

From Concept to Consistent Multi-View Renders

Integrating shape grammars with Stable Diffusion in architectural workflow

Yichao Shi¹, Jiaxiang Luo^{2✉}, Chunlan Wang³, Tzu-Chieh Kurt Hong⁴
and Tian Tian Lo⁵

^{1,3}Georgia Institute of Technology, ²ETH Zurich,

⁴University of Kansas, ⁵Hong Kong Polytechnic University.

¹yichao.shi@gatech.edu, ORCID: 0009-0000-7715-8277

²jialuo@ethz.ch, ORCID: 0000-0002-6982-3692

³cwang932@gatech.edu, ORCID: 0009-0004-8096-3217

⁴kurt.hong@ku.edu, ORCID: 0000-0002-7716-683X

⁵tt-sky.lo@polyu.edu.hk, ORCID: 0000-0002-1992-0777

Abstract. The incorporation of artificial intelligence (AI) in architectural design has progressed markedly. Nonetheless, current workflows are constrained. A major issue is the inconsistency in AI-generated renderings from different perspectives of the same architectural design. Moreover, there is the problem of insufficient integration with the conceptual design phase. This paper introduces a unified workflow that combines shape grammars with Stable Diffusion to enhance architectural visualization. The proposed methodology employs a three-stage computational process: first, upon inputting the design parti, it produces diverse and feasible floorplans via shape grammars implemented in the Shape Machine; next, it extrudes these floorplans into three-dimensional models; finally, it renders these models into high-quality, cohesive interior visuals from various perspectives utilizing Stable Diffusion, enhanced with ControlNet and LoRA. The results demonstrate that our autonomous and efficient workflow substantially reduces design and rendering time, while enhancing control and flexibility in design generation. This workflow streamlines the design process from initial concepts to final presentations, improving the quality and consistency of renders relative to conventional AI workflows.

Keywords. Shape Grammars, Stable Diffusion, Computational Design, Architectural Workflow, Automated Floorplan Generation

1. Introduction

Artificial intelligence (AI) in the field of architecture serves as a catalyst for innovation, accuracy, and creative expression. A precursor to the two primary design methodologies, subtraction and generative processing, is the ascendancy of generative AI and symbolic AI in the design realm. Generative AI, in conjunction with deep

learning, can facilitate the development of architectural designs, exploration of various styles, and planning of layouts (Li et al., 2024). Diffusion models, such as Stable Diffusion (SD), have emerged as leaders in representation, demonstrating proficiency in generating aesthetically pleasing text-to-image (T2I) and image-to-image (I2I) translations (Zhang et al., 2023). Conversely, symbolic AI is represented through shape grammars (SG), which rely on formalized knowledge integrated within structural exploration (Stiny, 1980).

As a generative AI model, SD has been extensively utilized in the creation of images that stimulate creative ideas in architecture. However, current SD exhibits significant limitations. This encompasses inadequate architectural feasibility, inconsistent renderings, limited output control, and inefficiency in producing practical designs. Contemporary SD workflows, encompassing T2I, I2I, and inpainting generative techniques, frequently fail to satisfy the intricate requirements of architecture. The designs produced by SD may appear appealing but can be impractical or inconsistent from various perspectives. Therefore, although these models can produce aesthetically pleasing images, they frequently lack the structural coherence essential for architectural significance. Consequently, there is a necessity for techniques that produce high-quality, consistent, and intelligible renderings from various perspectives. This will guarantee that designs are aesthetically pleasing and architecturally precise.

