Reimagining Chinese Garden Design: An Interactive Approach Using Stable Diffusion

Chunlan Wang[1], Yichao Shi ✉[2], Changda Ma[3], Hang Xu[4], Patrick Kastner[5]
[1,2,3,4,5]*Georgia Institute of Technology.*
[1]*cwang932@gatech.edu, 0009-0004-8096-3217*
[2]*yichao.shi@email.com, 0009-0000-7715-8277*
[3]*cma326@gatech.edu, 0009-0005-9355-8525*
[4]*hxu623@gatech.edu, 0000-0002-4063-6295*
[5] *patrick.kastner@gatech.edu@gatech.edu, 0000-0003-4940-341X*

Abstract. Traditional Chinese gardens (TCGs) possess considerable cultural and aesthetic significance and demonstrate a high level of complexity in their architectural expression. While large-scale generative models such as Stable Diffusion (SD) are routinely employed for creating architectural concept drawings, they often fail to capture the nuanced spatial arrangements and symbolic elements of TCGs. This paper presents a novel machine-learning-driven workflow, incorporating fine-tuned diffusion models and user-friendly controls, to generate Chinese garden masterplans that address these shortcomings. Specifically, we integrate DreamBooth to specialize a base SD model with TCG features, LoRA to infuse stylistic variations, and ControlNet to condition the generation on user-defined color schemes (which map to major landscape components). An interactive web interface allows designers to input high-level functional requirements (e.g., pavilion-to-greenery ratio) and styles, and to refine the generated layout through brushes. Results demonstrate that our system produces garden masterplans with greater spatial accuracy and aesthetic fidelity compared to off-the-shelf diffusion models. This research provides a new pathway for integrating user-centered design processes with generative AI in a culturally specific architectural context.

Keywords. Traditional Chinese Garden, Interactive Design, Stable Diffusion, ControlNet, LoRA, Landscape Architecture, Masterplan

1. Introduction

Traditional Chinese gardens (TCGs) represent centuries of architectural ingenuity and cultural philosophy, emphasizing harmony between humanity and nature (Liu & Chen, 2019). Design principles often hinge on nuanced spatial experiences, symbolic meaning in plants and rocks, and carefully orchestrated views (Smith, 2018). Despite their cultural importance, creating TCGs in contemporary practice remains challenging. Many historical gardens have been partially or completely lost, causing a scarcity of references (Liu & Chen, 2019). Moreover, the design expertise required involves deep knowledge of historical layout techniques, symbolic elements, and traditional craftsmanship, thereby limiting the number of professionals who can effectively engage in such projects (Zhang & Wang, 2012).

Advancements in AI-driven generative modeling (e.g., StyleGAN, DALL·E, Stable Diffusion) have the potential to ease some of these challenges by rapidly producing conceptual layouts. However, preliminary explorations with off-the-shelf diffusion models (e.g., Stable Diffusion v1.5) indicate that they tend to generate TCG images without meaningful spatial arrangements, as shown in Figure 1. These models lack the specialized training data and conditioning mechanisms necessary to capture the intricate interplay of elements—water, rocks, pavilions, corridors, and vegetation—unique to TCGs.

Figure 1. The drawings provided by Stable Diffusion fail to accurately depict the architectural layout of Chinese gardens.

To address these gaps, researchers have employed methods such as Pix2Pix for color-labeled layout generation (Isola et al., 2017; Yu et al., 2015; Liu et al., 2022; Lu et al., 2025). More recently, ControlNet (Zhang, Rao, & Agrawala, 2023) has been introduced to condition Stable Diffusion on external signals (e.g., segmentation maps). ControlNet, in combination with LoRA (Low-Rank Adaptation) fine-tuning, can yield high-fidelity and spatially consistent results while keeping computational overhead manageable. Although certain attempts (Zhuang et al., 2024) have shown that text-to-city generation can be guided quantitatively, they do not address the user interactivity and cultural-specific design principles embedded in TCGs.

