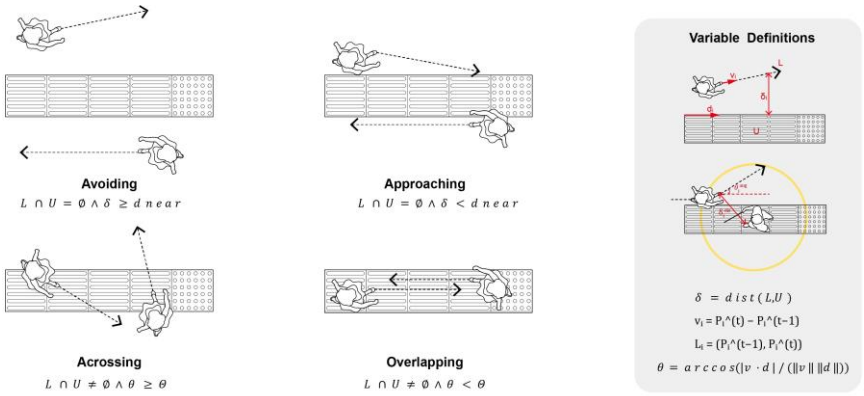


Figure 3. Classification and computation of pedestrian-tactile paving interaction types

Specifically, the classification proceeded in two steps. We first examined whether a pedestrian trajectory segment L intersected the polygonal area of the tactile paving U. If a segment did not intersect the paving, we calculated the minimum distance δ between the trajectory segment and the nearest point on the tactile paving. Segments with δ smaller than 0.5 m were regarded as approaching, whereas those equal to or greater than 0.5 m were regarded as avoiding. When an intersection did occur, we then measured the nondirectional angle θ between the pedestrian's movement vector and the local direction of the tactile paving. Intersections with θ less than 30° were classified as overlapping, while equal to or greater than 30° were classified as acrossing.

Finally, we distinguished two analytical conditions: (1) everyday flows, in which all trajectories within the study area were considered; (2) presence of persons with disabilities, in which we focused on periods when at least one blind or wheelchair user appeared. When persons with disabilities walked on tactile paving, the tactile paving itself was treated as the primary analysis zone. When they walked off, a 1 m lateral buffer was constructed around their trajectories. For each site and condition, we aggregated classified trajectory segments and computed the proportion of the four interaction types, enabling comparative analysis across space and time.

3. Results

3.1. CLASSIFICATION MODEL PERFORMANCE

The trained YOLO-based classification model achieved stable performance. The Top-1 accuracy was 0.93, and the Top-5 accuracy reached 1.00 (Figure 4a). Among the four pedestrian classes, the wheelchair category performed best, with a precision of 1.00. The blind, rider, and normal pedestrian classes also achieved high precisions of 0.97, 0.92, and 0.88, respectively (Figure 4b). Overall, the model provides robust discrimination for the pedestrian categories, sufficient for the subsequent trajectory interaction analysis.

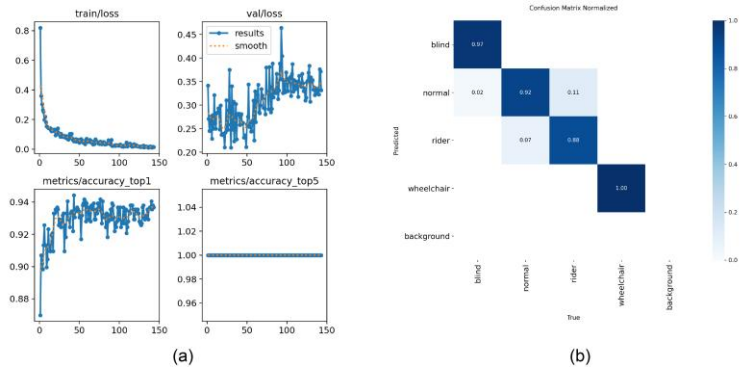


Figure 4. Training results. (a)Training and validation curves; (b)Normalized confusion matrix

3.2. INTERACTION PATTERNS WITH TACTILE PAVING UNDER EVERYDAY PEDESTRIAN FLOWS

Based on video detection from eight sites, we first examined how pedestrian flow intensity and organization shaped interactions with tactile paving across three peak periods. The results showed strong but time-varying protection of tactile paving under everyday flows (Figure 5).

During the morning commute, a relatively cool, "yielded" corridor emerged along the tactile paving, especially on sidewalks, suggesting that directional and purposeful movement naturally preserves the guidance path. By midday, as activities diversified around shopfronts and mixed-traffic zones, this cool corridor merged into surrounding hotspots. Waiting clusters appeared near intersections, and portions of the tactile path became embedded within higher-density zones. During the evening peak, scattered hotspots occurred directly along and around much of the tactile paving, implying that when flows are denser and more heterogeneous, tactile paving is more frequently used as part of the shared walking surface.

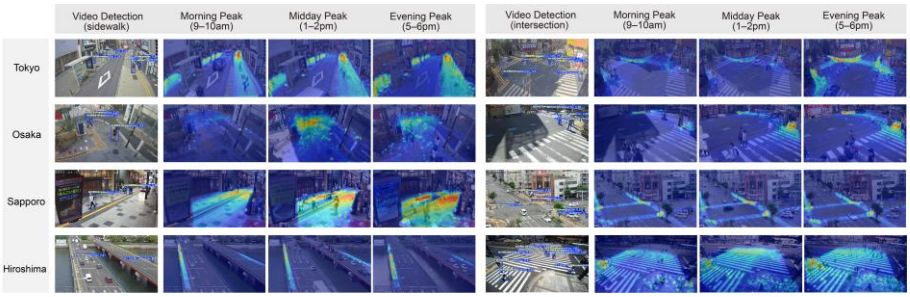


Figure 5. Video-based pedestrian detection and heatmaps in the peak period

Trajectory classification showed that sidewalks and intersections exhibit distinct interaction profiles. Spatial configuration and functional role further modulate interaction patterns. Sidewalks in commercial districts, such as those in Sapporo, Osaka, and Tokyo, displayed relatively high avoiding rates, reaching 63.1%, 54.8%, and 48.1% respectively (Figure 6). Although these streets are embedded in busy shopping environments, pedestrians still showed a pronounced tendency to leave the tactile corridor unoccupied. In contrast, the Hiroshima sidewalk, which functions as a pure passage space with limited commercial frontage or visual attractors, recorded the lowest avoiding rate at 19.5%. The tactile path there became part of an undifferentiated walking surface.

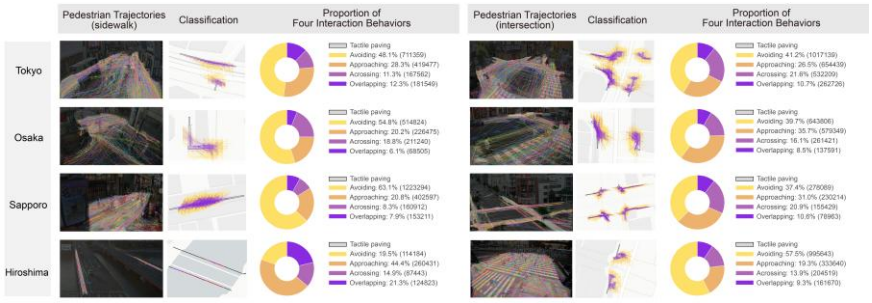


Figure 6. Trajectory clustering and proportions of four interaction behaviors across eight sites.

At intersections, patterns shifted. The Hiroshima intersection, linked to a transport hub, exhibited a higher avoiding rate than the other three cities, suggesting that clearer functional organization and signage could help preserve tactile corridors even under heavy flows. Across cities, intersections tend to register higher proportions of approaching and acrossing, reflecting the multifunctional pressures of waiting, turning movements, temporary parking, and street commerce.

Finally, we aggregated trajectory data from the eight sites by hour over the week to examine the spatio-temporal dynamics of interaction patterns, further confirming these visual patterns (Figure 7). Overall, avoiding was the predominant behavior, with a mean proportion of 47.9%, followed by approaching (26.6%), crossing (15.5%), and overlapping (10.0%). Specifically, avoiding generally declined from morning to evening, with the most notable dip around 13:00. And it was consistently higher on sidewalks than at intersections, indicating relatively safer tactile conditions in more linearly structured spaces during everyday flows. Approaching and acrossing were comparatively stable, with slight midday increases at intersections. Overlapping was rare and exhibited only modest temporal fluctuation, rising slightly on sidewalks at midday while remaining nearly flat at intersections.

Additionally, box plots of hourly proportions revealed that sidewalks exhibited greater variability, especially in avoiding and overlapping, while intersections displayed more concentrated distributions. This suggested that behavior in highly structured, signal-controlled spaces was more stable and predictable, whereas more flexible sidewalk environments were prone to fluctuating patterns of protection and encroachment.

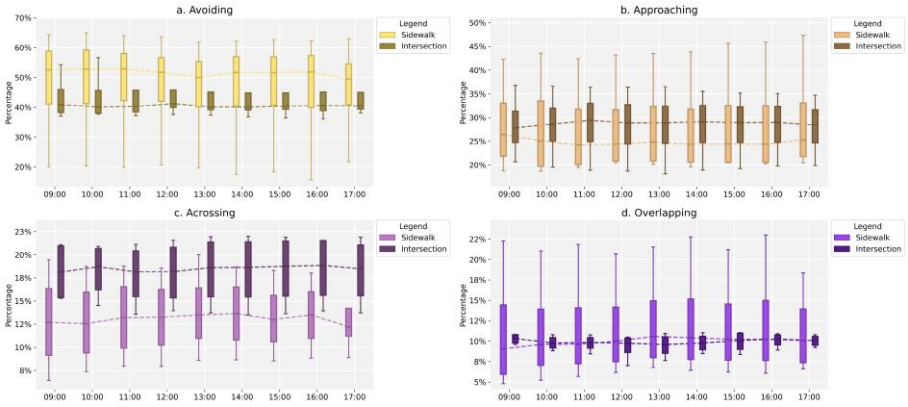


Figure 7. Hourly variation of four pedestrian interaction behaviors under Everyday Flows

Taken together, these results indicate that the tactile paving environment is relatively safer in the morning, while midday presents the highest risk of obstruction. Spatially, despite avoiding being the primary behavior, sidewalks are more prone to density-fluctuating patterns of avoiding and overlapping, whereas intersections more frequently witness functional occupation behaviors like approaching and acrossing.

3.3. INTERACTION PATTERNS IN THE PRESENCE OF PERSONS WITH DISABILITIES

We next examined how surrounding behavior changes when persons with disabilities are present. Using the Hiroshima intersection as an illustrative case, we distinguished two situations: blind pedestrians walking on the tactile path and those walking off it (Figure 8).

When blind pedestrians walked along tactile paving, 56.8% of nearby trajectories showed avoiding, with clear local detours and an average avoidance distance of 3.1 m. When they walked off the tactile paving, 49.2% of trajectories still exhibited avoidance, but routes were smoother and maintained more consistent spacing, with an average avoidance distance of 2.3 m. Avoidance was thus slightly stronger, and detours more pronounced, when the blind person followed the tactile guidance path.

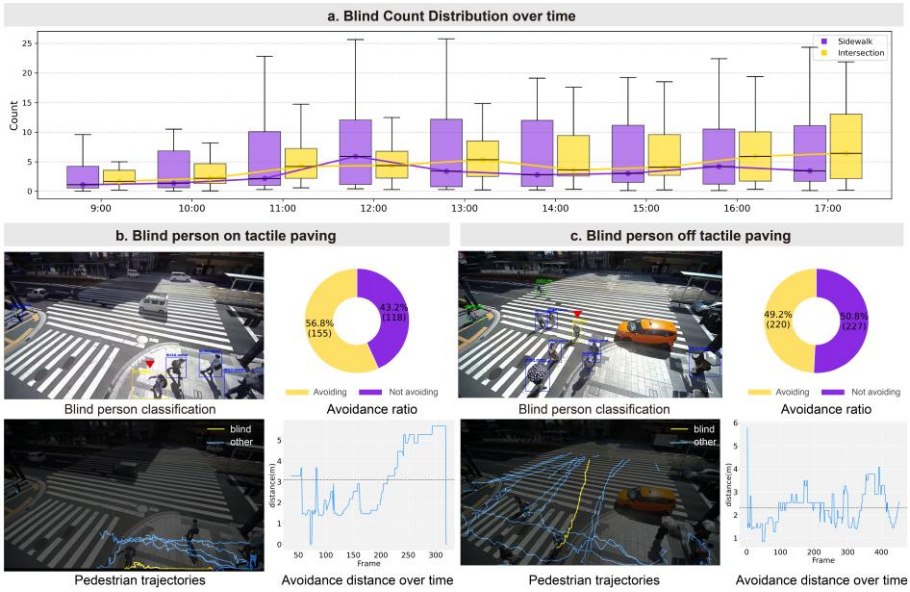


Figure 8. Temporal distribution of trips by people with disabilities and their walking position relative to tactile paving

Similarly, aggregating across all cities and hours when at least one person with disabilities was present, we observed that avoiding increased to an average of 52.0%, acrossing decreased to 9.6%, overlapping accounted for 8.2%, and approaching became the rarest behavior, at 3.8% (Figure 9).

Specifically, avoiding remained dominant throughout the day, with shallow troughs around 11:00 and 15:00, and was consistently higher at intersections than on sidewalks. Approaching showed low overall shares but became more volatile after 14:00, with a slightly higher proportion on sidewalks. Acrossing exhibited a mild midday rise and an evening dip with similar levels at sidewalks and intersections. Overlapping remained low but was more dispersed on sidewalks, including occasional off-peak spikes, while staying low and stable at intersections.

These patterns suggest that the presence of persons with disabilities enhances avoiding awareness, particularly in signal-controlled environments, yet sidewalks continue to exhibit heterogeneous and sometimes inconsistent interaction responses.

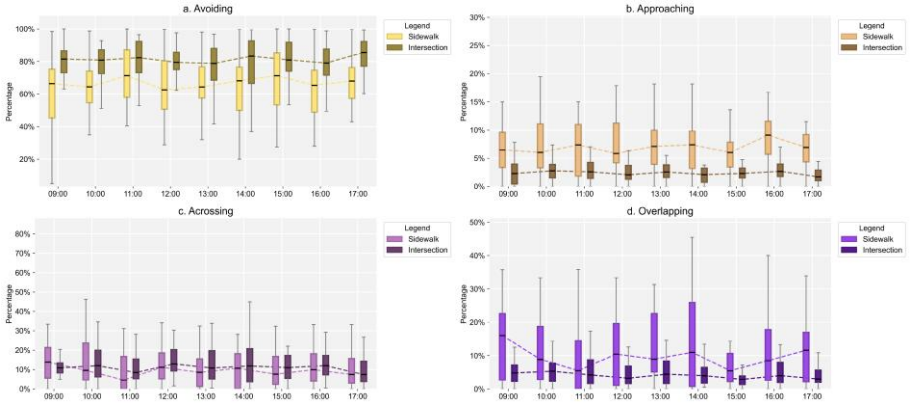


Figure 9. Hourly variation of interactions in the Presence of People with Disabilities

4. Discussion

Our findings reveal clear temporal dynamics in pedestrian–tactile paving interactions. Morning periods provide safer tactile conditions than evenings, with stronger avoidance and relatively cool heatmap corridors. This echoes evidence that some blind pedestrians deliberately avoid peak crowding periods to reduce turbulence and uncertainty in navigation (Henly & Brucker, 2019). Notably, when a person with disabilities is present during the evening peak, surrounding pedestrians exhibit higher avoidance than in the morning. This pattern suggests that people may simultaneously encroach more on tactile paving at busier times, yet behave more cautiously when they actually encounter vulnerable users.

The contrast between sidewalks and intersections further highlights how spatial configuration and functional roles shape accessibility risks. Sidewalks in active commercial streets show high average avoiding rates but large variability, suggesting that the same tactile corridor can shift from clearly preserved to heavily encroached depending on momentary crowd densities and local attractors. Intersections, by contrast, face more multifunctional pressures, such as waiting queues, turning movements, temporary parking, street vendors, and informal social uses (Barlow et al., 2005). These functions intensify competition for limited space and increase the risk that tactile paving is used as convenient waiting or stacking zones. Yet, our results also show that when persons with disabilities appear, avoidance rates at intersections surpass those on sidewalks, indicating that the combination of signalization, clearer spatial demarcation, and social awareness may trigger quicker yielding responses.

This study has several limitations. Although the blind and wheelchair categories achieved high recognition accuracy and were manually checked, real-world generalization remains imperfect, and the absolute number of disability cases in our dataset is limited. In addition, our camera locations are concentrated in Japanese commercial and entertainment areas. Future work should test transferability across

global contexts and include a broader range of disabilities and mobility aids.

5. Conclusion

This paper presented a novel framework for analyzing pedestrian–tactile paving interactions in dense urban spaces, using over 500 hours of video from eight sites in four Japanese cities. Combining computer vision and spatial analytics, we classified four interaction types and examined their spatio-temporal dynamics under everyday flows and in the presence of persons with disabilities. Our findings underscore that tactile paving is not a static intervention but part of a dynamic socio-spatial system, and large-scale video analytics provide a useful basis for evidence-based inclusive design and management.

Attribution Statement

ChatGPT (OpenAI, 2025) was used to (1) refine the language of the manuscript and (2) to improve flow.

References

Barlow, J. M., Bentzen, B. L., & Bond, T. (2005). Blind pedestrians and the changing technology and geometry of signalized intersections: Safety, orientation, and independence. *Journal of Visual Impairment & Blindness*, 99(10), 587-598.

Hall, E. T. (1966). *The hidden dimension.* Garden City.

Henly, M., & Brucker, D. L. (2019). Transportation patterns demonstrate inequalities in community participation for working-age Americans with disabilities. *Transportation Research Part A: Policy and Practice*, 130, 93-106.

Nimma, D., Al-Omari, O., Pradhan, R., Ulmas, Z., Krishna, R., El-Ebiary, T. Y. A. B., & Rao, V. S. (2025). Object detection in real-time video surveillance using attention based transformer-YOLOv8 model. *Alexandria Engineering Journal*, 118, 482-495.

OpenAI. (2025). ChatGPT Model: GPT-5 [Large language model].https://openai.com/chatgpt/

Pigeon, C., Baltenneck, N., Evennou, M., Uzan, G., & Galiano, A. R. (2026). Mobility in adults with visual impairment: Results from an online survey. *Transportation Research Part F: Traffic Psychology and Behaviour*, 116, 103429.

Rosa, M. P., de Mello, G. S., & Morato, S. (2021). Tactile paving surfaces at bus stops: The need of homogeneous technical solutions for accessible tourism. *Journal of Accessibility and Design for All*, 11(2), 259-294

van Ameijde, J., Cheng, S., Leung, K. S., Zhang, W., & Dong, J. (2025). Decoding micro-social interactions in public space: A computer-vision-based method. *Computational Urban Science*, 5(1), 55.

Watanabe, T., Yamaguchi, T., Koda, S., & Minatani, K. (2014). Tactile map automated creation system using OpenStreetMap. In K. Miesenberger, D. Fels, D. Archambault, P. Peňáz, & W. Zagler (Eds.), *Computers helping people with special needs: ICCHP 2014 (Lecture Notes in Computer Science, Vol. 8548, pp. 42-49)*. Springer.

Yang, J., & Chen, M. (2025). Assessing the impact of urban amenities on people with disabilities in London: A multiscale geographically weighted regression analysis. *Habitat International*, 161.

Zhao, R., Edelenbos, J., de Jong, M., Belabas, W., & Björner, E. (2025). Challenges of implementing inclusive city policies in three European port cities. *Npj Urban Sustainability*, 5(1), 47.

Revealing Post-Pandemic Social Interaction Shifts in Urban Squares:

A Deep Learning-Based Dynamic Monitoring Framework

Shuyu Huang[1], Nanshuo Mu[2], Shuyao Ren[3], Qi Qin[4], Xintong Lyu[5], Waishan Qiu[6], and Chongyang Ren[7]
[1]*College of Architecture and Urban Planning, Tongji University*
[2]*Melbourne School of Design, The University of Melbourne*
[3,4]*School of Architecture, Southeast University*
[5,6]*Department of Urban Planning and Design, The University of Hong Kong*
[7]*Civil, Environmental and Geomatic Engineering, University College London*
[1]*hsh4yu@tongji.edu.cn, 0000-0001-7978-6795*
[2]*nanshuo.mu@student.unimelb.edu.au, 0009-0006-1762-6079*
[3]*sy.ren@seu.edu.cn, 0009-0003-9318-1567*
[4]*qinqi@seu.edu.cn, 0009-0003-2445-7814*
[5]*lyuxt24@connect.hku.hk, 0009-0002-3236-8291*
[6]*waishanq@hku.hk, 0000-0001-6461-7243*
[7]*chongyang.ren.25@ucl.ac.uk, 0009-0008-3330-834X*

Abstract. The COVID-19 pandemic reshaped social interaction through public-space closures and distancing measures. However, quantitative evidence on its longterm effects on social interaction remains limited. Grounded in Proxemics theory, this study integrates computer vision techniques, combining YOLO object detection with DeepSORT tracking, to dynamically quantify social interaction in urban squares. Based on 240 minutes of pre- and post-pandemic video data from four Italian squares, over 300,000 pedestrian trajectories were identified and analyzed. Five indicators were then calculated, including Interpersonal Distance (IPD), Approach Event Rate (AER), Group Ratio (GR), Collision Event Rate (CER), and Dynamic Avoidance Intensity (DAI), to measure different states of social interaction. The results show that (1) IPD increased overall in the post-pandemic period; (2) social interaction patterns exhibited polarization, with some spaces showing intensified aggregation (higher AER, GR, and CER) and others maintaining social restraint;(3) changes in DAI revealed spatial variations in avoidance strategies. The study demonstrates the long term and spatially heterogeneous impacts of the pandemic on social interaction and emphasizes the key role of spatial form in shaping micro-scale behavioral adaptation.

Keywords. Social Interaction, Urban Squares, Post-pandemic, COVID-19, Computer Vision, Deep Learning, Urban Resilience

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 695-704.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

1. Introduction

Urban squares are often regarded as "containers" of civic life and social activity (Gehl, 2006; Jacobs, 1989), where gathering and conversation form the core of spatial vitality (Whyte, 1980). Environmental psychology further highlights that face-to-face social interaction in public spaces can significantly enhance individual well-being, promote social cohesion and urban vitality, and strengthen urban resilience and productivity (Chen et al., 2025). However, sudden public health emergencies can fundamentally alter people's risk perception and behavioural patterns, disrupting urban social order and testing the resilience of social networks and support systems. Among these, the COVID-19 pandemic has been the most globally influential in recent years. Worldwide lockdowns and distancing measures silenced urban squares, institutionally constraining collective activities and reshaping behavioural patterns in public spaces. Everyday interactions were abruptly interrupted, while prolonged caution and avoidance fostered a "restrained sociability" as the new norm (Ersoz & Yildirim, 2023). Some studies also observed that post-lockdown park visits exceeded pre-pandemic levels, reflecting a strong desire to revive social life. This raises a key question: in the post-pandemic era, are social interactions in public spaces suppressed, revitalised, or reconfigured?

Empirical studies on transformations of social interaction in public spaces remain methodologically limited. Traditional observations and surveys lack sufficient spatiotemporal scale and precision. The rise of computer vision (CV) and deep learning technologies has brought significant breakthroughs. Object detection models such as YOLO (You Only Look Once) combined with DeepSORT tracking algorithms can automatically identify and follow pedestrians, generating accurate trajectory data from video footage (Wojke et al., 2017). Although some studies have attempted to recognise conversational groups through pedestrian position and orientation (F-formations) (Setti et al., 2015), a systematic framework for classifying social interaction relationships in public spaces is still lacking.

Edward Hall's theory of proxemics suggests that interpersonal distance reflects behavioural intentions and social relationships, providing a theoretical foundation for micro-scale observation of social interaction (Hall, 1990). Recent advances in computer vision (CV) techniques enable fine-grained measurement of interpersonal distance, movement direction, and speed. Based on this theoretical and technical foundation, this study proposes a CV-based analytical framework that takes interpersonal distance as the core indicator and derives four supplementary metrics from pedestrian trajectories to capture interaction states such as approach, aggregation, and avoidance. Four historically significant Italian squares with high social and touristic activity were selected as empirical cases. Approximately 240 minutes of pre- and post-pandemic video footage were analysed, identifying around 300,000 pedestrian trajectories to compare social interaction patterns before and after the pandemic. The study examines whether pandemic-related behavioural changes in public spaces are suppressed, revitalised, or reconfigured over time, providing empirical evidence to support the design of resilient, socially supportive.

2. Methodology

2.1. RESEARCH FRAMEWORK

The analytical framework of this study consists of five steps (Figure 1). Step 1: High-resolution webcam videos were collected from four Italian squares, covering both the pre-pandemic (2018–2019) and post-pandemic (2025) periods. Step 2: The original footage was enhanced using video super-resolution techniques to improve the accuracy of pedestrian detection. Step 3: Pedestrians were identified using the YOLO model and tracked through the DeepSORT algorithm. The resulting trajectories were then converted from camera coordinates to real-world geographic coordinates. Step 4: Based on the trajectory data, five key indicators were calculated to characterise different states of social interaction. Step 5: The results of these indicators were integrated to reveal the changes and resilience characteristics of social interaction in the post-pandemic era.

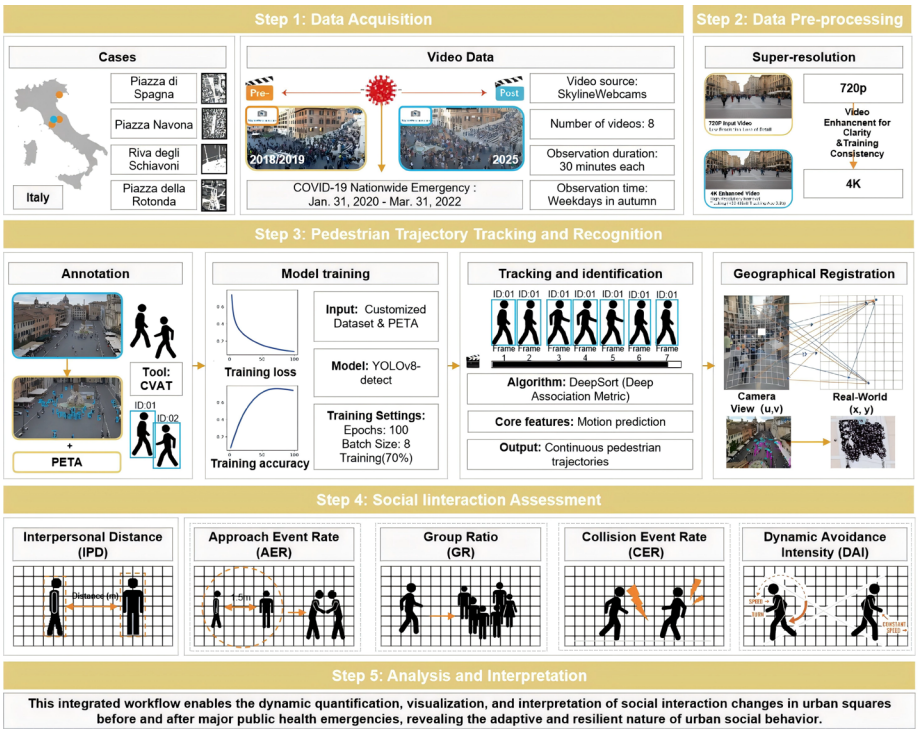


Figure 1. Research framework

2.2. CASE STUDY

This study selected four representative urban squares in Italy as empirical research sites: (1) Piazza Navona, (2) Piazza di Spagna, (3) Riva degli Schiavoni, and (4)Piazza della Rotonda. These squares are high-density pedestrian environments and important venues for informal social interaction, serving as vibrant nodes of gathering, leisure, and tourism. Each square remained continuously active across pre- and post-pandemic periods, ensuring temporal comparability.



Figure 2. Study areas and pre/post COVID-19 Video

2.3. VIDEO DATA CURATION

All video data were obtained from the publicly accessible platform SkylineWebcams (https://www.skylinewebcams.com), which continuously records urban activities in public spaces using fixed-position cameras. Video segments were selected under similar environmental conditions (see Table 1). All identifiable pedestrian information was fully anonymized during processing.

Table 1. Overview of the video dataset

Square	City	Year(Pre/Post COVID-19)	Season	Day Type	Weather	Duration (min)
Piazza Navona	Rome	2018 / 2025	Autumn	Weekday	Clear	30 each
Piazza di Spagna	Rome	2019 / 2025	Autumn	Weekday	Clear	30 each
Riva degli Schiavoni	Venice	2019 / 2025	Autumn	Weekday	Clear	30 each
Piazza della Rotonda	Rome	2018 / 2025	Autumn	Weekday	Clear	30 each

2.4. YOLO-BASED PEDESTRIAN TRAJECTORY DETECTION

2.4.1. Annotation and Training

To ensure model robustness, this study combined a general public dataset (PEdesTrian Attribute, PETA) with a customised dataset, each contributing half of the total samples. The customised dataset was constructed by frame sampling from video sequences: from each one-minute segment (about 1,800 frames), three clips of ten frames each were randomly selected to capture representative pedestrian features while maintaining temporal continuity. The combined dataset was divided into training and validation sets in a 7:3 ratio, with all validation samples drawn from the customised dataset to ensure generalisability.

The YOLOv8 model was initialised with COCO-pretrained weights, using a batch size of 8, an initial learning rate of 1×10^{-4}, and 100 training epochs. Model performance was evaluated using three key metrics: training accuracy (the proportion of correctly identified pedestrian instances), training loss, and validation loss, which respectively measure the deviation between predicted and ground-truth bounding boxes in the training and validation datasets.

2.4.2. Detection and Geospatial Registration

The DeepSORT algorithm was applied to maintain consistent pedestrian identities

across consecutive frames. Four fixed feature points within each square were manually matched to real-world coordinates to convert pixel coordinates into geographic space (Wojke et al., 2017). All positional data were standardised to the WGS 1984 / UTM Zone 46N coordinate system.

2.5. QUANTITATIVE INDICATORS OF SOCIAL INTERACTION

2.5.1. Macro Scale: Interpersonal Distance

Interpersonal Distance (IPD) measures the physical spacing between individuals, reflecting cultural norms, social relations, and psychological comfort. It is classified into three types: Intimate (0–0.45 m), Personal (0.45–1.20 m), and Social (1.20–3.60 m). In YOLO-based detection, pedestrians were represented as vector points without body dimensions. To approximate actual spacing, a 0.7 m shoulder-width correction was applied following Salazar-Miranda et al. (2025). The median pairwise distance between all pedestrians per frame was defined as the IPD metric.

2.5.2. Micro Scale: Spatiotemporal Motion Metrics

To reveal the micro-level mechanisms of social interaction, four trajectory-based motion metrics were developed (Table 2). These indicators characterise the dynamics of approach, aggregation, and avoidance between individuals.

Table 2. Social Interaction Metrics

Metrics Name	Measurement definition	Significance
Approach Event Rate	Counts events where the interpersonal distance between two individuals continuously decreases from above 1.9 m to 1.9 m or less, and remains within this threshold for at least 2 seconds within a 120-frame window (Salazar-Miranda et al., 2025).	Indicates initiation of engagement or affiliative tendency.
Group Ratio	Proportion of clustered pedestrians per frame (DBSCAN,ε = 1.9 m, minPts = 5). GR = number of clustered individuals / total detections per frame, averaged across frames.	Reflects group cohesion and collective presence.
Collision Event Rate	A collision occurs when two individuals sequentially enter and occupy the same 1.15 m-diameter area within ≤1 s and both leave within ≤1 s. CER = number of collision events / total person-frames, averaged across windows (Arun et al., 2021).	Represents spatial crowding and potential conflicts.
Dynamic Avoidance Intensity - velocity(v) / direction(θ)	Avoidance is identified when two pedestrians are predicted to approach within 1.15 m in 1 s but subsequently diverge. Directional and velocity changes are then averaged to derive as DAI_θ and DAI_v(Hoffmann and Mortimer, 1994).	Captures proactive adjustment and situational awareness.

3. Results

3.1. EVALUATION OF MODEL PERFORMANCE

Loss values steadily decreased and stabilised, while accuracy progressively improved during training (Figure 3a–c). The final accuracy reached 0.87, with low losses (training < 0.2, validation ≈ 0.4), demonstrating the model's strong robustness in pedestrian detection .

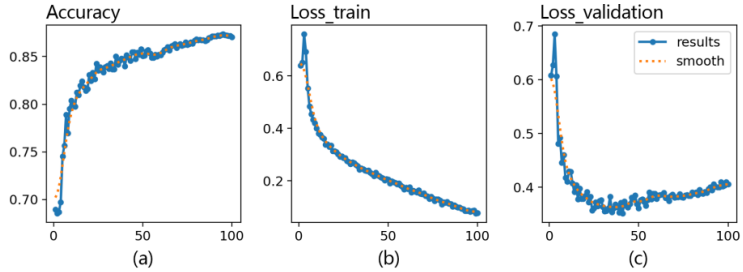


Figure 3. Training performance of YOLOv8 model. (a) Training accuracy curve, (b) Training loss curve, and (c) Validation loss curve.

3.2. FOUNDATIONAL SHIFT OF IPD PATTERNS

Figure 4 compares IPD before and after the pandemic. Both the overall results (Figure 4a) and those for individual squares (Figure 4b) show more concentrated IPD distributions, with the median increasing by about 10–15%. This reflects an elevated threshold of personal space and more uniform spacing patterns, indicating that individuals maintained greater social distances after the lockdown.

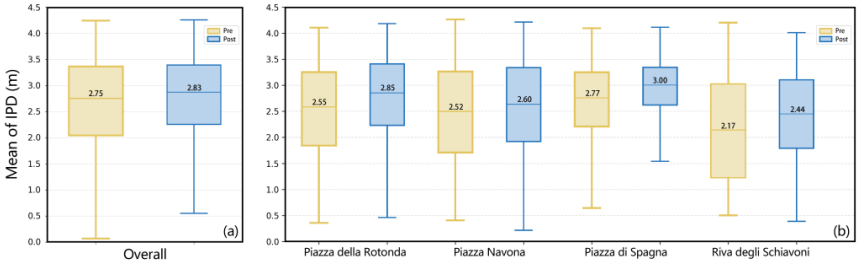


Figure 4. Box plots of median Interpersonal Distance (IPD) among pedestrians: (a) overall, (b) across different urban squares.

