From Sketches to Renderings:

A Comparison of Rapid Visualization Methods for Sketches Based on ControlNet (CN)

Sucheng Yao
School of Architecture and Design
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
sucheng.yao1@kmutt.ac.th

Kanjanee Budthimedhee*
School of Architecture and Design
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
budthime@yahoo.com

***Abstract*—The use of the ControlNet (CN)-controlled Stable Diffusion (SD) model increases the potential for artificial intelligence to enhance efficiency in architectural design. This study compared the results of using different CN processors (Canny, MLSD, Lineart, Scribble, Softedge) to process various styles of design sketches and generate architectural renderings. The aim was to explore the potential of each CN processor to visualize architectural design sketches. The results indicate that with appropriate CN control, architectural hand-drawn sketches can be quickly visualized, yielding satisfactory outcomes. The comparisons revealed that: (1) CN performs better with orderly hand-drawn sketches; (2) CN weights between 0.5 and 1.0 are more effective for image control; and (3) minimizing conceptual lines in hand-drawn sketches reduces the likelihood of rendering failures. This study provides a foundation for future research in architectural and planning image generation and offers a more efficient and cost-effective design method for architectural professionals.**

Keywords-generative design, design method, ControlNet, Stable Diffusion, architectural

I. INTRODUCTION

Artificial Intelligence (AI) is a discipline that studies the use of computers to simulate certain human thought processes and intelligent behaviors, such as learning, reasoning, thinking, and planning. This field has been developing since the 1950s. Machine Learning (ML) is a subfield of AI that focuses on developing techniques for recognizing data patterns. Generative Adversarial Networks (GANs) represent an emerging area within deep learning [1]. Most research in the fields of architectural design and urban planning uses GANs to generate building facades [2]and layouts[3]. However, the volume of training samples for GANs has always been a critical obstacle[4]. Although unsupervised models like DCGAN can train on large datasets, handling downstream tasks remains challenging[5]. Additionally, designers in the application sector rarely use the Python language, making it challenging for non-computer science professionals to leverage AI for enhancing design efficiency.

The Stable Diffusion (SD) model is designed for the task of generating images from prompts. In contrast to generative models [6]like Midjourney [7]and DALLE2 [8], which produce uncontrolled drawings, ControlNet (CN) is a neural network architecture that adds spatial conditioning control to large, pre-trained text-to-image diffusion models[9]. Extensive research has shown that CN facilitates control over image diffusion models in applications. When used in conjunction with CN, SD can control the generated results, producing more precise design drawings. For ease of model loading and image generation, mainstream research currently employs the Web UI of Stable Diffusion as the control system. The Web UI provides a more intuitive interface for users, making it more favorable for practical applications.

SD offers designers a simpler creative tool. Based on the practical application of SD and CN, this paper compares five linear processors in CN (Canny, MLSD, Lineart, Scribble, Softedge) and explores a method for transforming architectural design sketches into design renderings. This method can effectively intervene in the preliminary design phase, addressing the time-consuming issue of rapid visualization in conventional design stages.

II. METHOLOGY

A. Data Set

On the Pinterest platform, search for the keywords: "hand-drawn, design, architecture, environment"; filter the results to exclude colors, focusing on pen line drawings featuring architectural environments in a 45-degree human perspective. This search will yield three common types of hand-drawn sketches, including: one simple schematic sketch without precise drafting tools (Sketch 1), one schematic design sketch that combines precise drafting and freehand drawing (Sketch 2), and one schematic design sketch created entirely with drafting tools (Sketch 3). These sketches simulate three different hand-drawing methods that designers might use for rapid design visualization.

B. Experimental Process

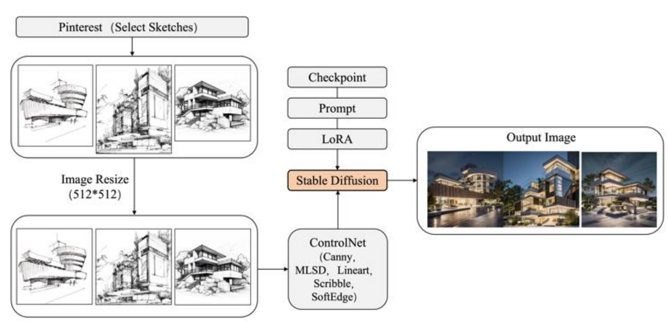

Fig. 1 Experimental Process

For the study, the Checkpoint model (laowang_ARCH_MIX.V0.5Fp16Fix.safetensors) was selected as the base model. This model has undergone extensive iterative training and produces realistic-style outputs. Each group will use the same LoRA model (<lora:UIAXD4.0>) and

2022 Suzhou Polytechnic Institute of Agriculture of Young Teachers Scientific Research Ability Enhancement Program Project (QN[2022]05)

T.I. Embedding model (Juggernaut Negative Embedding _v1.0), with consistent weight controls. The online experiments will be conducted on the Alibaba Cloud AI platform PAI, employing a fast, creative, and high-quality sampling method and scheduler type: DPM++ 2M Karras, iterating for 25 steps[4]. The CFG Scale will be set to 7.

Image control comparisons will be based on ControlNet v1.1.449 linear processors (Canny, MLSD, Lineart, Scribble_xdog, Softedge). Both input and output images will be 512x512 pixels. The same semantic generation parameters (architectural design, modern style, full frame, landscape, surrealism, iso400) will be used in SD to compare the impact of different preprocessors and models on the generated results. The study will also compare the effects of different ControlNet control weights (0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0) on model output quality(Fig. 1).

III. RESULTS AND ANALYSIS

A. Canny

Under the same parameter settings, the images generated based on the Canny processor did not yield realistic drawings for any of the three sketches. For Sketch 1, the hand-drawn contour lines remained prominent. Sketch 2 contained many extraneous lines, resulting in a rather chaotic overall image. Sketch 3, which had richer sketch details, produced the best results; however, the enhanced contours caused the vegetation in the image to appear two-dimensional. Comparing the three generated images with the hand-drawn sketches, it was observed that the Canny processor exerts a strong control over the hand-drawn lines. This control can enhance the representation of architectural or landscape forms in the sketches, but the presence of unrecognizable lines can degrade the quality of the generated images(Fig. 2).

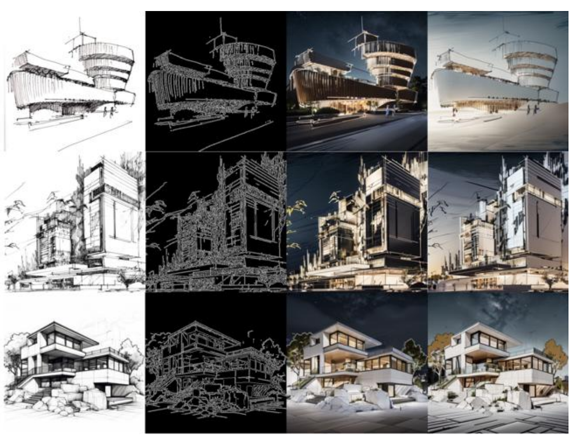

Fig. 2 Effect Picture Generate by Canny.
Prompt：<lora:UIAXD4.0:0.6>, architectural design, modern style, full frame， landscape,surrealism,iso400. Negative prompt: Juggernaut Negative Embedding _v1.0（Scale=7.0， W=512,H=512， ContolNet Model：control_v11p_sd15_canny_fp16， Weight：1.0）.
From top to bottom: Sketch1,2,3

Figure 3 Effect Picture Generate by Different Weight of Canny.
The weight from left to right：0.25， 0.5， 0.75， 1.0， 1.25， 1.5， 1.75， 2.0

The CN weight limits the machine's inference capability[4]. When the control weight is reduced (0-0.75), the generated images appear more realistic. However, when the overall weight reaches 1.0 and above, especially beyond 1.25, the line elements in the images are intensified. Among the three styles of sketches, Sketch 3 produced the best results based on the Canny processor, given its detailed representation（Fig. 3）.

B. M LSD

The images generated using the MLSD processor generally exhibited low accuracy. For Sketch 1, which features curved architectural forms, the processor failed to recognize the building contours, resulting in generated images with insufficient correlation to the original sketch. However, for Sketch 2 and Sketch 3, which contain more straight lines, the processor was better able to recognize these elements and produce similar images. Nonetheless, details such as window openings and staircases in the building facades showed significant alterations. Overall, the images generated using MLSD introduced considerable changes to the original sketches, although the resulting images tended to appear more realistic(Fig. 4).

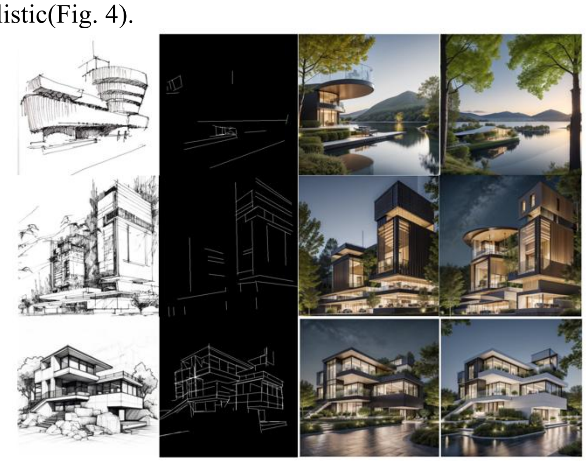

Fig. 4 Effect Picture Generate by MLSD.
Prompt：<lora:UIAXD4.0:0.6>, architectural design, modern style, full frame， landscape,surrealism,iso400. Negative prompt: Juggernaut Negative Embedding _v1.0（Scale=7.0， W=512,H=512， ContolNet Model：control_v11p_sd15_mlsd_fp16， Weight：1.0）.
From top to bottom: Sketch1,2,3

Fig. 5 Effect Picture Generate by Different Weight of MLSD.
The weight from left to right：0.25，0.5，0.75，1.0，1.25，1.5，1.75，2.0

According to Fig. 6, it was found that within the 0-2.0 weight range, there was no loss of style. The generated images from different style sketches remained relatively stable in terms of their effects（Fig. 5）.

C. *Lineart*

The images generated using the Lineart processor showed good results in terms of accuracy and realism for Sketch 2 and Sketch 3. However, Sketch 1 exhibited some instances where the generated images lacked material details and retained prominent hand-drawn lines. Observing Sketch 1, Sketch 2, and Sketch 3, it is evident that the Lineart processor struggles to accurately recognize unclear lines in the sketches, such as the lines at the top of the building in Sketch 1 and the background lines in Sketch 2 and Sketch 3(Fig. 6).

Fig. 6 Effect Picture Generate by Lineart.
Prompt：<lora:UIAXD4.0:0.6>, architectural design, modern style, full frame， landscape,surrealism,iso400. Negative prompt: Juggernaut Negative Embedding _v1.0（Scale=7.0，W=512,H=512，ContolNet Model：control_v11p_sd15_lineart_fp16，Weight：1.0）.
From top to bottom: Sketch1,2,3

Fig. 7 Effect Picture Generate by Different Weight of Lineart.
The weight from left to right：0.25，0.5，0.75，1.0，1.25，1.5，1.75，2.0

Overall, the CN weights between 0.25 and 1.0 maintain a good balance of overall style and image texture. However, within the 1.0 to 2.0 range, the architectural forms in the three hand-drawn sketches increasingly resemble the hand-drawn effect. Meanwhile, the realism of the images, including aspects like material, lighting, and environment, progressively deteriorates as the weight increases(Fig. 7).

D. *Scribble*

Scribble has two linear processors: Scribble_pidinet and Scribble_xdog. The Scribble_pidinet processor only captures simple outer contours and converts them into a doodle-like form, making it unsuitable for controlling architectural design, so its generated effects were not explored in this study. The output effects based on the Scribble_xdog processor were relatively stable, with almost no image breakdowns in the generated images from the three sketches(Fig. 8). The CN effectively controlled the design forms. Although the control over sketch shapes varied with changes in CN weight, the Checkpoint model's style remained stable. Even when the weight exceeded 1.0 and Sketch 1 began to show unclear distinctions between interior and surrounding environments, such as some images blending building interiors with other ancillary structures, the overall images did not break down(Fig. 9).

Fig. 8 Effect Picture Generate by Scribble.
Prompt：<lora:UIAXD4.0:0.6>, architectural design, modern style, full frame， landscape,surrealism,iso400. Negative prompt: Juggernaut Negative Embedding _v1.0（Scale=7.0，W=512,H=512，ContolNet Model：control_v11p_sd15_scribble_fp16，Weight：1.0）.
From top to bottom: Sketch1,2,3

Fig. 9 Effect Picture Generate by Different Weight of Scribble
The weight from left to right：0.25，0.5，0.75，1.0，1.25，1.5，1.75，2.0

E. *Softedge*

Using the Softedge processor, the generated images for Sketch 1 and Sketch 2 resemble marker-filled drawings, with very prominent contour lines and a lack of realism. Although Sketch 3 produced relatively realistic textures, the overall details were compromised, such as the ground tiles and foreground rocks(Fig. 10). When the CN weight varied, Sketch 1 and Sketch 2 began to show partial material loss at weights greater than 0.5. At weights above 1.0, the images were almost entirely uncontrolled by the Checkpoint model. When the CN weight exceeded 1.5, Sketch 3 also lost all realistic details, including the sky, textures, and vegetation(Fig. 11).

Fig. 10 Effect Picture Generate by Softedge.
Prompt：<lora:UIAXD4.0:0.6>, architectural design, modern style, full frame， landscape,surrealism,iso400. Negative prompt: Juggernaut Negative Embedding _v1.0（Scale=7.0， W=512,H=512，ContolNet Model：control_v11p_sd15_softedge_fp16， Weight：1.0）.
From top to bottom: Sketch1,2,3

Fig. 11 Effect Picture Generate by Different Weight of Softedge
The weight from left to right：0.25， 0.5， 0.75， 1.0， 1.25， 1.5， 1.75， 2.0

IV. DISCUSSION

Based on the research results, the following findings were observed:

1. When using CN to control SD model for generating effect pictures, hand-drawn sketches that use precise drafting tools and have fewer extraneous lines result in more accurate CN-controlled images. Freer linework leads to more varied architectural results. However, the MLSD processor is not suitable for designs with few straight-line elements. The Canny and Softedge processors are not suitable for highly abstract conceptual sketches. To generate more reference-worthy drawings rather than conceptual images, clear hand-drawn expressions are essential during the sketching stage.

2. Among the CN weights ranging from 0 to 2.0, the Scribble_xdog processor retains the Checkpoint model's style better than the other four processors. For accurate control over the images, it is recommended to keep the processor weights between 0.5 and 1.0. This range maintains the Checkpoint model's style while ensuring CN's control over the images.

3. When using the SD model for generating schematic drawings, it is advisable to minimize the use of overly artistic or abstract expressions, as the model cannot recognize the meaning of unknown lines, which can negatively impact the generated results.

However, this method has certain limitations. First, the SD model-generated drawings lack a sense of scale, which is crucial for the implementation of a project. Therefore, this method is currently only applicable to the conceptual phase, particularly for visualizing ideas and concepts. Additionally, it is necessary to consider integrating this approach with other traditional design software such as Photoshop and SketchUp. For instance, using the MLSD processor to generate line drawings that are then refined in Photoshop before generating more accurate results can lead to more precise design outcomes.

V. CONCLUSIONS

This study compared methods for generating architectural renderings from hand-drawn sketches using SD model controlled by CN. Five linear processors—Canny, MLSD, Lineart, Scribble, and Softedge—were used for generation comparisons. Additionally, the control weights of CN were adjusted to compare the results of each processor and model under different weights. This process enables designers and design students to more easily visualize and refine their ideas during the design process, facilitating better communication with clients and peers. Overall, the use of artificial intelligence can enhance the efficiency of designers' work. However, the application of foundational design thinking and the creativity of designers will become increasingly important in future design work.

ACKNOWLEDGMENT

None

REFERENCES

[1] R, Robin, B, Andreas., L, Dominik, E, Patrick, O, Bjorn, High-resolution image synthesis with latent diffusion models, IEEE/CVF Conf. Comput. Vision Pattern Recogn. (CVPR) 2022, pp. 10674–10685.

[2] I, Phillip, Z, Junyan, Z, Tinghui, E, Alexei.A, Image-to-image translation with conditional adversarial networks. IEEE Conf. Comput. Vision and Pattern Recogn. (CVPR) 2017, pp. 5967–5976.

[3] H, Weixing, Z, Hao, Architectural drawings recognition and generation through machine learn-ing. Proceedings of the 38th Annual Conference of the Association for Computer Aided 、 Design in Architecture, Mexico City, Mexico, 2018, pp. 18–20.

[4] M, Haoran and Z, Hao, Text Semantics to Image Generation: A method of building facades design base on Stable Diffusion model, Proceedings of the 5th International Conference on Computational Design and Robotic Fabrication (CDRF 2023), 2024, pp. 24-34.

[5] R, Robin, B, Andreas., L, Dominik, E, Patrick, O, Bjorn, High-resolution image synthesis with latent diffusion models, IEEE/CVF Conf. Comput. Vision Pattern Recogn. (CVPR) 2022, pp. 10674–10685.

[6] R, Jian, X, Mengzhen, A New Art Design Method Based on AIGC: Analysis from the Perspective of Creation Efficiency, 2023 4th International Conference on Intelligent Design (ICID), 2023, PP.129-134.

[7] A, Borji, Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2, 2022, arXiv:2210.00586.

[8] R, Robin, B, Andreas., L, Dominik, E, Patrick, O, Bjorn, High-resolution image synthesis with latent diffusion models, IEEE/CVF Conf. Comput. Vision Pattern Recogn. (CVPR) 2022, pp. 10674–10685.

[9] Z, Lvmin, R, Anyi, A, Maneesh, Adding Conditional Control to Text-to-Image Diffusion Models, 2023, ArXiv, abs/2302.0554