This paper proposes a three-step interactive workflow for TCG design: a.Color Scheme Generation: Users input desired area proportions of key design elements (pavilions, rocks, water features, greenery, etc.), and the system generates a color-coded segmentation map; b.Color Scheme Refinement: Users can paint or erase segments of the layout to refine the arrangement; c.Masterplan Rendering: The refined color scheme is passed to a fine-tuned diffusion model (with ControlNet and LoRA) to produce a stylized TCG masterplan image. Users can add text prompts to specify drawing style, era references, or aesthetic preferences.

We evaluate our system by comparing the generated plans against those from a vanilla Stable Diffusion model and by reviewing the outputs with professional architects and researchers in Chinese garden design. Results suggest our specialized approach yields richer and more spatially coherent TCG layouts, demonstrating the promise of interactive generative AI for culturally sensitive design tasks.

2. Background

Traditional Chinese gardens are designed to embody ideas of harmony and balance, often featuring intricate rockeries, winding corridors, framed views, water ponds, and pavilions that serve as vantage points (Yu et al., 2015; Yang, 2018). Their layered spatial experiences—where paths meander through partitioned "scenes," each offering a distinct perspective (Yu, Ostwald, & Gu, 2016)—demand nuanced control over arrangement and proportions, a requirement that can challenge generic diffusion-based image generation models (Zhang & Wang, 2012; Dan et al., 2024). In particular, the meandering circulation and complex layering of TCG layouts defy simplistic grid-like planning; key elements such as rocks, water, and plants carry deep cultural and philosophical meaning, which must be reflected in the final design; and the limited availability of high-quality plan drawings, given the scarcity and historical nature of TCG references, further hinders model training. While these issues are not exclusive to TCGs, the gardens offer a distinct lens through which to examine the broader challenge of producing culturally and spatially coherent designs using artificial intelligence.

Color-coding plans is a common strategy in both the landscape architecture domain and computer vision (semantic segmentation) for clarifying spatial relationships. Researchers studying TCGs (Yu et al., 2015; Liu et al., 2022) frequently employ color-coded diagrams to delineate water features, vegetation, pavilions, corridors, and other elements. This helps highlight design patterns such as adjacency, enclosure, and transition. In generative AI workflows, color-coded segmentation maps serve as strong conditioning inputs (Isola et al., 2017; Zhang et al., 2023), ensuring that the arrangement of objects remains consistent with user intent.

Recent diffusion-based methods (Ho et al., 2020; Dhariwal et al., 2021) have outperformed GANs on image synthesis tasks, but controlling layout details remains a challenge. ControlNet (Zhang, Rao, & Agrawala, 2023) extends Stable Diffusion by injecting external conditions such as depth maps, canny edges, or semantic segmentation maps, thereby providing a more reliable mechanism to control the spatial composition of generated images. LoRA (Low-Rank Adaptation) further allows partial fine-tuning of large diffusion models at lower computational cost (Ruiz et al., 2022). DreamBooth (Ruiz et al., 2022) is another fine-tuning approach aimed at personalizing diffusion models to specific "subjects" or styles. In the design of TCGs, these methods can be complementary: DreamBooth can inject domain knowledge of TCG elements, LoRA can handle style adaptation, and ControlNet can maintain structural accuracy from the color schemes.

3. Methods

The process initiates with masterplans of Chinese gardens, which are delineated with color schemes and divided into 512×512 pixels squares, as illustrated in Figure 2. The labeled data is structured into datasets of masterplan squares and color schemes, utilized to fine-tune two distinct checkpoint models with Dreambooth. LoRA and ControlNet undergo training on these datasets to enhance spatial representation and rendering quality. ControlNet conditions the generation process, whereas LoRA controls the stylization of masterplans. AI models function as the backend of the user interface, allowing users to input color ratios via sliders to obtain the generated color scheme initially. Users can modify this schema and utilize ControlNet and LoRA to obtain region-specific masterplan designs.