Using the inverse of IPD as a weight to represent social activity intensity, spatial heat maps were produced (Figure 5). Clear spatial differences emerged: after the pandemic, Piazza di Spagna and Piazza della Rotonda retained multiple active clusters, while in Piazza Navona and Riva degli Schiavoni overall social intensity declined and became more concentrated. In the former, activity no longer centred around the iconic fountain, whereas in the latter, interactions shifted toward building edges.

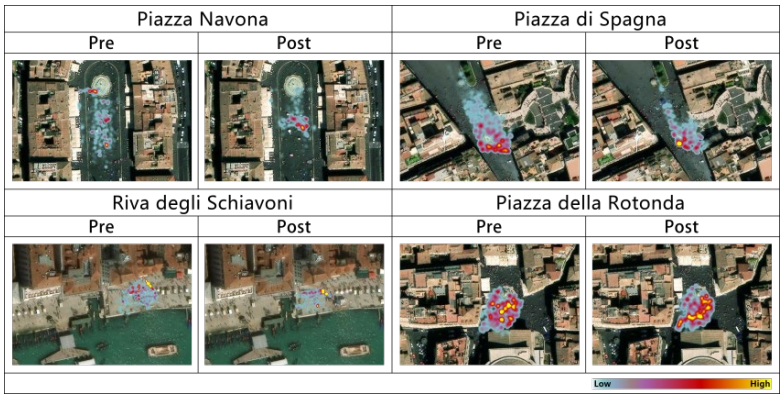


Figure 5. Spatial distribution of social interaction hotspots across urban squares, visualized using IPD-weighted density.

Figure 6 shows changes in IPD across different distance categories. Overall (Figure 6a), intimate and social distances increased after the pandemic, while personal distance decreased slightly, indicating a general tendency to maintain greater physical spacing in public spaces. Interpersonal distances increased across all squares, but the magnitude and direction of change varied by distance type, likely influenced by differences in spatial form and functional structure (Figure 6b).

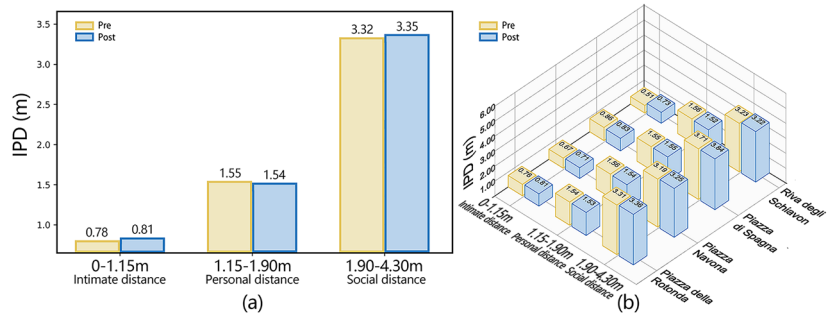


Figure 6. Changes in median IPD across different distance categories: (a) overall results; (b) comparisons across urban squares.

3.3. DYNAMIC EVOLUTION OF SOCIAL INTERACTION

Figure 7 shows the variations in AER, GR, and CER across the overall and four urban squares before and after the pandemic. Overall (Figure 7a, 7c, 7e), all three indicators increased, indicating that aggregative social behaviour and overall social activity in public spaces strengthened significantly after the pandemic. Distinct spatial differences were observed across the squares (Figure 7b, 7d, 7f). Piazza della Rotonda and Piazza di Spagna showed simultaneous increases in all three indicators, suggesting stronger group cohesion and higher social activity intensity. In contrast, Piazza Navona and Riva degli Schiavoni showed declines across all three indicators, reflecting more restrained and less dynamic social behaviour.

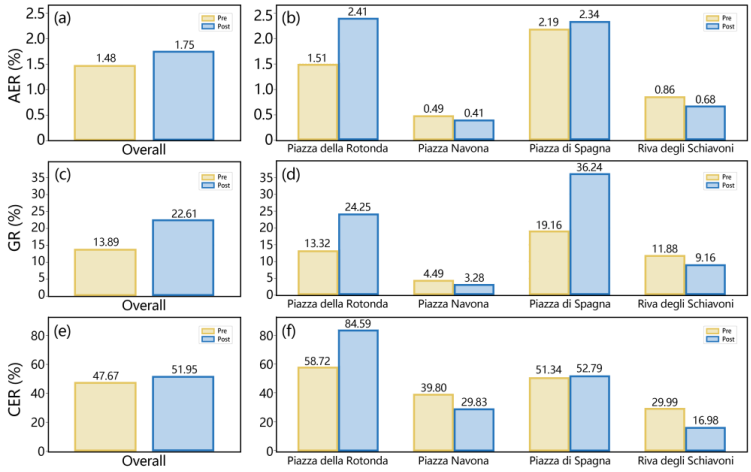


Figure 7. Variations in AER, GR, and CER between the pre-pandemic and post-pandemic periods. Subplots (a, c) show overall results, while (b, d) compare different urban squares.

Figure 8 shows changes in DAI_v and DAI_θ before and after the pandemic, reflecting pedestrians' adjustments in speed and direction when facing potential collision risks. Overall (Figure 8a, 8b), pedestrians reduced their speed less but turned at larger angles, indicating smoother yet more responsive avoidance strategies. However, both indicators exhibited wider extreme ranges, suggesting that avoidance behaviour became more polarised and dependent on specific spatial contexts.Different avoidance patterns were also observed across the squares (Figure 8c, 8d). In Piazza di Spagna and Piazza Navona, pedestrians showed larger turning angles and more evident speed adjustments, indicating stronger risk perception and adaptive responses. In contrast, pedestrians in Riva degli Schiavoni and Piazza della Rotonda generally shifted from deceleration to acceleration, with varying turning angles, revealing spatial differences in risk-response behaviour.

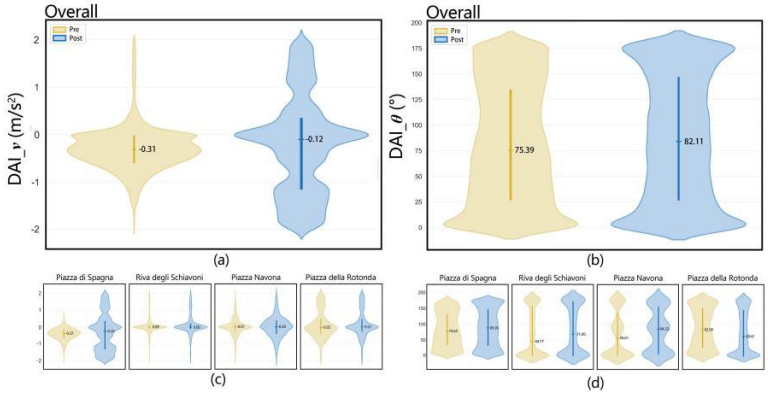


Figure 8. Variations in DAI_v and DAI_θ between the pre- and post-pandemic periods. Subplots (a, c) show overall results, while (b, d) compare different urban squares.

4. Discussion and Conclusion

4.1. DIFFERENTIAL CHANGES IN SOCIAL INTERACTION

This study revealed differentiated changes in social interaction within public spaces under the influence of a major public health event. The analysis of IPD showed that individuals generally maintained greater physical spacing in the post-pandemic period, reflecting a cautious recovery mechanism. Two distinct evolutionary paths were identified. In Piazza di Spagna and Piazza della Rotonda, AER, GR, and CER all increased significantly, indicating the recovery and intensification of aggregative interactions. This pattern aligns with Mehta's concept of new proxemics, suggesting that temporary spatial defensiveness nurtured a long-term desire for face-to-face contact. In contrast, Piazza Navona and Riva degli Schiavoni showed declines across all three indicators, reflecting a more restrained behavioural mode. These findings reveal the dual impact of the pandemic on social interaction: on one hand, the resurgence of social activity in some spaces reflected the compensatory release of suppressed social demand as restrictions eased and psychological security returned (Ersoz & Yildirim, 2023); on the other, persistent social restraint and avoidance in certain spaces weakened social connectedness.

4.2. SPATIAL HETEROGENEITY OF SOCIAL INTERACTION

The four squares analysed in this study are located in the historic cores of Italian cities and designated as UNESCO World Heritage sites, characterised by highly concentrated social and touristic activities. Although they share similar user compositions and functional attributes, post-pandemic social interaction patterns showed notable differences, reflecting the micro-regulatory influence of spatial environments on social behaviour. Previous studies have shown that spatial morphology and functional configuration largely determine the form and intensity of social interaction. Piazza di Spagna and Piazza della Rotonda, with their semi-enclosed or circular layouts, high spatial inclusiveness, and diverse mixed-use surroundings, promote incidental encounters and group gatherings, exhibiting stronger sociopetal characteristics after the pandemic. In contrast, Riva degli Schiavoni and Piazza Navona, with open and linear forms lacking enclosure and clear social focal points, displayed sociofugal tendencies (Loo & Fan, 2023), confirming Gehl's (2006) observation that overly open or large-scale spaces tend to suppress spontaneous interaction.

4.3. PLANNING AND DESIGN INSIGHTS AND FUTURE DIRECTIONS

This study overcomes the temporal and spatial limitations of traditional observational methods by proposing an integrated analytical framework that combines CV with Proxemics theory, enabling the dynamic quantification and visualisation of social interaction and revealing its differentiated changes and spatial heterogeneity under public health crises. The results indicate that post-pandemic social recovery in public spaces does not occur uniformly, but is strongly mediated by micro-spatial form. Semi-enclosed or spatially articulated layouts, clear social focal points, and mixed-use edges facilitate the re-emergence of sociopetal behaviour. In contrast, spaces characterised by

limited enclosure, the absence of spatial anchors, and open or linear configurations are more likely to sustain sociofugal interaction patterns. From a planning and design perspective, these findings suggest that post-pandemic public-space design should move beyond an exclusive emphasis on openness or crowd-dispersal efficiency. Greater attention should instead be given to spatial legibility, human-scale enclosure, and the provision of diverse behavioural choices. Creating environments that allow individuals to flexibly modulate interpersonal distance, levels of engagement, and avoidance behaviours may help restore everyday social interaction while maintaining a sense of safety, thereby enhancing the social resilience of public spaces.

Despite these contributions, several limitations remain. The temporal representativeness of the findings is constrained by the amount and duration of available video data, and the focus on Italian historical squares may limit generalisability. Future research should extend this framework to broader urban and cultural contexts. In addition, quantitative modelling of the influence of spatial morphology on social interaction would further support the adaptive design of resilient public spaces for future public health emergencies.

Attribution Statement

ChatGPT (OpenAI, 2025) was used to assist in language refinement and proofreading. All conceptualization, analysis, and interpretation were conducted by the authors.

References

Arun, A., Haque, Md. M., Washington, S., Sayed, T., & Mannering, F. (2021). A systematic review of traffic conflict-based safety measures with a focus on application context. Analytic Methods in Accident Research, 32, 100185.

Chen, W., Natapov, A., & Ali, Y. (2025). Can urban vitality be seen? Video analytics of social interaction and land use. Land Use Policy, 159, 107818.

Ersoz, N. D., & Yildirim, Y. (2023). Sociopetality or sociofugality? The effects of post-pandemic on public open space. URBAN DESIGN International, 1–10.

Gehl, J. (2006). Life Between Buildings: Using Public Space. Danish Architectural Press.

Hall, E. T. (1990). The Hidden Dimension. Anchor.

Hoffmann, E. R., & Mortimer, R. G. (1994). Drivers' estimates of time to collision. Accident Analysis & Prevention, 26(4), 511–520.

Jacobs, J. (1989). The Death and Life of Great American Cities. Knopf Doubleday Publishing Group.

Loo, B. P., & Fan, Z. (2023). Social interaction in public space: Spatial edges, moveable furniture, and visual landmarks. Environment and Planning B: Urban Analytics and City Science.

OpenAI. (2025). ChatGPT Model: GPT-5. https://openai.com/chatgpt/

Salazar-Miranda, A., Fan, Z., Baick, M., Hampton, K. N., Duarte, F., Loo, B. P. Y., Glaeser, E., & Ratti, C. (2025). Exploring the social life of urban spaces through AI. PNAS, 122(30), e2424662122.

Setti, F., Russell, C., Bassetti, C., & Cristani, M. (2015). F-formation detection: Individuating free-standing conversational groups in images. PloS One, 10(5), e0123783. Whyte, W. H. (1980). The Social Life of Small Urban Spaces. Conservation Foundation.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. 2017 IEEE International Conference on Image Processing (ICIP), 3645–3649.

Urban Morphology and Regeneration

Just Emotional Geographies:

Intersecting Decolonial and Post-Gender Design in Garden Cities with AI Audits

Yuxuan Liu[1] and Trishla Chadha[2] and Noor Mubesher[3] and Aya Meskawi[4] and Provides Ng[5]
[1] *College of Architecture and Urban Planning, Tongji University.*
[2,3,4,5] *Bartlett School of Architecture, University College London.*
[5] *School of Architecture, The Chinese University of Hong Kong.*
[1]*yuxuanliu@tongji.edu.cn, 0000-0003-3160-5750*
[2]*trishla.chadha.21@ucl.ac.uk, 0000-0001-7011-1644*
[3]*noor.mubesher.21@ucl.ac.uk, 0009-0004-8696-5686*
[4]*aya.meskawi.22@ucl.ac.uk, 0009-0004-4689-1248*
[5]*provides.ng.19@ucl.ac.uk, 0000-0001-6975-4642*

Abstract. The concept of Garden City advocates for an urban design approach that balances community and development needs. However, such a concept was operationalised as an instrument of segregation in colonial eras, evident in dense urban contexts like Hong Kong, New Delhi, and more. Such urban morphology could contribute to dataset bias that perpetuates racialised, gendered spatial inequities. This study critically interrogates the capacity of AI-driven spatial audits to diagnose such exclusions. Employing a decolonial framework and emotional geography methods, the study integrated computational analysis of street-view imagery and social media data with critical urban theory. Our findings reveal how algorithmic indices, rooted in Eurocentric principles, misread urban realities. They valorised sterile, elite enclaves with high 'Safe' scores, overlooking low female presence and negative local sentiment. These dissonances expose how computational epistemologies reproduce colonial logics of legibility and control. We argue that AI must be repositioned as a critical interlocutor; its failures to align with lived experience are not errors but analytical evidence, illuminating the limits of algorithmic reasoning and charting needs toward more situated, postcolonial urban metrics.

Keywords. Digital Geography, Algorithmic Bias, Gendered Urban Exclusion, Postcolonial Garden City, Affective Dissonance

1. Introduction

Urban design in Hong Kong and New Delhi continue to reproduce spatial hierarchies rooted in colonial-era Garden City ideologies (Chu, 2022). While Ebenezer Howard's Garden City model promoted balance between community, nature, and urbanity, its

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 707-716.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

colonial application served imperial aims that created segregation, privileged European interests, and displaced local communities (Li, 2025). Instead of egalitarian green towns, these imported models materialised as instruments of racial, gendered, and classed differentiation.

In Hong Kong, Kowloon Tong epitomised this spatial bifurcation via its low dense prestige architecture adjacent to high dense districts that housed labouring locals: the urban form enabled racialised servitude systems, gendered labour practices, and segregated sex work economies (Chan, 2012; Yu & Wong, 2013). In New Delhi, Lutyens' monumental Garden City formalised imperial dominance through expansive boulevards, elite civil lines, and bungalow zones that encoded caste and class hierarchies (King, 1976).

Recent advances in AI promise to quantify urban inequalities using computer vision and geospatial analysis. However, the epistemic assumptions embedded in AI models, particularly their alignment with Western CPTED (Crime Prevention Through Environmental Design) frameworks, raised fundamental questions (Richardson et al., 2019). In postcolonial contexts, such metrics risk reproducing colonial ways of seeing privileged order, visibility, and sterility over communal presence, shaping unequal access, mobility, and safety in the city to the present day. To what extent can AI-driven spatial audit capture colonial continuities of gendered exclusion, and where do computational methods fail to align with lived, emotional geographies of safety and justice?

2. Garden City & Gender Inequality

Urban theorists have long examined how colonial planning models produced enduring socio-spatial inequalities across postcolonial cities. Scholars such as Chu (2022) and Li (2025) highlighted contradictions inherent in colonial Garden City projects, where utopian narratives of green, orderly, "modern" townships obscured the displacement, segregation, and spatial compression experienced by local communities. Studies from New Delhi, Quezon, Ipoh, Canberra, Zanzibar, and more illustrated how transplantations of European ideals routinely generated hierarchical landscapes that privileged colonial settlers and formalised racial and class divisions (Bigon, 2014; O'Neill, 2017; Boquet, 2016).

The case of Hong Kong revealed acute intersections between gender, labour, and spatial ordering. British legal reforms expanded certain formal rights for women, yet these reforms coexisted with, in some cases intensified, gendered systems of servitude such as Mui-tsai domestic slavery (Carroll, 2009; Du, 2023). The colonial military and economic demand for sexual and domestic labour proliferated one-woman brothels (一樓一鳳) and port-linked prostitution markets (Howell, 2004; Berney, 2017), contributing to the moral stigmatization of Chinese women more broadly (Yu & Wong, 2013). Spatial studies of Hong Kong's dense urban fabric revealed how brothel locations and domestic worker enclaves mapped onto colonial racial hierarchies, shaping differential patterns of visibility and mobility (Taikwun, 2023). These insights underscored that "female presence" in public space must be interpreted through the lens of gendered and racialised labour, not merely demographic representation.

Recent technological approaches including AI-driven spatial audits (Pan et al.,

2025), perception-based safety modelling (Wu et al., 2022), and GIS-based inequality mapping (Zhand et al., 2024) offered tools for analysing urban inequities at scale. However, scholars cautioned how computational systems often inherit Western epistemy, particularly regarding safety, order, and visual legibility (Bottazzi, 2024). These models risk misinterpreting dense, informal, or culturally specific spatial practices as "disorder," while valorising elite exclusionary landscapes as "safe."

Overall, global indices attempt to measure gender equality and intersectional inclusivity, significant gaps remain in linking urban morphology to culturally situated gendered experience, particularly in post colonial contexts where labour geographies, caste/class dynamics, and racialised histories shape how women navigate and inhabit space. The conflicting outputs of AI-driven metrics in this study underscored these concerns, demonstrating how algorithmic audits can reproduce colonial biases unless critically embedded within decolonial, qualitative, and community-centred frameworks.

3. Methodology

This study employed a mixed-methods approach rooted in critical data science and decolonial praxis, integrating computational, spatial, and qualitative analyses. Its tripartite framework (emotion analysis, machine vision, and human adjudication) provided convergent lenses to interrogate the gendered colonial legacies within the urban forms of New Delhi and Hong Kong. These sites are selected as morphological archetypes to test the feasibility of the audit across opposing planning logics; while highly specific, they validate the methodology before its expansion to the broader, heterogeneous urban fabric. These were designed specifically to surface contradictions between algorithmic readings and lived experience, rejecting assumptions of computational neutrality.

The objectives were: (1) to identify perpetuated spatial hierarchies by comparing Garden City planning in Hong Kong and New Delhi; (2) to test whether the proposed AI-driven spatial audits inherited Eurocentric or colonial traits; and (3) to examine where AI-derived indices diverged from affective and experiential accounts of safety and gendered presence, with the hypothesis:

H1: in India, Garden City areas have more gender equity

H2: in HK, Garden City areas have less gender equity

H3: in India, lower income areas have less gender equity

H4: in HK, lower income areas have more gender equity

3.1. RESEARCH APPROACH & ETHICAL GROUNDING

The method integrated AI-driven spatial audits with human-centred interpretation, explicitly acknowledging that AI systems and spatial indices were embedded within specific epistemologies, often shaped by Eurocentric planning paradigms and colonial histories of legibility. Computational findings were consequently treated as one evidential layer among many, to be triangulated and contested by historical archives, feminist scholarship, and community narratives across four key dimensions.

Gendered Presence & Use drew on Jacobs (1961) and Hayden (1997) to analyse

how urban form shaped gendered occupation, informed by Pain (2001) and Valentine (1989) on geographies of fear. It critically assessed how computer-vision 'presence' could mask unequal labour relations rather than signal equitable access.

Spatial Security & Infrastructure employed CPTED principles (Jeffery 1971; Skogan 1990), revealing contradictions between high 'safe infrastructure' scores in elite, low-activity enclaves and low perceived safety by women.

Colonial Urban Morphology, informed by King (1976) and Home (1997), assessed how colonial planning logics persisted through spatial hierarchies of density, access, and segregation.

Public Perception & Social Emotions, guided by Foucault (1977), Lefebvre (1996), and feminist geographers (Koskela & Pain 2000), interrogated emotional responses and civic discourse on safety and justice. It specifically analysed how fear, comfort, exclusion, and resistance were expressed through language and location, investigating where these affective geographies clashed with algorithmic assessments of 'safety' or 'risk'.

3.2. DATA COLLECTION & ALGORITHMIC FRAMEWORK

A systematic multimodal dataset was assembled from three primary sources: Street-view imagery captured urban morphologies, infrastructures, and street activity, detecting pedestrians, furniture, lighting, and built form. Geotagged social media posts, collected via gender- and safety-related keywords, provided data on affective expressions, risk perceptions, and public space discussions. Historical colonial maps and contemporary geospatial datasets provided the baselines for tracing continuities and ruptures between colonial planning and contemporary spatial hierarchies..

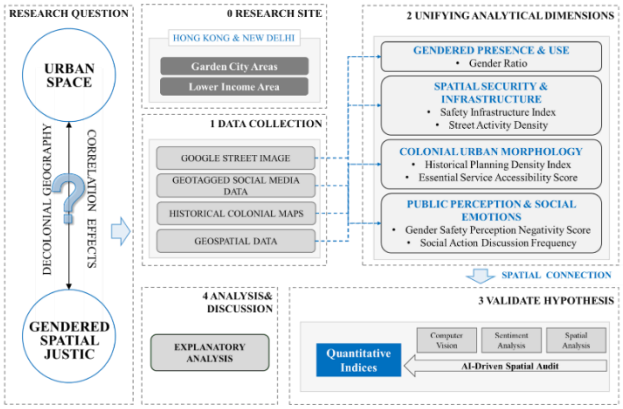


Figure 1. The overall Research Framework

A suite of machine-learning and GIS-based methods translated the theoretical dimensions into measurable indices standardised to a percentage scale (0–100%) for this feasibility test. Recognising that quantitative analysis alone risked reproducing bias, the study embedded a Human-in-the-Loop (HITL) validation process: algorithmic outputs were critically examined against historical and qualitative

evidence. Attention was given to model bias, particularly gender misclassification and racial skew (Buolamwini & Gebru 2018), and to the partial temporality of street-view imagery.



Figure 2. Study Area

4. AI-driven Audits: Preliminary Results

The empirical data yielded a nuanced response to the initial hypotheses. While findings in the New Delhi context aligned with H1 and H3—showing higher female presence and better service accessibility in Garden City areas—the results in Hong Kong directly contested H2 and H4, where the planned districts exhibited significantly higher gender equity metrics than their lower-income counterparts. In New Delhi, the Garden City area exhibited higher female presence (18.2%) than lower-income area (3.9%), yet this was still markedly low. Conversely, in Hong Kong, the Garden City area demonstrated a significantly higher female proportion (60.0%) than its lower-income counterpart (43.8%). This inverted relationship between the cities undermined any simplistic correlation between planning typology and gender equity.

A stark paradox emerged in safety perception. Despite its high score in Safety Infrastructure (69.2%), New Delhi's Garden City area had a considerably negative Gender Safety Perception (24.6%) than both Hong Kong sites. This indicated that superior physical infrastructure did not translate into a greater sense of safety, exposing a critical failure of purely metric-based security audits.

The Colonial Garden City Paradox. Findings complicated the legacy of colonial morphology. The planned Garden City areas in both Hong Kong and New Delhi consistently showed higher Essential Service Accessibility Scores (86.8%, 82.7%), validating one intention of such planning. However, the social outcomes diverged sharply.

In New Delhi, low-density, planned morphology correlated with low street activity and low female presence, suggesting spatial exclusion. In Hong Kong, the same planning coincided with highest female presence and lowest safety negativity perception, demonstrating that its social impact was mediated by intense local context.

Table 1. AI-Driven Spatial Audit and Quantitative Index Calculation

Dimension	Index	Method	Data	Theoretical Grounding
1. Gendered Presence & Use	Gender Ratio in Public Space	Utilise computer vision to detect pedestrians in street-view images, followed by a gender classification model. The ratio is calculated per image or street segment as Ratio = Nfemale / Ntotal. Index values are standardised to a 0–100% scale.	Google Street View Imagery	Jacobs (1961); Hayden (1995) on gendered occupation; Pain (2001); Valentine (1989) on geographies of fear.
2. Spatial Security & Infrastructure	"Safe" Infrastructure Index	Employ computer vision to detect and quantify positive safety features (e.g., streetlights, benches, transparent storefronts) and negative features (e.g., graffiti, litter, decay). An index is created by weighting and aggregating these feature counts and standardising to a 0–100% scale.	Google Street View Imagery	CPTED principles (Jeffery 1971; Skogan 1990) to operationalise infrastructure-safety relationships.
	Foot Traffic Density & Street Activity	Use computer vision object detection to count the number of pedestrians and vehicles within street-view images as a proxy for the level of street activity and "eyes on the street." Values are standardised to a 0–100% scale.	Google Street View Imagery	Jacobs (1961) on "eyes on the street."
3. Colonial Urban Morphology	Historical Planning Density Index	Georeference and digitise colonial-era maps to calculate planned building coverage or floor area ratios. Compare with contemporary data to identify enduring spatial hierarchies. Values are standardised to a 0–100% scale.	Colonial era Historical Maps, Modern GIS & Satellite Data	King (1976); Home (1997) on the persistence of colonial planning logics.
	Essential Service Accessibility Score	Using GIS Network Analyst, calculate the average travel time or distance from residential block centroids to the nearest essential services (e.g., healthcare, food markets, public transit). Scores are inverted and standardised to a 0–100% scale (higher score = better access).	GIS Street Network Data, Residential & Service Location Data	Critical urban theory to assess how colonial hierarchies persist through segregated access.
4. Public Perception & Social Emotions	Gender Safety Perception Negativity Score	Filter geotagged social media text for keywords related to gender and safety. Apply the BERT model to calculate a compound sentiment score (-1 to +1) for each post. The index is the proportion of posts with a negative score (e.g., < -0.05) in a given area, standardised to a 0–100% scale.	Geotagged Twitter (X) Data	Foucault (1977); Lefebvre (1996); feminist geographers (Koskela & Pain 2000) on emotional responses and discourse.
	Social Action Discussion Frequency	Quantify the frequency of posts containing keywords related to social justice and urban rights (e.g., "protest," "right to the city," "community") to map areas of civic engagement discourse. Frequency counts are standardised to a 0–100% scale.	Geotagged Twitter (X) Data	Lefebvre (1996) on the "right to the city"; explores discourses of resistance.

The higher Historical Planning Density and lower service accessibility in both lower income areas confirmed their morphological history, which in Karol Bagh correlated with the most negative gendered safety perceptions (39.3%).

Divergence Between Algorithmic and Social Indices. The analysis revealed where algorithmic readings diverged from affective and discursive realities. For instance, Sham Shui Po displayed a high Safety Infrastructure Index (83.9%) but a moderate

Street Activity Density (37.4%) and a mid-range female presence (43.8%). Its very low Negative Perception on Gender Safety (13.1%), however, suggested that factors beyond measurable infrastructure fostered a perceived sense of safety. Furthermore, the higher Social Action Discussion Frequency in Hong Kong sites (2A: 43.2%, 2B: 31.7%) versus the lower frequencies in New Delhi indicated a more active public discourse around gendered space, which itself constituted a form of resilience not captured by spatial audits alone. This triangulation of data confirmed that computational indices required contestation by experiential and narrative evidence to avoid misleading conclusions.



Figure 3. Data processing pipeline and visualization of intermediate results

Index	Site	1A New Delhi	1B Karol Bagh	2A Kowloon Tong	2B Shamshuipo
Gendered Presence & Use	Gender Ratio (Female Proportion)	18.2%	3.9%	60.0%	43.8%
Spatial Security & Infrastructure	Safety Infrastructure Index	69.2%	18.8%	96.9%	83.9%
	Street Activity Density	34.2%	67.1%	17.2%	37.4%
Colonial Urban Morphology	Historical Planning Density Index	11.7%	25.1%	30.2%	41.6%
	Essential Service Accessibility Score	86.8%	60.5%	82.7%	73.9%
Public Perception & Social Emotions	Gender Safety Perception Negativity Score	24.6%	39.3%	16.5%	13.1%
	Social Action Discussion Frequency	28.6%	17.3%	43.2%	31.7%

Figure 4. Comparative Index Results for Case Study Sites

5. Computational Constraints & Mitigation

This study identified critical constraints inherent to the pipeline, outlining a systematic roadmap for mitigation in the next phase and serving as a guide for future researchers.

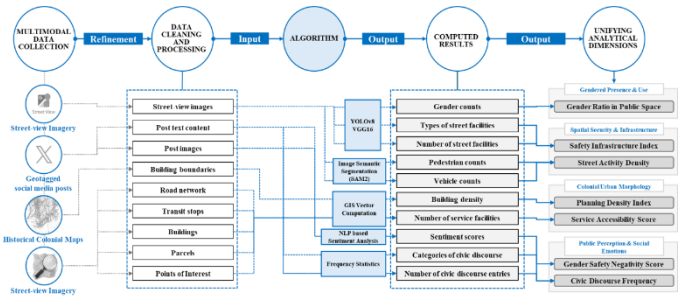


Figure 5. Final proposed data flowchart to be cross validated with human adjudication and qualitative triangulation in the next phase.

Table 2. Strategies for Mitigating Methodological Challenges

Constraints	Mitigation	Rationale / Method
Data Scarcity & Voids (GSV, Social)	Data Diversification & Ground-Truthing	Supplement API data with on-site qualitative fieldwork (photo-surveys, interviews) and participatory GIS (PGIS) mapping with local communities.
Algorithmic Bias (CV - Gender)	Re-training & Re-framing	Fine-tune models on locally sourced, diverse image datasets. Shift from binary "gender classification" to "presence counting" to avoid harmful classification.
Algorithmic Bias (NLP - Sentiment)	Custom Model Training (BERT)	Develop a custom, context-aware sentiment model (e.g., fine-tuning a BERT model) trained on locally annotated social media text from HK and Delhi.
Temporal Mismatch	Temporal Filtering & Longitudinal Analysis	Apply strict date filters to all datasets (e.g., only use data from 2020-2023). Where possible, collect data annually to build a longitudinal, rather than static, model.
Epistemic Mismatch (e.g., CPTED)	Co-designing Metrics (Decolonial AI)	Use workshops and qualitative interviews before analysis to co-design safety indices with community members, defining positive/negative features based on lived experience.
Standardisation Obscurity	Contextualised Indices	Report standardised scores alongside raw data (e.g., "50% [N=200]"). Develop composite indices that retain, rather than erase, the underlying data's complexity.
Historical Data Inaccuracy	Critical Cartography	Treat historical maps as qualitative data (texts) revealing colonial intent, rather than as precise geometric baselines. Used for historical context, not for 1:1 spatial comparison.

6. Contradictions & Conclusions

The contradictions uncovered by the audit highlight a broader methodological insight: AI-driven spatial metrics reproduce colonial logics of legibility, order, and surveillance, misinterpreting gendered safety in postcolonial contexts.

The Kowloon Tong Paradox: The model interpreted the high visibility of South Asian domestic workers as evidence of gender equity. However, these women occupy the space as underprivileged labourers, not as agents of public belonging. This reveals how computer vision collapses presence into empowerment, erasing labour hierarchies.

The Lutyen's Delhi Discrepancy: New Delhi scored high on infrastructural safety but low on female presence and sentiment. Its elite aesthetic (wide boulevards, gated compounds, restricted access) creates algorithmically "ideal" but socially exclusionary landscapes.

Failures of Eurocentric CPTED Metrics: The model rewarded empty streets, open sightlines, lack of informal commerce, and "cleanliness" in New Delhi; yet these correspond to elite, surveilled, exclusionary spaces (Jeffery 1971; Skogan 1990). However, lived safety articulated via social emotions, corresponded more closely to vibrant communal presence (Koskela & Pain 2000), which New Delhi failed in.

High-Density "Unsafe" Districts as Actually Safe: Sham Shui Po's high density and "messiness" were flagged as unsafe by the algorithm but interpreted positively by local communities: aligning with feminist and affective geographies emphasising eyes

on the street, familiarity, and communal presence (Jacobs, 1961).

Rooted in decolonial geography, this research synthesised computational social science, geographic information science, and critical urban theory to interrogate how the spatial designs of Garden Cities in Hong Kong and New Delhi reproduced colonial patterns of gendered exclusion. The empirical findings complicated initial hypotheses by revealing profound tensions between algorithmic and lived spatial readings. Specifically, the data demonstrated how computational metrics systematically misinterpreted social realities, such as conflating racialised labour mobility in Hong Kong with equitable access and misreading elite spatial exclusion in New Delhi as mere lack of occupancy, thereby exposing the epistemological limitations of importing uncritical quantitative indices into postcolonial urban contexts.

The analysis further established that high-activity districts, which algorithmic audits frequently flagged as 'disorderly', correlated with higher perceived safety among women in both cities. This directly contravened Eurocentric CPTED principles, which traditionally equated "messiness" as risks, instead indicating that street-level vitality and informal communal oversight were significant for generating a sense of safety, more than a mere presence of prescribed infrastructure. Consequently, the pronounced dissonance between quantitative indices and qualitative experience was reframed not as algorithmic failure, but as a central analytical insight, revealing the enduring coloniality embedded within standard urban assessment tools.

As a feasibility study, this investigation ultimately advocated for a decolonial computational approach to integrate lived experience and local knowledge with machine learning. Acknowledging the reliability challenges of using visual cues for gender classification in diverse cultural contexts, the final proposed framework explicitly mandates that algorithmic outputs be cross-validated with human adjudication and qualitative triangulation in the subsequent phase. Furthermore, to mitigate the temporal staticity of street-view imagery which constrains the Street Activity Density index, future research will integrate time-stamped geotagged social media or Location-Based Service (LBS) data as dynamic proxies for street vitality. Rather than discarding AI audits, the research repositioned it from an authoritative assessor to a critical muckracker: a methodological pathway for generating post-gender, intersectional, and historically grounded metrics appropriate for the complex realities of postcolonial contexts.