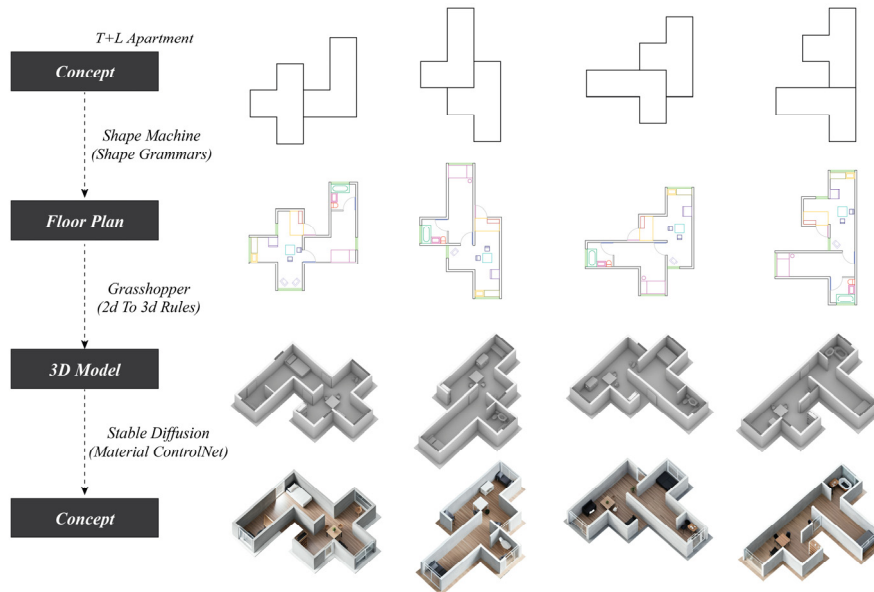


Figure 1. The proposed workflow incorporating SG and SD in interior design for apartments

Rule-based systems like SG offer solutions to the limitations of generative AI models. The SG interpreter, Shape Machine (SM), enables systematic investigation of design variations within defined parameters. This approach maintains control and architectural integrity (Economou et al., 2021). Architects are allowed to work with shape rules with

SG and SM to create diverse floorplan designs that balance creativity and functionality (Hong, 2021). Moreover, recent advancements in supplementary networks of the SD models exhibit potential to overcome the limitations of the original workflows. ControlNet and LoRA improve output quality by enhancing control and refining the generation process (Hartley et al., 2024). ControlNet ensures consistency of models with specified input patterns. LoRA facilitates the efficient adaptation of large models to new tasks without requiring extensive retraining. However, a cohesive process for integrating SG and SD to generate consistent, high-quality interior renders from multiple perspectives is presently lacking, which limits design optimization and obstructs communication between designers and stakeholders. Thus, we proposed a balance workflow between SG and SD as shown in Figure 1.

This study aims to present a cohesive workflow that integrates SG with SD, as shown in Fig 1. The approach integrates the regulatory and structural integrity of rule-based shape computation with the computational capabilities of SD. It starts with generating varied and practical floor plans, then transforms them into 3D models, and produces uniform interior visuals from multiple viewpoints. This workflow seeks to augment design efficiency, enhance render quality and consistency, as well as improve communication between designers and stakeholders.

2. Explorations on Rationalizing AI Rendering

Stable Diffusion (SD), a popular artificial intelligence tool, is already used by architects to produce detailed and lifelike representations. People utilize it to experiment with design concepts and implement different art styles (Li et al., 2024; Cao et al., 2024; Zhang et al., 2024). However, SD has distinct limitations. Its output frequently shows a deficiency in control and logic. Numerous outcomes appear favorable but lack practical values. This is partly due to the lack of consistency across various perspectives. Technologies such as ControlNet and Material-ControlNet enhance precision (AARG-FAN, 2024). However, these efforts still cannot make a coherent workflow for the architect to obtain renders from their initial concepts in a rational way.

Shape Grammar (SG) functions differently. It employs rules to produce designs that are logical and consistent. Stiny (1980) introduced this system to define forms via shape rules. Economou et al. (2021) created the Shape Machine (SM), which enables users to operate SG within software such as Rhino. This tool assists designers in accurately analyzing and generating designs (Economou & Hong, 2023). People have used SM to recreate historic architectural designs such as the Ionic Porches and Chinese brackets (Shi et al., 2024; Shi & Economou, 2024). The designs created through SG are typically precise and logical. In "Herds of Architecture II," Hong (2021) illustrates the automation of the architectural design process through Shape Machine. This project enables designers to visually develop from concepts to refined designs, which lays a solid reference for this research. However, its outcomes lack the visual profundity necessary for effective idea presentation in architectural practice. Shi and Wang (2024) combine SD and SG to address the randomness of AI, using SG to generate an augmented dataset for training SD models. However, they were unable to create a coherent workflow capable of producing a manageable architectural design with multi-view consistency. This highlights the gap in the project's research.

3. Methodology

The methodology begins with the input of the initial design concept, utilizing fundamental T and L shapes as the primary elements for the architectural floor plan of the apartment. These elements are then processed through a Shape Grammar (SG) system, following predefined rules to generate the floor plan. Upon execution of the SG rules, the system produces a detailed floor plan, incorporating architectural features and furniture layouts. The use of distinct colors and information layers allows for clear identification of walls, windows, and furniture within the design, facilitating the creation of a three-dimensional model. The resulting 3D model, along with its associated channel map representing material properties, is configured in Rhino using layer-based organization. This model is then imported into Stable Diffusion (SD) for rendering. To ensure consistency across perspectives, ControlNet is employed to refine the rendering process, maintaining coherence in material representation and object details. This integrated approach enables precise visualization and efficient design iterations.

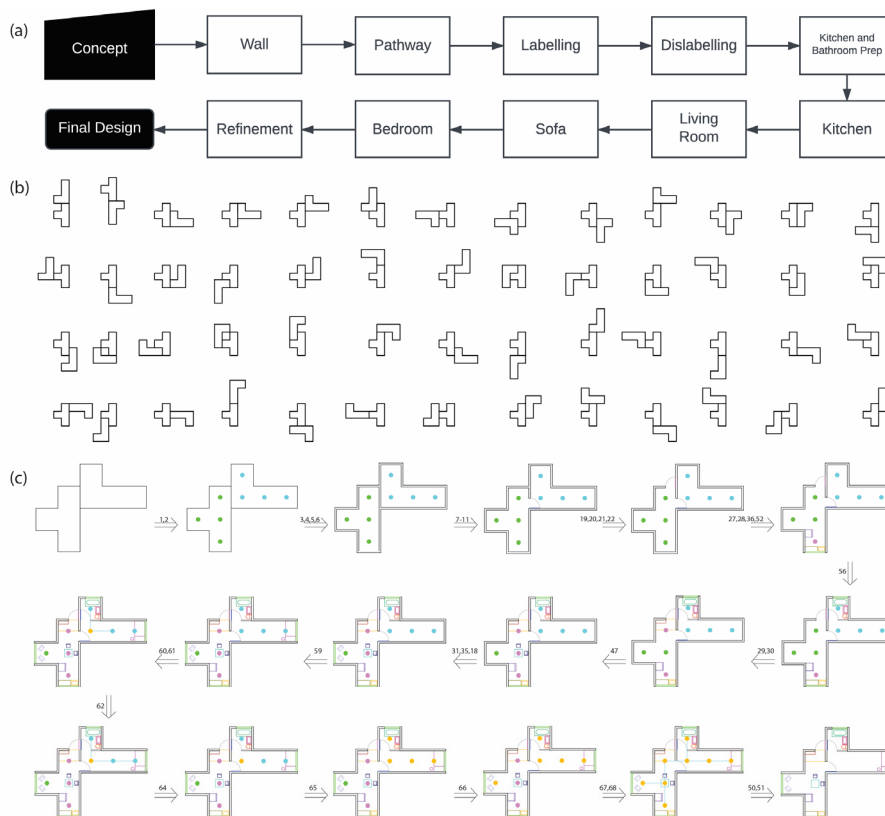


Figure 2. The shape rules used for floor plan configurations generation

3.1. STAGE 1: FLOORPLAN GENERATION WITH SHAPE GRAMMARS

As shown in Fig 2, SG functions as a rule-based production system, enabling the derivation of design configurations through the application of various shape rules or by altering the sequence of rule applications. The use of SM enables the compilation of rules using Rhino, a CAD software based on NURBS modeling. Conditions, jumps, and loop logic can be established to automate the application of all shape rules within SM. In SM, the rules applicable to a specific program can be categorized into blocks that authors can designate by name.

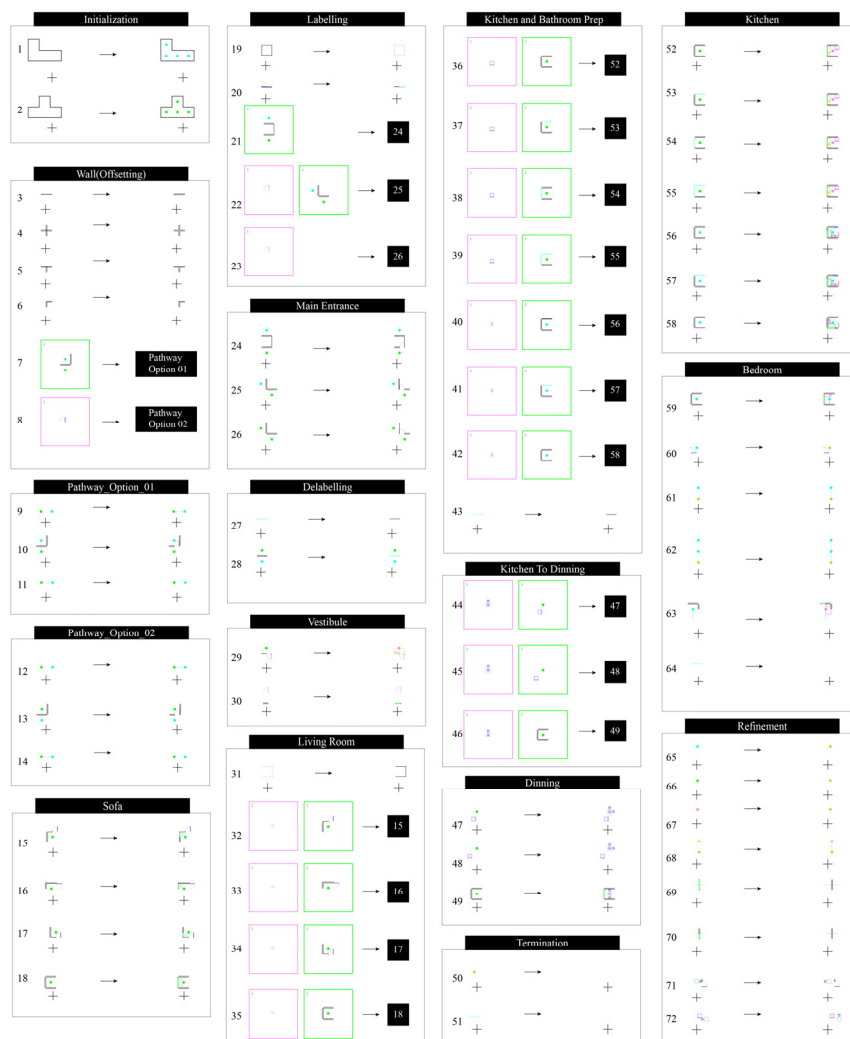


Figure 3. The operations of SG in SM: (a) Control flow (b) Initial shapes (c) productions

In SM, the implementation of shape rules relies on a control flow incorporating jumps, conditions, and loops. The program's workflow is primarily linear, executing judgments sequentially as depicted in Figure 3(a). This program's design concept integrates two geometries, T and L, employing different ways of geometric composition as illustrated in Figure 3(b). Each of these concepts can be used as an input to the SM program. SM subsequently establishes the rules depicted in Figure 2 and employs them for these conceptual shapes. It will autonomously recognize the applicable rules for every step and progressively generate the floor plan designs.

3.2. STAGE 2: 3D MODELING WITH SHAPE PRODUCTIONS

Given that SM can compile points and lines of various colours from distinct layers in Rhino3D, we can utilize the data from specific layers and colours to represent the three-dimensional information of each building component. For instance, black polylines can represent the boundaries of walls, whereas each furniture item will be indicated by an individual layer and colour. Consequently, the shape representations from Figure 3(c) can be transformed into three-dimensional models. We utilize a Grasshopper plug-in named Elefront, which enables the extraction of objects directly into the parametric design environment based on colour and layer data from the Rhino3D interface. Subsequently, we can create wall and glass models utilizing the extrude command, or position the appropriate furniture model accurately, thereby obtaining the 3D model corresponding to each type of floor plan configuration. This represents a method of transforming the SG from two-dimensional to three-dimensional information, as shown in Figure 4.

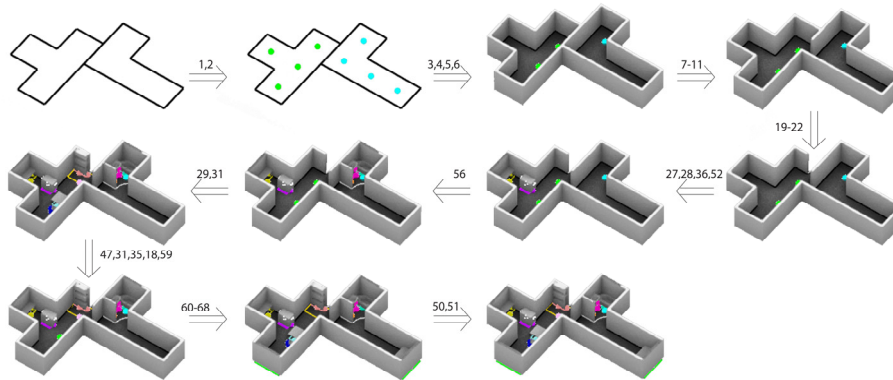


Figure 4. 3d Building model productions with SG and Elefront

3.3. STAGE 3: GENERATIVE 3D MODEL RENDERING

Utilizing the 3D models produced via SG to delineate the spatial configurations of the floor plan designs, Generative AI was subsequently employed to maintain stylistic coherence and uniform material rendering across various perspectives. Instruments such as SD, LoRA, and ControlNet were utilized to generate these models. This method evaluated the feasibility of combining generative AI with SG. We assessed the model's capacity to sustain consistent materials and styles across varying conditions by employing uniform prompts and parameters.

We incorporated images from multiple axonometric perspectives to study material specifics, color discrepancies, and furniture elements. The rendering effects have been evaluated utilizing various setups:

- a. Prompts employing a large-scale model: This configuration showed the fundamental materials and styles, revealing the generative AI's responses to the prompts.
- b. The integration of a large-scale model with LoRA enhanced detail and produced a more sophisticated representation of each style.
- c. Prompts utilizing a large-scale model and ControlNet: This configuration offered enhanced control over the 3D model. It enabled precise assessment of material variations while maintaining uniformity in materials and structures across perspectives.
- d. Multi-view input testing: Axonometric images from various angles were utilized to assess the consistency of results across all configurations.

The experiments demonstrated that the integration of generative AI and SG produces consistent, high-quality multi-view renderings. The methodology preserved consistent materials and styles while providing enhanced control and precision as shown in Figure 5.

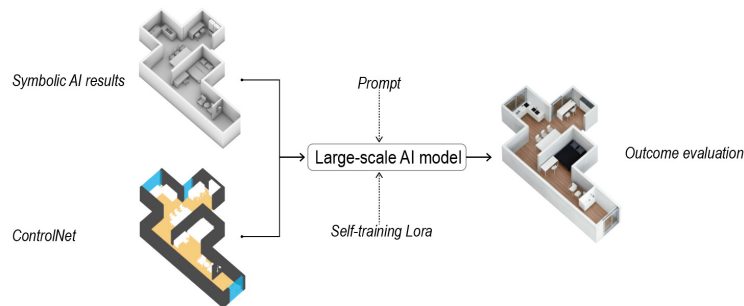


Figure 5. The workflow from SG outcomes to SD

4. Results

Figure 6 illustrates the outcomes of generative AI training from multiple axonometric viewpoints. Group A, positioned horizontally in the illustration, commences with A1 as the foundational background image for the training map. The prompt utilized for this phase was: "3D floor plan, contemporary residence, white furnishings, timber flooring, concrete wall." An authentic architectural model was utilized for the training. The outcomes were inadequate. The images displayed edge outlines; however, the content was disordered and exhibited perspective distortions, as illustrated in A2, A3, and A4 of Figure 6.

Group B presented a self-trained LoRA model integrated with ControlNet's outline functionality. The larger model and prompts were unmodified. B2 produced better results, featuring more defined shapes and enhanced color alignment with the prompts. However, B3 and B4 exhibited reduced stability. B3 introduced conjectural elements, resulting in deviations of walls and floors from the initial design. B4 exhibited a more

precise shape but did not conform to the timber floor specifications, as all materials were displayed as concrete.

Group C adhered to the identical prompts and model style, employing a self-trained LoRA model in conjunction with a pre-trained ControlNet. This ControlNet functioned similarly to semantic segmentation but was specifically designed for architectural models. It associated material types with distinct colors, including yellow for wood, blue for glass, and dark grey for concrete. The floor material was designated solely in yellow. Group C obtained uniform material color schemes, consistent designs, and stable furniture arrangements from various viewpoints.

Group D incorporated blue as an indicator for glass windows, in addition to yellow for wood. The findings indicated that Group D effectively produced glass textures in the designated areas. It preserved uniform material colors, styles, and furniture arrangements throughout all perspectives.

The concrete wall material was unlabelled in the tests for Groups C and D. This resulted in the concrete material appearing sporadically across the six rendering outputs. Controlling its frequency proved challenging. The outcomes were excessively

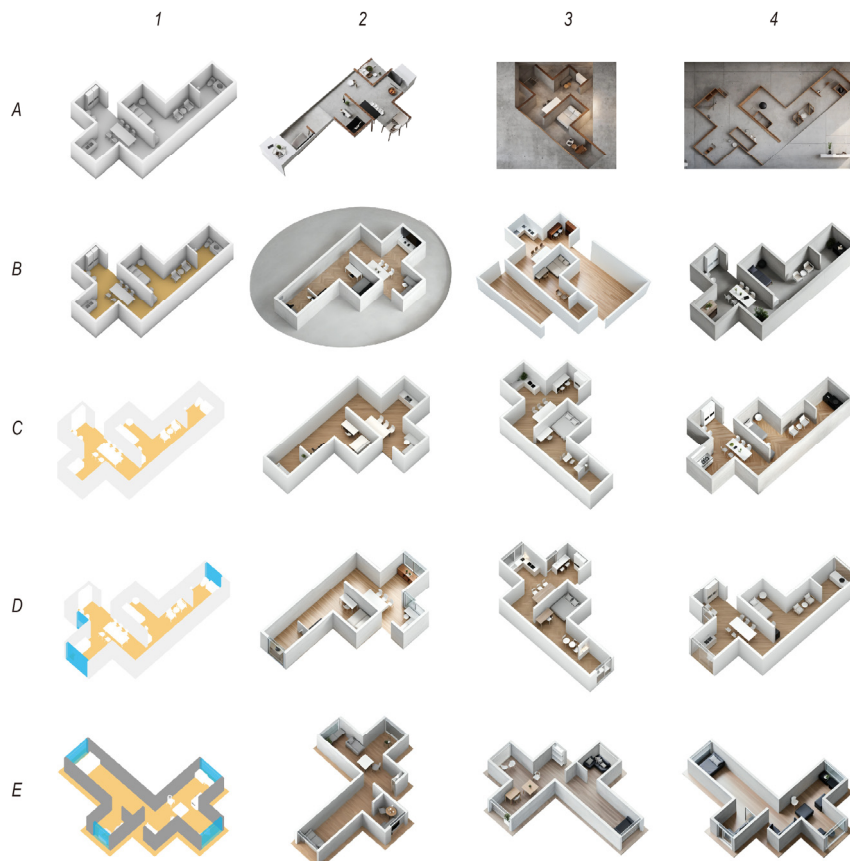


Figure 6. Productions from SD based on the outcomes of SG

erratic as they depended solely on prompt words. In Group E, an alternative floor plan was utilized as the training subject. This three-dimensional house plan was developed utilizing symbolic artificial intelligence. It featured a more straightforward presentation of its furniture as shown in Figure 6.

The outcomes for Group E indicated superior performance. The rendering closely aligned with the input reference of E1. The concrete walls, wooden floors, and glass windows exhibited stability in the outputs. The furniture, nonetheless, remained devoid of uniform colors and styles.

5. Discussion

This integrated process amalgamates the advantages of SG and SD to achieve outcomes that conventional methods struggle to replicate. SG offers a systematic framework for producing designs. Generative AI enhances adaptability and expressive rendering proficiency. By integrating SG with SD's multi-view rendering technique, we establish an elaborate design process. This methodology facilitates new possibilities for innovation and visual coherence in architecture.

Some limitations persist in the results. Obtaining uniform material textures from various viewpoints is still challenging for the tiny elements in the floor plans. The rendering of small furniture models frequently exhibits suboptimal performance due to certain types being misidentified by SD models. Additionally, SG, as a rule-based system, depends on predefined shape rules, which limits its adaptability to specific configurations. For example, the current set of shape rules can only generate T+L apartment layouts.

Despite these challenges, the workflow exhibits significant adaptability. An individual proficient in SG can readily adjust the shape rules to integrate new design concepts. For example, transitioning from T+L to T+L+L layouts necessitate slight modifications to the shape rules. The framework can subsequently automate the complete process, from design generation to rendering.

The findings offer a significant benchmark for attaining uniformity in multi-view 3D renderings. They also create a basis for subsequent enhancements and implementations in architectural design.

6. Conclusion

The combination of Shape Grammars (SG) and Stable Diffusion (SD) into a unified workflow has greatly enhanced the coherence and utility of AI-generated designs. This workflow eliminates many drawbacks of conventional AI design tools by integrating rule-based methodologies with generative AI. SG offers a systematic and rational framework for examining designs. SD, working together with tools such as ControlNet and LoRA, enhances rendering expressiveness while preserving stylistic coherence across various perspectives. This approach enhances design efficiency and precision. It also facilitates more effective communication between designers and stakeholders.

Several challenges remain in this study. Consistently rendering tiny elements remains challenging, as well as dependent on established shape rules that limit adaptability. Still, SG is versatile, enabling the refinement and evolution of the workflow as required. This research demonstrates how the integration of symbolic and

generative AI may overcome conventional design limitations. It creates new opportunities for innovative, efficient, and visually cohesive architectural processes.

References

- AARG-FAN. (2024). Material-Controlnet [Computer software]. GitHub. <https://github.com/AARG-FAN/Material-Controlnet>
- Cao, Y., Abdul Aziz, A., & Mohd Arshard, W. N. R. (2024). Stable diffusion in architectural design: Closing doors or opening new horizons? *International Journal of Architectural Computing*, 14780771241270257. <https://doi.org/10.1177/14780771241270257>
- Economou, A., & Hong, T.-C. K. (2023). Back to the Drawing Board: Shape Calculations in Shape Machine. In J. S. Gero (Ed.), *Design Computing and Cognition'22* (pp. 549–567). Springer International Publishing. https://doi.org/10.1007/978-3-031-20418-0_33
- Economou, A., Hong, T.-C., Ligler, H., & Park, J. (2021). Shape Machine: A Primer for Visual Computation. In J.-H. Lee (Ed.), *A New Perspective of Cultural DNA* (pp. 65–92). Springer Singapore. https://doi.org/10.1007/978-981-15-7707-9_6
- Hartley, Z. K. J., Lind, R. J., Pound, M. P., & French, A. P. (2024). Domain Targeted Synthetic Plant Style Transfer Using Stable Diffusion, LoRA and ControlNet. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 5375–5383. <https://doi.org/10.1109/CVPRW63382.2024.00546>
- Hong, T.-C. K. (2021). Herds of Architecture II. Retrieved November 5, 2024, from <https://k9krnd.net/Research-Shape-Machine-Herds-of-Architecture-II>
- Huang, W., & Zheng, H. (2018). Architectural Drawings Recognition and Generation through Machine Learning. *ACADIA*, 156–165. <https://doi.org/10.52842/conf.acadia.2018.156>
- Li, C., Zhang, T., Du, X., Zhang, Y., & Xie, H. (2024). Generative AI for Architectural Design: A Literature Review. *arXiv*. <http://arxiv.org/abs/2404.01335>
- Shi, Y., & Economou, A. (2024). Dougong Revisited: A Parametric Specification of Chinese Bracket Design in Shape Machine. In J. S. Gero (Ed.), *Design Computing and Cognition'24* (pp. 233–249). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71918-9_15
- Shi, Y., Mamoli, M., & Wang, C. (2024). Automated Generative Design in Archaeological Reconstruction of Greek Architecture.
- Shi, Y., & Wang, C. (2024). Optimizing Ionic-style facade creation with Stable Diffusion text-to-image using shape grammar. Paper presented at the 6th International Conference on Computational Design and Robotic Fabrication (CDRF 2024), Shanghai, China.
- Stiny, G. (1980). Introduction to Shape and Shape Grammars. *Environment and Planning B: Planning and Design*, 7(3), 343–351. <https://doi.org/10.1068/b070343>
- 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. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 3813–3824. <https://doi.org/10.1109/ICCV51070.2023.00355>
- Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene Parsing through ADE20K Dataset. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5122–5130. <https://doi.org/10.1109/CVPR.2017.544>
- Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., & Torralba, A. (2019). Semantic Understanding of Scenes Through the ADE20K Dataset. *International Journal of Computer Vision*, 127(3), 302–321. <https://doi.org/10.1007/s11263-018-1140-0>