

Supporting Complex Design Processes with Agentic System

A Case Study on Streetscape Design

Meizhu Chen¹, Weihao Chen², Yuqian Li³, Yingzhou Gao⁴, and Weiguo Xu⁵

^{1,2,3,4,5}*Tsinghua University*

¹*cmz23@mails.tsinghua.edu.cn, 0009-0005-1030-3464*

²*chenwh20@mails.tsinghua.edu.cn*

³*liyuqian@sz.tsinghua.edu.cn*

⁴*gaoyz24@mails.tsinghua.edu.cn*

⁵*xwg@mail.tsinghua.edu.cn, 0009-0004-7594-2534*

Abstract. Generative AI has the potential to support complex design processes, but current approaches primarily focus on well-defined stages, limiting their ability to reveal challenges in human-AI collaboration. In this work, we present ChatStreet, a multi-agent system designed to support streetscape improvement design. ChatStreet integrates a central planning layer, a bottom tool layer composed of specialized agents, and a memory layer that tracks user interactions, providing both automated and manual controls through a chat interface and parametric panel. We conducted a case study with 20 architecture students. Our results demonstrate that ChatStreet effectively supports the design process and enhances aesthetic outcomes. Furthermore, the study provides practical guidance for key capabilities of future agentic systems: world understanding, flexible task orchestration, and intent completion. These findings advance the development of AI tools that facilitate human-AI symbiosis in creative workflows.

Keywords. agentic AI, streetscape design, human-AI collaboration, design cognition, generative AI

1. Introduction

In generative AI-assisted design, AI systems are expected to support the whole design process seamlessly. Compared to stage-wise assistance, achieving this kind of holistic support requires more alignments between designers and the AI. The reason is that the whole design process often involves more atomic design tasks and their dynamic dependencies. Due to the iterative and open-ended nature of design (Cheung et al., 2025), such tasks cannot be arranged mechanically in a linear sequence. Instead, they need to be flexibly organized. Recent research primarily focuses on supporting well-defined (structured) stages and processes (C. Wang et al., 2025; Z. Wang et al., 2025; Wei & Herr, 2025), thus unable to better reveal the potential issues in complex human-

AI collaboration.

Currently, agentic AI is seen as a potential approach to supporting long and complex design processes (Sapkota et al., 2026). It often adopts a multi-agent method, customizing specific agents to handle each atomized task, and automatically planning and calling these agents. However, designers and AI often have different chains of thoughts about how to define and orchestrate these atomic tasks, which may lead to unexpected conflicts. It is unknown how designers perceive, resolve, or build upon these differences.

Motivated by this, we build a multi-agent system, ChatStreet, and investigate whether it can effectively support the design process and how designers' cognition evolve during this process. We choose streetscape improvement design as a case study. Because this task involves multiple components, such as design concepts, analysis of existing conditions, and improvements in functional and aesthetic aspects. It also utilizes various modalities, including text, image, and analysis data, and is simpler than other architectural design tasks (such as 3D generation), ensuring a more reliable AI baseline for evaluating the agentic system.

Our system consists of three layers: a central planning layer, a bottom tool layer calling different tools, and a memory layer tracking user interaction history. The bottom tool layer is composed of agents corresponding to various atomic tasks in streetscape design, including: a design advice agent, an image prompt suggestion agent, two types of image generation agents, a semantic segmentation agent, a mask adjustment agent. For user collaboration, we provide a natural language-based chat interface and a parametric control panel, to support both automatic execution and manual control.

We conducted a case study with 20 architecture students, who used our system to complete streetscape improvement design tasks in a workshop. They produced 544 images in total. We evaluated the design outcomes. We also examined the sessions of system usage, which includes the sequences of use interaction and agent response, the most frequent operations, the strategies for combining multiple agents, the timing of switching between agent-driven and manual image generation, and the patterns of user prompting.

Our system not only effectively supports the streetscape improvement design process, but also as a lens to explore the intricate mechanisms of human-AI collaboration in complex design processes. By investigating how designers interact with and adapt to a practical system, we provide valuable insights into the future development of agentic AI that can support more complex and open-ended design tasks.

In summary, our main contributions are as follows:

- We implement a multi-agent system, ChatStreet, which effectively supports the streetscape improvement design task.
- The case study demonstrates designers' collaboration patterns and preferences, providing insights into the cognitive dynamics of human-AI design symbiosis.
- We discuss the implications for future agentic systems, emphasizing the need for stronger world understanding, flexible task orchestration, and intent completion to better support human-AI collaborative design.

2. Related Work

2.1. AGENTIC AI FOR DESIGN

Agentic AI systems represent a shift from traditional AI, focusing on multi-agent collaboration, dynamic task decomposition, persistent memory, and autonomy (Acharya et al., 2025; Sapkota et al., 2026). Typically, LLMs serve as core reasoning components in these systems, while tool-augmented AI agents expand functionality (Sapkota et al., 2026).

Such systems are ideal for complex processes that require both coordination among different AI agents and human intervention. For example, Archi-Agents integrates multi-agent LLMs with a RAG knowledge base and BIM to generate BIM models from natural language inputs (Tang et al., 2025). Cheung et al. (2025) propose a multimodal design system for early-stage architectural design, exploring three modes of AI collaboration. Chen et al. (2025) introduce an agent-driven approach for precise 3D model generation, aligning vague user inputs with parametric programs.

However, existing research lacks a trade-off between the practical applicability of case studies and the fine-grained construction of systems. This study addresses this gap by developing a multi-agent system with specialized agents and distinct interfaces, exploring how designers interact with different parts of the agentic AI in creative workflows.

2.2. AI-SUPPORTED STREETSCAPE RESEARCH

Recent research on AI-supported streetscape studies has predominantly focused on the analysis of streetscapes, rather than on generative applications. AI techniques, such as deep learning-based classification models trained on self-collected datasets, prediction models, general multimodal models like GPT-4o, semantic segmentation, PCA, and clustering analysis, have been applied to evaluate various aspects of streetscapes (Lee et al., 2025; Ma et al., 2025; Mushkani & Koseki, 2026; Ogawa et al., 2024; Xiao & Tang, 2025; Xu et al., 2023). These systems assess factors like subjective perception, quality, feature identification, and aesthetic styles, helping urban planners and designers better understand street environments. Additionally, Nguyen et al. (2025) have employed generative AI to modify specific elements of streetscapes, investigating which changes have the most significant impact. However, the role of generative AI in streetscapes design remains underexplored, particularly in scenarios involving human participation.

As streetscape design increasingly involves public participation, there is a growing need for AI systems that can support non-professionals in the design process. This research gap highlights the need for further exploration on generative AI supporting streetscape design.

3. Methodology

To bridge the research gap, we introduce **ChatStreet**, an agentic system that involves multiple agents and two kinds of interfaces to handle various stages of streetscape design and respond to users' multimodal needs. ChatStreet strikes a balance between designer agency and AI autonomy.

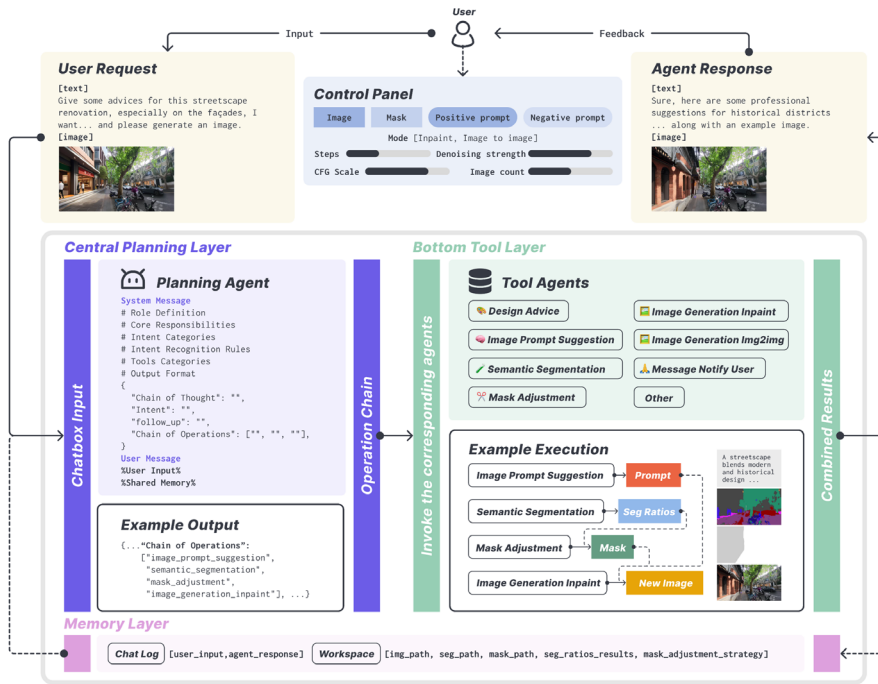


Figure 1. Agentic system framework.

As shown in Figure 1, ChatStreet consists of three layers: a central planning layer, a bottom tool layer that calls different tools, and a memory layer that tracks user interaction history. User requests can be submitted as either text or images. After processing by the agentic layers, ChatStreet generates a response for the user. All language reasoning components utilize the GPT-4o API. Additionally, users have access to a manual control panel that allows them to adjust the generated images in a fine-grained way. All image generation components call the API from Liblib, which is an online platform that provides access to a variety of AI-powered image generation models and tools, widely used for creative design.

3.1. CENTRAL PLANNING LAYER

Central planning layer manages task scheduling. It is responsible for understanding the user's intents and translate them into a sequence of executable sub-tasks. The core module of the central planning layer is the planning agent, which handles tasks such as parsing user input, reasoning user intent based on context, and planning the operation

chain. When user intent is unclear, the planning agent proactively requests clarification.

As shown in Figure 1, the system prompt of the planning agent includes role definition, core responsibilities, intent categories, intent recognition rules, tool categories, and output formats.

The main challenge in building the planning agent lies in enabling it to infer the most feasible operation chain within a limited context. During the design processes, a single intent may require multiple tools to work together to resolve the task. For example, when a user requests “increase the sky’s ratios by 10%”, the agent needs to perform semantic segmentation, generate an expanded mask, and then inpaint the image. To address this, the planning agent is set to engage in a chain of thought, and then generate the corresponding chain of operations, improving the accuracy of the reasoning process.

Additionally, the same intent may trigger different operation chains in different contexts. The context is essentially the shared memory between all agents and the user. Thus, the planning agent’s input should include not only the user input but also sufficient contextual information.

Finally, some tool agents may rely on the results of previous invoked agents. The central planning layer must also identify these dependencies and conflicts, ensuring smooth transitions between the design phases.

3.2. BOTTOM TOOL LAYER

Bottom tool layer is composed of eight agents, corresponding to atomic tasks in streetscape design. Their composition pattern is extensible and can integrate more tools.

The design advice agent facilitates conceptual discussions with designers by providing structured design strategies that help clarify current design concepts. It can also engage in discussions about streetscape images, leveraging GPT-4o’s multimodal capabilities to analyze the aesthetic characteristics of the original images and offer textual recommendations.

The image prompt suggestion agent helps with optimizing image generation prompts, providing assistance in design style, element adjustments, color schemes, and atmospheric details, improving the efficiency of users’ prompt writing.

The image generation agents include the inpaint agent and img2img agent. Both of them generate parameters and images through Liblib. They also have reasoning capabilities for generating image prompts when these have not been pre-defined. Additionally, the inpaint agent has a dependency on a pre-generated mask, using it as an input when generating prompts.

The semantic segmentation agent uses the pretrained DeepLabV3Plus-MobileNet model on Cityscapes, refer to commonly used models and datasets in streetscape research. It returns the top five elements with the highest proportion in the image, providing a quantitative metric for users to assess streetscape quality.

The mask adjustment agent generates masks on specific elements based on the results of segmentation. If the user says “expand the greenery area to 30%”, it can automatically execute a mask dilation algorithm to intelligently generate the area to be inpainted.

3.3. MEMORY LAYER

Memory layer is a crucial component for managing context, consisting of two aspects: the chat log records all chat history between the user and agents in the interface, while the workspace manages the current original image files, semantic segmentation image files, mask files, and text data that needs to be passed between agents. Each time the central planning layer and bottom tool layer are activated, they need to read the latest data from the memory layer. The results generated by tool agents are also saved promptly in the memory layer.

3.4. USER INTERFACE

ChatStreet's user interface is developed based on Gradio. Gradio provides an accessible framework for rapid prototyping of interactive machine learning interfaces. As shown in Figure 2, the interface is divided into two sections: on the left is a manual control panel, similar to WebUI, which provides simple parameter options that non-professionals can quickly understand; on the right is a chatbox, where users can interact with agents and receive design assistance. Users can independently choose which section to use. This design aims to investigate when users prefer to rely on agents and when they opt to control the process themselves.

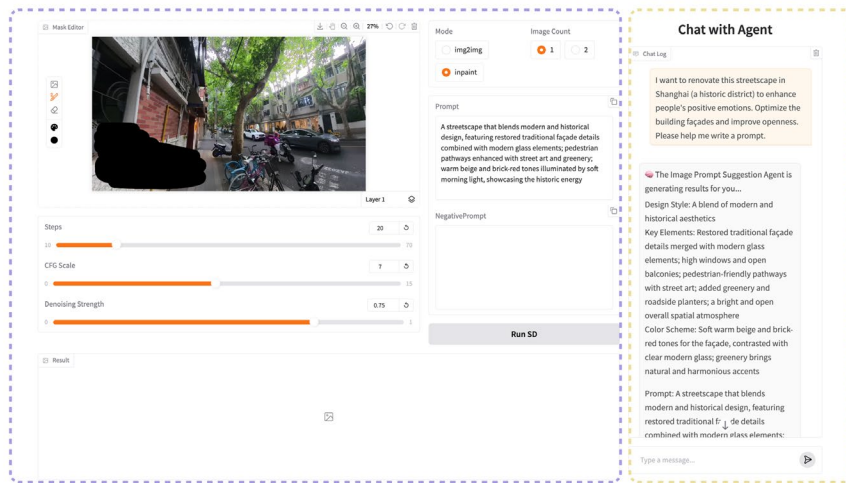


Figure 2. User interface.

3.5. CASE STUDY SETUP

We conducted a case study with 20 architecture students in a workshop. The students were tasked with using ChatStreet to complete streetscape improvement design tasks in one day.

The photos they used were all real streetscape data, collected by themselves in the early stage of the workshop using Insta360 cameras. They selected scenes with poor spatial quality. The streetscape content included three types: historic districts, urban parks, and residential areas.

4. Results

4.1. DESIGN OUTCOMES EVALUATION

In total, they produced 544 images. After removing duplicate data, we used LAION-Aesthetics V2 to evaluate the aesthetic gap between the modified images and the original images. LAION-Aesthetics V2 is a large-scale, pre-trained model designed to assess aesthetic quality in images. It offers a reliable and efficient means for evaluating visual content, and demonstrates strong generalizability across various types of images.

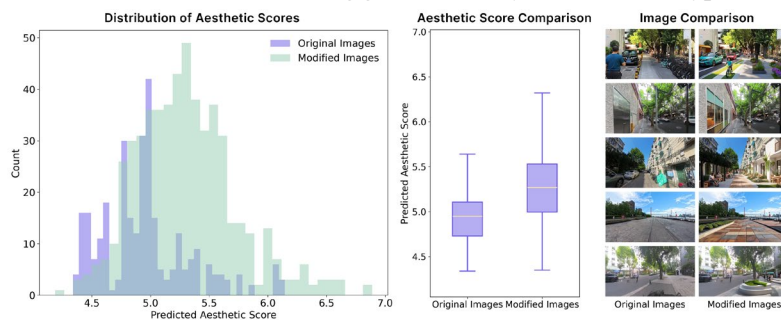


Figure 3. Aesthetic evaluation of original and modified Images.

As shown in Figure 3, the distribution of aesthetic scores shows differences. The aesthetic scores of original images are concentrated around 4.5 and 5.0, with none exceeding 6.3. Modified images have higher scores, with distributions still present above 6.5. After normality testing, neither category follows a normal distribution. Furthermore, due to the presence of multiple generated images corresponding to a single original image, the Mann-Whitney U test was employed, yielding a p-value smaller than 0.001. Therefore, modified images are significantly superior to original images in terms of aesthetics, demonstrating that ChatStreet's support for design tasks is effective.

4.2. HUMAN-AGENT COLLABORATION PATTERNS

We then analysed human-agent collaboration patterns during the design process by visualizing each user and agent interaction sequences and evaluating user prompt (Figure 4).

4.2.1. Usage of the manual control panel and the chatbox

Users generated images via the manual control panel far more frequently than through the agents. This tendency was pronounced in the middle and later stages of the design process, especially among more active participants. We observed that some users shifted to using the control panel for brush-based modifications, after failing to use agents to modify a specific region of the image via language. This indicates that designers have higher demands to make fine-grained modifications on the layout of the image, rather than the overall style.

4.2.2. Agent invocation

Overall, the frequency of single-agent calls far exceeded consecutive multi-agent invocations in each interaction round. Agents with strong dependencies, such as segmentation and mask agents, often appeared together and were the most common combination among multiple-agent chains. In contrast, combinations of weakly dependent agents were rare. For example, segmentation agent followed by design advice agent occurred only once. When complex requirements arise, planning long operation chains remains challenging, as confirmed by further observations.

In terms of agent invocation count, users most frequently utilized the image prompt suggestion agent. Users often input a simple design requirement and let the agent expand it into a more complete prompt. Meanwhile, users require relatively less support from agents regarding images and segmentation data.

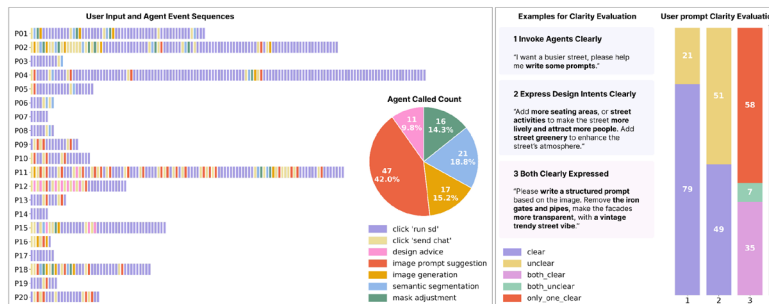


Figure 4. User interaction evaluation.

4.2.3. User expression clarity

Furthermore, text-based agentic system allows users to jointly express design intent and tool invocation, unlike traditional GUIs that separate tools into predefined functions. While this affords greater flexibility, it also poses a challenge, as user expressions can be ambiguous. So, we further analysed the user prompts. Experts encoded each user input text to determine whether it invoked the agent and expressed design intent clearly. Mentioning requests for the specific agentic operations was considered as “invoking the agent clearly”, such as “write a prompt”, “execute segmentation”, or “generate an image”. Mentioning specific design elements rather than just describing a feeling was considered as “expressing design intent clearly”, such as “greenery” or “iron gates and pipes”. We calculated the proportions of clear and unclear cases, as well as the proportion where both were clear, as shown in Figure 4.

The results show that the proportion of invoking agents clearly was higher, while the proportion of expressing design intent clearly was lower. The proportion where both were clear was relatively low. This suggests that it is not easy for users to both clearly express design intents and tool-use instructions, while expressing design intent clearly requires higher effort. This calls for future agentic systems to address the issue of understanding users’ ambiguous intentions.

Moreover, users’ clear requests were mostly concentrated on “write a prompt”, showing that the prompt suggestion agent can be better recognized within a non-

transparent agentic system.

5. Implications

We present insights from the case study and discuss their implications for future agentic systems that support design tasks. Following capabilities should be incorporated.

World Understanding. Our system highlights designers' strong demand for fine-grained modifications of streetscape images, evidenced by frequent references to specific regions and architectural elements. However, such requirements are difficult to interpret when expressed through natural language alone. This suggests that future agentic systems should rely on underlying AI models with stronger visual grounding and world understanding, enabling accurate localization and interpretation of design targets within complex visual scenes.

Flexible Task Orchestration. While the planning agent provides dynamic task orchestration, its outputs do not always align with user expectations. This indicates that future agentic systems should be capable of capturing the complex dependencies inherent in design tasks and coordinating multiple agents and tools through graph-based representations. In such systems, users primarily focus on evaluating the outputs and providing feedback, rather than manually managing task execution.

Intent Completion. Our evaluation reveals that users frequently express vague or underspecified design intentions, which are difficult to translate directly into executable design actions. In AI-assisted design systems, where interaction typically involves the system directly returning results to users, this challenge becomes especially pronounced. Future agentic systems should therefore be capable of inferring, completing, and operationalizing user intentions, bridging the gap between high-level expressions and actionable design operations.

6. Conclusion

This study introduces a multi-agent system, ChatStreet, and investigates whether it can effectively support the design process and how designers' cognition evolve during this process. The case study shows the system's effectiveness, with significant aesthetics improvement compared to the original streetscape images. We highlight several key considerations for future agentic systems in creative workflows. This work contributes to advancing design-oriented AI toward human-AI design symbiosis.

Attribution Statement

ChatGPT (OpenAI, 2025) was used to improve the flow of this paper.

References

- Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey. *IEEE Access*, 13, 18912–18936. <https://doi.org/10.1109/ACCESS.2025.3532853>
- Chen, M., Pan, S., Gao, W., Xu, W., & Lin, B. (2025). Agent-Driven Parametric 3D Building Generation via Natural Language Interface. *Proceedings of the 30th International*

- Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2025*, 1, 203–212.
- Cheung, L. H., Wang, L., & Lei, D. (2025). Conversational Application of Agentic Multimodal AI in Collaborative Architectural Design Environment: An architectural-focus AI design partner for early-stage design exploration. *Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2025*, 1, 173–182.
- Lee, M., Choi, B., & Hwang, S. (2025). Automatic method to predict visual pleasantness and unpleasantness of streetscapes and identify key microscale components for improving pedestrian environments. *Developments in the Built Environment*, 22, 100652. <https://doi.org/10.1016/j.dibe.2025.100652>
- Ma, H., Li, J., & Ye, X. (2025). Deep learning meets urban design: Assessing streetscape aesthetic and design quality through AI and cluster analysis. *Cities*, 162, 105939. <https://doi.org/10.1016/j.cities.2025.105939>
- Mushkani, R., & Koseki, S. (2026). Street review: A participatory AI-based framework for assessing streetscape inclusivity. *Cities*, 170, 106602. <https://doi.org/10.1016/j.cities.2025.106602>
- Nguyen, T., Micheli, S., & Luo, D. (2025). Rapid crowd-sourced Feedback of Generative AI Enhancements of Streetscape: A case study of Brisbane’s capabilities of hosting mega-events. *Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2025*, 4, 59–68.
- Ogawa, Y., Oki, T., Zhao, C., Sekimoto, Y., & Shimizu, C. (2024). Evaluating the subjective perceptions of streetscapes using street-view images. *Landscape and Urban Planning*, 247, 105073. <https://doi.org/10.1016/j.landurbplan.2024.105073>
- Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2026). AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges. *Information Fusion*, 126, 103599. <https://doi.org/10.1016/j.inffus.2025.103599>
- Tang, C., Tam, K.-M. M., & Herr, C. M. (2025). Archi-Agents: Approximating the architectural design process with collaborative specialized multi-agent system IIm and building information modeling. *Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2025*, 3, 141–150.
- Wang, C., Shi, Y., Ma, C., Xu, H., & Kastner, P. (2025). Reimagining Chinese Garden Design: An interactive approach using Stable Diffusion. *Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2025*, 1, 9–18.
- Wang, Z., Jin, H., Du, X., Ren, Y., Zhang, Y., & Xie, H. (2025). From Architectural Sketch to Conceptual Representation: Using structure-aware diffusion model to generate renderings of school buildings. *Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2025*, 1, 69–78.
- Wei, D., & Herr, C. M. (2025). Applying Multimodal Large Language Models in Ecological Facade Design: Enhancing local viability and ecological feasibility of design proposals. *Proceedings of the 30th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2025*, 1, 39–48.
- Xiao, Y., & Tang, Y. (2025). Can ChatGPT-4o assess the perceptions of streetscape change? Evidence from Shanghai, China. *Sustainable Cities and Society*, 130, 106674. <https://doi.org/10.1016/j.scs.2025.106674>
- Xu, H., Sun, H., Wang, L., Yu, X., & Li, T. (2023). Urban Architectural Style Recognition and Dataset Construction Method under Deep Learning of street View Images: A Case Study of Wuhan. *ISPRS International Journal of Geo-Information*, 12(7), 264. <https://doi.org/10.3390/ijgi12070264>