Acknowledgements

We acknowledge the Computational Urban Design Lab for their valuable help regarding street view image processing algorithms and essential service accessibility calculations. Special thanks go to Yue Zhong for her extensive support with social media sentiment analysis.

References

Berney, J. (2019). One woman's campaign: Stella Benson and the regulation of prostitution in 1930s colonial Hong Kong. Women's History Review, 28(6), 933–949.

Bigon, L., & Katz, Y. (Eds.). (2014). Garden cities and colonial planning: Transnationality and urban ideas in Africa and Palestine. Manchester University Press.

Boquet, Y. (2016). From Paris and Beijing to Washington and Brasilia: The Grand Design of Capital Cities and the Early Plans for Quezon City. Philippine Studies: Historical and Ethnographic Viewpoints, 64(1). https://doi.org/10.13185/2244-1638.4119

Bottazzi, R. (2024). Expanded Design. Technophany, 2(1).

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 77–91.

Carroll, J. M. (2009). A National Custom: Debating Female Servitude in Late Nineteenth-Century Hong Kong. Modern Asian Studies, 43(6), 1463–1493.

Chan, A. K. (2012). From 'civilising the young' to a 'dead-end job': Gender, teaching, and the politics of colonial rule in Hong Kong (1841–1970). History of Education, 41(4), 495–514.

Chu, C. L. (2022). Building Colonial Hong Kong: Speculative Development and Segregation in the City. Routledge.

Du, H. (2023). An evaluation of women's social status in colonial Hong Kong with a feminist lens: Case studies. Butler Journal of Undergraduate Research, 9(1), Article 3.

Foucault, M. (1977). Discipline and punish: The birth of the prison (A. Sheridan, Trans.). Pantheon Books. (Original work published 1975)

Hayden, D. (1997). The Power of Place: Urban Landscapes as Public History. MIT Press.

Home, R. (1997). Of Planting and Planning: The making of British colonial cities. Routledge & CRC Press.

Jeffery, C. R. (1971). Crime Prevention Through Environmental Design. American Behavioral Scientist, 14(4), 598–598. https://doi.org/10.1177/000276427101400409

King, A. D. (2012). Colonial Urban Development: Culture, Social Power and Environment. Taylor and Francis.

Lefebvre, H. (1996). Writings on cities (E. Kofman & E. Lebas, Eds. & Trans.). Blackwell.

Li, J. (2025). Housing colonialism: Regulating low density communities in 1950s Hong Kong. Journal of Historical Geography, 89, 137–147.

O'Neill, J. (2017, September 7–9). A garden city in Southeast Asia [Paper presentation]. Design History Society Annual Conference 2017: Making and Unmaking the Environment, Oslo, Norway.

Pain, R. (2001). Gender, Race, Age and Fear in the City. Urban Studies, 38(5–6), 899–913.

Pan, F., Huang, X., Bi, Y., Gao, Y., Ye, Y., & Wang, H. (2025). From Tools to Partners: How Large Language Models Are Transforming Urban Planning

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, And Justice.

Skogan, W. G. (with Internet Archive). (1990). Disorder and decline: Crime and the spiral of decay in American neighborhoods. New York : Free Press ; Toronto : Colleir Macmillan Canada. http://archive.org/details/disorderdeclinec00skog

Tai Kwun: Centre for Heritage and Arts. (n.d.). Sex workers in segregated spaces [Section 2.0 of Gender & Space].

Valentine, G. (1989). The Geography of Women's Fear. Area, 21(4), 385–390.

Wu, C., Ye, Y., Gao, F., & Ye, X. (2022). Using street view images to examine the association between human perceptions of locale and urban vitality in Shenzhen, China. Sustainable Cities and Society, 88, 104291. https://doi.org/10.1016/j.scs.2022.104291

Yu, M., & Wong, K. (2013). Tale of Shek Tong Tsui West: Curating an exhibition for Hong Kong Museum of History. Retrieved from City University of Hong Kong, CityU Institutional Repository.

Zhand, S., Varoudis, T., & Karimi, K. (2024). Decoding Tehran. In N. Charalambous, C. Psathiti, & I. Geddes (Eds), Space Syntax Symposium 14 (pp. 2037–2064). Gruppo editoriale Tab S.r.l. https://doi.org/10.36158/979125669032992

A Geometry-Based Computational Assessment of Parcel-Level Redevelopment Feasibility in Seoul

Sanghoon Byun[1], Dongjun Park[2] and Bumjoon Kang[3]
[1,2,3]*Seoul National University, Laus. (Lab. Architectural&Urban Space)*
[1]*bsh960flash@snu.ac.kr , 0009-0009-0741-1973*
[2]*gbjun7333@snu.ac.kr, 0009-0006-0959-0002*
[3]*bumjoon@snu.ac.kr, 0000-0002-8551-7652*

Abstract. Seoul's urban fabric is characterized by highly fragmented and historically inefficient parcel layouts that constrain redevelopment. In response, the city has promoted Small Residential Redevelopment Models (SRRMs) as an institutional strategy. However, systematic tools for assessing the geometric feasibility of these parcels remain underdeveloped, particularly those capable of integrating complex local building codes and economic land value standards. This study introduces an automated, geometry-based computational framework to evaluate parcel efficiency through a novel metric termed the Parcel Layout Efficiency (PLE) Index. The model quantifies four geometric indices—Shape, Topography, Buildable Area, and Road Access—each weighted according to user-defined parameters. Applied to more than 42,000 urban blocks across Seoul, the analysis reveals substantial quantitative differences between historically unplanned and modern planned areas, thereby validating the model's analytical robustness. This research provides a foundational framework for pre-project feasibility analysis, offering a scalable tool for data-driven evaluation of urban redevelopment potential and contributing to the advancement of computational methods in urban morphology and design decision-making.

Keywords. Geometry-Based Automation, Parcel Layout Evaluation, Feasibility Test, Urban Regeneration, Computational Design, Small Residential Redevelopment, Seoul, Rhino-Grasshopper

1. Introduction

1.1. BACKGROUND AND PROBLEM STATEMENT

Seoul's urban fabric is fragmented by historic organic developments, yielding irregular or inefficient parcel layouts, such as landlocked or flag-shaped parcels. These parcels are a major factor hindering development. In most cases, a development on such a single, irregular parcel may not be acceptable (Korean Building Act, Article 44). It is significantly inefficient to develop such parcels in terms of the systematic land use

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 717-726.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

management (Lee et al., 2009)., ultimately detracting from the parcel's value (Villanueva & Colombo, 2021).

To address this problem, "Small-scale Residential Redevelopment Models (SRRMs)" are being institutionally encouraged in Seoul (Seoul Metropolitan Government Ordinance on Urban and Residential Environment Maintenance; KRIHS, 2023). However, existing tools are inadequate for evaluating the parcel layout feasibility of these redevelopment project sites. There is a lack of objective indices that integrally reflect complex local building regulations and economic land value standards (Official Standard Land Price).

1.2. RESEARCH OBJECTIVES AND SCOPE

The objective of this paper is to develop an automated, geometry-based model, the Parcel Layout Efficiency (PLE) index, to evaluate the relative efficiency levels of these parcel layouts to apply SRRM projects. The research question is: "Can we computationally model parcel efficiency by integrating geometric constraints with Korea's specific legal (building regulations) and economic (official land appraisal) standards?"

To this end, we propose a computational framework developed using the Rhino Python API within the Grasshopper environments, linked with GIS. This framework focuses on evaluating the geometric aspects of parcel layouts, automating the assessment of four core indices—Shape, Topography, Buildable Area, and Road Access—which reflect the standards of the Official Standard Land Price and building regulations.

This study is a preparatory step toward the ultimate goal of reviewing the feasibility of specific "Small-scale Residential Redevelopment Model" projects before implementation. Therefore, prior to detailed verification of specific project models, we conduct a foundational analysis by applying the developed PLE model to over 42,000 blocks in Seoul to identify macro-trends. Specifically, we validate the model's core effectiveness by comparing the PLE indices of 'historically unplanned residential areas' and modern 'planned residential areas', utilizing the classifications from the Seoul Research Institute's (2009) <Study on the Urban Form of Seoul> report.

2. Related Work

2.1. PARCEL SHAPE AND LAND VALUE

The physical characteristics of parcels, especially their shape, are a key factor in determining land's economic value. Villanueva & Colombo (2021) argued that the geometric characteristics of parcels directly affect land use efficiency and development potential, which is ultimately reflected in land value.

Korea's Official Standard Land Price (OSLP) appraisal system is a representative case that institutionalizes this relationship. According to the OSLP survey and evaluation standards (MOLIT, 2021), parcel shape is reflected in land price calculation as one of the 'lot conditions' and is classified into eight types. Studies (e.g., Park, 2023) confirm that in Korean appraisal practice, regular shapes are systematically valued higher than irregular ones. Ham & Kim (2024) also noted this practice, pointing out

issues of subjectivity and proposing deep-learning models for classification.

This valuation perspective aligns with urban morphology research (Kropf, 1996), where landlocked or irregular parcels lead to construction constraints (Colwell & Scheu, 1998), causing land use inefficiency and value depreciation (Lin, Su & Li, 2023). This aligns perfectly with the concept of 'inefficient parcel layout' defined in the `Introduction`, highlighting the need for quantitative evaluation.

2.2. LIMITATIONS OF PRIOR WORK AND THIS STUDY'S CONTRIBUTION

Existing research on quantifying parcel layout efficiency shows three clear limitations in evaluating the parcel layout feasibility for the "Small-scale Residential Redevelopment Models" mentioned in the `Introduction`.

First, the content of the indices fails to reflect the real-world development context. Prior studies have relied on simple 2D Shape Indices (Demetriou et al., 2013; Arslan, 2021), failing to incorporate the physical and institutional context of building regulations or road access. This leads to structural limitations in assessing a parcel's actual value, such as rating a geometrically 'efficient' but practically unbuildable parcel as 'superior' (Arslan, 2021).

Second, the analysis scale is disconnected from policy reality. Existing research (Zhang et al., 2023) has been confined to fixed units of 'individual parcels' (Hong & Lee, 2022) or the 'entire block' (Li & Quan, 2024). This leaves a gap in analytical frameworks capable of measuring the efficiency of 'Groups of Parcels', which are flexibly formed based on real-world needs in projects like "Small-scale Residential Redevelopment Models".

Third, the evaluation method is stuck at an 'absolute' standard. Existing indices (Demetriou et al., 2013) attempt to measure absolute 'excellence' applied uniformly to all parcels, such as proximity to a predefined 'ideal shape'. However, parcel efficiency should be approached from the perspective of 'relative efficiency'—evaluating how efficiently a parcel is laid out compared to the maximum potential of its containing block, not against an 'ideal shape'. Such a relative comparison model is absent in prior research.

These limitations clearly show the need for an index that (1) reflects the real-world institutional context (value, regulations), (2) aligns with the policy scale (parcel groups), and (3) can evaluate 'relative efficiency' against the block's potential. This study proposes the PLE index to overcome these three gaps.

3. Method

This study develops the Parcel Layout Efficiency (PLE) evaluation system, which automates geometric evaluation, to quantitatively measure the 'inefficient parcel layout' (i.e., layouts that detract from parcel value) defined in the `Introduction`.

This system is based on the computational framework proposed in the `Introduction`. That is, it links a framework developed using the Rhino Python API in a Grasshopper environment with GIS, focusing on the geometric aspects of parcel layouts. It provides a framework that automatically evaluates efficiency across various

scales, from individual parcels to entire blocks, based on four core indices (Shape, Topography, Buildable Area, Road Access) that combine Seoul's GIS data with Korean building regulations and Official Standard Land Price standards.

3.1. EVALUATION FRAMEWORK OVERVIEW

The entire workflow of the PLE evaluation system proposed in this study consists of four stages, as shown in Figure 1

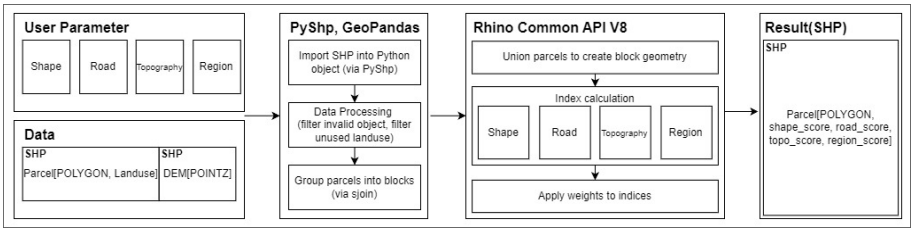


Figure 1. Flowchart of the PLE evaluation system

Rationales for the Four Indices in the PLE Framework

The computational logic of each index is grounded in four reference frameworks: (1) Korea's Official Standard Land Price appraisal criteria (MOLIT, 2025), (2) the Korean Building Act and its enforcement decree, (3) geometric shape parameters from prior research (Demetriou et al., 2013), and (4) tacit design knowledge from architectural practice. These text-based legal and economic standards were translated into geometric logic to reflect real-world development context.

1. Shape Index: The three sub-indices of the Shape Index reflect concepts from both the appraisal criteria and prior research. It implements the "available area ratio" used in the appraisal standard to identify "irregular shapes." It also adapts the Compactness parameter from Demetriou et al. (2013). Finally, it combines the appraisal standard's "regular shape" definition (aspect ratio near 1:1.1) with Demetriou et al.'s Boundary Points parameter (four vertices as optimal).

2. Topography Index: This index quantifies the slope by following the topography classification in the MOLIT (2025) guidelines. In the Korean appraisal system, land is categorized into "Flat," "Gentle Slope," or "Steep Slope," as these factors directly determine construction costs and earthwork feasibility. The model calculates the average gradient of the parcel to reflect these physical and economic limitations.

3. Road Access Index: This index integrates the "Road Contact" criteria from the appraisal guidelines (MOLIT, 2025) and Article 44 (Relationship between Building Site and Road) of the Building Act. It evaluates both the economic value derived from road width (e.g., Broad, Medium, or Narrow roads) and the legal requirement for development permission, which necessitates a minimum 2-meter contact with a road.

4. Buildable Area Index: Rather than simple gross area, this index calculates the net developable volume by parameterizing Article 46 (Designation of the Building Line) and Article 58 (Clearance within the Site) of the Building Act. The algorithm accounts for mandatory setbacks from the "Building Line" and "Adjacent Boundary Lines," which significantly reduce the usable area in fragmented or narrow parcels, thereby

reflecting the site's practical economic capacity.

3.2. FOUR CORE INDICES FOR PLE

The PLE index independently evaluates the efficiency of parcel layouts from four perspectives(1) Shape, (2) Topography, (3) Buildable Area, and (4) Road Access and calculates a weighted sum. Each index is normalized to a value between 0 and 1.

3.2.1. Shape Index, S_{shape}

Shape Index (S_{shape}) evaluates 'relative efficiency' as the ratio of the average shape score of individual parcels to the shape score of the entire block (undivided stat.

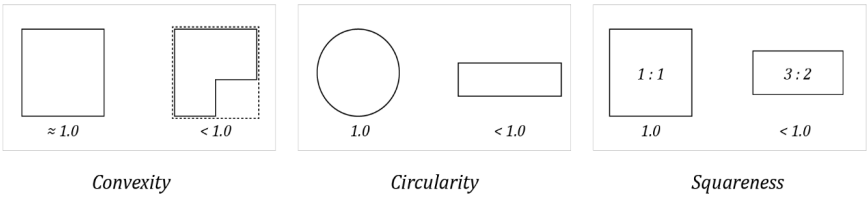


Figure 2. Example of Shape Index sub-elements.

$$S_{shape} = \frac{\frac{1}{n}\sum_{i=1}^{n} S_{shape}(i)}{S_{shape}(block)}$$

Where:

- n: Total number of parcels within the block
- $S_{shape}(i)$: Shape score of individual parcel i (weighted sum of sub-indices below)
- $S_{shape}(block)$: Shape score of the entire block (undivided state) (weighted sum of sub-indices below))

Each shape score ($S_{shape}(i)$ or $S_{shape}(block)$) is calculated as a weighted sum of three sub-indices:

$$S_{shape}(entity) = w_{conv} \cdot S_{conv}(entity) + w_{circ} \cdot S_{circ}(entity) + w_{sq} \cdot S_{sq}(entity)$$

(where, $w_{conv} + w_{circ} + w_{sq} = 1$)

- S_{conv}: Convexity score(Area/Convex Hull Area)
- S_{circ}: Circularity score ($4\pi \cdot \text{Area}/\text{Perimeter}^2$)
- S_{sq}: Squareness score (aspect ratio and vertex penalty)

3.2.2. Topography Index, S_{topo}

The Topography Index ((S_{topo})) evaluates how much lower the average relief of subdivided parcels is compared to the relief of the entire block.

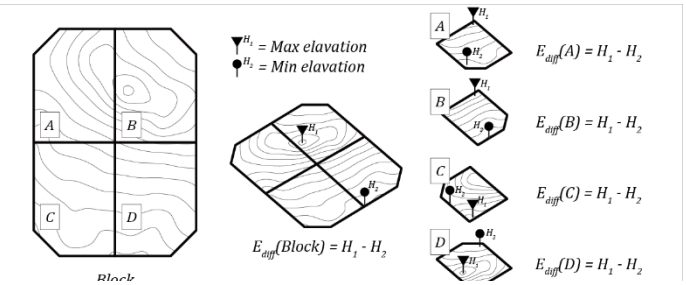


Figure 3. Conceptual diagram of the Topography Index (S_{topo}) evaluation.

$$S_{topo} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} E_{diff}(i)}{E_{diff}(block)}$$

Where:

- n: Total number of parcels within the block
- $E_{diff}(i)$: Relief of individual parcel i(max elevation - min elevation within parcel)
- $\frac{1}{n}\sum_{i=1}^{n} E_{diff}(i)$: Average relief of individual parcels within the block
- $E_{diff}(block)$: Relief of the entire block (max elevation - min elevation within the block)

3.2.3. Buildable Region Index, S_{region}

The Buildable Region Index (S_region) evaluates 'relative efficiency' as the ratio of the 'sum' of buildable areas of currently subdivided parcels to the total buildable area of the entire block (undivided state).

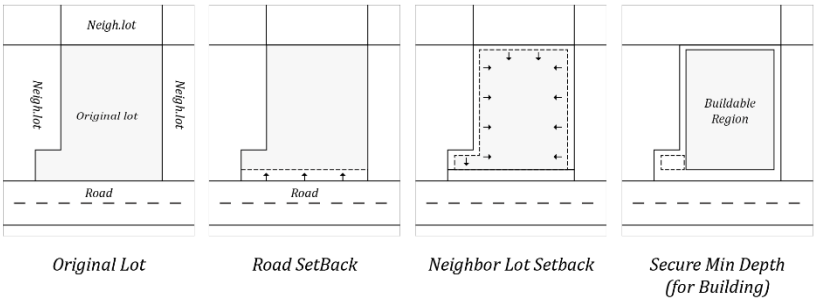


Figure 4. Example of Buildable Area calculation process.

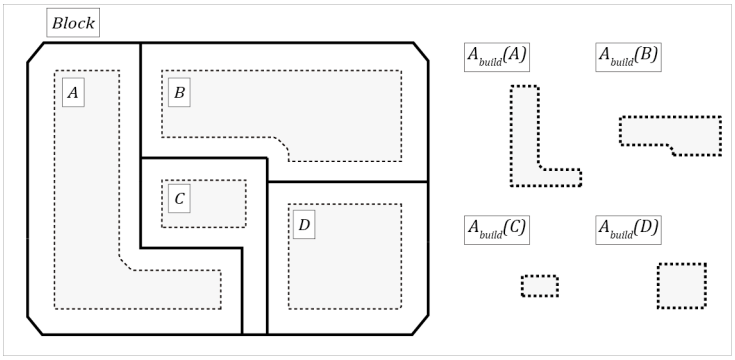


Figure 5. Conceptual diagram of the Buildable Region Index.

$$S_{region} = \frac{\sum_{i=1}^{n} A_{build}(i)}{A_{build}(block)}$$

Where:

- n: Total number of parcels within the block
- $A_{build}(i)$: Buildable area of individual parcel i (area after applying regulatory setbacks and minimum width requirements)
- $\sum_{i=1}^{n} A_{build}(i)$: Sum of buildable areas of all individual parcels within the block
- $A_{build}(block)$: Total buildable area of the entire block (undivided state)

3.2.4. Road Access Index, S_{road}

The Road Access Index (S_{road}) evaluates 'relative efficiency' as the ratio of the average road access score of currently subdivided parcels to the maximum possible road access score the block can achieve.

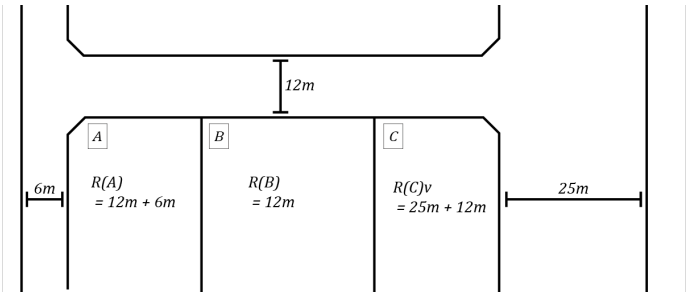


Figure 6. Conceptual diagram of the Road Access Index (S_{road}) evaluation.

$$S_{road} = \frac{\frac{1}{n}\sum_{i=1}^{n} R(i)}{R_{block}}$$

Where:

- n: Total number of parcels within the block
- $R(i)$: Road access score of individual parcels i (multiplier value assigned according to 12 types in the land price ratio table)
- $\frac{1}{n}\sum_{i=1}^{n} R(i)$: Average road access score of individual parcels within the block
- R_{block}: Maximum road access score the entire block (undivided state) can achieve

3.3. OVERALL PLE SCORE CALCULATION AND USER CUSTOMIZATION

The final Parcel Layout Efficiency (PLE) index is calculated through a 'weighted sum' of the four core indices (S).

$$PLE = (w_{shape} \cdot S_{shape}) + (w_{topo} \cdot S_{topo}) + (w_{region} \cdot S_{region}) + (w_{road} \cdot S_{road})$$

- $w_{shape}, w_{topo}, w_{region}, w_{road}$: User-defined weights for each index (constraint, $w_{shape} + w_{topo} + w_{region} + w_{road} = 1$)
- $S_{shape}, S_{topo}, S_{region}, S_{road}$: Four block-level relative efficiency indices calculated above (values between 0-1)

Importantly, as evident from the formula, the PLE index evaluates 'Relative Efficiency'. It is not an absolute value index measuring the worth of subdivided parcel outcomes (V_i), but rather a ratio index assessing how efficiently the current 'subdivision action' has been performed relative to the block's total potential (V_{block}) before subdivision.

The PLE index ranges between 0 and 1, where values closer to 1 indicate that the current parcel layout maximally utilizes the block's potential value, while values closer to 0 indicate severe layout inefficiency and high need for integrated development (re-subdivision).

3.4. APPLICATION TO ANALYSIS SITES

The developed PLE evaluation system was applied to 902,001 target parcels across Seoul. These parcels were grouped into a total of 42,000 blocks. We selected two residential zones (Type 1 and Type 2 General Residential), which are the most common in Seoul. Within these zones, eight 'historically unplanned residential areas' and eight 'planned residential areas' were selected to compare the PLE index values by land use type and development history.

4. Result

Of the 42,000 blocks initially identified, certain cases were excluded from the analysis. Single-parcel blocks (where PLE scores inherently converge to 1) and large-scale apartment complexes (inadequately represented in the data) were removed. This filtering process resulted in a final dataset of 27,491 blocks.

For this foundational analysis, all four indices were assigned equal weights

($w_{shape} = w_{topo} = w_{region} = w_{road} = 0.25$).

Furthermore, as this study is intended to be applicable to Small-scale Residential Redevelopment Models (SRM), the user-configurable parameters, including building regulations (e.g., setbacks), were set based on standards for residential use.

The analysis of these blocks confirmed a significant quantitative difference between historically unplanned and modern planned areas, validating the model's effectiveness. High-ranked blocks (high PLE) were concentrated in planned development districts (e.g., Gangnam-gu, Seocho-gu), while low-ranked blocks were predominantly located in historic urban cores and hillside residential areas (e.g., Gwanak-gu, Jongno-gu).

This divergence is not just in the final PLE score but is evident in the individual indices. For example, the Shape Index (S_{shape}) clearly visualizes this difference. Figure 7 (a) shows a planned area in Gangnam-gu, characterized by regular, rectilinear parcels that result in high shape scores (dark red). In contrast, Figure 7 (b) shows a historically unplanned area in Jongno-gu, where irregular and fragmented parcels result in very low shape scores (light/white).



Figure 7. Comparison of Shape Index (S_{shape}) scores. (a) High scores in a planned area (Gangnam-gu). (b) Low scores in an unplanned area (Jongno-gu).

While some blocks had uniformly low scores, many low-PLE blocks were characterized by critical deficiencies in one or two specific indices (e.g., S_{topo} or S_{road}), suggesting targeted areas for improvement rather than uniform poor performance across all metrics.

5. Conclusion

This study proposed an automated PLE evaluation framework, using Rhino/Grasshopper and GIS, to assess parcel layout efficiency in Seoul. It integrates Korean legal (regulations) and economic (Official Standard Land Price) standards by evaluating four indices (Shape, Topography, Buildable Area, Road Access).

Applied to 27,491 blocks, the analysis revealed a wide PLE score range (0.917 to 0.052). This confirmed a significant efficiency gap between planned (high-ranked) and historic/hillside (low-ranked) areas, with low scores often driven by specific index failures (e.g., S_{topo} or S_{road}).

This study provides a foundational analysis for identifying macro-trends in parcel layout efficiency across Seoul. The framework may serve as a tool for evaluating alternative configurations in future re-parceling simulations. Additionally, further

validation is needed to examine statistical correlations between PLE scores and actual land prices or development density such as floor area ratio (FAR).

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2025-00556250).

Attribution Statement

Gemini (Google, 2024) was used for manuscript refinement and translation.

References

Google. (2024). Gemini [Large Language Model]. https://gemini.google.com

Arslan, F., Degirmenci, H., Akkaya, S., & Jürgenson, E. (2021). A new approach to measure parcel shapes for land consolidation. Kahramanmaraş Sütçü İmam Üniversitesi Tarım ve Doğa Dergisi, 24(5), 1059-1067.

Colwell, P., & Scheu, T. (1998). Public land use constraints: Lot and house configuration. Journal of Real Estate Research, 16(2), 201-218.

Demetriou, D., See, L., & Stillwell, J. (2013). A parcel shape index for use in land consolidation planning. Transactions in GIS, 17(6), 861-882.

Ham, H., & Kim, K. (2024). A Study on the Classification of Land Parcel Shape Using Deep Learning. Journal of the Korean Association of Geographic Information Studies, 27(1), 1-13. [In Korean]

Hong, S., & Lee, Y. (2022). Typification of irregular shaped land parcels using machine learning. Journal of the Architectural Institute of Korea, 38(3), 189-198. [In Korean]

Korea Research Institute for Human Settlements (KRIHS). (2023). A Plan for Linking Small-scale Maintenance and Regeneration to Resolve Blind Spots in Old Residential Area Maintenance (Kukto Policy Brief, 926). [In Korean]

Kropf, K. (1996). Urban tissue and the character of towns. Urban Design International, 1(3), 247-263.

Lee, I., Lim, S., & Kim, C. (2009). Analysis of the effects of parcel shape on development density [필지형상이 개발밀도에 미치는 영향 분석]. Journal of the Urban Design Institute of Korea, 10(4), 151-162. [In Korean]

Li, N., & Quan, S. J. (2024). Discovering urban block typologies in Seoul: Combining planning knowledge and unsupervised machine learning. Cities, 150, 104988.

Lin, S. W., Su, H. J., & Li, H. R. (2023). The Relationship of Land Assembly and Prices of Cost Equivalent Land in the Equalization of Land Rights Act the Case Study of Taiwan. International Review for Spatial Planning and Sustainable Development, 11(2), 222-239.

Ministry of Land, Infrastructure and Transport (MOLIT). (2021). Standards for Survey and Appraisal of Official Standard Land Price. [In Korean]

Park, H. (2023). A Study on the Effect of Land Parcel Shape on Land Price in Appraisal Practice. Journal of the Korea Real Estate Analysts Association, 29(1), 45-58. [In Korean]

Seoul Research Institute. (2009). A study on urban form of Seoul [서울의 도시형태 연구]. Seoul Development Institute. [In Korean]

Villanueva, M. P., & Colombo, S. (2021). Impact of parcel fragmentation on the calculation of the real estate value of land belonging to farms. New Medit: Mediterranean Journal of Economics, Agriculture and Environment, 20(1).

Zhang, P., Ghosh, D., & Park, S. (2023). Spatial measures and methods in sustainable urban morphology: A systematic review. Landscape and Urban Planning, 237, 104776.

Building Act, Law No. 19745 (2024). [건축법]. [In Korean]

Between Numbers and Narratives

Integrating Qualitative and Quantitative Methods for Typological Analysis of Dutch Cross Buildings

Jess Chang[1], Casper van Engelenburg[2], Marija Mateljan[3], Seyran Khademi[4] and Georg Vrachliotis[5]
[1,2,3,4,5] *Delft University of Technology.*
[1]*H.Chang-1@tudelft.nl, 0009-0008-6237-774X*
[2]*C.C.J.vanEngelenburg@tudelft.nl, 0000-0002-9705-9644*
[3]*M.Mateljan@tudelft.nl, 0000-0003-4081-1433*
[4]*S.Khademi@tudelft.nl, 0000-0003-4623-3689*
[5]*G.Vrachliotis@tudelft.nl, 0000-0001-7590-845X*

Abstract. Floor plans are essential sources for morphological studies in architectural history, offering insight into design thinking while opening to quantitative analysis. Yet, key contextual dimensions still rely heavily on the researcher's interpretation. This study proposes a hybrid method that integrates graph-based similarity measures with human judgment to evaluate typological similarities in floor plans. We demonstrate the approach using a dataset of Dutch cross buildings, a rarely examined building type that served as community care centres. Our findings indicate that computational algorithms can capture nuanced similarity that exceeds human perceptual precision, while also revealing outliers to serve particular historical interests. Additionally, a flow-based algorithm produced numerical results that align closely with the architect's typological interpretation based on circulation. Together, these results offer new ways to understand building types, suggesting that quantitative insights can substantiate historical arguments, prompt new research questions, and extend to other building categories shaped by distinct spatial orders and use patterns.

Keywords. Computational analysis, architectural history, floor plan, building typology, Dutch cross buildings

1. Introduction

Architectural histories must reveal ideologies, policies and societal shifts that have shaped the built environment of today (Lucas, 2016). Developing a critical work of architectural historiography requires a profound understanding of various 'lenses' through which history can be examined. These include social, colonial, postcolonial, and gender perspectives on spatial production. Due to the interrelated nature of these frameworks, architectural history is not purely descriptive but highly analytical, relying

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 727-736.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

on the careful curation of materials to construct a coherent 'narrative' (Groat and Wang, 2013). Naturally, architectural historiography favours qualitative methods, including the analysis of archival documents and drawings to examine chronological and typological development. Floor plans, particularly, are widely used as a fundamental source of knowledge. As cultural artefacts, they reflect the cultural values, social norms, and architectural practices of their time through spatial configurations, proportions, and relationships. Because these aspects are grounded in geometry and topology, floor plans can also be analysed quantitatively (Koutamanis, 2019). With the increased digitisation of archival collections (Molenda, 2018) and the availability of powerful computing systems, computer-aided quantitative analysis offers new opportunities in historical architectural research. Yet, the main challenge remains integrating quantitative insights with nuanced interpretation, to capture the particularities of historical periods, geographic contexts and building types—factors that cannot be fully expressed through numbers alone.

Therefore, this study introduces a hybrid methodological framework for historical research that combines computational methods with architectural interpretation. It addresses the following research question: To what extent can graph-based similarity measures, validated against architects' perceptual judgments, advance typological analysis in historical research? The framework enables quantitative typological analysis and visualisation of architectural floor plans, combining human insight with computational analysis through structured visual representations. It also demonstrates scalability for comparative typological studies, exemplified by an analysis of Dutch 'cross buildings'—a type of historical community care architecture in the Netherlands.

2. Background

2.1. HISTORY AND RELEVANCE OF CROSS BUILDINGS

Cross buildings ("kruisgebouwen" in Dutch) are community care centres that supported the private healthcare services of the cross associations ("kruisvereniging") in the past. These associations used existing buildings as warehouses to store and rent out medical goods since their establishment in the late 19th century. As demand grew and more cities sought their own branches, cross buildings began to be constructed within communities nationwide (Merwijk, 2004). These buildings played an important role in improving public health and helped shape the modern healthcare system (Rensman & Schoor, 2013). Despite the historical significance of cross buildings, they have received little scholarly attention, likely because many have been converted into houses or offices, making direct empirical study difficult. This limitation, however, opens space for alternative approaches—like the one presented here—that combine archival research with computational methods to support research hypotheses with quantitative evidence and encourage further typological study.

2.2. CROSS BUILDING AS A UNIQUE BUILDING TYPE

The services offered by cross associations were dependent on the needs of the local community. Despite their diverse activities, two functional groups remained consistent in their internal spatial organisation. The first group comprised a warehouse and a

cleaning room for the medical supplies. The second focused on patient care and included a waiting room, box room and doctor's room, forming the child health clinic (or "consultatiebureau" in Dutch), which offered examinations and consultations to mothers and infants. Additional clinical facilities, such as isolation rooms, bath houses, laboratories and special treatment rooms were also often included. Among these, the *isolation room* is particularly significant for our study. According to the earliest found design guideline (IJsselmuiden, 1948), isolation rooms were deliberately incorporated to temporarily separate clinic visitors carrying infectious diseases. Because "isolation" is inherently a spatial concept, the room's placement in relation to other functions raises critical questions about the overall spatial organisation. These distinctive functional and spatial characteristics contributed to the emergence of the cross building as a recognisable building type, translating principles of modern hygiene and preventive healthcare into an accessible community setting. The recurring spatial and functional patterns across multiple cross buildings further exemplify the notion of architectural typology.