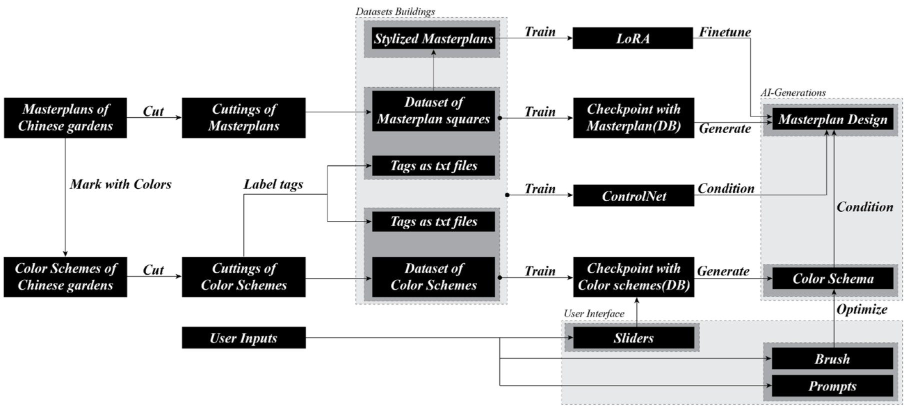

Figure 2. The overview workflow of the methods.

3.1. DATA PREPARATION AND PROCESSING

A primary objective was to build a training dataset that accurately reflects TCG layouts, incorporating water features, rockeries, pavilions, corridors, green space, and paths. We collected 68 historical and modern TCG masterplans (where available) from architectural archives and research articles. To augment this small dataset, we also included stylized TCG-like masterplan images sourced from publicly available architectural collections and design references. The rationale for incorporating stylized data was twofold: a. Visual Consistency: The stylized images adopt watercolor or ink-inspired aesthetics reminiscent of traditional Chinese art, bridging potential gaps in historical plan references; b. Diversity: These images complement the limited historical references and help the diffusion model generalize better by offering a broader range of styles.

Each masterplan was annotated into seven color segments (pavilion/building area, water bodies, corridors, paths, grass/plants, rockeries, and walls/boundaries). This color scheme broadly follows conventional TCG labeling in literature (Yu et al., 2015; Liu et al., 2022) while also aligning with standard segmentation approaches in computer vision.

After annotation, we sliced each masterplan into 512×512-pixel tiles to match Stable Diffusion's default input size and facilitate the generation of smaller "patches" that could later be recombined. Data augmentation (e.g., flipping, rotating) was applied to reduce overfitting. We discarded tiles that contained only a single color (e.g., entirely water or entirely greenery), leaving us with 548 paired images (original masterplan patches and their color-segmented counterparts).

To enrich semantic tags, we calculated each tile's color composition (e.g., 15% pavilions, 25% corridors, 35% greenery, 25% rockery) and grouped them into 26 "style categories." Some categories were assigned thematic labels (e.g., "Reverent Simplicity") referencing typical naming conventions in Chinese garden scholarship. We acknowledge that this labeling is partially heuristic; it helps cluster tiles exhibiting roughly similar design intentions—low built area, high greenery, minimal water, etc. A more systematic classification method (e.g., cluster analysis) remains a promising avenue for future work.

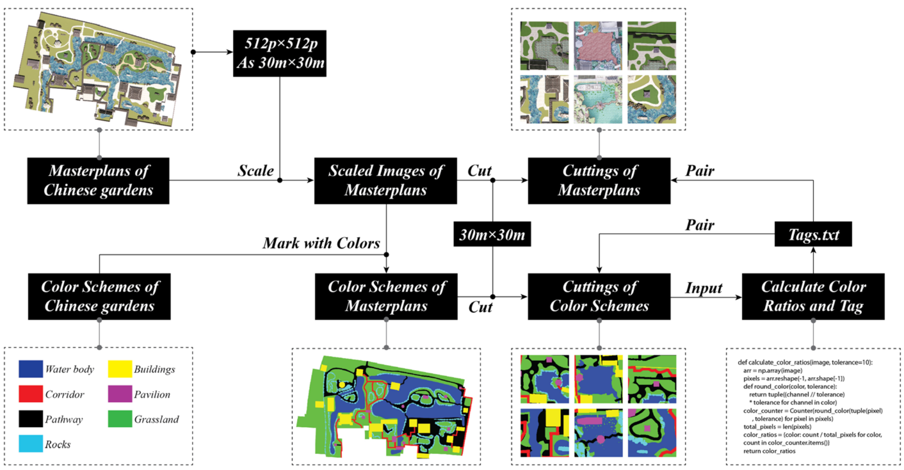

Figure 3. The pre-processing workflow of the dataset.

3.2. MODEL TRAINING

To leverage the strengths of Stable Diffusion v1.5 while adapting it to TCG design, we trained three specialized components: a. DreamBooth-based TCG Checkpoints; b. LoRA modules for style adaptation; c. ControlNet for structural conditioning.

3.2.1. *Checkpoint models finetune with Dreambooth*

We began with the publicly available SD-v1.5.ckpt as our base model and employed DreamBooth (Ruiz et al., 2022) to fine-tune it using two types of data: first, 548 color-labeled patches designed to teach the model how color segmentation maps to TCG semantics, and second, a subset of 344 high-quality TCG plan images annotated with tokens such as "Chinese Garden Plan," "Brick Pathway," and "Traditional Building Roof."

We set the resolution to 512×512, a learning rate of 1e-6, and trained for 8–20 epochs depending on the dataset segment. Models were saved every 2 epochs, and we selected the version with the lowest validation loss and minimal overfitting. Noise offset and multi-scale noise discount were set at 0.1 to preserve fine details.

3.2.2. *LoRA models training*

LoRA modules were trained on 60 stylized watercolor TCG images to capture specific artistic styles (e.g., ink wash, watercolor) without overwriting the base TCG semantic knowledge. We again used SD-v1.5.ckpt (or a DreamBooth checkpoint) as the backbone. Training was done at 512×512 resolution, 10 epochs, and a learning rate of 1e-4. By focusing only on LoRA parameters, we achieved a lightweight approach: the main model remains stable, while LoRA layers capture stylistic variations.

3.2.3. *ControlNet models training*

ControlNet (Zhang, Rao, & Agrawala, 2023) was introduced to preserve the structure indicated by the color segmentation map. We paired each of the 548 TCG patches with its color-labeled map and trained for 10 epochs (learning rate = 1e-4, batch size = 4) using an NVIDIA 4070Ti GPU. The final ControlNet checkpoints enable the system to condition on user-created color schemas, ensuring the generated output respects the input layout. The loss function was:

$$\mathcal{L} = \mathbb{E}_{z_0,t,\epsilon,c_t,c_f}[||\epsilon - \epsilon_\theta(z_t, t, c_t, c_f)||^2]$$

where c_f refers to the features extracted from the segmentation map, and c_t denotes the text tokens (e.g., "Chinese garden, water feature, pavilion").

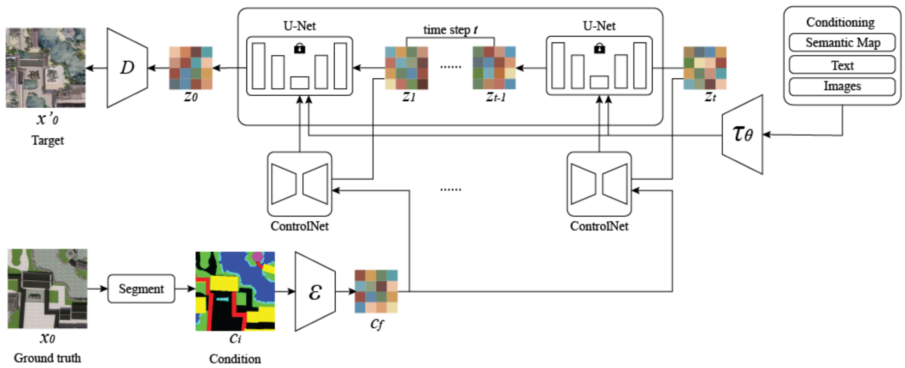

Figure 4. The flow of ControlNet training.

4. User Experience and Interface

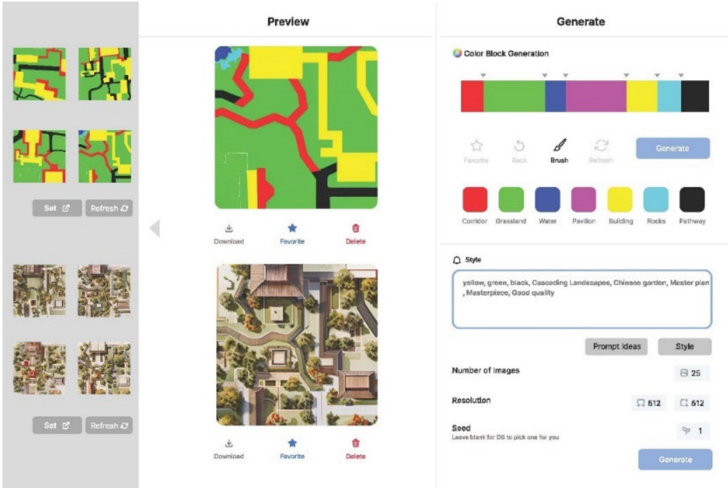

Figure 5. The user interface of the web app for Chinese garden generation.

We created a web interface (Figure 5) that streamlines garden design into three steps: color scheme generation, refinement, and masterplan rendering. Users first adjust sliders to set desired percentages of elements like pavilions and water features, generating initial color-segmented layouts. In the refinement phase, they can modify these schemes using brush tools to ensure design coherence, such as continuous water features. Finally, users add style prompts like "watercolor" or "Ming Dynasty aesthetic" to generate the final rendering. The system combines ControlNet for layout preservation and LoRA layers for style implementation, making traditional garden design accessible to both novices and professionals.

5. Results

We reintroduced original prompts into the DreamBooth-trained checkpoint to generate color schemes, then measured how well the top three-color segments match user-input proportions (Figure 6). On average, 71.7% of the generated tiles met the target distribution (±5% error margin). Certain color combinations exhibited slightly lower compliance, largely due to overlapping semantic elements (e.g., a pavilion partially over water). Improving the segmentation resolution or employing multi-stage refinement may further enhance fidelity.

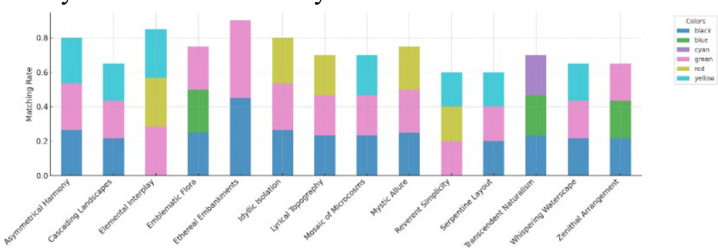

Figure 6. Matching rate of key Tokens with colour contributions.

Figure 7 illustrates the outcome of the user's input regarding color proportions and prompts, resulting in the color scheme and the final master plans. The ControlNet model we developed is integral to this process. This approach guarantees that the color scheme impacts the master plan's impact, making the drawings produced by SD more architecturally relevant.

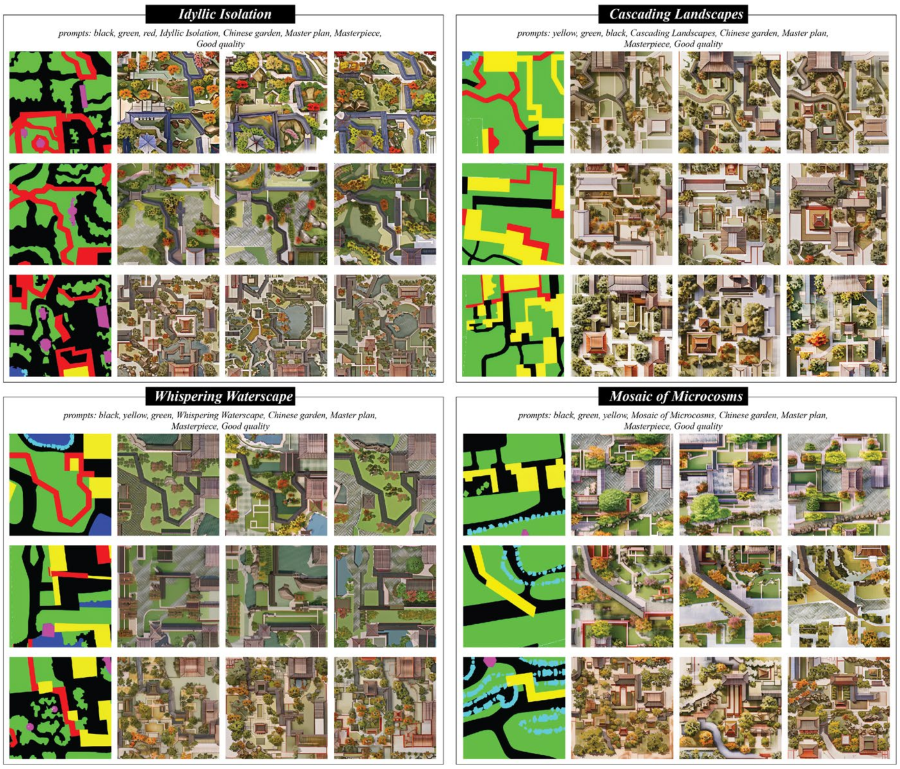

Figure 7. The generated masterplans under different colour schemes

6. Discussion

Our approach integrates user-interactive design processes with advanced diffusion-based generation, offering a novel way to produce TCG layouts. Color-coded plans not only serve as a universal language for architects but also provide strong constraints for diffusion models, ensuring the final output remains faithful to user-defined proportions and spatial organization. By infusing domain-specific knowledge through these color codes, the method proves broadly applicable to other cultural or historical site designs.

The synergy of DreamBooth, LoRA, and ControlNet underpins this pipeline. DreamBooth embeds TCG-specific features—such as rockeries and pavilion morphologies—into the base diffusion model, while LoRA enables style adaptations without retraining the entire network. ControlNet enforces layout coherence by conditioning on segmentation maps, thus preserving the designer’s intended plan. Although our primary focus is on TCGs, whose layered vistas and symbolic details are especially challenging to capture, the same principles can be extended to other architectural traditions—such as Persian gardens or Japanese zen gardens—that demand meticulous spatial composition.

A key question is whether TCGs are inherently more nuanced than other design styles. While many heritage sites are similarly complex, TCGs emphasize interconnected spaces and layered perspectives rather than simple, symmetrical forms, making them particularly difficult for off-the-shelf diffusion models to handle. Our findings suggest that curated data and interactive controls significantly enhance generative fidelity, and future work could apply these strategies to other culturally rich typologies. Further refinements might include integrating depth maps or corridor outlines for finer spatial control, gathering higher-resolution archival material to improve accuracy, and conducting user studies to evaluate aspects like design satisfaction and efficiency. Additionally, metrics such as Fréchet Inception Distance (FID) or spatial adjacency checks could help quantify layout plausibility and cultural correctness in subsequent iterations of this research.

7. Conclusion

This paper demonstrates that integrating Stable Diffusion with specialized fine-tuning (DreamBooth, LoRA) and structural conditioning (ControlNet) enables the generation of coherent, high-quality masterplans for Traditional Chinese Gardens. By offering color-based segmentation controls and an interactive refinement interface, we empower both experts and novices to participate in the generative design process. Qualitative comparisons confirm that our system better captures the symbolic elements and nuanced spatiality of TCGs than an off-the-shelf diffusion model.

Beyond TCGs, our methodology provides a robust template for using generative AI in other culturally rich or historically significant architectural contexts. Future enhancements will focus on refining dataset breadth, improving large-area consistency, and developing more fine-grained control modalities—ultimately promoting a richer collaboration between AI-driven tools and human designers.

References

Dan, H.-C., Huang, Z., Lu, B., & Li, M. (2024). Image-driven prediction system: Automatic extraction of aggregate gradation of pavement core samples integrating deep learning and interactive image processing framework. Construction and Building Materials, 453, 139056. https://doi.org/10.1016/j.conbuildmat.2024.139056

Dhariwal, P., Nichol, A., Ramesh, A., et al. (2021). Diffusion Models Beat GANs on Image Synthesis. Proceedings of the Neural Information Processing Systems.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. Advances in Neural Information Processing Systems, 33, 6840–6851.

Hugging Face. (2023, February 14). ControlNet: Adding conditional control to text-to-image diffusion models. Hugging Face Blog. Retrieved November 21, 2024, from https://huggingface.co/blog/controlnet

Hugging Face. (2023, March 24). Train your ControlNet. Retrieved November 22, 2024, from https://huggingface.co/blog/train-your-controlnet

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Liu, B. & Chen, X. (2019). Challenges in the Preservation of Traditional Chinese Gardens. Journal of Landscape Architecture, 14(2), 34-45.

Lu, B., Dan, H.-C., Zhang, Y., & Huang, Z. (2025). Journey into automation: Image-derived pavement texture extraction and evaluation. arXiv. https://arxiv.org/abs/2501.02414

Smith, J. A. (2018). The Lost Gardens of China: Documentation and Interpretation. Historical Gardens Review, 24(1), 56-67.

Yang, Y. (2018). Design Analysis of Chinese Classical Gardens, URL: https://conservancy.umn.edu/items/a043988e-b843-4ece-9683-8d09d085d91d

Yu, R., Amini Behbahani, P., Ostwald, M. J., & Gu, N. (2015). Wayfinding in traditional Chinese private gardens: A spatial analysis of the Yuyuan Garden. Proceedings of the Architectural Science Association Annual Conference, 22–33.

Yu, R., Ostwald, M. J., & Gu, N. (2015). Parametrically generating new instances of Traditional Chinese Private Gardens that replicate selected socio-spatial and aesthetic properties. Nexus Network Journal, 17(3), 807–829. https://doi.org/10.1007/s00004-015-0263-7

Yu, R., Ostwald, M. J., & Gu, N. (2016). Mathematically defining and parametrically generating Traditional Chinese Private Gardens of the Suzhou region and style. Environment and Planning B: Urban Analytics and City Science, 45(1), 44–66. https://doi.org/10.1177/0265813516665361

Zhang, J., Huang, Y., Li, Z., Li, Y., Yu, Z., & Li, M. (2024). Development of a Method for Commercial Style Transfer of Historical Architectural Facades Based on Stable Diffusion Models. Journal of Imaging, 10(7), 165. https://doi.org/10.3390/jimaging10070165

Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3836-3847).

Zhang, Y., Gong, Y., Cui, D., Li, X., & Shen, X. (2024). DeepGI: An automated approach for gastrointestinal tract segmentation in MRI scans. arXiv. https://doi.org/10.48550/arXiv.2401.15354

Zhang, Y., Leng, Q., Zhu, M., Ding, R., Wu, Y., & Song, J. (2024). Enhancing text authenticity: A novel hybrid approach for AI-generated text detection. arXiv. https://doi.org/10.48550/arXiv.2406.06558

Zhuang, J., Li, G., Xu, H., Xu, J. & Tian, R. (2024). TEXT-TO-CITY: Controllable 3D Urban Block Generation with Latent Diffusion Model. In 29th International Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA 2024 (pp.169-178). The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA).