GENERATIVE DESIGN INTUITION FROM THE FINE-TUNED MODELS OF NAMED ARCHITECTS' STYLE

Youngjin Yoo, Hyun Jeong & Youngchae Kim
Department of Interior Architecture and Built Environment, Yonsei University, South Korea
Seung Hyun Cha
Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology (KAIST), South Korea
Jin-Kook Lee
Department of Interior Architecture and Built Environment, Yonsei University, South Korea

ABSTRACT: *This paper suggests the potential application of generative artificial intelligence-based image generation technology in the field of architecture, for early phase shape planning, using the styles of renowned architects. The study employed the following approaches: 1) Intensive image generation based on the styles of 20 architects to test the AI's recognition ability and image quality. 2) Additional training was conducted for architects with low recognition rates to construct an enhanced learning model in the quality of image generation. 3) In addition to generating architectural visualization images using existing architects' design styles, alternative styles were proposed through design combinations, aiming to concretize ambiguous idea communication in the early stages of design and enhance its efficiency. The study sheds light on the future prospects of applying this generative AI model in the field of architecture.*

KEYWORDS: *Design Style of Architects, Generative AI, Image Generation, Fine-tuning*

1. INTRODUCTION

In the field of architecture, visualization plays a crucial role in comprehending and evaluating complex design alternatives and spatial qualities [Greenberg, 1974]. Especially in the early design stages, it allows clear expression of design ideas and spatial concepts, enabling the identification and resolution of potential issues and facilitating effective communication among stakeholders [Akin, 1978]. Ultimately, early-stage visualization defines the design direction, enhances collaboration, efficiency, and leads to better outcomes. However, creating high-quality visualization images, particularly during the abstract design phases, remains challenging. While advancements in 3D modeling and rendering have improved the realism of visualizations, the process still demands time and specialized skills [Fonseca, 2017]. Currently, the emergence of AI and machine learning-based image generation models offers the ability to create images from text in a short timeframe. Applying this technology in the field of architecture has the potential to expedite the design process and foster creative design solutions.

Building upon this, our research focuses on the feasibility of generating architectural visualizations using AI-based image generation method. In Chapter 3, we tested the performance of the image generation AI model based on architects' styles, and in Chapter 4, we conducted additional training based on the test results. Finally, Chapter 5 demonstrates the practical applications of the Image generation AI including trained model.

2. BACKGROUNDS

2.1 Architectural visualization generation methods

Architectural visualization has evolved significantly over the years, transitioning from traditional manual techniques to embrace the power of digital technology. Historically, architects relied on hand-drawn sketches, physical models, and paintings to communicate their design ideas [Kehir AI-Kodmany, 2001; Atilola et al., 2016]. These methods, though expressive, had limitations in terms of scale, precision, and the time-intensive nature of creation. As architecture moved into the digital era, Computer-Aided Design (CAD) emerged as a game-changer, enabling architects to produce accurate and editable digital representations of their designs [Chiu, 1995]. It marked the beginning of a transformative shift in architectural visualization, offering architects the ability to iterate rapidly, explore design alternatives, and create highly detailed virtual models.

As technology continued to advance, architectural visualization expanded its horizons to encompass photorealistic rendering, three-dimensional (3D) modeling, and immersive experiences [Koutamanis, 2000]. Sophisticated

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Youngjin Yoo, Hyun Jeong, Youngchae Kim, Seung Hyun Cha and Jin-Kook Lee, *Generative Design Intuition from the Fine-Tuned Models of Named Architects' Style,* pp. 917-925, DOI 10.36253/979-12-215-0289-3.91

rendering software, bolstered by powerful Graphics Processing Units (GPUs), enabled architects to create high-fidelity visualizations that realistically conveyed materiality, lighting, and texture. 3D modeling provided a comprehensive understanding of spatial relationships [Eastman, 1999], offering architects the ability to manipulate and analyze their designs in a virtual environment [David et al., 2022]. This progress in technology not only increased the efficiency of the design process, ultimately leading to better-informed design decisions and more visually impactful presentations.

2.2 Image generation artificial intelligence (AI)

In 2014, Generative Adversarial Networks (GANs) emerged as a dominant paradigm for image generation research. GANs showcase their prowess by creating realistic images through competitive training involving a generator and a discriminator [Goodfellow et al., 2014]. As the stability of GAN training methods improved, the focus shifted towards generating images with specific attributes and refining the generated outputs [Karras et al., 2020]. These techniques have been applied to comprehend the information conveyed in architectural drawings, making it interpretable for computers. [Kim et al., 2019; Kim et al., 2020]

Since 2020, within the diverse landscape of image generation AI platforms, several notable options have emerged. Midjourney [Oppenlaender, 2022] specializes in style blending, empowering users to influence the fusion of multiple styles within the generated images. DALL-E 2 [Ramesh et al., 2022] creates images from textual descriptions, showcasing the potential to transform words into visuals, despite occasional inconsistencies. In contrast, Stable Diffusion [Rombach et al., 2022] leverages a diffusion model, ensuring stability during training and providing the capacity to manage image quality and intricacy. It shows immense promise in bridging the gap between abstract architectural concepts and their visual manifestation.

Among these, Stable Diffusion holds particular promise for architectural visualization research, given its ability to handle complex image transformations, align well with architectural subtleties, provide stability during training, and offer control over output quality and detail [Oppenlaender et al., 2023; Borji, 2023]. This positions Stable Diffusion as a potent tool to bridge the gap between architectural concepts and visual representation, redefining how architects approach their work and streamlining the creative process.

2.3 Potential for architectural visualization automation

There has been extensive research in image generation AI; however, its full potential for architectural visualization has yet to be realized. This research introduces a novel approach to architectural visualization using image generation AI models, emphasizing their transformative impact on this field. By harnessing advanced machine learning techniques, the study explores innovative methods to enhance architectural visualization, including text-to-image generation, which creates images from textual descriptions [Saharia et al., 2022]. This capability enables the generation of highly realistic images, making it a versatile tool with significant potential for various architectural visualization applications.

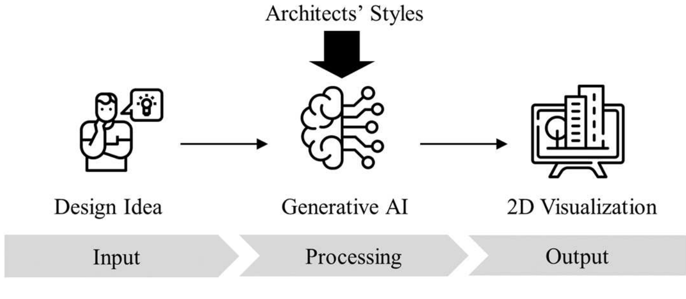


Fig. 1: Research approach: Image generation AI based architectural visualization

3. INTENSIVE TEST OF IMAGE GENERATION AI WITH ARCHITECTS' STYLE

3.1 Image generation test for architects' styles

Image generation artificial intelligence (AI), particularly Stable Diffusion (SD), involves two primary methods. The first method generates images from a text prompt, known as text-to-image generation. The second method, image-to-image generation, requires a seed image in addition to text prompts to generate images based on both inputs. In this paper, we focus primarily on text-to-image generation which generates images (Img_G) using the

"$generate()$" function, requiring a AI model (M), parameters ($Param$), and prompts (P_t).

$$generate(M,\ Param,\ P_t) = Img_G \tag{1}$$

$$Param = \{resolution, sampling\ method, sampling\ steps, CFG\ scale\} \tag{2}$$

$$P_t = \{SDP, RQP\} \tag{3}$$

The $Param$ consist of four components: $resolution$, determining the image size in pixel; $sampling\ method$, selecting method for image extracting from latent space; $sampling\ steps$, defining the number of extraction stages; and Classifier-free guidance scale ($CFG\ scale$), specifying the influence level of the prompt. The P_t consist of Scene Description Prompts (SDP), describing the target scene, visual composition, and graphic style, and Resolution Quality Prompts (RQP), adjusting the image's quality. Additionally, to prevent errors, each prompt composition includes negative prompts to specify what should be excluded. Table 1 provides example prompts corresponding to its composition.

Table 1: Prompt composition and its examples.

Composition of P_t	Positive Prompt example	Negative prompt example
Scene description prompts (SDP)	A residential house, professional photograph, photorealistic rendering, deep depth of field, high-key lighting, two-point perspective, etc.	Commercial buildings, painting, sketch, bird's-eye view, isometric, portrait, cropped view, etc.
Resolution quality prompt (RQP)	realistic shadows, enhance-detail, v ray rendering, full HD, masterpiece, highly detailed, high quality, 8k, etc.	low quality, too much noise, normal quality, watermark, blurry textured, blurry, noise, faint, text, etc.

In this section, we tested the performance of the text-to-image method defined earlier for generating architectural visualization. We randomly selected 20 architects who have received architectural awards or have had significant international influence, and generated images reflecting their styles. While additional descriptive keywords could enhance image quality by further delineating each architect's features, we excluded them for a clearer assessment of the default model's architect's style recognition capabilities. Instead, we used only the prompt "Architect's name-inspired residential house" and prompts associated with photorealistic rendering, commonly used in architectural visualization. We generated approximately 100 to 150 images for each architect in a local PC environment, with a resolution of 1024 by 512 pixels. The generated results are summarized in Figure 2.

None style
"A residential house"

Low Recognition
"Renzo Piano-inspired residential house"

High Recognition
"Frank Gehry-inspired residential house"

Fig. 2: Result of text-to-image generation test

3.2 Findings and ongoing inquiry in image generation AI

The generated results were assessed based on three criteria for their alignment with P_t. This assessment encompassed: (1) Style fidelity, which measures the accuracy of depicting the design characteristics of architects, (2) Domain fidelity, which verifies the representation of unique features for residential houses, and (3) Image quality, assessing the extent to which the photorealistic style rendering prompt was reflected in terms of graphic style, composition, and resolution.

The image generation test results indicated that the current SD model achieved a high level of domain fidelity and overall image quality. However, it exhibited low recognition for specific architects' styles, regardless of their prominence, resulting in lower quality and less detailed images of generic Western-style residential houses without any corresponding style features. As a result, the need for further additional training of the existing image generation model to address these limitations in recognizing certain architects' styles became evident. Motivated by this necessity, we conducted additional training, specifically targeting Architects' design styles, as depicted in Figure 3.

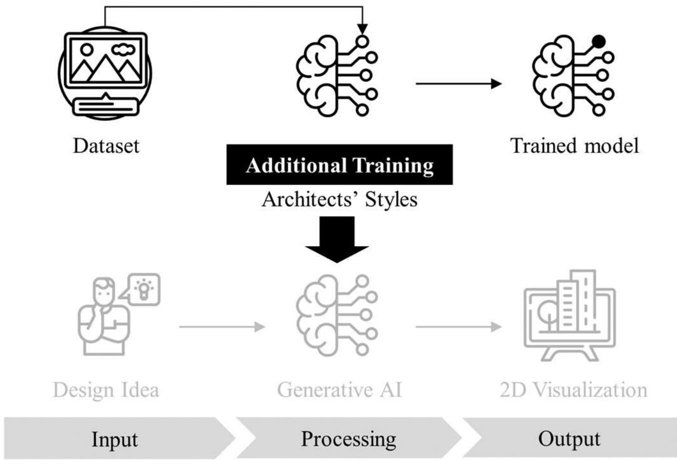


Fig. 3: Research Overview: Additional training for architects' styles

4. ADDITIONAL TRAINING FOR ARCHITECTS' STYLES

4.1 Additional training and data preparation

If the majority of generated images (Img_G) do not match the target image group (Img_t), it is required to replace the current model (M) with an alternative model (M'). This replacement can involve either substituting the model or enhancing it through further training. In this chapter, Low Rank-Adaptation (LoRA) approach [Hu et al., 2021] is employed for additional training, aiming to improve the recognition of specific architects' styles and to generate images that appropriately belong to the Img_t. The target model (M_t) is developed using the "$train()$" operator, based on the base model (M), hyperparameters ($Hyperparam$) and a target training dataset (D_t).

$$Most\ of\ Img_G \notin Img_t \Rightarrow M' \rightarrow M \quad (4)$$

$$train(M,\ Hyperparam,\ D_t) = M_t \in M' \quad (5)$$

Hyperparameters play a significant role in both the model's learning process and the subsequent performance of the M_t. We specifically focused on three crucial hyperparameters: the training batch size (BS_t), the number of epochs ($epoch$), and the learning rate (α). At the same time, the effectiveness of additional training relies on a high-quality dataset (D_t) containing image data (Img_D) along with corresponding annotation text files (Txt_D).

$$Hyperparam = \{BS_t, epoch, \alpha\} \quad (6)$$

$$D_t = \{Img_{D1}, Txt_{D1}, \ldots, Img_{Dn}, Txt_{Dn}\} \quad (7)$$

The additional training process, as depicted in Figure 4, involves two essential steps: (1) dataset preparation [Abdallah et al, 2017] and (2) training [Hu et al., 2021]. During the dataset preparation phase, meticulous training data collection is required to ensure alignment with P_t. Preprocessing phase aids in removing unnecessary content that might disrupt the training process. It is crucial to ensure content quality of training data, and the correspondence between Img_D and Txt_D. Following this, the Txt_D is paired with the respective Img_D, and the prepared D_t is then trained using the specified $Hyperparam$.

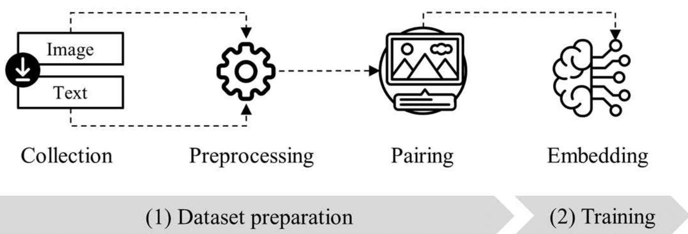


Fig. 4: Additional training process

4.2 Additional training of existing model with architects' styles

In this chapter, we provided additional training to architects who received low or no recognition in the image generation test discussed in Chapter 3. We conducted a few-shot learning using the previously defined training approach. By incorporating the trained LoRA model (M_t) into the image generation function, the possibility that generated images ($Img_{G'}$) closely resemble the designated Img_t is notably improved compared to the previous results. When utilizing M_t, in addition to M, it is crucial to input the application weight (W), a value ranging from 0 to 1, where 0 represents 0% and 1 represents 100%.

$$generate(M' \vee M(M_t, W),\ Param,\ P_t) = Img_{G'} \tag{8}$$

We compared the performance of the default model (M) with the trained model (M_t) by generating images with both. The image generation process followed equation (8), and the parameters (Param) and prompts (P_t) used for image generation remained consistent with those used in Chapter 3. As shown in Figure 5, the existing model had very low recognition rates for certain architects, so even with a full weight, the specific features of those styles were not represented. However, when using the trained model, these features are correctly displayed, and their application is proportional to the weight. The additional training allows us a wider range of style options that the original model could not achieve.

Default model 1.0 "SANAA-inspired residential house" | Trained model 0.5 "SANAA-inspired residential house" | Trained model 1.0 "SANAA-inspired residential house"

Fig. 5: Additional training results: SANAA style

5. DEMONSTRATIONS

Our investigation revealed that AI-driven image generation rapidly produces high-quality architectural visualizations from text prompts, empowering architects to easily create reference images and visualizations from the start of the design process. This chapter demonstrates the practicality of Image Generation AI, particularly Stable Diffusion, across various architectural styles. The three applications include: (1) building additional training models for desired architects' styles, (2) generating architectural visualizations applying an individual architect's style, and (3) generating style alternatives by combining more than two styles.

5.1 Implementation of different styles through additional training

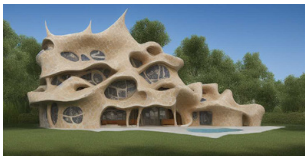
In this scenario, we employ image generation AI to incorporate diverse architects' styles, providing users with desired visualization outcomes through additional training. In this chapter, we conducted additional training following the process outlined in Figure 4, targeting five architects with very low recognition rates, aiming to enhance the model's level of detail. To ensure high-quality training images, we sourced project photographs from reputable sources, such as Architects' official websites, focusing on full facades in 1-point or 2-point perspective. Preprocessing involved image resizing and the removal of excessive information. Text data was constructed for each image, extracting from interviews with architects, expert analyses, and prior research about their styles. Each target style was trained with 15-25 datasets in average, with hyperparameters {1, 100, 0.0001}, and it took 8-15 minutes per each training.

The resulting model files, incorporated into the existing model, produce architectural exterior images closely mirroring architects' design styles, even when data is limited. In this chapter, we generated five M_t files, each representing the styles of different architects, capable of producing high-quality images comparable to those shown in table 2 of chapter 5.2.

5.2 Visualization of design alternatives from text prompts

This scenario describes how we acquired a diverse set of creative reference images representing different architects' design styles. In this chapter, we applied the M_t developed in the previous chapter to M in order to generate architectural visualizations based on the styles of 20 selected architects, using the same prompts as those used in the image generation test in chapter 3.1. We generated approximately 100 to 150 images for each architect based on equations (1) and (8), with the parameters {(1024, 512), Euler α, 20, 7}. These images, as demonstrated in Table 2, accurately reflect not only their respective styles but also maintain the essential characteristics of residential buildings, even for architects with little prior experience in residential projects. These generated outputs provide a rich source of diverse and concrete ideas and inspirations right from the initial stages of architectural design, streamlining communication and facilitating the design process.

Table 2: Resume of generated visualizations from text prompts

Input prompt	Output		Descriptive Keywords
I.M. Pei-inspired residential house, Photorealistic rendering prompt set			Modernist, minimalist, geometric, cultural fusion, monumental, symmetrical, glass and steel, iconic, etc.
Renzo Piano-inspired residential house, Photorealistic rendering prompt set			Lightness, Transparency, industrial materials, fluidity, civic and public focus, open spaces, etc.
Le Corbusier-inspired residential house, Photorealistic rendering prompt set			Modernism, functionalism, free façade, open floor plans, concrete, horizontal windows, etc.
SANAA-inspired residential house, Photorealistic rendering prompt set			Minimalist, subtle elegance, organic forms, conceptual simplicity, fine steel structure, white color, transparency, etc.
Shigeru Ban-inspired residential house, Photorealistic rendering prompt set			Sustainability, paper architecture, wooden modular structure, organic design, grid, organic forms, patterns, etc.
Frank Lloyd Wright-inspired residential house, Photorealistic rendering prompt set			Organic architecture, prairie style, horizontal lines, flat roofs, clerestory windows, cantilevered overhangs, etc.
Antoni Gaudi-inspired residential house, Photorealistic rendering prompt set			Curved lines, mosaic and tilework, nature-inspired design, whimsical details, unconventional forms, use of color, etc.
Mies van der Roe-inspired residential house, Photorealistic rendering prompt set			Minimalism, steel and glass, open floor plans, linear and geometric design, Bauhaus influence, international style, etc.

Ex) Photorealistic rendering prompt set = *Positive prompts:* professional photograph, photorealistic rendering, realistic, enhance-detail, v ray rendering, full HD, masterpiece, highly detailed, high quality, 8k, two-point perspective, exterior view, full shot, deep depth of field, f/22, 35mm, high-key lighting, natural lighting, realistic shadows; *Negative prompts:* low quality, bad proportion, awkward shadows, unrealistic lighting, pixelated textures, too much noise, unrealistic reflections, normal quality, watermark, bad perspective, confusing details, blurry textured, blurry, noise, cloudy, faint, text.

5.3 Combination between architects' styles

This scenario illustrates the creation of diverse image references by blending multiple architectural styles, resulting in novel and previously unseen styles. Users can expand their architectural image references using image generation AI by combining the styles of two or more architects. The P_t and $Param$ for these operations are the same as those in other image generation cases, except for the SDP (Scene Description Prompts), which is observable in Table 3. This setup allows for a comparison between the results of applying a single style and the application of multiple styles, facilitating an assessment of the progress of the operations.

Table 3: Example of combination of architects' styles using text-to-image method

Classification		Mono-style: SANAA style	Mono-style: Luis Barragan style	Multi-style: SANAA and Barragan
Input	Model	Trained model		
	Parameters	Resolution: 1024 × 512 / Sampling method: Euler a / Sampling steps: 20 / CFG scale: 7		
	Prompts	SANAA-inspired residential house, Photorealistic rendering prompt set	Luis Barragan-inspired residential house, Photorealistic rendering prompt set	SANAA and Luis Barragan-inspired residential house, Photorealistic rendering prompt set
Output				
Descriptive Keywords		Minimalist, elegance, sensitivity, fine steel structure, white color, simplicity, transparency, etc.	Minimalism, color, geometry, concrete, simplicity, play of light and shadow, etc.	Fine structures, colorful, rectilinear, concrete, simplicity, geometry, etc.

As shown in Table 3, the combination of two different styles is evident and noticeable. When the curvilinear style of SANAA is combined with the rectilinear style of Luis Barragan, the curvilinear aspect of SANAA becomes less pronounced. Additionally, the resulting style incorporates the color palette and materiality of Luis Barragan, along with SANAA's distinctive design feature of thin structures. These findings demonstrate that image generation AI can create new alternative styles based on existing ones, potentially generating a variety of additional alternatives.

6. CONCLUSION

This research marks the initial steps in exploring the potential of architectural visualization through image generation AI, with a specific focus on the Stable Diffusion model. The study underscores the significant impact of image generation AI, particularly in the field of architecture and its application in early-stage architectural visualization. Leveraging deep learning and image generation techniques, we trained the model to capture the distinctive styles of renowned architects, using this knowledge to visualize typical residential houses. Our testing revealed that while the default SD model generally produces high-quality architectural visualizations with domain fidelity, it does face limitations in recognizing the unique styles of architects. However, we demonstrated that these limitations can be improved through additional training, highlighting the powerful potential of image generation AI.

This approach plays a pivotal role in bridging the gap between abstract design concepts and tangible visual representations, empowering architects to effectively convey their creative ideas. Integrating AI technology into architectural visualization broadens creative possibilities, enabling architects to explore a diverse range of design alternatives. Looking ahead, further research is essential to develop comprehensive and refined methods for additional training, expanding beyond architects' styles to other targets. Additionally, the focus should be on enhancing the accessibility and utility of this technology by exploring other generation methods, such as image-to-image, and the development of user-friendly tools.

7. AKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MIST) (No. NRF-2022R1A2C1093310)

REFERENCES

Abdallah, Z., Du, L., & Webb, G. (2017). Data Preparation. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer. https://doi.org/10.1007/978-1-4899-7687-1_62.

Akin, O. (1978). How do architects design. Artificial Intelligence and Pattern Recognition in Computer Aided Design, 65-104.

Atilola, O., Tomko, M., & Linsey, J. S. (2016). The effects of representation on idea generation and design fixation: A study comparing sketches and function trees. Design studies, 42, 110-136. https://doi.org/10.1016/j.destud.2015.10.005.

Borji, A. (2023). Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. https://doi.org/10.48550/arXiv.2210.00586.

Chiu, M. L. (1995). Collaborative design in CAAD studios: shared ideas, resources, and representations. *In Proceedings of International Conference on CAAD Future* (Vol. 95, pp. 749-759).

David, A., Joy, E., Kumar, S., & Bezaleel, S.J. (2022). Integrating Virtual Reality with 3D Modeling for Interactive Architectural Visualization and Photorealistic Simulation: A Direction for Future Smart Construction Design Using a Game Engine. Second International Conference on Image Processing and Capsule Networks, 300. https://doi.org/10.1007/978-3-030-84760-9_17

Eastman, C.M. (1999). Building Product Models: Computer Environments, Supporting Design and Construction. (1st ed.). CRC Press: Florida, (Chapter 1).

Fonseca, D., Redondo, E., Valls, F., & Villagrasa, S. (2017). Technological adaptation of the student to the educational density of the course. A case study: 3D architectural visualization, Computers in Human Behavior, 72, 599-611. https://doi.org/10.1016/j.chb.2016.05.048.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Greenberg, D. P. (1974). Computer graphics in architecture. Scientific American, 230(5), 98-107. https://www.jstor.org/stable/24950079.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. https://ar5iv.labs.arxiv.org/html/2106.09685.

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8110-8119). DOI: 10.1109/CVPR42600.2020.00813.

Kheir Al-Kodmany. (2001). Visualization Tools and Methods for Participatory Planning and Design. Journal of Urban Technology, 8(2), 1-37. DOI: 10.1080/106307301316904772

Kim, J., & Lee, J. K. (2020). Stochastic detection of interior design styles using a deep-learning model for reference images. Applied Sciences, 10(20), 7299. https://doi.org/10.3390/app10207299.

Kim, J., Song, J., & Lee, J. K. (2019, January). Approach to auto-recognition of design elements for the intelligent management of interior pictures. *In Proceedings of the 24th International Conference on Computer-Aided Architectural Design Research in Asia: Intelligent and Informed, CAADRIA* (pp. 785-794). https://doi.org/10.52842/conf.caadria.2019.2.785.

Koutamanis, A. (2000) Digital architectural visualization. Automation in Construction, 9(4), 347-360. https://doi.org/10.1016/S0926-5805(99)00018-7.

Oppenlaender, J. (2022, November). The Creativity of Text-to-Image Generation. *In Proceedings of the 25th International Academic Mindtrek Conference* (pp. 192-202). https://doi.org/10.1145/3569219.3569352.

Oppenlaender, J., Visuri, A., Paananen, V., Linder, R., & Silvennoinen, J. (2023). Text-to-Image Generation: Perceptions and Realities. https://doi.org/10.48550/arXiv.2303.13530.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. https://doi.org/10.48550/arXiv.2204.06125.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684-10695). https://doi.org/10.48550/arXiv.2112.10752

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., ... & Norouzi, M. (2022). Photorealistic text-to-image diffusion models with deep language understanding. Advances in Neural Information Processing Systems, 35, 36479-36494. https://doi.org/10.48550/arXiv.2205.11487.