In architectural discourse, typology refers to the study of recurring patterns that define classes of buildings with shared formal and functional characteristics. It involves grouping architectural works by certain inherent structural similarities (Argan, 1963). The concept of type reveals a fundamental duality in architecture: each work embodies both uniqueness and repeatability. This tension also introduces the notion of the *prototype* (Moneo, 1978), which serves as both origin and reference: a concrete example through which typological principles are tested and refined. In the context of cross buildings, such prototypes played an important role. For instance, a 1948 cross association publication presented a 'prototype' example as an ideal design, illustrating the optimal spatial composition, dimensions, and functions of cross buildings (IJsselmuiden, 1948). It provided a tangible reference for constructing multiple cross buildings while allowing local adaptations, exemplifying the balance between typological consistency and spatial uniqueness. All of these spatial and functional characteristics make cross buildings ideal subjects for computer-aided typological study. Moreover, their relatively small scale and rigid spatial logic make them suitable for testing and refining our human-machine interactive approaches.

3. Related Works

3.1. COMPUTATIONAL APPROACHES TO STUDYING HEALTHCARE TYPOLOGY

Healthcare facilities—highly regulated spaces requiring controlled flows, hygiene, and optimised staff processes—have long served as testbeds for computational methods. Early work, emerging from 1960s–70s design automation (Cross, 1977), formalised functional and spatial relationships using geometric and topological models (March & Steadman, 1971; Steadman, 1983), followed by Evidence-Based Design in the 1980s and later expansions through operations management (Benitez et al., 2019) and AI (Zhao et al., 2021). Yet much of this research, largely by engineers and mathematicians (Jamali et al., 2020), has prioritised functional efficiency over architectural and historical interpretation—a gap this study addresses by integrating computational analysis with human-centred, typological reading.

3.2. COMPUTATIONAL METHODS IN FLOOR PLAN ANALYSIS

Understanding the logic of space and how it connects to the 'social performance' of buildings has been subject to computer-aided analysis, popularised through the theory of Space Syntax (Hillier et al., 1976). Space Syntax-derived methods result in some form of spatial measurement and operate on each input *individually*. The results of these measurements are compared statistically or visually across a set of inputs. In contrast, other more recent computational approaches (e.g., Dalton & Kirsan, 2008; Engelenburg et al., 2025) measure the structural resemblance *between* floor plans: hence, measures are *pairwise*. In most such works, floor plan similarity is cast as a comparison problem over graphs, where access graphs of floor plans are compared using e.g., graph edit distance (GED) (Sanfeliu & Fu, 1983). Other similarity measures over graphs, such as shortest-path graph kernels (GKs) (Borgwardt & Kriegel, 2005), have been only sporadically used to measure floor plan similarity (Engelenburg et al., 2025). Despite their increasing use, these methods treat GED or GK as abstract similarity metrics—reporting numerical scores without examining what aspects of spatial configuration such scores actually reflect. We systematically investigate how GED and GK correspond to perceived architectural features and whether they can aid in understanding type and typology.

4. Methods

4.1. DATA COLLECTION, PREPARATION AND REPRESENTATION

We collected data through archival research, focusing on sources that document the history and design of cross buildings. Key sources include magazines (Bouw, 1970; 1967; 1958; 1956; Katholieke Gezondheidszorg, 1948), and Dutch regional and national archives. Following the initial archival research, we selected 24 cross buildings designed and constructed between the 1940s and 1960s, of which 15 buildings were filtered that contain "isolation rooms". To standardise the digital format, floor plans were scanned and redrawn: rooms and openings were manually drawn as independent graphical objects ('segmented' images). The access graphs were automatically extracted using Python. The data curation pipeline is visualised in Fig. 1.

We defined a "functional unit" for child health clinics consisting of six key rooms: hallway, waiting room, box room, changing room, doctor's room, and isolation room. Using these functions and their topology, we automatically derived subgraphs from the initial dataset. Cross buildings were named after their city.



Figure 1. Data preparation and representation in the case of Groesbeek

4.2. FLOOR PLAN ANALYSIS

Instead of measuring spatial properties of plans individually, we compute pairwise similarities by comparing their access graphs, which naturally encode spatial components and their connections. Each score reflects the degree of spatial similarity between two plans, with higher values indicating a stronger spatial match. These scores are relative metrics, meaningful only in comparison to other plan pairs—for example, comparing the scores of plan A to plans B and C reveals which plan is more similar to A. The scores are then interpreted through historical and architectural lenses of function, circulation, and type; analysed through visual inspection and diagrammatic sketching. Insights from both approaches are compared iteratively, revealing differences and similarities, as well as opportunities for mutual refinement.

4.2.1. Similarity Measures on Graphs

To measure the similarity between two graphs $G1$ and $G2$, we consider (1) a similarity based on graph edit distance (GED) (Sanfeliu & Fu, 1983) and (2) a shortest-path graph kernel (GK) (Feragen et al., 2013). The GED measures the 'effort' of transforming $G1$ into $G2$, and is given by the minimal amount of edit operations (add/delete/change a node/edge) to do the transformation. The similarity is given as $\exp(-\mathrm{GED}/(\mathrm{N1}, \mathrm{N2})$, where Ni is the number of nodes in Gi. GK computes the similarity between $G1$ and $G2$ by how similar the set of all shortest-paths are in $G1$ compared to the set of $G2$. The sum over the similarities of all pairs of shortest-paths is normalised by the maximum self-similarity of $G1$ and $G2$. GED and GK, in our formulation, do not consider the room shapes.

4.2.2. Experiments

We conducted three sequential experiments:

1) **Measuring spatial uniqueness through GED.** As a reference, we used a prototype as the starting point of the comparative study (See 2.2). Given the standardised nature of child health clinics, we hypothesised that many cross buildings would closely resemble this prototype, thus having high similarity scores with respect to others. To test this, we use GED to quantify the degree of spatial similarity between floor plans.

2) **Comparing GED with human judgment of similarity.** To interpret the numeric GED results, we evaluated the scores with an architect's qualitative perspective, focusing specifically on functional organisation, spatial layout, and circulation patterns.

3) **Circulation-based typological analysis and GK approach.** Drawing on insights from previous steps, the concept of circulation guided the next stage of analysis. By sketching over the floor plans to map visitors' movements, we identified distinct typological categories based on sequence. To evaluate whether computational analysis can reflect these perceptual distinctions, we applied GK which is suited for representing structural and sequential connectivity in graphs, and reflected on the results from a historical perspective.

5. Results

5.1. MEASURING SPATIAL UNIQUENESS THROUGH GED

Numerical results of sGED are provided in Fig. 2. Each entry in the similarity matrix amounts to the score between the floor plans of the corresponding column and row. The higher the score, the more magenta it is; the lower the score, the more light blue it is. The matrix is symmetrical due to the nature of pairwise comparison. Note that, as previously stated (Sec. 4.2), the scores should be considered relative to each other.

Interestingly, similarity scores vary widely across floor plans (0.21–0.85), and no pair is ever topologically identical (1.0), contrary to our expectation of strong uniformity. In effect, *each plan is topologically unique*. This highlights how computational methods can reveal fine-grained differences with a level of precision that manual analysis can hardly match.

We also assessed prototypicality within the set. Surprisingly, the designated Prototype is not the closest match to the others; instead, Hengelo shows the highest average similarity, making it the most typical case. The Prototype appears less representative mainly because it lacks a "changing room", which most other examples have. The colour visualisation also identifies Voorburg and Vinkeveen as the most distinct cases, suggesting them as the most unique functional units. Their divergence may reflect innovation or atypical planning decisions, marking them as strong candidates for further historical investigation.

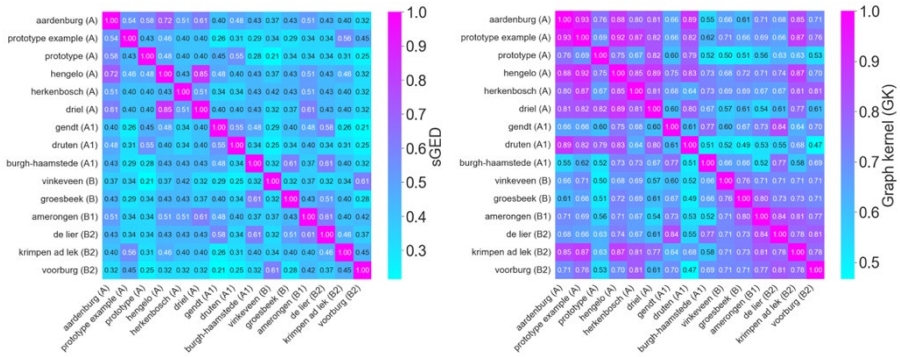


Figure 2. sGED (left) and GK (right) scores.



Figure 3. Floor plans of Hengelo, Driel and Aardenburg before and after the rotation

5.2. COMPARING GED WITH HUMAN JUDGMENT OF SIMILARITY

To understand whether the GED results accurately reflect the concept of similarity as perceived by architects, two pairs were selected for further interpretation: Hengelo-Driel (score 0.85) and Hengelo-Aardenburg (score 0.72). We noticed immediately that visual comparison of the floor plans was challenging, as the buildings' orientations differ in the drawings, even when their layouts are similar. In contrast, the GED is invariant to such transformations, and is, instead, computing the similarity based on the topology and room categories. To make a meaningful comparison, the floor plans of Driel and Aardenburg were rotated and mirrored to align with Hengelo (Fig. 3). After this adjustment, the three functional units are indeed highly similar, differing mainly in the location of the 'changing rooms'.

Through further visual inspection and interpretation, we found that despite the GED score between Hengelo and Driel, we consider Hengelo to be more similar to Aardenburg in terms of layout and circulation. In Hengelo, 'changing rooms' were located at the end of the box room, whereas in Driel, the 'changing rooms' are centrally located, creating a different circulation pattern. According to the GED calculation, however, this difference affected the score. This process highlights the distinction between human perception of spatial similarity that is sensitive to orientation, and computational measures, which evaluate similarity regardless of orientation. This difference should be carefully considered in future spatial studies that involve both human and machine actors.

5.3. CIRCULATION-BASED TYPOLOGICAL ANALYSIS AND GK APPROACH

Building on previous analyses, we observe that architects' interpretation of floor plans is intuitively tied to spatial sequence and circulation patterns of users. According to the archival study, visitors (mothers and infants) would experience the clinic in a specific order: (1) hallway (2) waiting room (3) box room (4) changing room (5) doctor's room (Etten, 1938) and (6) isolation room if infectious symptoms were detected. Considering this spatial sequence, we examined all 15 floor plans and found they can be grouped into several types based on circulation patterns: the main (A) loop and (B) branch types, along with several in-between subtypes that capture variations between them, namely A1, A2, B1 and B2 (Fig. 4).

In Type A (loop), visitors experience the clinic in a circular sequence; the isolation room sits next to the doctor's room, allowing easy separation during examinations. Type B (branch) introduces two path splits—one to the doctor's and one to the isolation room—creating a 'linear' experience. Here, the isolation room is placed apart from other spaces to limit infection risk. Several units also combine elements of both types, forming hybrid configurations.

To determine the alignment between GK's 'clustering' and the pre-determined types (A, B, etc.), we investigated GK's induced similarity matrix (Fig. 2). And, indeed, GK's clustering correlates with the architect's perceptual understanding of layouts, especially in the case of A and B: similarities between samples within each cluster (A-A, B-B) are significantly higher than those between the two (A-B, B-A). Computing GK, therefore, seems to correlate with the pre-determined types. This could be

explained by how GK computes similarity, which is predominantly based on similarities between paths, thus sequences. The results provide evidence that GK is well-suited for typological analysis. However, despite the alignment between the two main types (A, B), the intra-similarity scores of the subtypes (e.g., A1) are not always high, resulting in several outliers within the same subgroup. When examining the floor plans of these examples, the reasons behind the discrepancies between the numeric result and the actual spatial layouts are not immediately clear, representing a limitation of this study.

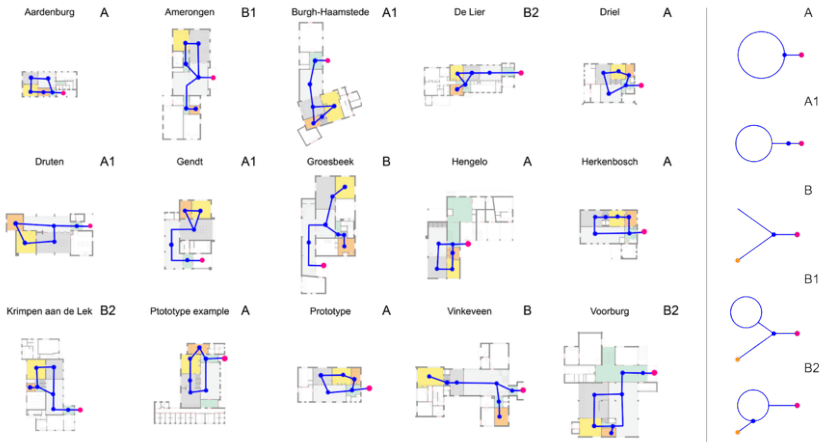


Figure 4. Typological grouping from the architect's perspective, sketched on a tablet.

6. Discussion

Through the three steps of the experiment, from GED, to human analysis, to GK, the study offered us new ways of understanding building typology. The first step revealed subtle variations among seemingly 'alike' functional units by computationally analysing all links and functions. In contrast, human understanding of type involves abstracting formal complexity in relation to social practices, with circulation being a crucial factor in identifying spatial types. This insight guided the third step, showing that GK effectively groups buildings by their circulational structure. Overall, the research shows that algorithmic analysis can enrich typological studies by expanding our definition and perception of types. Still, it should complement, not replace, architectural reading, fostering a productive dialogue between computational precision and human insights.

The numeric evidence highlights the 'uniqueness' of the core clinical function, demonstrating how local associations and architects adapted designs to local needs despite financial constraints (Huige, 2011), corroborating and extending insights from archival records. Computational analysis also helped identify exemplary cases for deeper study. For example, Vinkeveen shows low similarity to other examples in both GED and GK analysis, highlighting its originality and authenticity—qualities warranting further investigation. All of these findings complement our ongoing investigation of cross buildings.

While our framework has potential for diverse applications, the experiment is labour-intensive, from floor plan collection to data preparation, as noted by Ferrando et al. (2019). This type of research demands substantial tacit knowledge from trained architects, not only to process data, but also to interpret and link numerical results to architectural typological understanding. Although tools for automated graph extraction and semantic segmentation of floor plans have been developed within the field of computer vision, their specific focus on housing typologies makes them unsuitable for this study. Moreover, the small dataset may influence interpretation. As noted in the GK analysis (Sec. 5.3), the discrepancy between human judgment and the GK results in subtype categorisation was not apparent. We speculate that a larger dataset might reveal additional spatial patterns, enabling more accurate grouping. This illustrates how perceptions of types can be influenced by both the number of examples and algorithmic outcomes.

Future work could apply machine learning to scale up typological comparison and reveal latent spatial patterns in large datasets of historical floor plans. While our analysis used explicit graph similarity measures, data-driven models could learn higher-order links between circulation, function, and form. The goal is not to replace typological reasoning but to strengthen it, offering new ways for historians to search, compare, and interpret spatial data.

7. Conclusion

Computational analysis of historical building floor plans may appear as a purely 'technical' exercise, driven by automation, big data, and digital archives. In this paper, we go further, proposing a new approach in architectural history that integrates typological analysis and graph-based computational methods to generate meaningful spatial knowledge and foster a 'dialogue' between humans and machines. By framing typological comparison as a problem of graph similarity, our method shows how quantitative similarity measures can trigger important questions about the logic of spatial organisation. Although rooted in a study of Dutch cross buildings, this framework can be applied to other building types with rigid spatial order, such as factories, airports, courthouses and prisons. Ultimately, by integrating quantitative and qualitative approaches, this study demonstrates how computational analysis can reveal subtle spatial patterns, while human interpretation explains their architectural and socio-cultural significance, offering a versatile framework for comparative typological research across diverse geographic and temporal contexts.

Acknowledgement

Archival research, historical analysis and floor plan curation, including digitisation and retracing, were conducted by Jess. Computational programming and visualisation were conducted by Casper. Marija connected architectural and computational aspects of the study. Conceptualisation of the paper, literature review and result interpretation were conducted by Jess, Casper and Marija, who also co-wrote the manuscript. Seyran and Georg offered insights to frame and structure the project and provided feedback on the final manuscript.

References

Argan, G. C. (1963). On the Typology of Architecture. Architectural Design, December.
Benitez, G. B., Da Silveira, G. J. C., & Fogliatto, F. S. (2019). Layout Planning in Healthcare Facilities: A Systematic Review. HERD: Health Environments Research & Design Journal, 12(3), 31–44.
Borgwardt, K. M., & Kriegel, H.-P. (2005). Shortest-Path Kernels on Graphs. Proceedings of the Fifth IEEE International Conference on Data Mining, 74–81.
Cross, N. (1977). The automated architect. Pion; WorldCat.
Dalton, R. C., & Kirsan, C. (2008). Small-Graph Matching and Building Genotypes. Environment and Planning B: Planning and Design, 35(5), 810–830.
Engelenburg, C. van, Gemert, J. van, & Khademi, S. (2025, September 3). LayoutGKN: Graph Similarity Learning of Floor Plans. BMVC.
Etten, O. (Director). (1938). Het Zonnelied
Feragen, A., Kasenburg, N., Petersen, J., de Bruijne, M., & Borgwardt, K. (2013). Scalable kernels for graphs with continuous attributes. Advances in Neural Information Processing Systems, 26.
Ferrando, C., Dalmasso, N., Mai, J., & Cardoso Llach, D. (2019). Architectural Distant Reading: Using Machine Learning to Identify Typological Traits Across Multiple Buildings. 204–217.
Groat, L. N., & Wang, D. (2013). Architectural Research Methods (2nd ed). John Wiley & Sons, Incorporated.
Hillier, B., Leaman, A., Stansall, P., & Bedford, M. (1976). Space Syntax. Environment and Planning B: Planning and Design, 3(2), 147–185.
Huige, J. J. C. (2011). Van kruiswerk tot thuiszorg: De moeizame strijd voor erkenning van een boeiende maar complexe werksoort in de periode 1946-1990. Landelijke Stichting Beheer Kruiswerk.
IJsselmuiden, H. G. M. (1948). Het Bouwplan. Katholieke Gezondheidszorg, 29–33.
Jamali, N., Leung, R. K., & Verderber, S. (2020). A review of computerized hospital layout modelling techniques and their ethical implications. Frontiers of Architectural Research, 9(3), 498–513.
Koutamanis, A. (2019). Building Information - Representation and Management: Fundamentals and Principles. TU Delft OPEN Publishing.
Lucas, R. (2016). Research methods for architecture. Laurence King Publishing.
March, L., & Steadman, P. (1971). The Geometry of Environment: An Introduction to Spatial Organization in Design. WorldCat.
Merwijk, T. van. (2004). Kruisgebouwen: Categoriaal Onderzoek Wederopbouw 1940-1965.
Molenda, A. (2018). Between Creators and Keepers. How HNI Builds its Digital Archive.
Moneo, R. (1978). On typology. Oppositions, 13, 22–45.
Rensman, E., & Schoor, L. van den. (2013). Zuster over de vloer: Thuiszorg, toen en nu. Stichting Else.
Sanfeliu, A., & Fu, K.-S. (1983). A Distance Measure Between Attributed Relational Graphs for Pattern Recognition. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13(3), 353–362.
Steadman, P. (1983). Architectural morphology: An introduction to the geometry of building plans. Pion; WorldCat.
Zhao, C., Yang, J., Xiong, W., & Li, J. (2021). Two Generative Design Methods of Hospital Operating Department Layouts Based on Healthcare Systematic Layout Planning and Generative Adversarial Network. Journal of Shanghai Jiaotong University (Science), 26(1), 103–115.

Materialising Latent Hybridity

Interpolating the Natural and the Artificial for Post-Digital Urban Futures

Sheng-Yang Huang[1], Enriqueta Llabres Valls[2], Xiao Wang[3], Yongchun Wang[4], Hengyu Gong[5], Junyi Wu[6], Yanlin Bao[7]
[1,2,3,4,5,6,7]*Bartlett School of Architecture, University College London*
[1]*william.huang@ucl.ac.uk, 0000-0002-2748-5031*
[2]*e.llabress@ucl.ac.uk, 0000-0002-5593-8667*
[3]*x-wang.24@alumni.ucl.ac.uk , 0009-0000-3407-7047*
[4]*yongchun.wang.24@alumni.ucl.ac.uk, 0009-0001-3171-6728*
[5]*ucbvhgo@ucl.ac.uk, 0009-0009-6227-1598*
[6]*junyi.wu24@ucl.ac.uk, 0009-0006-8506-5226*
[7]*yanlin.bao@uclmail.net, 0009-0009-2679-6135*

Abstract. This study presents a computational design workflow that investigates non-dual relational hybridity between natural and artificial systems. Two contrasting datasets—3D scans of herbaceous plants and images of Brutalist façades—were curated to define the morphological spectrum of this continuum. A fine-tuned Stable Diffusion 1.5 model, using its internal CLIP encoder and VAE decoder, was employed within ComfyUI to interpolate between the two domains through a two-scale latent walk. The resulting embeddings were visualised in Grasshopper through PCA reduction and t-SNE optimisation, producing an animated three-dimensional latent cloud that revealed the manifold's internal topology. Selected mid-zone artefacts were translated into Trellis-meshed geometries and fabricated via Fused Filament Fabrication (FFF) with PLA. CLIP-based semantic evaluation confirmed that these artefacts occupied balanced positions along the natural–artificial continuum. The research positions computational form and material expression as post-digital mediators of new materialist thought, advancing an architectural reasoning that moves beyond the either/or logic of modernity toward a continuous process of entanglement, transformation, and becoming across digital and material domains.

Keywords. Generative AI, latent space, latent walk, bio-digital fusion, post-digital design, semantic evaluation

1. Introduction and Research Background

Contemporary architectural and urban design remain structured by modernism's enduring binaries—nature and culture, organic and artificial, human and non-human. These separations are not only conceptual but embedded in the material and social

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 737-746.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

organisation of cities, perpetuating inequality and exclusion. As Soja (2013) observes, such dualism limits spatial justice by framing the city as a purely human construct rather than a shared ecological field.

This study explores how generative artificial intelligence can help dissolve these inherited divisions by engaging the material and computational processes through which hybrid realities already take form. Taking inspiration from Latour's (2012) view that the modern condition has always been hybrid, the research positions design as a means of making such entanglements visible and material. Within the latent space of AI, natural and artificial attributes are no longer opposites but continuous variations within a shared manifold. The designer's role thus shifts from classification to navigation—curating transformations across a spectrum of possibilities. Here, AI is treated not as a passive instrument but as a collaborative agent that extends spatial and material reasoning.

The concept of the Novacene (Lovelock, 2019) situates this investigation within an ethical dimension, framing intelligence—human, biological, and artificial—as coexistent within a single planetary ecology. It directs the research toward understanding computation as an ecological practice, one that mediates between different forms of intelligence and matter rather than separating them.

Accordingly, the project adopts an exploratory stance, approaching computation as a site of discovery where form emerges through negotiation among data, material, and design intent. Latent space interpolation enables designers to trace subtle transitions between natural and artificial morphologies, revealing continuities that resist categorical separation. Underlying this is the conviction that digital and physical matter belong to a shared generative continuum—an interval where the organic and the constructed intertwine. Within this continuum, opposition gives way to collaboration, leading to the central question of the study: how might hybrid design processes materialise such continuity and reimagine the urban environment as an evolving field of relations rather than fixed typologies?

To explore this question, datasets of plant morphologies and Brutalist façades were curated and embedded within latent space. Guided "latent walks" (Chen & Stouffs, 2021; Huang, 2024; Karras et al., 2020) produced interpolated artefacts that were materialised through multi-material 3D printing. These prototypes act not as resolved designs but as probes—testing how generative processes can articulate a post-digital mode of hybridity that reconnects technological intelligence with ecological imagination.

2. Methodology

This research develops and evaluates a computational workflow designed to generate hybrid architectural forms that synthesise natural and artificial characteristics across both digital and material domains. The process unfolds through five interconnected stages—data curation, latent space embedding and interpolation, latent walk and Grasshopper visualisation, material transformation and 3D printing, and semantic validation (Figure 1). Each stage contributes to translating conceptual notions of hybridity into tangible spatial and material outcomes.

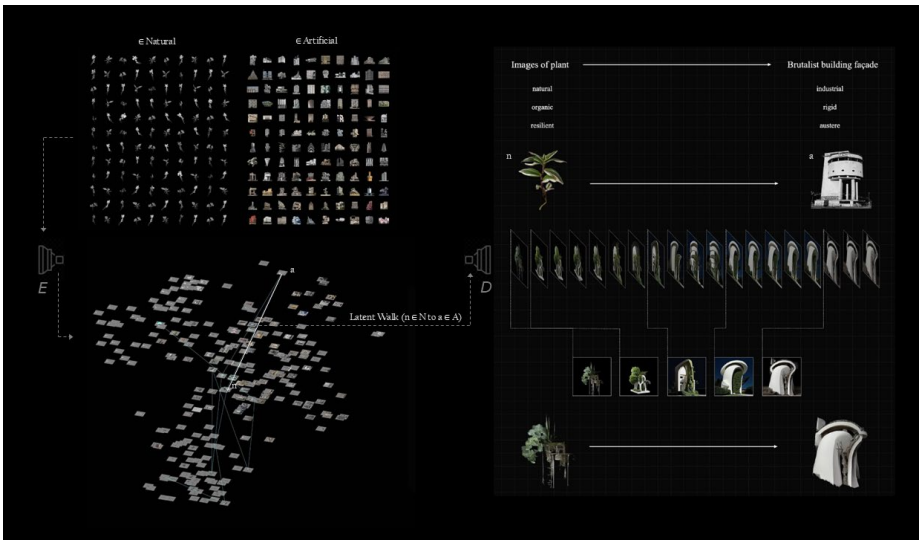


Figure 1. The generative workflow from data curation to latent walk, followed by 3D reconstruction and materialisation processes.

2.1. DATASET CURATION

To explore the continuum between the natural and the artificial, two morphologically divergent datasets were constructed to represent each domain. The goal was to define the boundary conditions for subsequent interpolation between contrasting typologies.

The natural dataset consisted of 295 three-dimensional scans of herbaceous plants captured through structured-light scanning, preserving fine surface textures and branching geometry. The meshes were cleaned and simplified in MeshLab to balance computational efficiency and morphological fidelity while maintaining diversity across species and form.

The artificial dataset contained 295 images of Brutalist façades collected via targeted web scraping. Because no specific geographic site was predetermined, the data were sourced globally through random sampling to avoid regional bias. Initially, 452 and 336 raw samples were obtained for the natural and artificial categories respectively. Each dataset was subsequently processed using a computational similarity filtering pipeline based on perceptual image hashing. A 64-bit difference hash (dHash) was computed for every image, and pairwise Hamming distances were calculated to identify near-duplicate samples. Images with a distance below a threshold of 5 bits were automatically removed, ensuring that only perceptually distinct samples were retained. This method provided an efficient and reproducible means of eliminating redundant or overly similar entries prior to manual semantic screening. Additional manual screening by the authors removed samples that presented semantically implausible or morphologically inconsistent matches.

To avoid stylistic redundancy, iterative filtering was applied using cosine similarity of ResNet-50 embeddings (He, Zhang, Ren, & Sun, 2016), ensuring both datasets maintained comparable levels of geometric and textural complexity. The curation

process aimed to maximise semantic divergence while retaining formal coherence, thereby establishing a strong morphological contrast that would serve as the foundation for interpolation and hybrid generation. After the full filtering and manual review process, both datasets were finalised with 295 samples each, representing the natural and artificial domains respectively. These curated collections formed the definitive input for the subsequent latent embedding and interpolation stage.

2.2. LATENT SPACE EMBEDDING AND INTERPOLATION

The curated datasets were next encoded into a shared latent manifold that enables continuous transitions between natural and artificial typologies. Feature embeddings were generated within ComfyUI (Comfyanonymous, 2023) using a fine-tuned version of Stable Diffusion 1.5 trained on architectural façade imagery (Figure 2). Within this model, the CLIP encoder produced high-dimensional embeddings, while the VAE decoder reconstructed images from interpolated latent states, allowing smooth traversal through latent space.

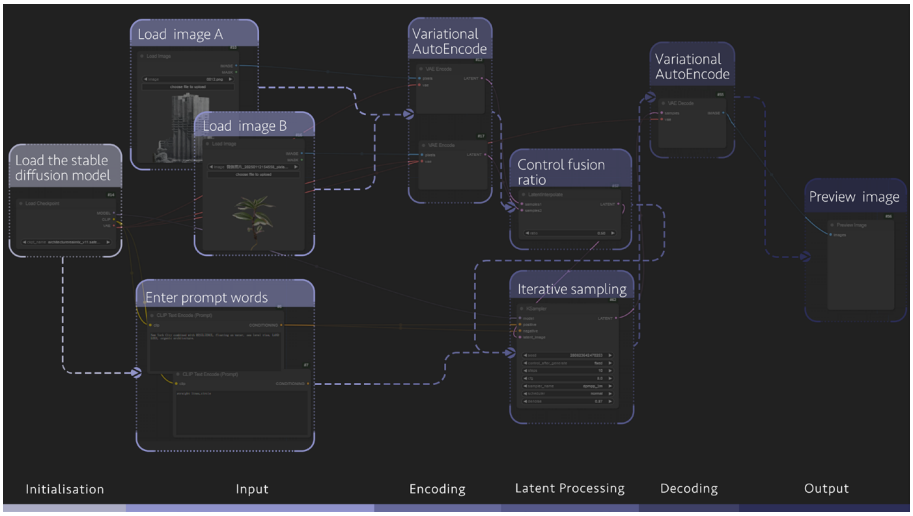


Figure 2. The latent walk workflow (one step) in ComfyUI.

Latent walks were performed by interpolating between two distant latent vectors z_A(plant) and z_B(façade). To capture both coarse and fine transitions, a two-scale interpolation scheme was implemented:

$$t_i = \{0.0, 0.1, 0.2, \ldots, 1.0\}, \quad t_{i+\frac{1}{2}} = \frac{t_i + t_{i+1}}{2}$$

which yields

$$t = \{0.00, 0.05, 0.10, 0.15, \ldots, 0.95, 1.00\}$$

each interpolated latent vector was computed as

$$z_t = (1 - t) \cdot z_A + t \cdot z_B$$

and decoded through the model's VAE to generate the corresponding hybrid image $I_t = \mathrm{D}(z_t)$.

This stage transforms categorical difference into a measurable continuum of form and texture. The interpolated embeddings produced here constitute the data foundation for both the spatial analysis in Grasshopper and the subsequent material translation.

2.3. LATENT WALK VISUALISATION IN GRASSHOPPER

Within ComfyUI, image sequences were generated through latent interpolation, after which the resulting frames were imported into Grasshopper 3D for spatial analysis. The ResNet50 embeddings described in Section 2.1 were imported into Grasshopper 3D and processed through a custom GH_CPython component to visualise their spatial relationships. The component applied Principal Component Analysis (PCA) to reduce the embedding dimensions to three principal components, followed by t-Distributed Stochastic Neighbour Embedding (t-SNE) optimisation to improve local clustering and maintain semantic proximity (Jolliffe & Cadima, 2016; van der Maaten & Hinton, 2008).

The resulting three-dimensional imagery cloud (see Figure 1, bottom left) visualised the latent manifold as a continuous morphogenetic field. Each interpolation step from ComfyUI was mapped to a corresponding coordinate within this distribution, enabling the animated visualisation of the latent walk as a trajectory through the manifold.

This workflow established a direct bridge between computational and spatial reasoning: ComfyUI defined the generative logic of interpolation, while Grasshopper provided a designer-ly interface through which the structure of latent space could be analysed, interpreted, and curated for materialisation.

Building upon this spatial cognition framework—where visual features operate as spatial metrics—the subsequent analysis focused on developing a strategy for the latent walk. The starting and ending points of each trajectory were determined according to the Euclidean distances between image embeddings within the latent space. This approach aimed to maximise the representational tension between natural and artificial domains: each interpolated segment was selected to exhibit high displacement diversity, while the overall trajectory maintained a substantial distance between its initial and terminal positions in the latent manifold.

2.4. MATERIAL TRANSFORMATION AND 3D PRINTING

Following visual exploration, selected latent outputs were transformed into physical artefacts to test their spatial and structural coherence. Two-dimensional image features were converted into three-dimensional geometries through a Trellis-based meshing process. Adaptive triangulation and gradient-driven surface relaxation preserved the balance between organic branching and architectural regularity, ensuring that the resulting geometries could be fabricated without distortion or collapse.

The decision to employ Fused Filament Fabrication (FFF) (Tymrak, Kreiger, & Pearce, 2014) using Polylactic Acid (PLA) was guided by the need to maintain precision, scalability, and ecological responsibility within architectural prototyping. Compared with resin-based methods such as SLA or DLP, FFF enables the production

of larger components at significantly lower cost while preserving geometric accuracy sufficient for spatial and volumetric studies. PLA, as a biodegradable thermoplastic derived from renewable sources, ensured material stability and environmental sustainability throughout the fabrication process (Farah, Anderson, & Langer, 2016; Wittbrodt & Pearce, 2015). Printing parameters were maintained within standard architectural tolerances—approximately 0.15 mm layer height, 200 °C extrusion temperature, and 55 °C bed temperature—to balance surface quality with production efficiency.

This configuration aligns with the conceptual intent of the study: to demonstrate how computationally derived hybridity can extend beyond digital simulation into physical expression. The fabrication phase therefore acts not merely as output but as a reflective test of the workflow's capacity to translate latent morphologies into material form, closing the loop between algorithmic generation and embodied architectural artefacts.

2.5. SEMANTIC VALIDATION VIA CLIP

The final stage evaluated the semantic balance between natural and artificial categories using CLIP (OpenAI ViT-B/32) with the textual prompts "natural" and "artificial". For each image I, cosine similarities were computed as:

$$s_{\text{nat}} = \cos\big(f(I), g(\text{natural})\big), \quad s_{\text{art}} = \cos\big(f(I), g(\text{artificial})\big)$$

and their difference

$$\Delta = s_{\text{nat}} - s_{\text{art}}$$

Values near 0 indicate hybrid identity. Artefacts satisfying $|\Delta| < 0.05$were defined as semantic mid-zone artefacts, confirming balanced positions within the natural–artificial continuum.

3. Results

Figure 3 presents a sequence of intermediary states that mark the most generative moments within the latent continuum. The forms reveal subtle negotiations between vegetal branching and architectural order, where curvature begins to stabilise into spatial frameworks. Rather than blending the two domains evenly, these intermediates expose points of tension—where growth-like proliferation meets structural alignment. Their spatial porosity and directional flow suggest an emergent tectonic logic, neither organic nor artificial but conditionally both. This visual transition captures the study's central concern: hybridity as a process of becoming, expressed through the gradual reorganisation of form rather than through symbolic fusion. Within this array, specific intermediates were intentionally selected for further development, reflecting the designer's interpretive agency. The act of selection situates human judgement within the generative process, resonating with new materialist thought that recognises design as co-produced by human and non-human agencies.

Figure 4 translates these digital artefacts into physical prototypes. The 3D-printed models preserve the layered gradients of the interpolated images yet introduce a new tactile register through PLA extrusion. Filament striations, slight surface roughness,

and local variations in translucency reveal how computational continuity acquires material presence. These imperfections are not defects but expressions of a hybrid materiality, where digital precision encounters the contingencies of fabrication.

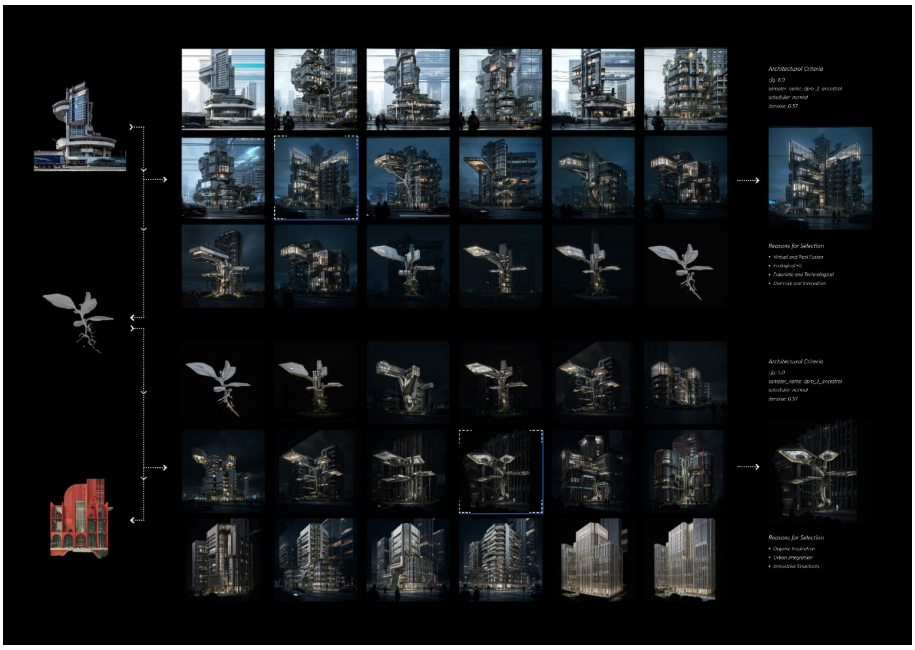


Figure 3. Sequence of intermediary states generated through latent interpolation between natural and artificial datasets. The selected intermediates highlight moments of formal negotiation where vegetal branching meets architectural order, demonstrating the designer's interpretive agency within the generative process.



Figure 4. The printed artefacts materialise hybrid morphologies with tactile striations and layered translucency, revealing how computational continuity acquires embodied material presence.

To quantitatively verify the hybrid identity of the selected prototypes, we applied the CLIP-based semantic validation defined in the Methodology section. Each prototype image $I$was encoded using CLIP's image encoder $f(I)$and compared against two textual prompts—"natural" and "artificial"—encoded by the text encoder $g(T)$. Cosine similarities were computed as $s_{\text{nat}} = \cos\left(f(I), g(\text{“natural”})\right)$and $s_{\text{art}} = \cos\left(f(I), g(\text{“artificial”})\right)$. Their difference $\Delta = s_{\text{nat}} - s_{\text{art}}$represents the relative semantic balance along the natural–artificial continuum, where values close to zero indicate hybrid identity.

As shown in Figure 4, the three prototypes selected for fabrication yielded Δ values of –0.021, +0.034, and –0.015. Since all absolute values fall below the mid-zone threshold ($|\Delta| < 0.05$), these artefacts can be regarded as semantically balanced. This threshold was derived empirically from the overall Δ distribution across all interpolated samples, corresponding roughly to one standard deviation around zero. To ensure robustness, alternative prompt pairs such as "organic/synthetic" and "vegetal/constructed" were tested, confirming consistent classification.

These results provide quantitative evidence that the workflow generates semantic mid-zone artefacts—design outcomes balanced between the natural and the artificial across both latent and material domains. Together, Figures 3 and 4 complete the conceptual loop between computation and embodiment, showing that hybridity is not merely visual but also material and experiential. Through their shifting geometry and tangible grain, the artefacts return the argument to its core question: how design mediates between natural and artificial systems within a single evolving continuum.

4. Discussions

Earlier computational design studies have addressed hybridity primarily as a question of biomimicry, performance optimisation, or material integration (Oxman, 2010; Menges & Reichert, 2015). Such approaches, while innovative, often re-inscribe the dualism between the organic and the artificial by seeking to replicate natural efficiency through artificial systems. In contrast, the present study repositions hybridity as a relational condition within the latent space, where distinctions between nature and artifice are continuously negotiated rather than reproduced.

Within the broader discourse on connectionist design (Huang, 2024), latent space represents not merely a technical substrate but an epistemic field in which design reasoning becomes distributed across data, model, and designer. This differs from earlier parametric or rule-based systems (Carpo, 2017; Terzidis, 2006), where knowledge was encoded explicitly. Here, form emerges through implicit negotiation among datasets and embeddings. The study thus contributes to an ongoing theoretical shift—from deterministic parametrisation to probabilistic emergence—aligning with the new materialist emphasis on becoming rather than being (DeLanda, 2011).

Translating latent interpolations into material artefacts through FFF extends the computational into the tangible. This act of fabrication is not an afterthought but a critical interface between digital speculation and ecological accountability. Rather than reiterating shape (or image)-centred generative practices that privilege representation over material process—the very tendency critiqued within Rivka Oxman's discourse on new structuralism (2010)— this workflow foregrounds the entanglement of material

agency, energy, and sustainability. The use of biodegradable PLA reflects a small but concrete gesture towards what Lovelock (2019) terms a "symbiotic intelligence," where design negotiates coexistence between biospheric and machinic processes.

By operationalising hybridity through latent space traversal and semantic validation, the study offers an alternative framework for design reasoning—one that is interpretive rather than instrumental. For pedagogy, this workflow suggests how designers might engage AI as a collaborator, learning to read and curate rather than merely command computational systems. This aligns with the ongoing redefinition of design as a mediated practice of sense-making rather than a purely productive act (Cross, 2006; Candy & Edmonds, 2018).

Extending this framework to the urban scale outlines a model of post-digital urban futures, where morphological hybridity becomes a principle of spatial organisation rather than a merely aesthetic condition. In such futures, digital and ecological systems co-produce the city as a living manifold—responsive, adaptive, and continuously negotiated among human, machinic, and biospheric intelligences. The latent continuum explored here prefigures not isolated artefacts but a relational, generative paradigm where porosity, layering, and branching logics offer transferable principles for post-digital urban form.

While exploratory, the framework opens paths for extending computational hybridity—by incorporating urban ecological morphologies, multimodal data (e.g., LiDAR + text), and integrated metrics linking semantic and material performance. These directions advance hybridity as both design method and epistemic model.

5. Conclusion

This study demonstrated how generative AI can materialise non-dual hybridity between natural and artificial systems through a reproducible computational workflow integrating dataset curation, latent interpolation, and material fabrication. The experiments revealed that hybrid artefacts occupy balanced positions within the natural–artificial continuum, both computationally and physically, indicating that latent space serves not only as a technical interface but as an epistemic field for mediating continuity across digital and material domains. These findings highlight the potential of AI-driven design to move beyond representational synthesis toward ecological negotiation, where computation operates as an agent within planetary material processes. The implications extend to post-digital architectural practice and pedagogy, suggesting new models of collaboration between human, machinic, and biospheric intelligence. Future research could expand this approach to urban and multispecies contexts, advancing a more relational and ecologically attuned framework for design intelligence.

Acknowledgements

The authors would like to thank all members of MArch Urban Design RC18 (2024–25) at the Bartlett School of Architecture for their collaboration and inspiration, as well as Mr Martyn Carter, B-made Operations Manager, for his technical assistance with 3D printing.

References

Candy, L., & Edmonds, E. A. (2018). Practice-based research in the creative arts: Foundations and futures from the front line. *Leonardo, 51*(1), 63–69. https://doi.org/10.1162/LEON_a_01471

Carpo, M. (2017). *The second digital turn: Design beyond intelligence.* MIT Press.

Chen, J., & Stouffs, R. (2021). From exploration to interpretation: Adopting deep representation learning models to latent space interpretation of architectural design alternatives. In *Proceedings of the 26th International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2021)* (pp. 131–140). The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).

Comfyanonymous. (2023). *ComfyUI: The most powerful and modular visual AI engine and application* [Computer software]. GitHub. Retrieved 13 April from https://github.com/comfyanonymous/ComfyUI

Cross, N. (2006). *Designerly ways of knowing.* Springer.

DeLanda, M. (2011). *Philosophy and simulation: The emergence of synthetic reason.* Continuum.

Farah, S., Anderson, D. G., & Langer, R. (2016). Physical and mechanical properties of PLA, and their functions in widespread applications — A comprehensive review. *Advanced Drug Delivery Reviews, 107*, 367–392. https://doi.org/10.1016/j.addr.2016.06.012

Haraway, D. J. (2016). *Staying with the trouble: Making kin in the Chthulucene.* Duke University Press.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Huang, S.-Y. (2024). *On the architectural rationality of connectionist design* [Doctoral dissertation, University College London]. UCL Discovery.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A, 374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). Training generative adversarial networks with limited data. In *Advances in Neural Information Processing Systems*, 33, 12104–12114.

Latour, B. (2012). *We have never been modern*. Harvard University Press.

Lovelock, J. (2019). *Novacene: The coming age of hyperintelligence*. MIT Press.

Menges, A. (2012). Material computation. *Architectural Design, 82*(2), 14–21. https://doi.org/10.1002/ad.1373

Oxman, N. (2010). *Material-based design computation* [Doctoral dissertation, Massachusetts Institute of Technology]. MIT DSpace.

Oxman, R., & Oxman, R. (2010). The new structuralism: Design, engineering and architectural technologies. *Architectural Design, 80*(4), 14–23. https://doi.org/10.1002/ad.1101

Soja, E. W. (2013). *Seeking spatial justice*. University of Minnesota Press.

Tymrak, B. M., Kreiger, M., & Pearce, J. M. (2014). Mechanical properties of components fabricated with open-source 3-D printers under realistic environmental conditions. *Materials & Design, 58*, 242–246. https://doi.org/10.1016/j.matdes.2014.02.038

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine *Learning Research, 9*, 2579–2605.

Wittbrodt, B. T., & Pearce, J. M. (2015). The effects of PLA color on material properties of 3-D printed components. *Additive Manufacturing, 8*, 110–116. https://doi.org/10.1016/j.addma.2015.09.006

Urban Blocks Regionalization By Integrating Spatial Proximity And Morphological Similarity

Ruize Wu[1], Jiachen Bu[2] and Ziyu Tong[3]
[1,2,3]*School of Architecture and Urban Planning, Nanjing University.*
[3]*Key Laboratory of Urban AI and Green Built Environment of Provincial Higher Education Institutes, Nanjing University*
[1]*502024360024@smail.nju.edu.cn*
[2]*jiachenbu31@nju.edu.cn, 0000-0001-5545-9589*
[3]*tzy@nju.edu.cn, 0000-0002-5872-0890*

Abstract. Urban morphological regionalization, the process of identifying urban areas that are both morphologically similar and spatially contiguous, is crucial for urban analysis and planning. However, existing methods commonly struggle to balance internal morphological homogeneity with spatial continuity, often leading to fragmented regionalization outcomes. To address this challenge, this study proposes a dual-channel graph neural network, named Morph-GNN. The model learns block-level embeddings sensitive to both "spatial proximity" and "morphological similarity" by constructing a graph that encodes these dual relationships, which are then utilized for final clustering. Empirical results from a case study on 4,364 urban blocks in Nanjing demonstrate that Morph-GNN can strike a balance between spatial continuity and feature similarity. Compared to the selected baseline models, the proposed method generates regions that achieve a favourable trade-off between spatial continuity and morphological homogeneity—significantly improving spatial continuity without notably compromising internal morphological homogeneity. This proves that the proposed method effectively resolves the fragmentation problem inherent in existing approaches and achieves a significant advancement in balancing the core challenges of morphological homogeneity and spatial continuity, thereby providing a more refined and scientific tool for urban planning.

Keywords. Urban Morphology, Graph Neural Network (GNN), Regionalization, Spatial Proximity, Morphological Similarity

1. Introduction

In the process of rapid urbanization, understanding urban morphological patterns through the classification of neighbourhood typologies has become a central approach to uncovering intrinsic urban regularities and guiding sustainable development. Urban morphology focuses on the study of the structure and patterns of the city's physical form (Yang et al., 2022). Unlike classification methods that view elements in isolation,

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 747-756.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

morphological regionalization emphasizes partitioning based on spatial homogeneity and heterogeneity, aiming to reveal local similarities within regions and overall differences between them (Yang et al., 2022). An ideal regionalization outcome should consist of spatially continuous and internally morphologically homogeneous areas, thereby providing effective guidance for macro-scale planning. This approach is crucial for addressing challenges brought by rapid urbanization, such as functional zoning, traffic congestion, and environmental equity.

Urban morphology analysis has evolved from subjective judgment to a data-driven paradigm. Early expert-driven methods, though insightful, were limited by their strong subjectivity, labor-intensiveness, and difficulty in generalization (Fleischmann et al., 2021). Subsequent quantitative methods, such as K-Means-based clustering, could analyze geometric features to identify urban patterns (Bobkova et al., 2019) and were widely regarded as a key technical path to achieving regionalization. However, they introduced new limitations: while capturing attribute similarity, these algorithms ignore the crucial geographical adjacency in urban space, leading to spatially fragmented partitioning results. This outcome, "aggregated in attributes but fragmented in space," accomplishes the "classification" of blocks but falls far short of the regionalization goal of forming coherent "regions," thereby greatly diminishing its practical value. The field of urban morphology generally recognizes three fundamental elements: buildings, streets, and blocks (Fleischmann et al., 2021), and the definition of measurable morphological features—including categories like dimension, shape, and spatial distribution—is key to clustering analysis. In the digital and intelligent era, applying unsupervised clustering to urban big data has become increasingly popular, offering objective insights, for instance, by identifying morphological patterns and block types through clustering or new metrics (Li and Quan, 2023). Nevertheless, traditional methods continue to neglect the interaction of geographical adjacency between blocks, thereby undermining the accuracy of the regionalization results.

These deficiencies collectively point to a research gap: existing methods have failed to jointly model the two driving relationships of morphological similarity and spatial proximity within a unified framework. On the one hand, simple classification methods struggle to characterize the complexity of urban form by ignoring the interactions between blocks (Li and Quan, 2023). On the other hand, some graph-theory algorithms that adopt the first law of geography (such as Louvain) tend to overemphasize spatial connectivity, thereby grouping together adjacent blocks with vastly different morphologies and sacrificing the fidelity of the morphological analysis. These two extremes jointly reveal a central challenge: how to strike a balance between morphological homogeneity and spatial compactness. Although some Graph Neural Network (GNN) methods have attempted to integrate spatial information to improve clustering performance (Liang et al., 2025; Lu and Tong, 2025), they remain insufficient in coupling the complex relationship between urban morphology and geographical location, especially at the fine-grained block level. Based on this gap, our research question (RQ) is: Can integrating node similarity (morphological proximity) and geographical adjacency within a graph network enhance the practical applicability of urban morphological regionalization, thereby providing more scientifically coherent guidance for block-scale planning? To address this RQ, we propose an unsupervised clustering method centered on a dual-channel graph neural network applied to an urban

block graph network. Specifically, we model blocks as nodes with morphological attributes and incorporate a dual-channel structure: (1) a spatial adjacency channel for geographical proximity, and (2) a morphological similarity channel for feature-based connections. This structure enables joint representation learning, producing high-dimensional embeddings that capture both types of relationships. Subsequently, a classic clustering algorithm is applied to these embeddings to delineate the regionalization typologies. We will demonstrate the effectiveness of our proposed method in identifying urban morphological regions and provide a comparative analysis against baseline models using multiple evaluation metrics.

2. Case study and Methods

2.1. CASE STUDY

The case study for this paper is the main urban area of Nanjing, the capital of Jiangsu Province, China, with its scope defined according to the Nanjing Municipal Land and Space Master Plan. Nanjing is a city with a 3,000-year history of city-building. Its long history of urban development has given rise to a rich diversity of urban morphological spaces distributed across eleven administrative districts. In the course of modern urbanization, Nanjing has produced a multitude of urban blocks. The attributes of these blocks, such as their size and shape, significantly influence the form and configuration of urban buildings, streets, and public spaces (Shpuza, 2023), and they are considered the fundamental elements of urban form. Therefore, the subject of this study is the 4,364 urban blocks within Nanjing. The area is shown in Figure 1.

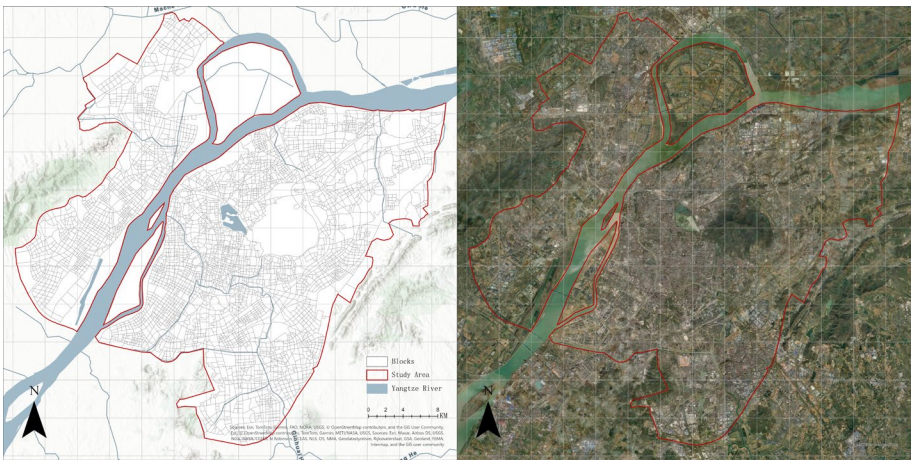


Figure 1. Case Study Area

2.2. DATA DESCRIPTION AND PROCESSING

This study utilizes vector geographical geometry data of urban blocks and urban building data as its primary data sources. The block data were generated by segmenting the urban boundary data with the road network data, both collected from Tianditu (https://www.tianditu.gov.cn/), and the geometric attributes were subsequently

computed in GIS. Urban buildings, distributed within each block, are a crucial component of the block's morphology. The building data, which includes building footprints and heights, was sourced from 3D-GloBFP. Consistent with multiple previous studies that assume a standard single-story height of 3 meters, the number of floors can be estimated from the building height.

2.3. GRAPH MODEL CONSTRUCTION

This study constructs 13 urban morphological indicators to represent the characteristics of urban blocks, encompassing metrics for block form (P, A, Comp), shape (GSI, OSR, FAR, SC, AH, Po), variation (UH, UA), and thermal environment (NDVI, NDBI). Previous research has demonstrated that these indicators are sufficient to represent the morphological features of urban blocks. We model the blocks as nodes, use their morphological indicators as attributes, and construct initial edges based on adjacency relationships to build the Morph-GNN.

Our objective is to identify urban morphological regional typologies. Nodes within the same urban morphology region must satisfy the following conditions: (1) their attributes should be as similar as possible, and (2) they should be geographically adjacent as much as possible to form contiguous areas. To achieve this goal, we construct the following graph neural network.

First, the urban blocks are represented as an undirected graph where each node signifies an urban block. The node features are composed of block form, arrangement, shape factor, and variation. There is a total of 4,364 nodes. Second, to meet our conditions, we construct two types of edges. The first type is the spatial adjacency edge, which is directly based on the geographical proximity of the blocks and is implemented using the spatial join toolbox in GIS. The second type is the morphological similarity edge. For this, we calculate the cosine similarity of the morphological indicators between blocks and add the top k-most similar non-adjacent nodes for each node as similarity edges (with a default k=3). After the graph is constructed, a dual-edge graph incorporating both spatial topology and morphological similarity is formed, which serves as the input for the GNN.

After constructing the dual-channel GNN model, the objective is to extract deep features from the graph. Therefore, the next step is representation learning: the original node features undergo a linear transformation, ReLU activation, batch normalization, and dropout to obtain the initial hidden representation. To process the spatial adjacency edges, GraphSAGE is employed to aggregate features from the node itself and its neighbors. The representation for the morphology channel uses a single-head Graph Attention Network (GAT), which incorporates attention weights.

Finally, the output embeddings from the two channels are concatenated and fused to generate the final node representation, which captures both local spatial structure and global morphological patterns. Based on this embedding, we utilize the silhouette coefficient to determine the optimal number of clusters and then employ the K-Means algorithm to obtain the final urban morphological regionalization results for subsequent comparative analysis.

2.4. BASELINE ALGORITHMS

We use a variety of baseline algorithms to compare their performance against our proposed graph neural network.

(1) K-Means: The K-Means clustering algorithm groups nodes with similar attributes and assigns each node to the nearest cluster, measuring distance in the feature space. However, the K-Means algorithm only uses node attributes as input and cannot consider edge connections.

(2) Agglomerative Clustering: Agglomerative clustering is a bottom-up clustering method. It begins by treating each data point or object as an individual cluster and then iteratively merges the most similar pairs of clusters to build a hierarchy. This process continues until all objects are merged into a single large cluster or a predefined number of clusters is reached. The core of this algorithm lies in defining inter-cluster similarity (e.g., single linkage, complete linkage, or average linkage), and a dendrogram is often used to visualize the clustering process.

(3) Louvain: The Louvain algorithm first generates small-scale communities by maximizing local modularity gain and then recursively aggregates the identified communities into nodes. The input for the Louvain method is the block network. In this study, only adjacency connections are used to detect communities.

(4) DeepWalk: DeepWalk is a random walk-based graph embedding model. It generates random paths composed of short walks to connect nodes. It takes the spatial adjacency matrix as input. In this study, after obtaining the node embeddings, the K-Means clustering algorithm is applied to generate the final communities.

(5) LINE: LINE is a network embedding method designed to preserve both local and global network structures. In this study, we apply it to the spatial adjacency matrix. The resulting node embeddings are then clustered using the K-Means algorithm to generate a regionalization scheme.

(6) Graph Convolutional Network (GCN): The Graph Convolutional Network (GCN) is a foundational GNN model that learns node representations by aggregating feature information from neighbours. It exhibits local isotropy, treating all neighbours equally. In this study, we use a GCN to learn feature representations of the urban blocks, followed by K-Means clustering to produce the final regionalization. A comparison with GCN allows for an evaluation of the performance gains achieved by more complex aggregation mechanisms.

2.5. EVALUATION METRICS

We evaluate all methods using the following metrics.

(1) Moran's I: A measure of global spatial autocorrelation. A positive value indicates spatial clustering (similar blocks are geographically close), a negative value indicates dispersion, and a value near 0 suggests a random distribution. In urban morphology analysis, this index is often used to assess the overall spatial structure of elements like buildings or blocks to determine their degree of clustering or dispersion.

(2) Local Spatial Cohesion (LSC): Used to evaluate the ultra-local morphological homogeneity within partitions. It is calculated by averaging the cosine similarity between the feature vector of each block and those of its immediate geographical

neighbours. A higher LSC value signifies a smoother morphological transition and stronger homogeneity within a region. This metric adheres to the First Law of Geography, aiming to quantify feature continuity at an ultra-local scale.

(3) Graph Gradient Index (Gi): Employed to assess the morphological smoothness and boundary clarity within clusters. It quantifies the "morphological gradient"—the ratio of morphological difference (Euclidean distance) to geographical distance—between adjacent blocks within the same cluster. A lower Gi value indicates more consistent morphology and smoother spatial transitions within a cluster, thus signifying sharper and clearer boundaries. We use the average of all cluster scores (Gi_Mean) as the global metric.

(4) Cosine Similarity: Measures the morphological consistency among members within a cluster. It calculates the similarity in direction between feature vectors; a higher value represents greater uniformity of morphological characteristics within the cluster. In the positive context of urban features, a higher positive value indicates stronger morphological coherence.

(5) Inequality: To assess the inequality of feature distribution within detected communities, the Inequality metric, as proposed by Liang et al. (2025), is used. It quantifies the degree of variation for a single feature within a community. A value close to 0 indicates perfect equality (uniform distribution), which is ideal for cohesive clustering, while lower values reflect superior clustering performance by minimizing intra-community differences.

(6) Adjusted Join Count Ratio (AJCR): To quantitatively evaluate the spatial continuity of community partitions, the Join Count Ratio is employed (Liang et al., 2025). This metric measures the proportion of all geographically adjacent block pairs that are assigned to the same community. However, this metric can be inflated by trivial partitioning methods that produce very few large communities. To correct this bias, we introduce an Adjusted Join Count Ratio (AJCR). By applying a penalty to methods that produce an insufficient number of partitions, the AJCR provides a fairer assessment of a model's ability to generate spatially continuous regions. A high AJCR value implies that the community boundaries align well with geographical adjacency. Since some methods generate only two regional classes, which lacks practical significance in urban studies, they are excluded from the AJCR calculation; only methods producing four or more classes are considered. The equation is shown as below:

3. Results

3.1. MORPHOLOGICAL REGIONS

After constructing the Morph-GNN and performing clustering, we visualized the results using GIS tools. This method significantly overcomes the problem of result fragmentation. According to the regional statistics, the 4,364 block units in Nanjing were classified into 568 spatially contiguous and homogeneous regions composed of blocks of the same category. The right panel of Figure 2 presents the statistical summary of the regionalization results. For instance, block categories such as 0 and 1 statistically form significantly larger, coherent developed areas. In contrast, category 5 is composed of numerous small micro-plots, presenting a highly fragmented pattern

spatially, scattered among the larger sectors. This co-existing pattern of macro-structures and micro-infill objectively reflects the complex reality of urban spatial organization. The histogram of overall plot areas corroborates this, indicating that the entire region is composed of a vast number of small plots and a small number of large ones. To a certain extent, Morph-GNN, while learning the morphological features of the blocks, successfully incorporated the constraint of geospatial proximity, thereby enabling the identification and partitioning of macro-scale urban structural units. The result is shown in Figure 2.

Morpho-GNN identifies four typical morphological regions in Nanjing, reflecting distinct development strategies from different periods, as illustrated in Figure 2. Region A, located in the Jianye District, is a more recently developed area dominated by high-density, high-rise residential blocks. Region B, situated in the Qinhuai District, is part of the old city center and is characterized by high-density, mid-to-low-rise residential buildings. Regions C and D are both in the Qixia District, on the city's outskirts, and primarily consist of low-density, mid-to-low-rise structures. Notably, Region D is largely composed of agricultural and industrial land.

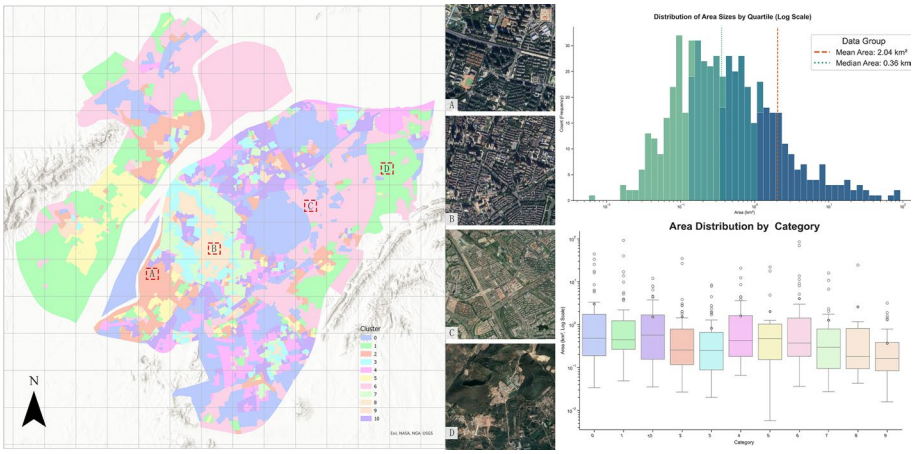


Figure 2. Morph-GNN Regionalization Results

In summary, the Morph-GNN method proposed in this study demonstrates significant advantages in urban morphological regionalization. Its core strength lies in the simultaneous optimization of block morphological similarity and geographical proximity through information aggregation on the graph structure within a unified representation learning framework. Consequently, its results ensure the spatial integrity of the regional divisions at a macro-scale while preserving the micro-level sensitivity to local morphological differences. This outcome represents a more precise and realistically interpretable expression of urban morphological structure, laying a solid foundation for subsequent quantitative analysis and in-depth urban studies.

To ensure that all methods are compared under optimal conditions, we first need to determine the optimal number of clusters (K). Specifically, we use the Silhouette Score to determine the optimal K value for K-Means (10 clusters), Agglomerative Clustering (2 clusters), DeepWalk (5 clusters), LINE (10 clusters), and GCN (10 clusters). The Louvain method automatically partitions the data into 26 classes. We present the classification results of each method visually using GIS, arranged on a single layout, as shown in the Figure 3.

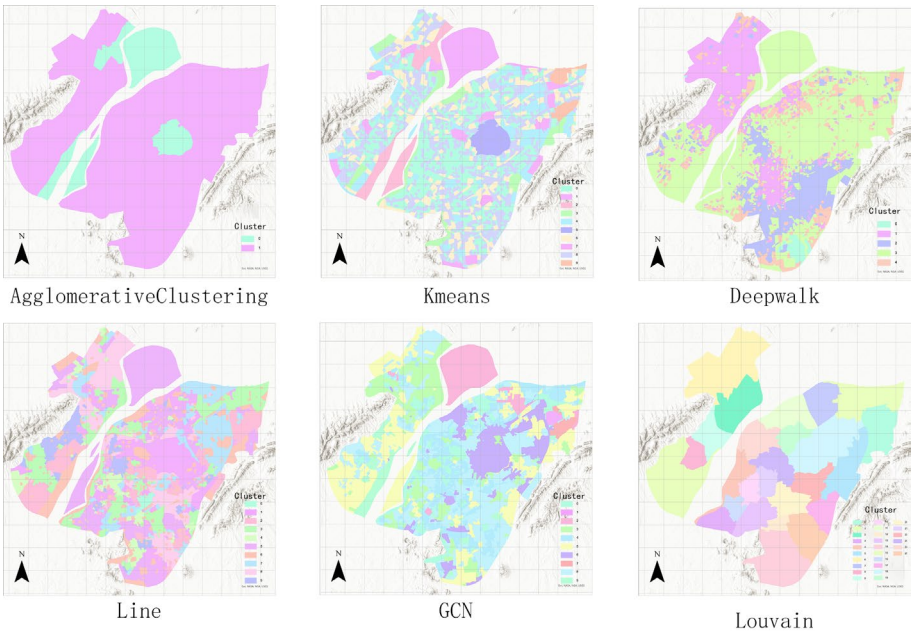


Figure 3. Baseline results

3.2. RESULTS OF EVALUATION METRICS

In this section, the proposed Morph-GNN method is compared against a series of baseline models to evaluate their performance in morphological partitioning within the Nanjing urban block network. Based on this, we comprehensively quantify the morphological fidelity of the partitioning results from each model using a range of metrics: an ideal partition should exhibit a high Moran's I, high LSC and cosine similarity, a low Gi_Mean, and a low inequality index. A comprehensive review across these metrics reveals that Morph-GNN demonstrates consistent superiority over most baseline models, with particularly significant advantages in enhancing local consistency and mitigating internal heterogeneity. This highlights the structural advancement of its dual-channel graph neural network architecture in urban morphology analysis. To systematically evaluate the model's performance, we conducted a comprehensive comparison of Morph-GNN against baseline models across three dimensions: spatial structure, boundary clarity, and internal homogeneity. The results are shown in Table 1.

Table 1. Features and Calculation Formula

Methods	Moran's I	LSC	Gi_Mean	CosineSimilarity	inequality	AJCR
Agglomerative Clustering	0.314	0.897	0.0866	0.981	0.4451	-
K-means	0.159	0.861	0.0276	0.985	0.0006	0.1518
DeepWalk	0.534	0.643	0.0866	0.865	0.0863	0.1178
Line	0.548	0.711	0.0863	0.952	0.0762	0.1623
GCN	0.510	0.929	0.0120	0.849	0.0711	0.1610
Louvain	0.945	0.937	0.0090	0.968	0.0972	0.3661
Morph-GNN	0.593	0.998	0.0101	0.972	0.0014	0.1685

In terms of spatial structure, we employed the global Moran's I and the Local Spatial Cohesion (LSC) index. The results show that Morph-GNN achieved the highest LSC score (0.998), demonstrating optimal morphological homogeneity at the local scale. Concurrently, its Moran's I (0.593) was second only to the Louvain algorithm, which purely optimizes for spatial adjacency. This indicates that Morph-GNN successfully forms macroscopically clustered and coherent regions without sacrificing morphological similarity. For boundary clarity, as evaluated using the Gi_Mean, Morph-GNN obtained the lowest score (0.0101). This confirms that our model effectively constrains sharp morphological transitions to narrow boundaries, thereby producing regions with the smoothest internal morphology and the clearest delineations. Regarding internal homogeneity, assessed by combining Cosine Similarity and the Inequality Index, our analysis revealed that while methods like K-Means exhibit high similarity, their high inequality scores expose a reliance on generating numerous, imbalanced "fragment" clusters. In contrast, Morph-GNN maintained a high Cosine Similarity (0.972) while achieving an extremely low Inequality Index (0.0014), proving its ability to generate regions that are not only highly uniform in morphology but also more balanced in scale and meaningful. In summary, the baseline models exhibit a "specialization" bias, failing to balance spatial structure with morphological homogeneity. Morph-GNN is the only model that demonstrates outstanding performance across all evaluation dimensions. It does not sacrifice a single performance metric but rather strikes a near-optimal balance among multiple objectives. By simultaneously achieving high spatial clustering, high morphological homogeneity, and the clearest regional boundaries, Morph-GNN yields a superior urban morphological regionalization scheme among all compared methods.

4. Conclusion and discussion

This study introduces Morph-GNN, a dual-channel graph neural network designed to resolve the inherent conflict between morphological homogeneity and spatial continuity in urban regionalization. Our core contribution is the expansion of "proximity" from a singular geographical dimension to a dual construct that integrates both geographic adjacency and morphological similarity. We contend that ignoring morphological proximity—similarities in development patterns, architectural styles, or densities—yields a biased understanding of urban structure.

The proposed Morph-GNN architecture directly encodes these two proximity types into a unified framework. Experimental results demonstrate that our model surpasses

traditional clustering and graph-based methods by achieving a superior balance among multiple objectives. Specifically, it avoids the common trade-off where internal morphological consistency is sacrificed for geographical continuity. The resulting partitions exhibit a high degree of correspondence with the city's intrinsic morphological fabric, providing a more scientific basis for fine-grained urban policies, such as urban renewal and heritage conservation. Methodologically, this work showcases the effectiveness of embedding domain-specific geospatial knowledge directly into the GNN architecture, offering a new paradigm for solving geographic problems that moves beyond conventional feature engineering.

However, this study has limitations. First, the geometric data used is a simplified representation of the urban environment, and any data-reality mismatch could impact analytical precision. Second, due to computational constraints, the model's performance may not have reached its theoretical optimum, with issues like blurred micro-scale boundaries requiring further refinement. Future work will focus on incorporating multi-modal data (e.g., Points of Interest, street view imagery) to transition from a purely physical analysis to a comprehensive functional and perceptual understanding of urban regions. We will also explore advanced deep learning models capable of revealing the multi-scale, hierarchical structures inherent in cities.

Acknowledgement

This research was supported by National Natural Science Foundation of China (52578026).

References

Bobkova, E., Berghauser Pont, M., & Marcus, L. (2019). Towards analytical typologies of plot systems: Quantitative profile of five European cities. *Environment and Planning B: Urban Analytics and City Science*, 48(4), 604-620.

Fleischmann, M., Feliciotti, A., Romice, O., & Porta, S. (2021). Methodological foundation of a numerical taxonomy of urban form. *Environment and Planning B: Urban Analytics and City Science*, 49(4), 1283-1299.

Li, N., & Quan, S. J. (2023). Identifying urban form typologies in Seoul using a new Gaussian mixture model-based clustering framework. *Environment and Planning B: Urban Analytics and City Science*, 50(9), 2342-2358.

Liang, Y., Zhu, J., Ye, W., & Gao, S. (2025). GeoAI-enhanced community detection on spatial networks with graph deep learning. *Computers, Environment and Urban Systems*, 117, 102228.

Lu, T., & Tong, Z. (2025). Graph-Based Typology Identification and Analysis of Urban Form A case study of Nanjing main district. *In 30th International Conference on Computer-Aided Architectural Design Research in Asia: Intelligent and Informed, CAADRIA 2025* (pp. 39-48). The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).

Shpuza, E. (2023). The shape and size of urban blocks. *Environment and Planning B: Urban Analytics and City Science*, 50(1), 24-43.

Yang, L., Yang, X., Zhang, H., Ma, J., Zhu, H., & Huang, X. (2022). Urban morphological regionalization based on 3D building blocks—A case in the central area of Chengdu, China. *Computers, Environment and Urban Systems*, 94, 101800. https://doi.org/https://doi.org/10.1016/j.compenvurbsys.2022.101800

From Cells to Strategies Game

Learning Block Layouts through Policy Networks and Game-Inspired Rewards

Han Xie[1], Mark Balzar[2], Hans Schartner[3], Zeynep Aksöz Balzar[4] and Linxue Li[5]
[1,2,3,4]*TU Wien*
[1,5]*Tongji University.*
[1]*e12402265@student.tuwien.ac.at, 0009-0005-0228-4101*
[2] *mark.balzar@tuwien.ac.at, 0009-0002-2035-9638*
[3] *hans.schartner@tuwien.ac.at*
[4] *zeynep.balzar@tuwien.ac.at, 0000-0002-3634-0248*
[5] *lilinxue@tongji.edu.cn*

Abstract. This study presents a hybrid framework integrating cellular automata (CA) and deep reinforcement learning (DRL) for adaptive urban block design under wind-driven environmental conditions. CA generates base-layer layouts that self-organize toward ventilation-oriented configurations through local interaction rules, establishing initial airflow paths and block–courtyard relationships. DRL extends this logic into three-dimensional stacking, learning strategies for courtyard openness, wind-tower enclosure, and cross-layer spatial connectivity. A Go-inspired reward function based on liberties ("qi") guides the agent to balance spatial coherence with ventilation performance. The Barcelona Eixample district is selected as the testbed due to its regular grid, repetitive courtyard structure, and high thermal vulnerability during Mediterranean summers—conditions that make it ideal for evaluating rule-based and learning-based ventilation strategies. Unlike iterative Computational Fluid Dynamics (CFD)-driven optimization, the proposed method relies on lightweight geometric heuristics, enabling rapid generation of spatially coherent and well-ventilated block layouts at a fraction of the computational cost. Compared with random stacking baselines, the DRL agent produces more stable block distributions and higher reward scores, indicating the effectiveness of combining simple local rules with learned spatial strategies.

Keywords. Thermodynamics Design, Deep Reinforcement Learning, Cellular Automata, Policy Networks, Go-Inspired Rewards

1. Introduction

Designing urban blocks that respond to wind flow, solar access, and thermal comfort remains a core challenge in computational design. Conventional generative methods

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 757-767.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

rely on fixed heuristics or Computational Fluid Dynamics (CFD) simulations, which are either too rigid or too computationally expensive for iterative exploration (Runchal & Rao, 2020). Meanwhile, games like Go illustrate how simple local rules can generate emergent spatial complexity and long-term strategy (Silver et al., 2016).

This study uses the Barcelona Eixample district as the testbed due to its high thermal vulnerability identified in preliminary analysis and its increasing need for heat-mitigation and natural ventilation under Mediterranean summer conditions. The Eixample's regular grid and repetitive courtyard structure make it especially suitable for rule-based modeling, aligning well with the procedural nature of cellular automata (CA) and the strategy-learning capabilities of deep reinforcement learning (DRL).

Within this context, the research investigates how a rule-based agent—developing from CA to DRL—can learn spatial strategies through interaction rather than explicit programming, enabling the generation of ventilation-responsive block layouts that balance environmental performance with coherent design logic.

2. Research Methodology

This study develops an AI-based layout agent that generates ventilation-responsive urban block configurations without using CFD at every iteration. The framework integrates CA, policy-gradient reinforcement learning, and a Go-inspired reward function, enabling a transition from simple local rules to learned spatial strategies.

The system operates with three consistent spatial types:

Building block: provides massing, enclosure, and airflow resistance.

Courtyard: offers low-resistance lateral ventilation space and defines openness.

Wind tower: enables vertical airflow through stack effect and can serve as a structural or circulation core.

These three elements form a minimal vocabulary for ventilation-oriented morphology and are shared across both the CA ground-floor generation and the RL stacking process.

2.1. CELLULAR AUTOMATA FOR BASE-LAYER GENERATION

A rule-based CA framework is developed to generate and evolve ground-floor spatial layouts under prevailing wind conditions. The CA follows the fundamental mechanisms of classical cellular automata, including discrete cell states, a regular grid lattice, neighbourhood-based interactions, and iterative state-transition rules(Wolfram, 1983). The specific transition rules employed in this study are originally designed by the authors to reflect ventilation-oriented spatial logic.

2.1.1. Basic rules

Each building cell is evaluated against three ventilation criteria:

(a) Bidirectional ventilation: If there are open/courtyard cells along both the upwind and downwind the presence of open or courtyard spaces both upwind and downwind, ensuring cross-ventilation potential;

(b) Single-direction ventilation:at least one courtyard and one wind tower within its eight-cell neighborhood.

(c) Dead block: If none of the above conditions are met and there are no other towers within a three-grid radius, the cell can be converted into a wind tower to serve as an auxiliary ventilation point; otherwise, it will be converted into a courtyard.

Through iterative updates, this rule set drives the CA toward a ventilation-oriented spatial equilibrium, where open spaces and wind towers align naturally with airflow paths.

2.1.2. Iteration strategies

To explore the dynamic stability of this system, three update strategies were tested.

(a) Global synchronous update (Figure 1a) produced strong oscillations between distinct patterns, revealing instability and poor convergence.

(b) Incremental asynchronous update (Figure 1b), where each change immediately triggered neighborhood recalculation, yielded more organized patterns with courtyards forming linear strips perpendicular to the wind, though wind towers tended to cluster along the downwind boundary.

(c) Stochastic asynchronous update (Figure 1c), which randomized the update order at each iteration, effectively reduced clustering and produced stable, well-distributed configurations.

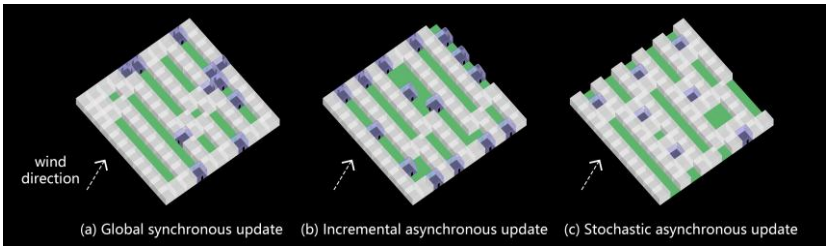


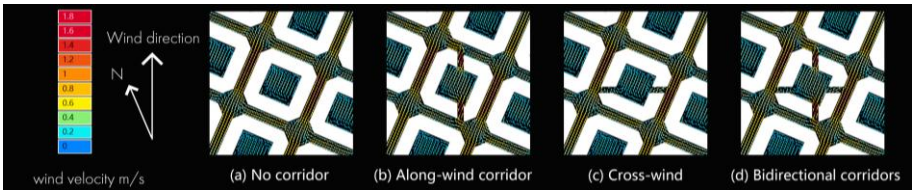
Figure 1. CA iteration strategies

This comparison shows that randomized asynchronous updating provides the best balance between order and adaptability (Figure 1c). Through iterative local interactions, each cell autonomously adjusts to meet ventilation criteria, resulting in courtyard strips aligned perpendicular to the wind and a more consistent distribution of towers and open spaces. The emergent stability of these configurations demonstrates that the CA system self-organizes in a bottom-up manner, effectively coordinating the ventilation needs of individual cells while achieving overall spatial coherence and environmental performance. These stable layouts serve as the initial conditions for subsequent courtyard connection and reinforcement learning experiments.

2.2. COURTYARD CONNECTION PATH ALGORITHM

By adjusting initial parameters, multiple CA-generated layouts meeting basic ventilation criteria were produced. However, the resulting courtyards were often isolated and oriented perpendicular to the prevailing wind, limiting their contribution to airflow pathways. Previous studies (Huang et al., 2019; Mishra et al., 2025) have shown that aligning streets or alleys with the wind direction significantly improves

ventilation, while cross-wind arrangements can create stagnation zones. And CFD simulations in the Eixample morphology compared four alley configurations—none,



along-wind, cross-wind, and bidirectional—showing that along-wind and bidirectional corridors substantially enhanced courtyard ventilation (**Error! Reference source not found.**).

Figure 2. Airflow patterns under four street-corridor configurations

To extend this principle to CA-generated layouts, we developed a Courtyard Connection Path Algorithm. Individual courtyards are first identified using DFS and then ordered along the wind vector. Adjacent courtyards are linked via BFS along shortest paths, and a weighted Dijkstra search extends the network toward upwind and downwind boundaries, accounting for traversal costs. This hierarchical, wind-oriented connection transforms fragmented courtyard patches into continuous airflow corridors, improving ventilation organization across the block layout (Figure 3).

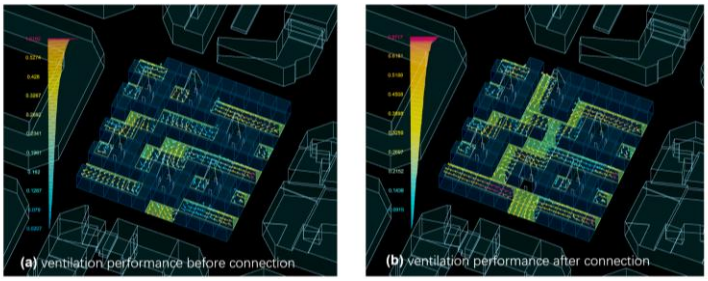


Figure 3. courtyard connection algorithm ventilation study

2.3. VENTILATION EVALUATION

The original plan was to combine CFD evaluation with a genetic algorithm (GA) to select optimal layouts. However, repeated experiments revealed two major challenges:

(1) Synchronization issues — CFD simulations are computationally intensive, making iterative optimization impractical.

(2) Nonlinear CA evolution — The final layout often deviates significantly from the initial configuration due to the nature of CA system (Wolfram, 1983), preventing the GA from reliably improving solutions; highly scored initial candidates can evolve into very different patterns.

As a result, CFD was used only for one-time evaluation of candidate layouts. Pedestrian-level (1.7 m) wind speed was converted into a comfort score using a Gaussian mapping (Feng et al., 2004) centered at 1.5 m/s ($\sigma = 0.7$), with out-of-range wind speeds penalized. The resulting score provided a lightweight metric to select

promising initial configurations. This approach demonstrates that rule-based, learning-driven strategies are more effective in highly nonlinear CA systems than attempting to couple GA or CFD iteratively.

2.4. DEEP REINFORCEMENT LEARNING: POLICY GRADIENT

2.4.1. Overview

The stable base-layer layouts generated by the CA engine serve as initial conditions for DRL stacking. A convolutional policy network, inspired by AlphaGo (Silver et al., 2016) , is trained via policy gradient methods to sequentially place blocks on a layered grid, taking an $n \times n \times 1$ occupancy tensor as input and outputting a probability distribution over feasible positions. Each episode runs until a pre-defined layer height (50), with trajectories collected for REINFORCE updates (Williams, 1992).

The reward combines courtyard quality and wind tower ventilation efficiency, integrating connectivity, openness, area, and a "liberty"-inspired enclosure measure (Fairbairn, 2010), analogous to the number of adjacent free spaces in Go, enabling multi-objective optimization of spatial quality and environmental performance using lightweight geometric heuristics rather than costly CFD simulations.

2.4.2. Stacking Logic

Although stacking actions are chosen by the RL agent, basic deterministic rules are defined to ensure structural coherence and to establish a finite and feasible action space.

In this program, cell types are defined as:

1 = block 2 = courtyard 3 = wind tower 4 = removed space 0 = empty

For each grid position (i, j), its six adjacent neighbors are denoted as U, D, F, B, L, R . Stacking rules are summarized as follows:

Table 1. Staking rules

RULES	DOWN CELL (D)	STACKBILITY CONDITION
No.1	1	Always stackable
No.2	2	Unstackable
No.3	4	Stackable if any of D's neighbors (L,R,B,F) = 1 or 3
No.4	0	Stackable if any of D's neighbors = 1 or 3
No.5	3	Stackable if neighbor exists 3 or 0 (Neighbor update (0 → 3) allows the steering of the wind tower)

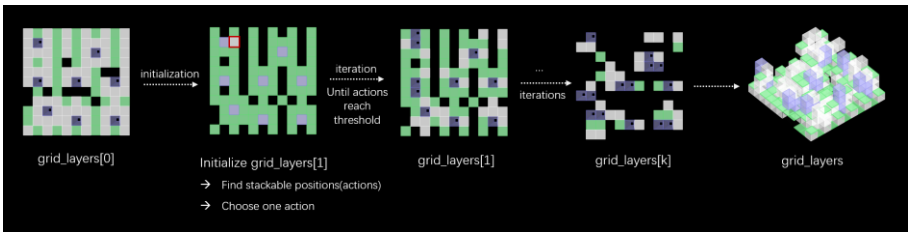


Figure 4. Stacking process

The stacking process is starting from a fixed base layer (Figure 4), then a new layer is initialized as: 1 and 4 to 2, 2 to 0, 3 and 0 remain as themselves. Then, the program start to calculate all valid positions based on stacking rules in . All valid stacking positions form the action set $\boldsymbol{\mathcal{A}}(s)$. At each step, the policy network outputs $\boldsymbol{\pi_\theta}(a|s)$; a grid position $(\boldsymbol{i},\boldsymbol{j})$ is sampled and updated to 1. When the number of stacked actions exceeds a dynamic threshold ($\boldsymbol{\theta} = \boldsymbol{a} \cdot |\boldsymbol{\mathcal{A}}(s)|$), the environment initializes a new layer and continues. The process stops after ($\boldsymbol{k}_{max}$) layers or when no valid actions remain.

2.4.3. Courtyard Reward Mechanism

The courtyard reward evaluates the environmental quality surrounding each building block. Courtyards are classified as major or minor depending on their size, with an area threshold of four cells distinguishing the two (as shown in Figure 5).

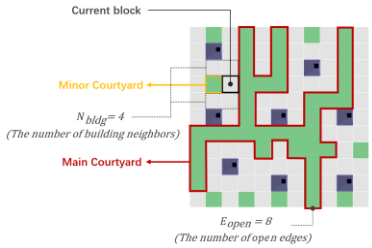


Figure 5. Diagram of courtyard types and parameters

(1) Major Courtyard Reward

Major courtyards emphasize spatial accessibility and ventilation. Their overall score is determined by connectivity and openness:

$$S_{\text{main}} = 0.5\, S_{\text{conn}} + 0.5\, R_{\text{open}},$$

Where

$$S_{\text{conn}} = \frac{N_{\text{bldg}}}{N_{\text{block}} + 10^{-6}}, \qquad R_{\text{open}} = \frac{E_{\text{open}}}{N_{\text{bldg}} + 10^{-6}}.$$

(2) Minor Courtyard Reward

Minor courtyards emphasize privacy and localized airflow. Their scores integrate the number of connected courtyards, their average area, and connectivity:

$$S_{\text{minor}} = 0.1\, N_{\text{minor}} + 0.1\, \bar{A}_{minor} + 0.8\, S_{\text{minor,conn}}.$$

(3) Combined Block Reward

The total block reward is computed as a weighted combination of the best major-courtyard score, minor-courtyard score, and area contribution:

$$R_{\text{block}} = \max(S_{\text{main}}) + S_{\text{minor}} + 0.2\,S_{\text{area}},$$

with

$$S_{\text{area}} = \min\left(\sum \frac{A}{N_{\text{bldg}} + 10^{-6}},\ 1.0\right).$$

An additional penalty term balances the proportion between total block number and courtyard area:

$$P_{\text{balance}} = \max!\,(0,\ 1 - (r_{\text{actual}} - r_{\text{ideal}})^2).$$

This reward structure balances the openness and accessibility required by major courtyards with the privacy and ventilation roles of minor ones.

2.4.4. Wind tower Reward Mechanism

Ventilation performance is known to increase when wind towers are enclosed by sufficient thermal space. Inspired by the concept of liberties in the game of Go, we evaluated each cluster of connected wind towers based on the number of surrounding "liberties" (air cells) and their enclosure conditions. Unlike Go—where a stone with no liberties is considered dead—in our case, a wind tower with no surrounding liberties represents a fully enclosed and thus most efficient configuration.

For each tower cluster, four reward conditions are defined (examples are shown in Figure 6):

(1) No liberties: highest efficiency, reward = 1.

(2) With liberties: reward = 0.5 - n / perimeter, where n is the number of liberties and perimeter is the cluster boundary length.

(3) Enclosed liberties: if all liberties are internally enclosed and their total area ≤ 8, reward = 0.5.

(4) Liberties crossing the grid boundary: lowest efficiency, reward = –0.5.

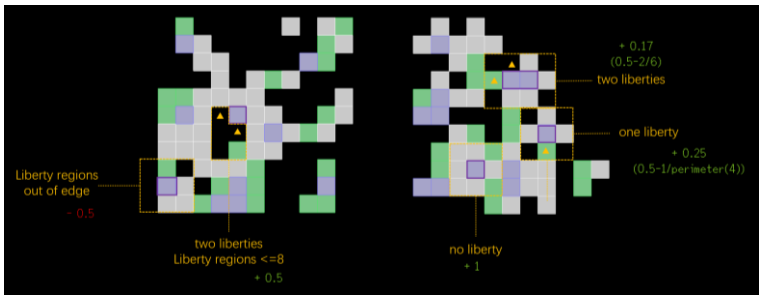


Figure 6. Wind tower Reward Mechanism

The total wind tower reward of all clusters was then normalized by the number of wind tower cells, yielding the average enclosure efficiency per tower.

3. Results and Discussion

3.1. STACKING RULES AND COURTYARD REWARD OPTIMIZATION

In early experiments, flaws in the stacking rules 3 and 4 led to rule 5 becoming dominant as stacking progressed and available blocks decreased, resulting in a stepped, tower-like structure along the wind tower's torsion direction (Figure 7a). Extending the support conditions to include both block and wind tower units (1 or 3) produced a denser, more uniform distribution. Introducing an agent revealed issues with vertical wind tower alignment, horizontal block alignment, and excessively large interlayer spacing (Figure 7b), due to repetitive action selection at each layer. After correcting these errors, the overall block density increased, and the layout became more balanced.

The next round of experiments tested different stacking thresholds. Increasing the threshold from 0.3 to 0.5 raised block density, enhancing spatial enclosure and linking external corridors to internal wind-tower courtyards (Figure 7c–d). However, excessive stacking reduced courtyard rewards, as many blocks lost direct courtyard contact across layers. Lowering the threshold to 0.3 improved courtyard connectivity but did not fully resolve inefficient stacking. To address this, stacking-related rewards were reduced and a balance penalty based on the block-to-courtyard ratio was introduced, resulting in clear improvement—average courtyard rewards increased by 1.5–4.2 (Figure 7e–f). At a higher threshold (0.5), rewards slightly improved, indicating that the agent began balancing courtyard openness with overall block density.

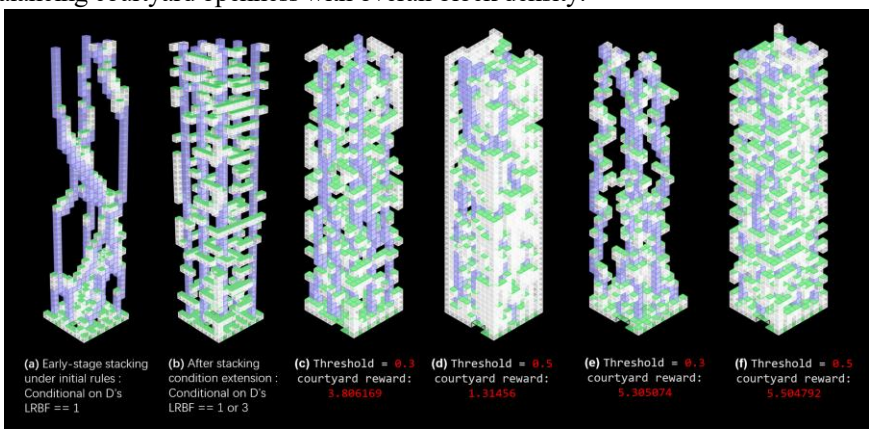


Figure 7. Courtyard reward mechanism study

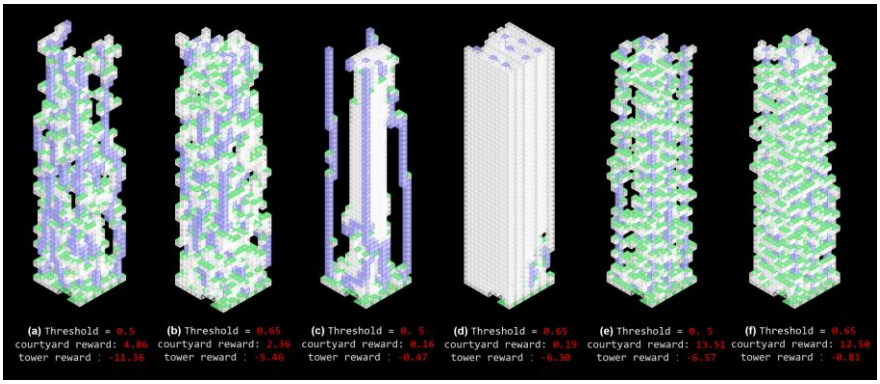
3.2. WIND-TOWER REWARD OPTIMIZATION

After optimizing courtyard quality, the wind tower reward was evaluated. In the first phase (Figure 8a), courtyard rewards remained stable but wind tower rewards were very low; increasing the threshold improved tower enclosure slightly but reduced courtyard performance, indicating poor balance between objectives (Figure 8b).

In the second phase (500 epochs), training converged but produced abnormal layouts: central towers became over-enclosed while others clustered at the corners due to a bug in boundary detection. This “liberty”-based mechanism, inspired by Go’s qi,

mistakenly treated open edges as blocked, yet demonstrated that the agent had internalized corner-prioritization strategies (Figure 8 c). After correcting the boundary detection, wind towers were enclosed more uniformly. However, recognizing out-of-bound liberties caused both tower and courtyard rewards to drop, suggesting that the learned policy still struggled to balance enclosure and openness. As training progressed, the model occasionally fell into suboptimal strategies, generating rigid grid-like patterns where wind towers were trapped between linear block arrays (Figure 8 d).

Introducing an early-stopping mechanism effectively prevented such degeneracy and improved overall balance, with courtyard rewards increasing substantially and tower efficiency stabilizing under moderate thresholds (Figure 8 e-f). Notably, when



the threshold was set to 0.65, both wind tower and courtyard rewards approached near-optimal values observed across all experiments (Figure 8 f).

Figure 8. Wind tower reward mechanism study

3.3. EMERGENT SPATIAL ORGANIZATION

Observing the final stacking results reveals several consistent spatial tendencies. Most blocks tend to cluster around wind towers, forming petal-like groupings centered on ventilation cores, while courtyards primarily occupy the grid peripheries, with some extending inward. Between adjacent layers, blocks exhibit interleaved arrangements, ensuring that each level maintains sufficient courtyard connection and airflow. Only a small number of blocks extend vertically, while the overall form develops a terraced morphology, where wind towers act as vertical anchors surrounded by layered block formations. This indicates that the agent implicitly learned a hierarchical organization that balances vertical ventilation efficiency with horizontal spatial continuity.

To further examine whether such organization truly emerges from the learned policy rather than from the stacking rules alone, a random-action baseline was introduced for comparison. Despite sharing the same placement constraints, the random strategy frequently collapsed before reaching comparable height levels: threshold-induced spatial saturation rapidly depleted available positions, preventing it from maintaining consistent block counts across layers. In contrast, the RL agent learned to preserve future buildability—strategically distributing blocks to avoid local dead-ends while maximizing reward. Even without full CFD validation, this behavioral

divergence provides a structural form of evidence: only the learned policy is capable of sustaining multi-layered, ventilated, and courtyard-connected formations. The contrast highlights that the resulting spatial coherence is not an artifact of the rules, but a direct outcome of the agent's decision-making.

4. Conclusion

This study developed a hybrid framework combining CA and DRL to generate adaptive urban block layouts under ventilation constraints. The CA provides a robust base-layer logic, ensuring airflow connectivity and initial courtyard distribution, while the DRL agent learns stacking strategies that balance spatial density and ventilation efficiency. Iterative reward optimization—particularly the Go-inspired liberty evaluation—enables the model to internalize spatial heuristics, efficiently enclosing wind towers and arranging blocks in terraced, hierarchical patterns.

However, some environmental design objectives remain difficult to quantify with unified and computationally efficient evaluation metrics, and agents trained under specific reward formulations may require retraining when design targets change. A feasible direction for future work would be to adopt multi-indicator learning frameworks combining neural evaluators and self-play-style reinforcement learning to enable reusable agents across objectives; however, ventilation-oriented training would demand large-scale CFD simulation data, which is currently computationally prohibitive. Therefore, this study adopts a geometrically lightweight surrogate evaluation system in place of CFD-based assessment. Even under this simplified framework, the DRL agent demonstrates clear strategy formation and spatial reasoning, confirming the feasibility and extensibility of the proposed approach.

Attribution Statement

ChatGPT (OpenAI, 2025) was used to refine the clarity and language flow. All analytical results, model implementations, and interpretations were solely conducted and verified by the authors.

References

Fairbairn, J. (2010). Invitation to Go. Dover Publications. https://books.google.com/books?id=ndsoAwAAQBAJ

Feng, Y.-R., Wang, A.-Y., & Zhen-Guo, L. (2004). ON THE DESIGN OF A NEW HUMAN COMFORT INDEX. In *JOURNAL OF TROPICAL METEOROLOGY* (Vol. 10, Issue 1).

Huang, Y. D., Hou, R. W., Liu, Z. Y., Song, Y., Cui, P. Y., & Kim, C. N. (2019). Effects of wind direction on the airflow and pollutant dispersion inside a long street canyon. *Aerosol and Air Quality Research*, 19(5), 1152–1171. https://doi.org/10.4209/aaqr.2018.09.0344

Mishra, N., Patra, A. K., & Penchala, A. (2025). Integrated influence of wind direction and street length on pollutant dispersion and pedestrian exposure in idealized street canyons: A CFD approach. *Journal of Wind Engineering and Industrial Aerodynamics*, 267, 106247. https://doi.org/10.1016/J.JWEIA.2025.106247

Nikolopoulou, M., & Lykoudis, S. (2006). Thermal comfort in outdoor urban spaces: Analysis across different European countries. *Building and Environment*, 41(11), 1455–1470. https://doi.org/10.1016/J.BUILDENV.2005.05.031

Runchal, A. K., & Rao, M. M. (2020). CFD of the Future: Year 2025 and Beyond. In A. Runchal (Ed.), *50 Years of CFD in Engineering Sciences: A Commemorative Volume in*

Memory of D. Brian Spalding (pp. 779–795). Springer Singapore. https://doi.org/10.1007/978-981-15-2670-1_22

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. https://doi.org/10.1038/nature16961

Wolfram, S. (1983). Statistical mechanics of cellular automata. *Reviews of Modern Physics*, 55(3), 601–644. https://doi.org/10.1103/RevModPhys.55.601

Colour Representations of Urban Space

MOHAMMED MAKKI[1], RAFAEL LUNA ZELAYA[2], LINDA MATTHEWS[3], LING KIT CHEUNG[4], NIMISH BILORIA[5], JORDAN MATHERS[6], JAMES MELSOM[7]
[1,2,3,4,5,7] *University of Technology Sydney.*
[6] *SJB*
1mohammed.makk@uts.edu.au, 0000-0002-8338-6134
[2]rafael.lunazelaya@uts.edu.au, 0000-0001-6354-3422
[2]linda.matthews@uts.edu.au, 0000-0002-2804-940X
[2]jckira@hotmail.com.hk, 0000-0001-5040-8686
[2]nimish.biloria@uts.edu.au, 0000-0003-3152-5187
[2]jmathers@sjb.com.au, 0000-0003-3803-2131
[2]james.melsom@uts.edu.au, 0000-0001-6703-6743

Abstract. The quality of urban space significantly impacts the sustainability and well-being of cities; however, assessing what defines a "good" or "bad" urban environment remains challenging due to the complexity of how these spaces are perceived and experienced. This research examines the role of colour in shaping such perceptions by analysing four urban spaces, categorised as either having high or low urban quality. Using image-based colour sampling and k-means clustering, each scene is reduced to a small palette of dominant colours that captures its chromatic character. These palettes are used to generate pixelated reconstructions of the original views, removing all typological and morphological cues so that colour becomes the sole remaining variable. Each reconstruction is translated into a pavilion that recreates the urban space through colour alone. The developed workflow links image processing, colour quantification, and fabrication, which offers a precise and repeatable method for exploring the impact of colour in urban design. The pavilions allow the public to respond to colour-only versions of familiar or unfamiliar spaces, supporting an investigation into how colour alone contributes to judgments of urban quality.

Keywords. Urban Quality, Colour Perception, K-means Clustering, RGB Proximity, Pavilion

1. Introduction

The quality of urban space plays a crucial role in shaping sustainability, well-being, and the everyday behaviour of city dwellers. Yet, as scholars have long noted, the complexity of urban environments makes their experiential qualities difficult to measure or represent (Lynch, 1960). Modern streets and public spaces are interpreted

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 769-778.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

through multiple sensory and perceptual channels, and the subjective nature of these impressions complicates any evaluation of what is considered a "good" or "bad" urbanscape. Research in environmental psychology and architectural theory has highlighted that colour is one of the most immediate and powerful visual cues influencing spatial perception, emotional response, and environmental preference (Behbudi et al., 2012; Gomes, 2019). Colour, therefore, represents a critical, but often under-examined, variable in determining how urban space is experienced and perceived.

The widespread use of cameras in contemporary cities enables viewers to engage actively with diverse urban conditions through the image. As these visual records circulate online, the pixel becomes the fundamental unit of representation, establishing new standards for indexing and quantifying the city's operations and physical characteristics. Complex urban scenes captured by Internet webcams are translated into numerical values and mapped onto a picture-plane grid of unique array patterns (Klette & Rosenfeld, 2004). Because these values describe form in terms of colour and brightness rather than line, they foreground the city's qualitative properties and support new ways of analysing its material "viewed" surfaces. In this context, digital image data has become a powerful means of collecting and interpreting urban information, with images now central to understanding and shaping the factors that define urban space quality.

Building on this understanding, the present study examines how colour alone, isolated from form, materiality, and spatial configuration, contributes to public perceptions of urban quality. This work extends a previous study that used public surveys and perception-based scoring to identify urban scenes that were consistently evaluated as high-quality or low-quality. By selecting two positively perceived and two negatively perceived urban spaces, the current research focuses on their chromatic composition and investigates whether the colours intrinsic to each scene encode recognisable signals associated with perceived urban quality. The project positions colour not merely as a surface attribute, but as a potential carrier of spatial meaning and atmosphere.

2. Background and Context

2.1. THE PERCEPTION OF URBAN QUALITY

The sustainability and longevity of the urban environment, as well as the satisfaction, comfort, and sense of belonging of its inhabitants, determine the quality of urban space (Carmona, 2010). These features of urban design encompass both physical and abstract aspects, such as auditory, visual, and tactile elements (Amin, 2002). Wayfinding, image identity, community, public life, physical fabric, and their intricacies are all included in a "good" urban environment, and visual identity is dispersed across paths, nodes, edges, districts, and landmarks (Amin, 2002). The psychological perception of space and the experiences of its inhabitants also play a role in the development of urban quality (Amin, 2002). Individuals and communities have varying opinions on urban space, rendering it necessary to examine experiences from multiple perspectives in order to learn how to enhance quality (Lynch, 1960).

Urban space quality involves multiple interconnected aspects that go beyond basic functions to include sensory, morphological, and experiential features. The physical aspects of urban form, including density, connectivity, and spatial layout, determine how people move through and perceive their environment. For instance, 'enclosure', which is shaped by the ratio of building heights to street widths, defines space and influences feelings of safety and comfort (Harvey, C., et al. 2015). Similarly, the concept of 'human scale' refers to the size, texture, and articulation of physical elements that match the size and proportions of humans and correspond to the speed at which humans walk (Long & Ye, 2019). 'Transparency', or visual permeability, is how well pedestrians can see activity beyond the street through windows and open fences (Ewing, R., et al. 2006), promoting social interaction and safety. 'Complexity' reflects visual richness from diverse architectural styles, materials, colours, and activities, engaging the senses and making places memorable. The textures and materials of surfaces, ranging from facades to pavements, provide tactile and visual cues that enhance recognition and aesthetic value (Spence, C., 2020). Ultimately, 'Imageability' is what makes a place distinctive and memorable, created by the interaction of these features, which evokes emotions and aids mental mapping (Lynch, K., 1960; McCunn, L. J., & Gifford, R., 2018). Among these qualities, 'colour' is critical as a key visual element (Qu, C., et al. 2025), as it is not only an essential part of urban aesthetics but also plays a significant role in shaping spatial perception, enhancing emotional experience, fostering social interactions, and influencing behavioural responses.

2.2. COLOUR AND ITS IMPACT IN THE URBAN SPACE

Colour plays a critical role in shaping the experiential and qualitative potential of urban environments (Mahnke, 1996). It is not merely an aesthetic attribute but a dynamic indicator of a city's materiality, activity, and ambience. The representation of a location's colour profile and its distribution across time provides valuable insights into how spaces are perceived and inhabited (Jaglarz, 2023). By analysing colour temporally, designers can identify patterns that influence comfort, safety, and character, enabling interventions that align with desired spatial qualities (Dabbagh, 2019; Gorzaldini, 2016).

Advanced imaging tools allow for precise extraction and consolidation of colour data from image stacks captured over time. This process transforms static visual impressions into predictive colour palettes that reflect the evolving chromatic conditions of a site (Matthews, 2022). For example, compressing multiple slices of an image stack into a single composite image reveals cumulative colour trends, while comparing individual slices highlights momentary variations. Such temporal mapping of colour informs decisions about material selection, façade treatments, and lighting strategies, ensuring that design responses resonate with the lived experience of urban space.

Colour analysis also extends to understanding the interplay between brightness and chromatic intensity. Variations in colour saturation and luminosity can accentuate or diminish the prominence of architectural elements, influencing how buildings relate to their surroundings (Bellia et al., 2013). This capability is particularly relevant in high-density urban contexts, where subtle shifts in colour and brightness can recalibrate spatial hierarchies and visual legibility (Minah, 2008).

Furthermore, colour serves as a proxy for activity and programmatic function. By redefining colour occurrences as quantifiable data, designers can trace correlations between chromatic patterns and urban behaviours (Chen et al., 2025; Manav, 2017). This approach reframes colour from a subjective aesthetic choice to an objective metric for spatial performance.

Ultimately, the systematic study of colour in urban environments offers a new paradigm for design and planning. It enables cities to be understood not only through their physical forms but through their temporal chromatic signatures, patterns that reveal how spaces adapt, evolve, and communicate identity. In this sense, colour becomes a powerful instrument for constructing urban atmospheres, guiding interventions that enhance both functional efficiency and experiential richness.

3. Method

The presented study builds on existing research (citation redacted for peer review stage) that identified high quality and low quality urban spaces based on the public's perception of these spaces. An online survey collected community evaluations of streetscapes, where participants rated images of 11 urban locations on 4 key qualities: beauty, comfort, safety, and ambience. Each scene was tagged "good," "neutral," or "bad" for each quality, thus translating the participants' subjective impressions into a quantified dataset. The results showcased clear correlations between the highest rated images, and the lowest rated images (figure 1). These highest-rated scenes tended to share favourable features such as varied materials and greenery, human-scale amenities like seating and street-front activations, and minimal traffic; whereas the lowest-rated scenes often lacked those qualities, instead showing monotonous surfaces, sparse vegetation, expansive open sky, and high vehicle quantities.



Figure 1. The highest rated images (left) and lowest rated images (right).

Building directly on these insights, the current study uses the highest and lowest rated images to expand the line of inquiry of urban quality, the colour composition of these images. The premise is that the palette of a place may be an important factor in how people experience urban space. Thus, the colour content of the highest and lowest

scoring images now becomes the entry point to investigate how visual colour cues contribute to perceived urban quality. By identifying the dominant colours in each group and analysing their potential influence, the project extends the original framework into the chromatic dimension of urban space.

In this context, a streamlined methodology was developed, consisting of two primary stages: First, colour sampling and palette extraction from the selected "good" and "bad" urban spaces identified in the prior study; and second, the fabrication of a pavilion that materialise those palettes in built form. Together, these stages allow the project to isolate colour as an experiential dimension and examine how it contributes to perceptions of urban quality when removed from identifiable forms and urban context.

3.1. URBAN QUALITY AND ITS PERCEPTION

The first stage focuses on identifying and distilling the dominant colours present in each of the four selected urban scenes. A colour-reduction process, through image pixelation and k-means clustering, was applied to the selected images to convert the full photographic spectrum into a small, manageable set of representative colours for each image, reflecting the tones that visually define the scene while eliminating minor variations. This approach allows the palettes to function as the sole representation of the character of each space, forming the only connection between the original urban spaces and their physical reconstructions. The method assumes that colour is not merely decorative but an important indicator of material, atmosphere, and activity within the city.

3.2. MANIFESTING COLOUR INTO A PHYSICAL SPACE

The second stage translates these extracted colour palettes into physical structures that the public can engage with. Each structure was designed as a three-dimensional "pixelation" of its corresponding image, composed of plywood pieces that collectively recreate the spatial distribution of colours in abstracted form. After each pixel was assigned a colour from the palette, the corresponding plywood elements were cut, primed, and painted using commercially available paint products whose hues match the palette colours as closely as possible. The pieces were then assembled according to their pixel positions, with no indication of the original urban forms, materials, or geometries; ensuring that the resulting structure communicates the ambience of its source image solely through colour relationships and their spatial arrangement.

4. Experiment

4.1. COLOUR PROCESSING AND PALETTE EXTRACTION

The colour composition of each of the four selected images was analysed through a pixelised reduction of each image into a regular grid of discrete colour samples (figure 2). The R, G, and B values of each sample were converted into X, Y, and Z coordinates within the CIELAB colour space (de L'Eclairage, 1976), which provides a perceptually uniform 3D representation of colour. This conversion allows colour similarity to be examined using Euclidean distances, offering a more reliable measure of how colours

relate to one another. S a result, a point cloud of chromatic data was generated for each image.

Once the colours were spatialised, the dataset was analysed using K-means clustering. Colours were grouped based on euclidean proximity, grouping similar colours to each other. For each image, the process produced clusters that reflected dominant chromatic tendencies within the represented urban scene. The cluster centre (the averaged position of all colours within each group) served as the representative colour for that cluster, and all original pixel colours within each group were replaced by their corresponding averaged colour (figure 3).



Figure 2. Pixelisation and colour sampling of the two highest rated images (left) and lowest rated images (right).

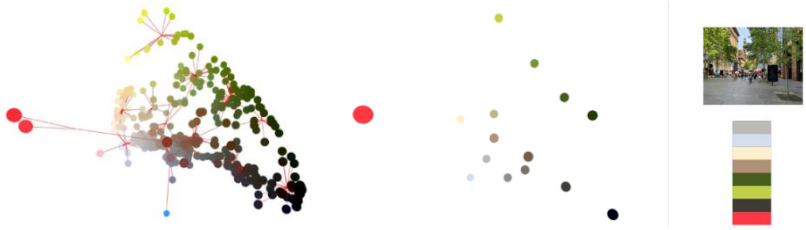


Figure 3. Plotting of the sampled colours in 3d space, and clustering to select the key colours

The K-value was determined based on two criteria, the first was to ensure the clusters allowed for outlier colours to form independent clusters, and second, the limitation imposed by the fabrication process - in which the number of colours was manageable to procure and use in fabrication. In this context, the K value was limited to a maximum of 9 for each image.

These simplified colours were then mapped back onto the original pixel grid, reconstructing the image with a reduced palette of 8-9 colours per scene. The result preserves the overall chromatic atmosphere of each urban space while removing fine-grained variations. This process yields a manageable, analytically coherent palette suitable for physical translation, and forms the conceptual foundation for the structure's colours in the fabrication stage (figure 4).



Figure 4. The colour reduction of each image into the key colours to form each image's palette

4.2. PAINT MATCHING

To translate the digital palettes into material form, commercially available colours from Dulux and Wattyl were selected due to their international commercial availability and comprehensive standard colour catalogues. Using the manufacturers' published datasheets, the RGB values for each manufacturer's catalogue were plotted in the 3d space through the same RGB to CIELAB conversion utilised in the previous sections, positioning each paint sample within the same 3D colour space used for the pavilion palettes. Each colour in the final colour palettes for each image was cross referenced against the manufacturer's colours, and the most similar colour (based on Euclidean proximity) was identified for fabrication. Once the final paint selections were confirmed, the required paint volume was calculated by measuring the surface area of each colour across all plywood panels and converting this to litre quantities based on manufacturer coverage rates (figure 5).

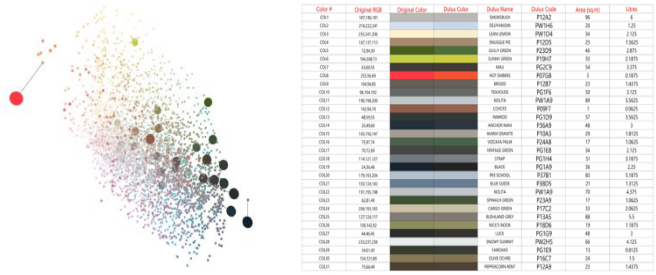


Figure 5. Proximity analysis / matching between the palette and the manufacturer's standard paints

4.3. STRUCTURE FABRICATION

The fabrication strategy centred on producing four identical pavilions constructed entirely from CNC-cut plywood components. The form of each pavilion aimed to create a simple enclosure, allowing for the user to enter the space and be surrounded by the colour composition representing each urban space (figure 6).

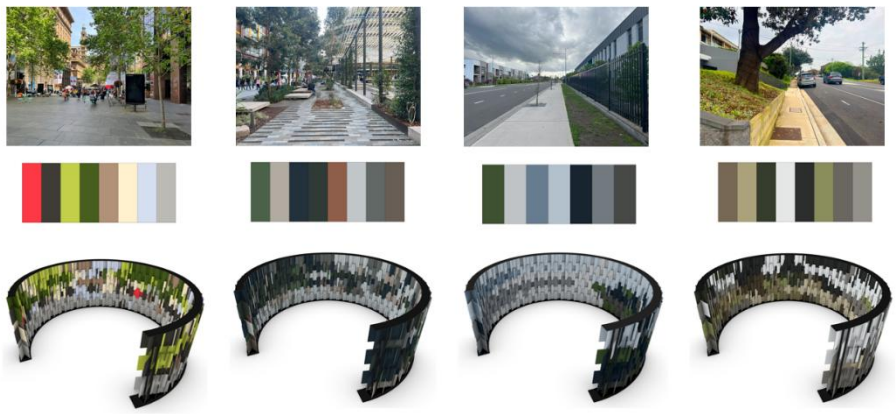


Figure 6. The mapping of each colour palette onto the pavilion panels

Each pavilion comprises a fully modular system of panels that can be assembled and disassembled without screws, nails or adhesives. The structural logic relies on interlocking grooves pre-cut into each panel, enabling rapid on-site construction, deconstruction and relocation.

The pavilion panels were organised into four groups: Base panels, Cap panels, Long wall panels and Short wall panels. The base consists of multiple pieces which, when assembled, form a circular footprint with a single opening to guide circulation. Into the cut-outs on the base, the first layer of long wall panels is inserted. The short wall panels are then slotted into both the long panels and the remaining base cut-outs, locking the first layer into a stable ring. Both long and short panels contain additional grooves on their upper edges, allowing the next layer of panels to be placed without shifting.

This stacking procedure is repeated for a total of six layers, gradually forming the full height of the pavilion. Once the wall assembly is complete, the cap panels are placed onto the final set of grooves, securing the structure. As a final reinforcement step, slender external columns are attached to the outer surface of the pavilion, increasing rigidity while maintaining the modular system's inherent reversibility. (figure 7).

5. Discussions and Conclusion

The findings of this research illustrates the potential of colour as an important factor in how urban spaces are perceived and evaluated. By reducing each selected scene to a limited palette of dominant colours, they demonstrate that chromatic information alone carries an unexpectedly strong imprint of the original environment's perceived quality. Although the pixelation process removes all morphological and typological cues, the resulting pavilion atmospheres still convey distinct tonal, emotional, and environmental characteristics. Positively perceived urban spaces tend to produce warm, varied, or verdant palettes, while negatively perceived ones often collapse into muted, grey–blue or desaturated tones.

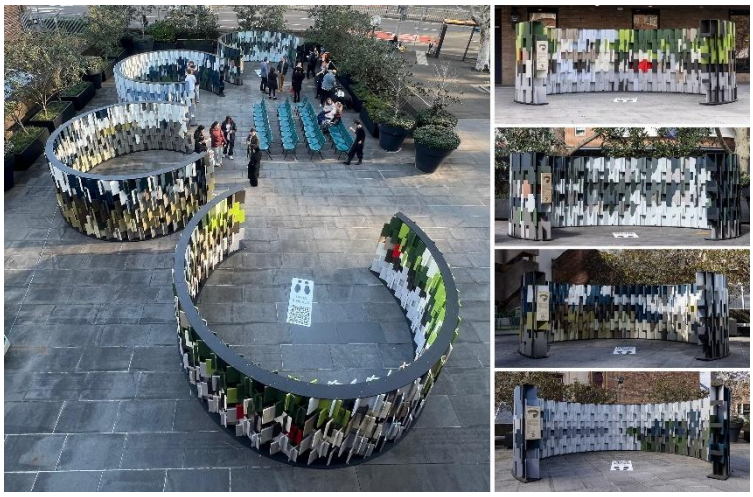


Figure 7. The constructed pavilions, one for each image

However, the study also encountered an important limitation: the available range of fabricated colours was constrained by budget, time, and reliance on commercially pre-mixed paints. While the clustering process produced precise representative colours, matching these to real-world paint catalogues required compromises, as some hues identified in the digital palette had no exact equivalent in the commercially available Dulux or Wattyl ranges.

Future research will gather visitor responses on how they perceive and react to these structures, requiring them to fill in a survey that mimics the original survey conducted to identify the best and worst urban spaces (reference redacted for peer review). This will provide insight on whether the chromatic content of an urban scene consistently encodes perceptual qualities. If strong correlations emerge, the work has the potential to inform colour-aware design strategies for streets and public spaces, contribute to government planning tools that integrate chromatic analysis, and support new frameworks for evaluating and improving urban quality (figure 8).

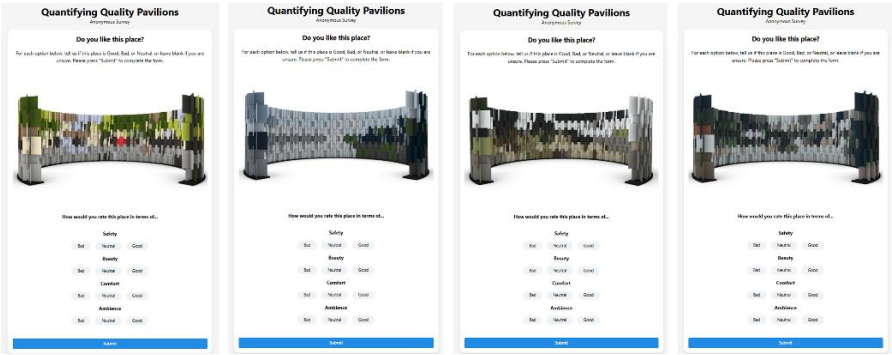


Figure 8. The survey that visitors to the pavilions will be requested to fill.

6. References

Amin, K. (2002). Urban quality and designing of spaces Case study for Nasr City, Cairo. *Proceedings of the 38th International Planning Congress*, Athens, Greece, 21–26.

Carmona, M. (2010). Contemporary public space: Critique and classification, part one: Critique. *Journal of Urban Design*, 15(1), 123–148.

Behbudi, R., Szasa, I., & Torabi, M. (2012). The Functions of Color in Urban Setting. *Aceh Development International Conference (ADIC2012)*, 26.

Bellia, L., Agresta, N., & Pedace, A. (2013). The colour of light for urban spaces. *In Colour and Colorimetry (pp. 311–317).* Editore Maggioli Editore Luogo pubbl. Rimini.

Chen, N., Wang, L., Xu, T., & Wang, M. (2025). Perception of urban street visual color environment based on the CEP-KASS framework. *Landscape and Urban Planning,* 259

Dabbagh, E. (2019). The effects of color and light on the beautification of urban space and the subjective perception of citizens. *Int. J. Eng. Sci. Invent.(IJESI)*, 8, 20–25.

de L'Eclairage, C. I. (1976). Colorimetry, *official recommendations of the international commission on illumination.* Publication CIE, 15, E1.

Ewing, R., Handy, S., Brownson, R. C., Clemente, O., & Winston, E. (2006). Identifying and measuring urban design qualities related to walkability. *Journal of Physical Activity and Health*, 3(s1), S223-S240.

Gomes, C. C. (2019). Colour functions in Urban Design: Communication, identity and user behaviour. *AIC Midterm Meeting "Color and Landscape."*

Gorzaldini, M. N. (2016). The effects of colors on the quality of urban appearance. *Mediterranean Journal of Social Sciences,* 7(5), 225–231.

Harvey, C., Aultman-Hall, L., Hurley, S. E., & Troy, A. (2015). Effects of skeletal streetscape design on perceived safety. *Landscape and Urban Planning*, 142, 18-28.

Jaglarz, A. (2023). Perception of Color in *Architecture and Urban Space*. *Buildings*, 13(8).

Klette, R., & Rosenfeld, A. (2004). *Digital geometry: Geometric methods for digital picture analysis.* Elsevier.

Long, Y., & Ye, Y. (2019). Measuring human-scale urban form and its performance. *Landscape and Urban Planning*, 191, 103612.

Lynch, K. (1960). *The image of the city* MIT Press. Cambridge MA, 208.

Mahnke, F. H. (1996). *Color, environment, and human response: An interdisciplinary understanding of color and its use as a beneficial element in the design of the architectural environment.* John Wiley & Sons.

Manav, B. (2017). Color-emotion associations, designing color schemes for urban environment-architectural settings. *Color Research & Application,* 42(5), 631–640.

Matthews, L. (2022). *Design Strategies for Reimagining the City: The Disruptive Image.* Routledge.

McCunn, L. J., & Gifford, R. (2018). Spatial navigation and place imageability in sense of place. *Cities*, 74, 208-218.

Minah, G. (2008). Colour as idea: The conceptual basis for using colour in architecture and urban design. *Journal of the International Colour Association*, 2.

Qu, C., Dong, J., & Wang, W. (2025). Bibliometric analyses of factors influencing color preferences in urban environmental spaces. *Frontiers* in Psychology, 16, 1588644

Spence, C. (2020). Senses of place: Architectural design for the multisensory mind. *Cognitive Research: Principles and Implications,* 5, 46.

Application-Oriented Automatic Identification of Urban Fabric

A Graph-Theoretic Approach to Morphological Unit Delineation

Xinru Wang[1], Yu Meng[2] and Chengyu Sun[3]
[1,2,3]*Tongji University; Shanghai Key Laboratory of Urban Regeneration and Spatial Optimization Technology*
[1]*wangx_ru@126.com, 0009-0004-1316-5581*
[2]*mengyu@tongji.edu.cn, 0009-0007-0905-2885*
[3]*ibund@126.com, 0000-0002-5686-5957*

Abstract. Urban morphology analysis in high-density cities requires spatial units that are both morphologically coherent and spatially continuous, yet many existing automatic delineation methods rely on single factors or clustering schemes that ignore spatial structure. This study asks how to define and generate Tissue Units (TUs) that better capture real urban fabric under such conditions. Our aim is to embed morphological unit theory into a graph-based, data-driven pipeline that links individual buildings to block-scale urban tissue. We construct a mixed building–plot–block graph, perform within-plot splitting with a graph neural network that combines geometric features and learnable edge affinities, and then apply hierarchical adjacency-constrained merging at the block scale to obtain contiguous TUs. An empirical study in the inner ring of Shanghai evaluates TU, plot, and regular grid partitions using multi-dimensional morphological indicators, within-unit mean distance W_mean, and variance-explained ratio R^2. Results show that TUs achieve lower W_mean and higher R^2 than the two baselines, indicating higher intra-unit consistency and clearer inter-unit separation. The proposed method offers a practical technical route for constructing boundary-explicit, attribute-stable units to support urban form analysis, renewal zoning, and policy design in complex high-density environments.

Keywords. Urban morphology, urban fabric, graph-based segmentation, graph neural network(GNN), high-density city

1. Introduction

One type of urban tissue is commonly defined as a building–space aggregate that exhibits relatively homogeneous morphological characteristics and spatial arrangement (Kropf, 2017); different types of urban tissue are layered and collaged across the city, jointly forming spatial structural units that carry urban heritage and economic activities. These units provide a fundamental basis for understanding urban

HUMANISTIC COMPUTATION AND INTELLIGENCE, Proceedings of the 31st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026, Volume 3, 779-788.
© 2026 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

spatial patterns and their evolution, and serve as key references for urban design and governance.

As urban space becomes denser and morphologically more complex, traditional tissue-identification practices based on manual interpretation and expert judgement face growing challenges in efficiency, consistency, and transferability. In response, numerous automated delineation methods have been proposed, drawing on building morphology, street networks, or combinations of multiple spatial features to enable more objective and quantitative analyses of urban tissue (Shen, Xu, & Liu, 2021; Colaninno, Roca, & Pfeffer, 2018; Guyot et al., 2021; Berghauser Pont et al., 2017).

Nevertheless, when applied to real-world urban environments, existing approaches still exhibit several structural limitations. Many methods focus on a single spatial element, limiting the integration of spatial relationships among buildings, plots, and streets. Even multi-element approaches often rely on feature-based clustering without explicit boundary constraints, making it difficult to generate spatial units that are both contiguous and clearly edge-bounded. Moreover, performance tends to degrade in highly heterogeneous areas with layered development patterns, constraining their applicability for fine-grained urban governance.

To address these issues, this study embeds Conzen's morphological unit theory into an analysis pipeline built on a building–plot mixed graph, and proposes a graph-theoretic framework for the automatic identification of TUs. Building on the block–plot structure, the method performs building-level splitting within plots and group merging at the block level in two stages, thereby generating TUs that are spatially contiguous, bounded by clear physical edges, and morphologically coherent. The resulting TU system is intended to provide a more precise analytical and managerial basis for spatial data analysis, redevelopment zoning and policy delineation in high-density urban environments.

2. Method

2.1. GRAPH-BASED FRAMEWORK FOR AUTOMATIC TU IDENTIFICATION

The notion of the Tissue Unit (TU) originates from Conzen's theory of morphological units (Conzen, 1969), which emphasizes groups of buildings that are relatively homogeneous in plan form and development history. Such units have long been considered valuable for urban form analysis and planning practice. However, in contemporary high-density cities characterized by complex plot morphology and layered redevelopment, the traditional one-to-one correspondence between TU and plot has become increasingly inadequate.

To address this limitation, we reconceptualize TU identification within a hierarchical block–plot structure. Instead of assuming that each plot corresponds to a single TU, we adopt a two-stage strategy that separates fine-grained differentiation from higher-level consolidation. Specifically, we first perform building-level splitting within each plot, and then merge the resulting groups at the block scale. This "split + merge" design aims to preserve intra-unit morphological consistency while enhancing inter-unit contrast and spatial separation.

Within this framework, TU generation is uniformly formulated using a graph-based representation. Building instances or initial building groups are modeled as nodes, spatial adjacency or shared boundaries as edges, and geometric attributes such as height, volume, and shape as node or edge features. This formulation offers two key advantages. First, graph partitioning naturally enforces spatial contiguity through connectivity constraints, avoiding the fragmentation often observed in conventional clustering approaches. Second, the explicit encoding of adjacency relations, shared boundaries, and multidimensional morphological attributes enables more stable characterization of both intra-TU consistency and inter-TU separability in dense and heterogeneous urban fabrics.

Building on this idea, we propose a two-stage, graph-based framework for TU generation, which is elaborated in Sections 2.2 and 2.3, respectively:

- Split within Plots: Within each plot, we construct a geometry-aware building relation graph, use a GNN to predict pairwise TU affinity between buildings, and apply an MST-based graph partition with an explicit no-split option to obtain building groups belonging to the same TU.
- Merge within Blocks. At the block scale, we construct a region adjacency graph over initial building groups and apply a hierarchical adjacency-constrained grouping strategy that jointly considers attribute similarity, boundary connectivity, and intra-group heterogeneity, so as to obtain spatially contiguous and semantically coherent TUs.

2.2. STAGE I: SPLIT WITHIN PLOTS

At the plot level, TU identification is formulated as a learnable graph partitioning problem with strong geometric priors. The primary objective of this stage is to determine whether buildings within the same plot belong to a single TU or to different TUs, and to automatically partition the plot in the latter case, while avoiding unjustified over-segmentation.

For each plot, all building instances are extracted as graph nodes, forming a node set V. Edges are constructed based on spatial adjacency between buildings, yielding an intra-plot building graph G = (V, E). Each node and edge is associated with a set of geometric features (Table 1) that explicitly encode building morphology and plot-level constraints.

Given the complexity and noise inherent in high-density urban environments, simple shape descriptors alone are often insufficient to capture meaningful grouping patterns. To strengthen the learning process, we incorporate height similarity and footprint area similarity as additional scalar features, which are injected into the loss function as soft geometric priors. These constraints encourage buildings with similar physical characteristics to be grouped together, while still allowing data-driven decisions.

On top of this graph representation, we adopt a graph neural network (GNN) to model the building relation graph in an end-to-end manner. Node and edge features are first embedded into a latent space using multi-layer perceptrons (MLPs), followed by a three-layer graph convolution that aggregates local contextual information within

each plot. Based on TU annotations, pairwise labels indicating whether two buildings belong to the same TU are automatically derived, and the network is trained using a cross-entropy loss. The model outputs, for each adjacent building pair, the probability of belonging to the same TU or to different TUs.

At inference time, these predicted probabilities are used to construct an affinity graph within each plot. Final building groups are obtained through MST-based graph partitioning with an explicit no-split option. When the predicted confidence for all candidate cut edges is insufficient to support a reliable split, the entire plot is retained as a single unit. This design balances learnable grouping with strong geometric constraints, improving the robustness and stability of within-plot TU identification in complex urban settings.

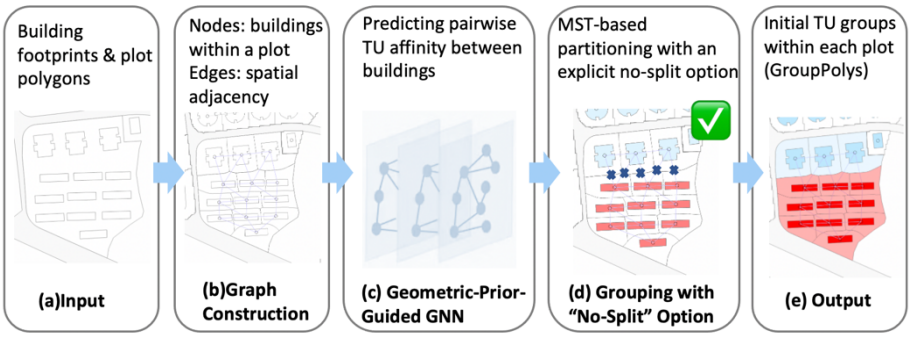


Figure 1. Technical workflow of Split within Plots

Table 1. Indicators and definitions (building level and unit level)

Indicators	**Definitions**
Area_building	Footprint area of the building polygon
Volume_building	Total built volume of the building
Height_building	Building height
Length_building	Perimeter of building footprint
Ch_compactness_building	Convex hull compactness
Cir_compactness_building	Circular compactness
Elongation_building	Plan elongation ratio
Orientation_building	Main orientation angle of the building footprint
Floor Space Index (FSI)	Sum_ volume_building /Area_unit
Ground Space Index (GSI)	Sum_ area_building /Area_unit
(Berghauser Pont & Haupt, 2010)	
Height_unit	FSI/ GSI
Length_unit	Sum_length_building /Area_unit

2.3. STAGE II: MERGE WITHIN BLOCKS

At the block scale, Stage II consolidates the initial building groups generated in Stage I in order to recover complete TUs that may span multiple plots. Rather than clustering all buildings globally, this stage is formulated as a hierarchical grouping process with explicit spatial adjacency constraints, ensuring that merging operations remain spatially continuous and morphologically meaningful.

For each block, a Region Adjacency Graph (RAG) is constructed, in which nodes represent initial building groups (GroupPolys) and edges indicate shared boundaries between adjacent groups. Each node is associated with an aggregated feature vector describing the morphology and development intensity of the group, including average building height, total built volume, floor space index (FSI), ground space index (GSI), and related attributes (Table 1).

Based on this representation, a merge scoring function tailored to TU identification is designed by jointly considering three factors. First, attribute similarity encourages the merging of groups with comparable physical characteristics. Second, boundary connectivity, measured as the proportion of shared boundary length relative to the smaller perimeter, captures the geometric continuity and strength of spatial contact between groups. Third, the increase in intra-group heterogeneity after merging is used as a penalty term, discouraging the fusion of groups with markedly inconsistent attributes.

The merging process follows a bottom-up hierarchical strategy. At each iteration, the adjacent group pair with the highest merge score exceeding a predefined threshold is selected and merged, after which group attributes and adjacency relations are updated accordingly. The process terminates when no candidate merge satisfies the threshold condition. This Hierarchical Adjacency-Constrained Grouping (HAGC) framework does not assume a predefined number of TUs; instead, it allows TUs to grow adaptively within each block based on local adjacency relations and rich morphological attributes, resulting in spatially contiguous, internally coherent, and mutually distinct units.

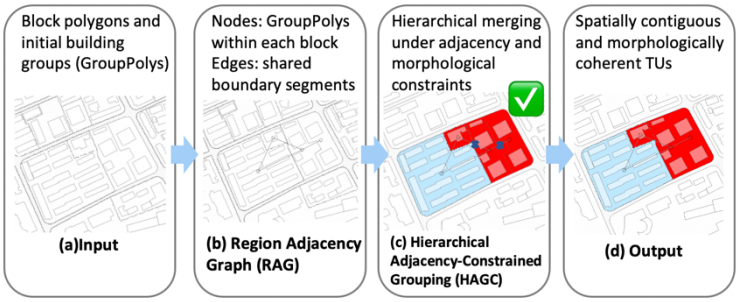


Figure 2. Technical workflow of Merge within Blocks

3. Case Study

3.1. STUDY AREA

This study uses the area within the Inner Ring Road of Shanghai as the study area (Fig.

3). It is a typical high-density central urban district, characterized by high development intensity, strong functional mixture, and a superposition of heterogeneous urban tissues formed across multiple development periods, leading to complex built forms and spatial structures.

On a harmonized base map, we construct three core spatial layers: building polygons with height and other morphological attributes, block polygons delineating street-block boundaries, and plot polygons nested within blocks. Together, they form a hierarchical block–plot–building structure. The study area combines rich tissue diversity with high parcel complexity, providing a representative testbed for evaluating the proposed TU generation framework in a real high-density urban context.

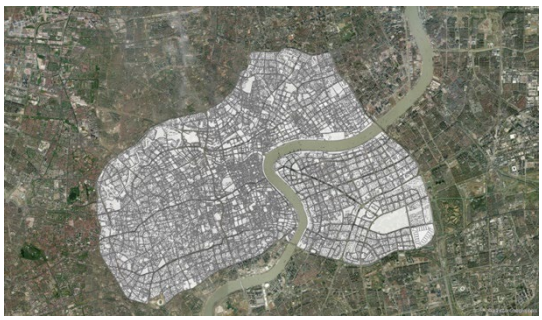


Figure 3. Study area

3.2. RESULT

3.2.1. Performance of the Within-Plot Splitting Model

This experiment evaluates the effectiveness of the proposed within-plot splitting model in Stage I, with a particular focus on its ability to correctly identify whether adjacent buildings belong to the same TU. To this end, we compare three edge-affinity modeling strategies under a unified experimental setting.

All methods operate on the same intra-plot building graph G=(V,E) and share an identical MST-based partitioning back-end, ensuring that performance differences can be attributed solely to the affinity prediction strategy. The rule-based baseline relies on heuristic thresholds on geometric attributes such as height and footprint area. The EdgeMLP baseline introduces learnable pairwise modeling based on geometric features but does not incorporate graph-based neighborhood aggregation. In contrast, the proposed GNN explicitly leverages local graph structure to model contextual building relationships.

Performance is evaluated at both the edge level and the grouping level. At the edge level, we report the area under the ROC curve (AUC) to assess the model's ability to distinguish same-TU and different-TU building pairs. At the grouping level, we report F1-score and Adjusted Rand Index (ARI) to evaluate the quality of the resulting within-plot building groups.

As shown in Fig. 4, the GNN-based model consistently outperforms both baselines. It achieves the highest AUC, indicating more reliable pairwise affinity prediction, and yields superior F1 and ARI scores after MST-based partitioning. Compared with the rule-based approach, the learnable models substantially improve robustness under complex geometric conditions, while the performance gap between EdgeMLP and GNN further highlights the importance of explicitly modeling local graph context. These results demonstrate that the proposed GNN-based splitting model provides more reliable initial building groups for subsequent block-level merging.

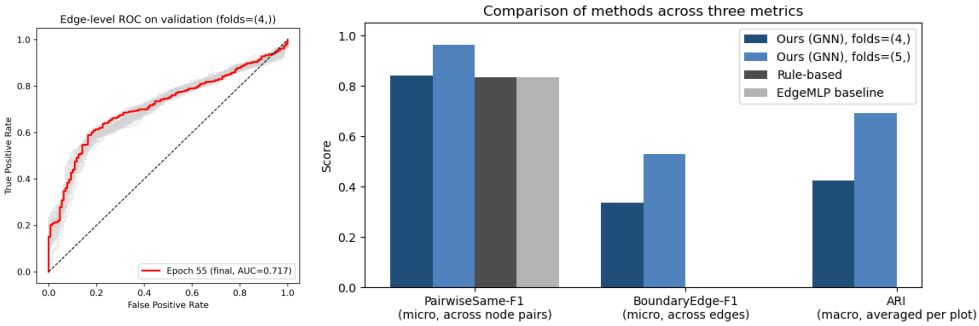


Figure 4. Performance of within-plot splitting under different edge affinity models.
(a) Edge-level AUC on the validation fold.
(b) Clustering quality in terms of pairwise F1 (micro), boundary-edge F1, and ARI (macro) .

Figure 5 presents qualitative results of the proposed within-plot splitting method. Across plots with varying densities and morphologies, the predicted TU groups (highlighted footprints) form coherent partitions that align with the underlying urban structure. Crucially, the model not only separates clearly differentiated building clusters, but also retains "no-split" configurations in homogeneous tissues such as sports complexes or university campuses, avoiding spurious boundaries where buildings are diverse individually but unified as a single urban ensemble.

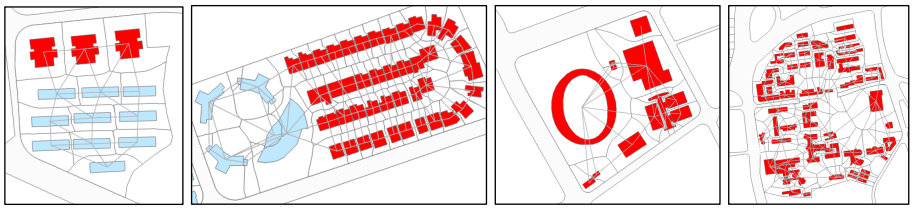


Figure 5. Qualitative examples of within-plot TU splitting results.

3.2.2. Performance of the TU Merging Algorithm

Beyond the aggregate metrics reported above, we further analyze the behavior of the proposed HAGC during the merging process by inspecting score–step curves for representative blocks. In different blocks, these curves exhibit two typical patterns. For block 2536, the curve contains only two merges with clearly high scores, after which all candidate edges fall below the threshold τ. This indicates that, given the initial partition, the interior of this block already shows strong internal consistency, and the algorithm can terminate after performing only a few "obviously necessary" merges. In contrast, the curve for block 3351 shows that the first three merges achieve markedly higher scores than the subsequent ones and can be regarded as high-confidence operations; the scores of the following merges gradually approach τ and correspond to "second-best" merges under threshold control(Fig. 6). Taken together with the visual examples, these patterns suggest that HAGC prioritizes highly beneficial merges while using the threshold mechanism to suppress marginal merges, thereby adapting the partition granularity across blocks and maintaining the stability of the final grouping.

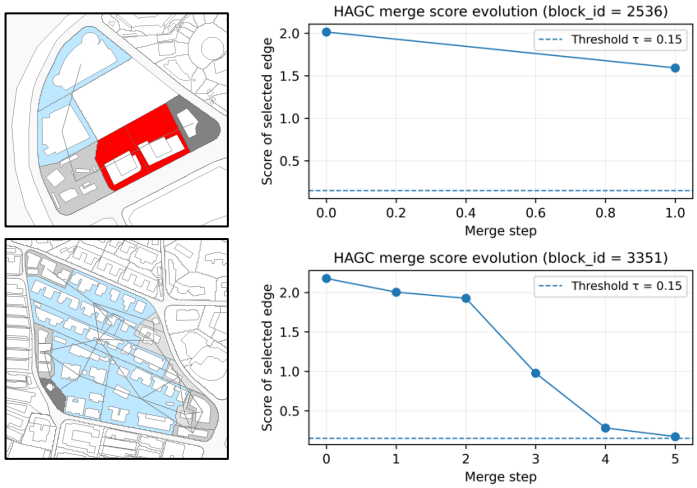


Figure 6. Spatial results and merge-score evolution for two representative blocks

3.2.3. Intra-Cluster Consistency and Inter-Cluster Separability

To evaluate the effectiveness of different spatial units in organizing building morphology, we partition the set of buildings according to three units—TU, Block, and Grid(200m)—in a seven-dimensional, standardized morphological feature space (height, volume, area, ch_compactness, elongation, cir_compactness, shape_length). For each unit, we compute the average Euclidean distance from buildings to the morphological centroid of that unit, denoted as W_mean , as a measure of within-cluster dispersion (smaller W_mean indicates higher intra-cluster similarity). In parallel, we perform between–within variance decomposition for each morphological attribute and obtain the proportion of variance explained, R^2, where larger R^2 indicates that the partition better explains morphological differences among buildings.

In terms of within-cluster dispersion (Fig. 7(a)), the distribution of W_mean for

TUs is markedly shifted downward, with a median of 1.16, compared to 1.92 and 1.85 for Blocks and Grids, respectively. This implies that, in typical cases, the average morphological difference among buildings within a TU is only about 60% of that within a Block, indicating that TUs exhibit higher intra-cluster consistency in the multi-dimensional morphological space.

Regarding variance explanation (Fig. 7(b)), the comparison of R^2 across the three partitions shows that TUs achieve the highest explained proportion for all morphological attributes, including height, volume, footprint area, and shape descriptors. Averaging R^2 over the seven attributes yields an average of about 0.46 for TUs, substantially higher than 0.23 for Blocks and 0.09 for Grids. In other words, under the TU partition, roughly half of the variance in building morphology can be attributed to differences between TUs, whereas Blocks and regular Grids explain only about one quarter and one tenth, respectively. Taken together, these results support the validity and advantage of TU as a spatial unit for organizing buildings, from both intra-cluster consistency and inter-cluster separability.

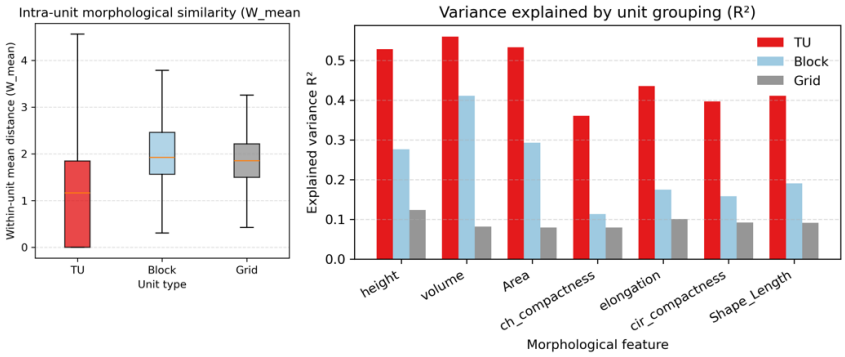


Figure 7. Intra-unit morphological similarity and variance explained by unit grouping

4. Conclusions and Future Work

In this study, targeting the problem of automatic fabric recognition in high-density urban contexts, we embed Conzen's morphological unit theory into a building–plot mixed-graph analysis framework and propose a two-stage TU generation method composed of "Split within Plots" and "Merge within Blocks." At the plot scale, we perform learnable graph partitioning on the building relation graph by combining graph neural networks with strong geometric priors; at the block scale, we adopt a Hierarchical Adjacency-constrained Grouping (HAGC) strategy that adaptively controls the merging strength while preserving spatial continuity. This pipeline enables the automatic construction of TUs from individual buildings and provides a feasible approach to generating tissue units with clear boundaries and stable attributes in real-world urban environments.

Within the study area and under the adopted indicator system, TUs exhibit lower within-unit W_mean and higher variance-explained ratio R^2 on multi-dimensional morphological indicators than conventional plot units and regular grids, indicating

higher intra-cluster consistency and stronger inter-cluster separability. It should be emphasized that these analyses do not deny the generality of the modifiable areal unit problem. Rather, given a fixed study area and multiple candidate zoning schemes, they compare which partitioning yields more stable and compact patterns in commonly used morphological indicators, thereby providing quantitative evidence for selecting the spatial units that are more appropriate for morphological analysis.

Future work can be extended in three directions: (i) scaling up to larger areas and more diverse plot types to validate the generalizability of the TU generation framework across different urban districts; (ii) further enriching graph features and prior constraints, and introducing more expressive graph models and uncertainty representations to improve robustness in highly heterogeneous areas; and (iii) coupling TUs with multi-source data such as population, land-use, and transportation information to explore their potential in renewal zoning, policy unit delineation, and urban design regulation.

Acknowledgements

This study is supported by a project of National Natural Science Foundation of China (Grant No. 52478030)

References

Berghauser Pont, M., Stavroulaki, G., Gil, J., Marcus, L., Serra, M., Hausleitner, B., Olsson, J., Abshirini, E., & Dhanani, A. (2017, July). Quantitative comparison of cities: Distribution of street and building types based on density and centrality measures. In Proceedings of the 11th International Space Syntax Symposium (SSS11), Lisbon, Portugal.

Berghauser Pont, M. Y., & Haupt, P. (2010). Spacematrix: Space, density and urban form. Rotterdam, The Netherlands: NAi Publishers.

Colaninno, N., Roca, J., & Pfeffer, K. (2011, September). Urban form and compactness of morphological homogeneous districts in Barcelona: Towards an automatic classification of similar built-up structures in the city. In ERSA conference papers (No. ersa11p769). European Regional Science Association.

Conzen, M. R. G. (1969). Alnwick, Northumberland: A study in town-plan analysis. London, England: Institute of British Geographers.

Guyot, M., Araldi, A., Fusco, G., & Thomas, I. (2021). The urban form of Brussels from the street perspective: The role of vegetation in the definition of the urban fabric. Landscape and Urban Planning, 205, 103947.

Kropf, K. S. (2017, July). Bridging configurational and urban tissue analysis. In Proceedings of the 11th International Space Syntax Symposium (SSS11), Lisbon, Portugal.

Shen, Y., Xu, Y., & Liu, L. (2021). Urban texture analysis from the perspective of network percolation. Urban Planning Forum, 2021(5), 40–48.

INDEX OF AUTHORS

A

Abdelmohsen, Sherif 2-421
Abe, Michihiro 2-277
Adhikary, Gautam 1-51
Adjeisah, Micheal 1-395
Agay-Shay, Oren 3-329
Ahmad, Lina 2-359
Ahmed, Mazen 2-421
Ahmed, Mohamed Aboeloyoon 2-71
Ahmed, Tarek MF 2-71
Akbarian, Shadi 1-357, 2-543
Akiyama, Tomoki 3-357
Aksöz Balzar, Zeynep 3-757
Al-Qattan, Emad 1-571
Alaçam, Sema 3-655
Alkadri, Miktha Farid 2-533
Aman, Jayedi 3-235
Andaji, Ali 1-357
Ang, Yu Qian 2-61
Anton, Ana 2-687
Arlene, Vania 3-635
Aş Çemrek, Handan 3-71
Asaf, Ofer 2-471
Ataman, Cem 3-245
Aung, Ye Yint 3-267
Awaji, Hiroki 2-277, 2-503
Ayaz Erdag, Ayca 1-71
Aydin, Burak 2-153

B

Bai, Zishen 1-175
Balzar, Mark 3-757
Bao, Ding Wen 2-625, 3-31
Bao, Luming 1-337
Bao, Yanlin 3-737
Barna, Gergely Peter 2-143
Bassing, Stefan 1-113
Bawaskar, Devashree 2-215
Bektas, Yagmur 2-205
Bhagat, Puja 3-409, 3-645
Bi, Kang 2-257
Biloria, Nimish 3-769
Boim, Anna 1-539
Bokos, Georgios-Marios 3-277
Bongwirnso, Umaru Mohammed 2-123
Booranamaitree, Warissara 2-11
Booth, Peter 2-585
Bo, Tianyi 1-377, 3-307
Bu, Jiachen 3-747
Burke, Adam 1-295
Búš, Peter 1-559, 3-153
Byun, Sanghoon 3-717

C

Cai, Charles Yushi 1-549
Cai, Wenqin 3-93
Cai, Yushu 2-11
Calixto, Victor 1-103
Cao, Miaoying 3-41
Cao, Peiling 1-395
Cao, Shuqi 1-93
Carpo, Mario 1-205
Carrasco, Oriol 2-595
Çekmiş, Aslı 3-543
Çetin, İsmail 3-543
Chadha, Trishla 3-707
Chai, Hua 2-645
Chang, Ching-Chih 1-457
Chang, Jess 3-727
Chang, Jian 1-395
Chang, Teng-Wen 1-457, 3-143
Chang, Tianxin 3-387
Chang, Yun-Tsui 1-713
Chan, Ka-Chun 2-697
Chef, Quentin 2-595
Chen, Ching Yen 2-369
Chen, Feixue 3-163
Cheng, Sifan 3-501
Cheng, Tsung Wei 2-369
Chen, Han-Mei 2-307
Chen, Jiajian 3-439
Chen, Jielin 1-1, 2-1, 3-1
Chen, Meizhu 1-315

Chen, Sihua 1-611
Chen, Tzu-Ying 2-595
Chen, Weihao 1-315
Chen, Wen-Xin 3-133
Chen, Xi 3-297
Chen, Xing 1-305
Chen, Zhongyao 1-601
Chen, Zi-Ru 3-133
Chen, Zijie 1-81
Chen, Ziyi 2-21, 3-93
Cheraghi, Negin 3-397
Cheung, Ling Kit 3-419, 3-769
Chien, Sheng-Fen Nik 1-1, 2-1, 3-1
Chien, Wei Jiun 2-185
Chi, Hao-Yun 3-317
Choudhury, Pradipta Roy 3-51
Chung, Yu-Hua 2-657
Codarin, Sara 1-477
Cole, Laura B 3-235
Colletti, Marjan 2-717
Crolla, Kristof 2-707
Cui, Qinyu 1-233

D

Dai, Anni 2-81
Dai, Donghui 1-467, 3-307
Dai, Shiqi 3-613
Danahy, Patrick 1-243
De Wolf, Catherine 2-267
Deng, Yuan 1-693
Deng, Yuxuan 3-93
Dillenburger, Benjamin 2-687
Ding, Peixue 1-601
Dong, Zhiyong 3-153
Dritsas, Stylianos 2-411
Duan, Chengbi 1-509
Duvenhage, Gert 2-389
Du, Xusheng 1-417, 1-427, 2-11
Du, Yuqing 3-175

E

Eidner, Fabian 2-81
Elsgaard, Niklas 2-605
Enzmann, Philipp 2-575
Erhan, Halil 2-41
Ernst, Cameron 1-407
Estrada, Rebeca Duque 2-595
Etoh, Minoru 3-491

F

Fahmy, Liza 2-71
Fang, Zheng 1-367, 3-553
Fan, Zhirui 2-645
Feng, Yunning 2-553
Finkelstein, Ofer 2-471, 2-481
Foged, Isak Worre 2-605
Fritsch, Lena 2-133
Fukuda, Hiroatsu 2-257
Fukuda, Tomohiro 1-265, 1-611, 3-491
Fu, Rong 3-185

G

Gambardella, Camilo Cruz 1-103
Gan, Yee-Siang 2-287
Gao, Naixiang 1-143
Gao, Tianyi 2-615
Gao, Wen 1-21
Gao, Yingzhou 1-315
Gaudillière-Jami, Nadja 2-133
Gebhard, Lukas 2-687
Geijer, Cecilia 2-205
Ghadge, Supriya 2-215
Gholamifard, Milad 2-563
Giridharan, Karthick Bangaru 2-103
Glaros, Alesandros 2-41
Globa, Anastasia 3-175
Goepel, Garvin 2-707
Golan, Lilu 2-339, 3-329, 3-449
Gomez, James 3-469
Gondo, Tomoyuki 2-503
Gönenç Sorguç, Arzu 1-71
Gong, Hengyu 3-737
Gonzalez, Edith Anahi 2-81
Goto, Hiroaki 3-357
Grobman, Yasha 2-339, 3-449
Grobman, Yasha J 3-329
Guerra, Daniel 1-243
Guimbretiere, Francois 1-295
Guo, Hanzhe 1-305
Guo, Kening 1-175
Guo, Xiaoxi 3-685

Guo, Yanan 2-595
Guo, Yuxuan 3-41
Guo, Zhaonan 2-625
Guo, Zhe 1-305
Güzelci, Orkan Zeynel 3-655

H

Haiman, Eduard 3-429
Han, Aize 3-93
Hannam, Marlin 1-447
Han, Yunsong 3-163
Han, Ziyang 3-521
Haque, Md Obidul 1-641
Harsono, Kevin 2-287, 3-267
Hasada, Hiroyuki 3-357
Hayashi, Sei 2-503
He, Haoyang 3-563
He, Jingmei 1-631
He, Pan 2-123
Herr, Christiane M 1-81, 2-123, 3-113
He, Yuhao 3-103
He, Zhengcheng 1-205, 1-275, 1-489, 2-615
Hildebrandt, Harrison 2-595
Hoberman, Chuck 1-367
Holloway, Adam 2-93
Hong, Chen 1-467, 3-307
Hong, Tzu-Chieh Kurt 1-407, 1-519
Honma, Yudai 1-215, 3-357
Hosseini, Seyyed Sahand 2-543
Hou, June-Hao 1-31, 1-581
Hou, Zheng 1-467, 3-307
Hsieh, Yu-Tse 1-581
Hsu, Hao 1-255
Hua, Hao 1-509
Huang, Hao 1-591
Huang, Jing 3-307
Huang, Jiqian 3-685
Huang, Lee-Su 1-477, 2-317
Huang, Sheng-Yang 3-737
Huang, Shuyu 3-695
Huang, Siyi 1-143
Huang, Weixin 1-601
Huang, Yerong 2-431
Huang, Yihan 3-367
Hu, Anqi 1-265
Hughes, Eric 2-389
Hung, Kwong Lau 3-665
Hu, Yusong 1-559

I

Ichniowski, Carolyn-Nelle 2-133
Ikeda, Yasushi 1-631, 2-51, 2-277, 2-379
Imai, Hiroki 3-357
Imam, Mohamed 2-41
Im, Su Been 1-663
Işıkdağ, Ümit 3-71

J

Jamil, Laiba 3-479
Jangid, Chethana M 3-51
Jeng, Taysheng 1-255
Jia, Muxin 1-703, 3-573
Jiang, Leyuan 2-635
Jiang, Maowei 3-153
Jiang, Xincheng 2-615
Jian, Yi 3-583
Ji, Guohua 1-93
Jin, Shitao 1-693
Jin, Yanzi 2-21, 3-93
Jipa, Andrei 2-687
Joffe, Narkis 3-329, 3-449
Joyce, Sam 1-653
Jun, Han Jong 1-185, 1-663

K

Kahlon, Yuval 3-113
Kahnamouei, Zohreh Hassanpour 2-41
Kang, Bumjoon 3-717
Kannenberg, Fabian 2-595
Kavak, Ayşe Ebrar 3-655
Kaynarkaya, Sena 3-543
Kazi, Armaan 1-243
Kepczynska-Walczak, Anetta 3-225
Khademi, Seyran 3-727
Khajehee, Arastoo 2-277
Kiany, Kiarash 3-593
Kim, Frederick Chando 1-61, 2-61, 3-459
Kim, Hanjun 2-635

Kim, Jong Bum 1-641, 3-235
Kimm, Geoff 1-103
Kim, Myoung Jin 1-185
Kimura, Toshiaki 2-503
Kirschnick, Lukas 2-575
Kiuchi, Toshikatsu 2-143
Knippers, Jan 2-81, 2-235, 2-595
Kobayashi, Hiroto 2-493
Kobayashi, Kosuke 3-257
Koehler, Daniel 1-285
Koh, Immanuel 1-195
Kolhhaussen, Gustav 2-133
Kongkaeo, Athiwat 1-417
Konstantinou, Alexandros 3-277
Körner, Axel 2-81, 2-235
Kotnik, Toni 3-665
Kotrozou, Athina 2-235
Kwang, Kai Jie 2-235
Kyriazis, Apostolos 3-479

L

Lam, Kahong 2-297
Lam, Katey 3-235
Larianovsky, Pavel 2-461, 2-471, 2-481
Laz Banti, Anindita 3-235
Leder, Samuel 2-349
Lee, Guan 2-717
Lee, Jieun 1-621
Lee, Keith J. 3-11
Lee, One Jae 2-327
Lee, Yocheng 2-687
Lei, Zhen 1-499
Leong, Sheu Wen 2-61
Leung, Pok Yin Victor 2-399
Liang, Kai-Hsiang 3-133, 3-205
Liang, Li Yee 1-549
Liao, Ryan 2-245
Li, Guanhong 1-437
Li, Hongyu 1-223
Li, Jia-Rong 1-457, 3-143
Li, Jiayi 1-437
Li, Jingxuan 1-275, 2-615
Li, Jingya 3-357
Li, Juntao 1-559
Li, Li 2-441
Li, Linxue 3-757
Lim, Lynus 1-195
Lin, Bo-Young 2-225
Lin, Che-Wei 1-387
Lin, Chia-Tsan 2-657
Lin, Chieh-Jen 1-529
Lindeman, Robert W. 3-367
Lindstam, Otto 2-595
Lin, Sheng-Yang 2-31
Lin, Sihui 1-163
Lin, Xiaojing 1-113
Li, Peilin 3-521
Li, Peiwen 1-133
Lipkowitz, Gabriel 1-549
Li, Shuyang 3-563, 3-625
Li, Suming 2-553
Li, Te 3-93
Liu, Jie 1-1, 2-1, 3-1
Liu, Jingyang 2-523, 3-21, 3-349
Liu, Jinzhou 1-489
Liu, Minjuan 1-123
Liu, Pengyuan 3-625
Liu, Shien 1-41
Liu, Xin 1-223
Liu, Xun 1-437
Liu, Yiming 2-123
Liu, Yu Qian 3-31
Liu, Yu-Jie 3-133
Liu, Yuan 3-61
Liu, Yunlong 3-625
Liu, Yuxi 2-175
Liu, Yuxuan 3-707
Liu, Zidong 1-285
Li, Xuechen 3-553
Li, Yuhan 2-307
Li, Yuhao 3-297
Li, Yujiao 3-81
Li, Yuqian 1-315
Li, Zhixian 3-215
Llabres-Valls, Enriqueta 3-737
Lloret-Fritschi, Ena 2-267
Loh, Paul 2-195
Løkke, Simonee 2-133
Lombardi, Davide 3-21
Long, Miao 2-523
Lo, Tian Tian 3-103,

3-287, 3-297, 3-583
Loy, Wei Win 2-195
Lu Chiang, Ting 3-317
Lu, Bo-Xuan 1-31
Lukosch, Heide K. 3-367
Luo, Jianing 2-93
Lu, Shuai 1-123, 3-521
Lu, Xinlin 1-295
Lyu, Xintong 3-613, 3-685, 3-695

M

Ma, Haoran 1-233
Mai, Inka 2-513
Makki, Mohammed 1-1, 1-447, 2-1, 3-1, 3-409, 3-419, 3-469, 3-479, 3-769
Ma, Luyao 2-523
Man, Haotian 2-635
Manjunath, Abhijeet 3-377, 3-531
Mao, Mengdi 2-717
Ma, Ruiqi 1-123
Maruyama, Junya 3-357
Mateljan, Marija 3-727
Mathers, Jordan 3-769
Mattei, Gabriele 1-387
Matthews, Linda 3-769
Ma, Zanwen 3-613
Ma, Zimeng 2-297
Mehrotra, Abhishek 2-327
Melsom, James 3-769
Memis, Ilayda 3-113
Menges, Achim 2-81, 2-163, 2-235, 2-349, 2-431, 2-595
Meng, Yu 3-675, 3-779
Meraki, Mehmet Efe 2-153
Merten, Aaron 2-575
Meskawi, Aya 3-707
Miao, Zhuang 1-591
Mitropoulou, Ioanna 2-267
Mitsuda, Eisuke 2-503
Mi, Xue 1-395
Mizuta, Yukika 1-621
Mottaghi, Esmaeil 2-41
Mubesher, Noor 3-707
Mueller, Caitlin 3-11
Mukai, Ayumi 1-215, 3-357
Mu, Nanshuo 3-695
Mutis, Sergio 2-389

N

Nakayama, Kazunori 2-503
Nakazawa, Jin 2-493
Narahara, Taro 1-703, 3-573
Nathansohn, Nof 3-11
Ng, Provides 3-635, 3-707
Nguyen Tran, Yen Khang 3-511
Nicholas, Paul 2-133
Nishimura, Yasuki 2-143
Nizam, Danikh 1-549
Noda, Sakiko 2-513
Noel, Vernelle AA 2-103
Nypelö, Tiina 2-205

O

O'Brien, Delaney 3-235
O'Neill, Sophia 2-245
Ogasawara, Masatoyo 3-257
Önalan, Beril 2-267
Ono, Kotaro 1-347, 1-621
Onuma, Goro 2-277
Or, Man Kit 2-287
Ottevaere, Olivier 2-697

P

Padda, Shahzadi 2-103
Pantopoulou, Artemis 3-123
Papadimitraki, Maria 2-163
Papanikolaou, Dimitris 3-123, 3-277
Park, Dongjoon 3-717
Parkinson, Thomas 3-175
Park, Se Ho 1-185
Park, Sujie 2-431
Patel, Rushil 2-153
Patle, Prakhar 2-153
Pearce, Thomas 2-575
Pei, Wanyu 3-563
Peng, Jiamin 2-697
Peng, Jiaxian 2-175

Petzold, Frank 3-113
Pigram, Dave 1-447

Q

Qin, Qi 3-695
Qin, Zihao 1-467
Qiu, Waishan 3-603, 3-613, 3-685, 3-695
Qi, Yi 2-625

R

Rafique, Usman Khalid 3-297
Rahbar, Morteza 2-543
Ramsgaard Thomsen, Mette 2-133
Rane, Sonali 3-377, 3-531
Ran, Ran 1-395
Rao, Fenggui 3-103
Rasneur, Sylvain 2-645
Ratoi, Lidia 1-559
Ren, Chongyang 3-613, 3-685, 3-695
Ren, Shuyao 3-695
Ren, Yuxiao 1-427
Rhee, Jinmo 3-593
Rian, Iasef Md 2-307
Rochanavibhata, Nuttapong 1-427
Rode, Rutuja 3-377, 3-531

S

Sabin, Jenny 1-295
Sader, Christine Khouri 2-245
Sahin, Ekin Sila 2-81, 2-431
Şahin, Yusuf Hüsayin 3-543
Sakai, Yusuke 2-503
Sato, Ayaka 2-133
Schartner, Hans 3-757
Schaumann, Davide 3-113, 3-429
Schnabel, Marc Aurel 3-185
Seifert, Clay 1-243
Sekardini, Heidy 2-533
Selvan, Surayyn 3-449
Sepehri, Pegah 3-235
Sfara, Josef 3-469
Shaked, Tom 2-481
Shay, Oren 3-449
Shen, Chen 1-123
Sheng, Yu-Ting 2-287
Shen, Jie 3-603
Shen, Pengyuan 1-21
Shen, Yang-Ting 2-31
Shen, Yanting 1-11
Shih, Shen-Guan 2-369, 2-657, 3-267
Shi, Jianing 1-113
Shinohara, Hiroyuki 3-439
Shi, Xinyu 2-257
Shi, Yichao 1-519
Showkatbakhsh, Milad 2-153, 3-377, 3-409, 3-419, 3-531, 3-645
Siriwardena, Lasath 2-235
Skånberg Dahlstedt, Toste 2-205
Slavina, Anastasia 3-329, 3-449
Song, Yang 2-113, 2-635, 3-349
Song, Zhicheng 1-21
Sosa, Marco 2-359
Spaw, Gregory Thomas 2-317
Spivak, Noam 2-461, 2-471
Sprecher, Aaron 1-539, 2-461, 2-471, 2-481
Sterman, Yoav 2-339, 2-461
Stouffs, Rudi 3-563, 3-625
Su, Huang 2-687
Sumitomo, Eri 2-113
Sun, Cheng 1-673
Sun, Chengpeng 2-257
Sun, Chengyu 1-223, 3-675, 3-779
Sun, Jiapan 3-81
Sun, Nancy 3-479
Sun, Yiting 2-615, 2-645
Su, Siwei 1-223
Su, Xinyu 1-285

T

Tageldin, Reem 2-51
Taghavifard, Reza 3-397
Tahavvori, Shaqayeq 2-235

Tahmasebi, Farhang 3-367
Tai, Hung-Chen 1-123
Takahashi, Ryohei 3-357
Takeshita, Shiro 3-357
Tam, Kam-Ming Mark 1-81
Tanaka, Masaya 1-347
Tang, Chao 1-81
Tang, Peng 2-451
Tanino, Yudai 3-491
Tan, Miaomeng 1-395
Tan, Yizhou 1-367
Tarek, Shaimaa 2-421
Tay, Tiffany Feng Lin 1-549
Tessmann, Oliver 2-175
Thompson, Peter 3-367
Tian, Xinle 2-523
Tibbits, Skylar 3-11
Tong, Ziyu 3-747
Topçu, Beyza Gül 3-655
Tran, Trong Tin 3-511
Triantafyllidis, Eleftherios 2-267
Tsaih, Lucky Shin-Jyun 3-267
Tsai, Ting-I 2-667
Tseng, Yu-Tzu 3-133
Tsiliakos, Marios 1-113
Tsou, Jin Yeu 1-591
Tu, Huijun 1-693
Tung, Ping-Ju 2-93
Tu, Shaotsu 1-133

U

Ünal, Gözde 3-543
Urroz, Celso 2-533, 2-563

V

van Ameijde, Jeroen 3-501
van Engelenburg, Casper 3-727
Varoudis, Tasos 3-635
Vrachliotis, Georg 3-727

W

Wagiri, Felicia 2-657
Wagner, Hans Jakob 2-163
Walczak, Bartosz Marek 3-225
Wang, Bohan 3-367
Wang, Dasong 1-509, 2-441, 2-451
Wang, Likai 1-1, 1-327, 2-1, 3-1, 3-339
Wang, Liuhan 2-493
Wang, Luying 1-601
Wang, Mengdie 3-195
Wang, Menghan 3-613
Wang, Qi 3-153
Wang, Shaohan 2-379
Wang, Shih-Yuan 1-1, 2-1, 2-185, 2-287, 3-1
Wang, Shiliang 3-685
Wang, Shitong 2-615
Wang, Sining 3-41
Wang, Suoao 1-631, 2-51, 2-379
Wang, Tiansu 1-467, 3-307
Wang, Tsung-Hsien 3-317
Wang, Xiao 3-737
Wang, Xin 1-653
Wang, Xinru 3-675, 3-779
Wang, Yanqing 1-11
Wang, Yiqing 1-377, 2-297
Wang, Yiwei 2-411
Wang, Yixuan 3-185
Wang, Yongchun 3-737
Wang, Yuankai 3-603
Wang, Yuxiao 3-163
Wang, Zhengyang 1-427, 2-11
Wang, Zilu 3-297
Wei, Zhaoxi 1-41
White, Dustin 2-563
Willmann, Jan 2-575
Wong, Jonathan 3-409, 3-645
Wong, Joseph Jing Hymn 2-677
Wong, Pak-Hei 1-153
Wong, Ueki 2-553
Wortmann, Thomas 2-81
Wu, Dongdong 1-233
Wu, Hao 3-81
Wu, Haopu 2-441, 2-451
Wu, Junyi 3-737
Wu, Ruize 3-747
Wu, Yujing 3-603, 3-613
Wu, Yulinhan 2-645
Wu, Yuxuan 1-93

X

Xia, Xinyue 3-41
Xie, Han 3-757
Xie, Haoran 1-417, 1-427, 2-11
Xie, Jinhao 1-305
Xie, Zhihao 1-41
Xin, Yangpeng 1-683
Xu, Hanwen 1-467
Xu, Weiguo 1-315
Xu, Xinhui 1-205, 1-275
Xu, Yilei 1-205, 1-275

Y

Yamada, Satoshi 1-347, 1-621
Yan, Chao 2-21, 3-93
Yang, Aries YingBo 1-61
Yang, Hengbo 1-223
Yang, Hongye 1-41
Yang, Ruyi 3-93
Yang, Yean 3-685
Yang, Yidan 1-21
Yan, Peiqi 1-81
Yan, Xiheng 2-399
Yao, Jiawei 1-11
Yao, Jiaxin 3-93
Yao, Yuhan 3-603
Yen, You-Wei 2-605
Ye, Yu 3-563
Yi, Hwang 2-327
Yin, Hongxi 1-631
Yin, Jun 3-521
Yin, Zhang 3-459
Yoshikawa, Masataka 1-477
Yoshikawa, Ryo 2-503
Yousif, Shermeen 2-533, 2-563
Yuan, Philip F 1-205, 1-275, 1-489, 2-553, 2-615, 2-645, 3-81
Yu, Boyuan 2-93
Yue, Yang 1-683
Yu, Jiaqi 1-175
Yu, Junwang 2-553
Yu, Lei 2-195

Yu, Yang 3-31
Yu, Zizheng 3-41

Z

Zachariah, Deepti 2-133
ZamaniGoldeh, Erfan 1-51
Zangori, Laura 3-235
Zastavni, Denis 2-645
Zboinska, Malgorzata A. 2-205
Zechmeister, Christoph 2-595
Zelaya, Rafael Luna 3-769
Zhang, Cheng 3-185
Zhang, Di 3-163
Zhang, Jian Jun 1-395
Zhang, Jiaren 1-673
Zhang, Jiaxi 1-489
Zhang, Jingbo 2-697
Zhang, Kaihan 1-233
Zhang, Kaiheng 1-703, 3-573
Zhang, Tong 1-499
Zhang, Wenting 3-501
Zhang, Ye 1-417, 1-427, 2-11
Zhang, Yixian 1-223
Zhang, Yu 3-625
Zhang, Yuanxi 3-583
Zhang, Yueying 3-287
Zhang, Yuxin 3-521
Zhan, Qiang 2-625
Zhao, Jiangyang 3-21, 3-349
Zhao, Yuting 2-451
Zhong, Jing 3-521
Zhou, Jianjia 1-205
Zhou, Ying 1-683
Zhou, Yuxin 3-491
Zhuang, Dian 1-673
Zhuang, Weimin 1-133
Zhuang, Zhi 3-387
Zhu, Junlin 1-11
Zhu, Xinjie 1-143
Zhu, Yalei 2-615, 2-645
Zlatanova, Sisi 3-215

CAADRIA 2026

Humanistic Computation and Intelligence
26 April - 2 May 2026
National Yang Ming Chiao Tung University
TAIWAN

This volume contains the peer-reviewed papers accepted for the 31st International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA) 2026

ISBN: 978-9-88713-590-6

©2026 All rights reserved and published by
The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong