PLANNING ALTERNATIVE BUILDING FAÇADE DESIGNS USING IMAGE GENERATIVE AI AND LOCAL IDENTITY

Hayoung Jo, Sumin Chae, Su Hyung Choi & Jin-Kook Lee
Department of Interior Architecture and Built Environment, Yonsei University, Seoul, Korea.

ABSTRACT: *This paper describes an approach utilizing Generative AI to support diverse design alternatives for building facades based on the local identity. Extensive research is currently being conducted for exploring the applications of LLM-based generative AI models to diverse kinds of visualizations. By applying generative AI to facade design, the study aims to develop additional training models that generate alternative design options reflecting local identity, facilitating the acquisition of remodel design images from multiple texts and images. Building facades in cities and regions are essential for people's aesthetic perception and understanding of the local environment, enabling the recognition and differentiation of specific areas from others. Therefore, implementation method of the additional training model based on generative AI in this study, reflecting this, can be summarized as follows: 1) collection and pre-processing of image data using Street View, 2) pairing text data with image data, 3) conducting additional training and testing with various inputs, 4) proposing relevant application methods. This approach can be expected to enable efficient communication of design at an early stage of the architectural design process beyond traditional 3D modeling and rendering tools.*

KEYWORDS: *Building facade, Generative AI, Local identity, Design alternative, Additional Training Model*

1. INTRODUCTION

Recently, platforms such as 'Midjourney,' 'Dreamstudio AI,' and 'Stable Diffusion' have been developed and used alongside Large Language Model (LLM) based platforms like 'ChatGPT' (OpenAI, 2022) to generate images using Diffusion models. These platforms are provided in accessible forms for the public, and their interfaces and functionalities are consistently updated. These platforms are based on generative artificial intelligence, allowing users to easily create desired images creatively by providing prompts and adjusting settings. This generative AI-based image creation approach is not only applied in design and art fields but also in various other domains. It is also being employed in architecture, generating images of diverse buildings and spatial designs in various styles, contributing to applied research.

In this study, the aim is to apply the image generation capability of generative artificial intelligence to obtain facade images of buildings. Furthermore, this involves creating building images with regional design identities, aiming to establish an approach for more efficient utilization during the initial building planning and design stages (Relph, 1976). This approach focuses on commercial buildings, allowing for the swift acquisition of creatively designed facade images in the early architectural phases by adjusting the degree of regional identity incorporation.

The research follows the following methodology: Initially, to evaluate the effectiveness of the image generation model, a repetitive process of image generation was conducted, resulting in the creation of a substantial number of images for testing. Based on these results, it was evident that additional training of the basic generative AI model was necessary. Subsequent steps for this additional training were carried out as follows: 1) Constructing a training dataset, 2) Conducting additional training and generating model files, 3) Confirming and utilizing result images incorporating the additional training model files. This was executed in the form of additional training utilizing the Diffusion-based model. The additional training was built upon LoRA (LoRA: Low-Rank Adaptation of Large Language Models), and by adjusting hyperparameters, it was ensured that high-accuracy images were generated. Following this, the generated additional training model files were applied to generate and confirm result images, suggesting an approach to visualize these images in the early architectural stages.

2. BACKGROUND

2.1 Image Generative AI

Since 2020, diffusion process-based techniques have gained prominence in the arena of deep learning-driven image synthesis. These approaches iteratively update pixel values to progressively generate images (Ho, Jain, & Abbeel, 2020). Concurrently, scholars have immersed themselves in artificial intelligence models that facilitate the

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Hayoung Jo, Sumin Chae, Su Hyung Choi, Jin-Kook Lee, *Planning Alternative Building Façade Designs Using Image Generative AI and Local Identity,* pp. 926-932, DOI 10.36253/979-12-215-0289-3.92

transformation of textual data into visual representations, marking significant progress in the domain of image generation (Ramesh, Dhariwal, Nichol, Cuy, & Chen, 2022; Saharia, Chan, Sawena, Li, Whang, Denton, … & Norouzi, 2022; Rombach, Blattmann, Lorenz, Esser, & Ommer, 2022).

While considerable scholarly inquiry has been devoted to deep learning-assisted image synthesis, its potential in the realm of architectural design visualization remains largely untapped (Kim, & Lee, 2020). This investigation introduces an innovative proposition for architectural design visualization, harnessing the capabilities of AI-driven image synthesis models and recognizing their transformative impact in the landscape of image generation. Through the application of these advanced machine learning techniques, this section aims to explore novel pathways to enhance architectural design visualization via AI-powered image training models.

With the advancement of the LLM model and the image synthesis technology, the feasibility of producing architectural visualization images based on provided textual input has become achievable. Termed as text-to-image synthesis, this process possesses the ability to generate highly realistic images, making it a versatile instrument for generating a diverse range of architectural visualization content. As AI technology continues its evolution, the role of text-to-image synthesis is expected to play a crucial role in the architectural domain. Consequently, the integration of AI-driven image synthesis enhances the potential for imaginative exploration beyond traditional methodologies.

2.2 New opportunities for Architectural Visualization

Architectural visualization, such as photorealistic images, plays a crucial role in enhancing communication within the field of architecture (Lee, Lee, Kim, & Kim, 2023). Firstly, photorealistic renderings transcend mere geometric massing, enabling architects to vividly convey their design intentions to clients. These images serve as intermediaries between architectural drawings and experiential aspects of architectural spaces by presenting architectural concepts in a reality-like manner (Kim, & Lee, 2022). Such visualizations facilitate shared understanding among stakeholders. Secondly, visualization empowers not only architectural professionals but also stakeholders, clients, and the public to grasp architectural visions that transcend architectural terminology and technical complexity. Visualized images like photorealistic renders enable individuals to comprehend the interaction between planned architectural attributes, ambiance, and the surrounding environment, enabling informed decision-making based on information. Transitioning from geometric massing to photorealistic render images allows for a more universal and comprehensive communication of intricate architectural concepts, thus promoting smoother communication.

In summary, integrating visualization images like photorealistic renderings into the architectural design process enables efficient communication in the early stages of architecture, induces information-based decision-making, and enhances creative design. While traditional architectural visualization relied on complex technical processes and necessitated GPUs and specialized hardware, leveraging generative AI, as discussed earlier, allows for obtaining numerous detailed visualization images effectively without the need for separate GPU renderers.

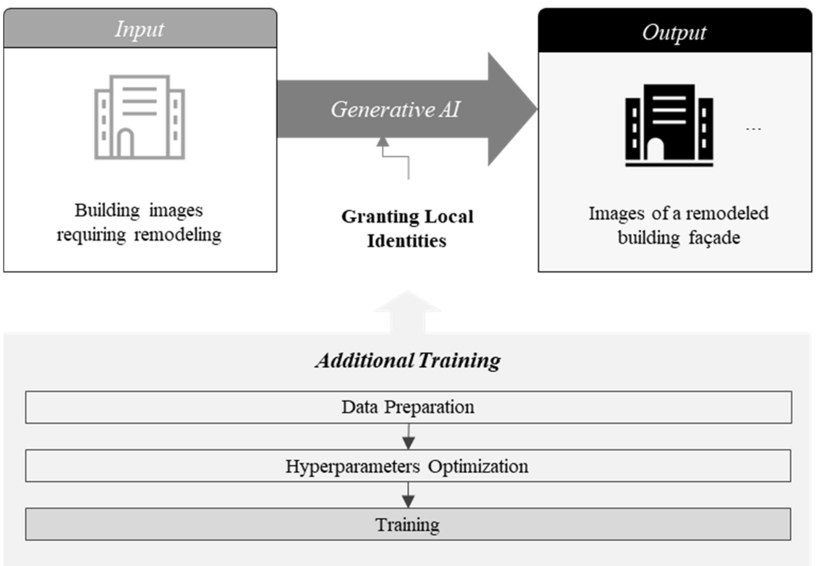


Fig. 1: Overview of the approach proposed in this study.

The following section examines the application of such generative artificial intelligence to architecture, exploring the potential of generating architectural images. This investigation, as outlined in the introduction, focuses on the design aspect of building facades within the realm of architectural elements (Kier, 1984). Specifically, this inquiry aims to determine the feasibility of effectively generating architectural visualization images by emphasizing regional identity as a pivotal design consideration within building facade design.

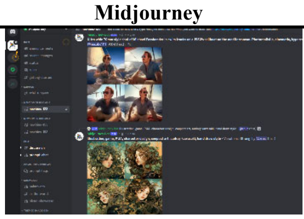
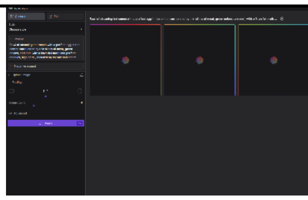
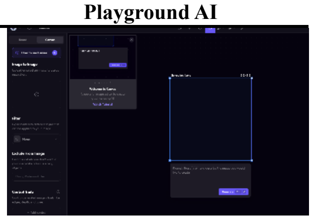
3. TEST ON BASIC IMAGE GENERATION MODELS

3.1 Test Generative AI Platforms

Various platforms are being developed using generative artificial intelligence to make it easily accessible for the public. These platforms utilize different interfaces and base models, resulting in a range of image generation platforms that cater to various user requirements such as freedom of generation, design style of images, sizes, and image quality. In this paper, we utilized the commonly used platforms 'Midjourney,' 'Dreamstudio AI,' and 'Playground AI' to understand their respective interfaces, directly engage with them, and explore their features and specific functionalities.

Among these three platforms, the latter two platforms, excluding 'Midjourney,' offer partial free usage for image generation, with subscriptions or purchases required for more extensive usage. Each interface provides common features including the option to select various image styles like 'Enhance,' 'Anime,' 'Photographic,' 'Comic book,' as well as the ability to create Positive and Negative prompts. All platforms also offer the functionality to adjust specific settings to generate images. Additionally, they provide an "Image-to-Image" feature wherein users can input desired images to generate text based on the images, resulting in the creation of different images. By utilizing these functionalities, one can quickly generate images tailored to specific requirements. For instance, when aiming to acquire building facade images as shown in Table 1, it becomes possible to generate images that incorporate more creative ideas. The following section will proceed with an examination of building facade image generation through detailed testing, utilizing prompts that encompass greater specificity and domain knowledge.

Table.1: Investigation of the interfaces of prominent platforms for image generation models and examples of generated images (The generated images from Midjourney and Dreamstudio AI are provided by openart (https://openart.ai/), while the examples generated by Playground AI are based on similar prompt-based approaches).

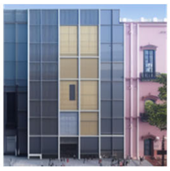
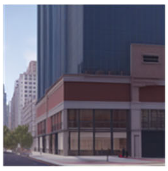
		Midjourney	Dreamstudio AI	Playground AI
Web Interface				
INPUT	Key Prompt		Building Façade Image	
OUTPUT	Generated Images			

3.2 Testing of Façade Image Generation Reflecting Local Design Identity

In this section, we aim to investigate whether it is possible to generate facade design images that reflect regional identity using generative artificial intelligence. To achieve this, we conducted image generation tests based on text prompts using the existing basic model grounded in Diffusion. The tests were divided into three main categories: facade images of buildings without region-specific text input, facade images of buildings reflecting Korean style, and facade design images of commercial buildings in Manhattan. The goal was to compare the generated images for these three categories. For each category, we utilized key prompts such as "Building Facade," "Building Façade reflects Korean style," and "Building Façade reflects Manhattan style." Additionally, we employed prompts to enhance image quality to generate results like those in Table. 2.

By utilizing the existing generative artificial intelligence-based model, it was observed that when region-related text prompts were input, corresponding images could generally be generated. However, this primarily resulted in localized images, and it was found that the generated facade design images did not exhibit diverse variations reflecting the unique images associated with each region. For instance, in the case of Korean facade images, predominantly images of buildings featuring traditional Eastern style hanok architecture were generated. Therefore, in the subsequent section, we proceed to construct a model through fine-tuning of the existing generative artificial intelligence model, aiming to determine if image generation with a focus on regional facade design identity can be achieved.

Table. 2: Example of generating building facade images with regional names using the basic generative AI model

No.	Key Prompts	Generated Images				
1	Building Facade					
2	Building Façade reflects Korean style					
3	Building Façade reflects Manhattan style					

4. CONSTRUCTION AND UTILIZATION APPROACHES OF THE ADDITIONAL TRAINING MODEL

4.1 Additional Training and Testing of Local Facade Design Identity Model

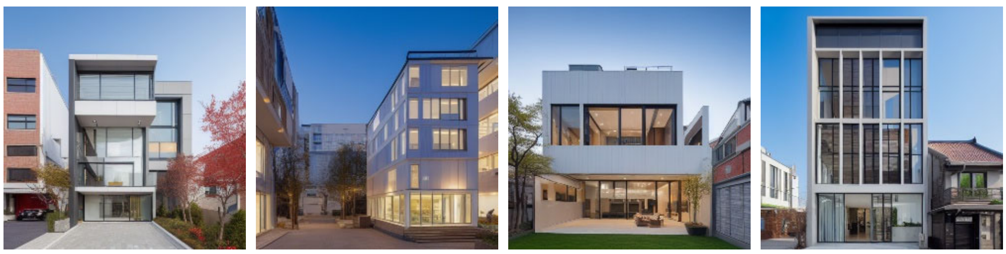
In this section, we aim to investigate the generation of facade design images that reflect regional identity by conducting additional training of a generative artificial intelligence model within the scope of the target region. Model construction utilized the Diffusion-based model implemented on the foundation of LLM (Large Language Model) for additional training. This additional training process can be summarized into three main stages: 1) Data Preparation, 2) Model Training, and 3) Image Testing and implementation. Data preparation involved pairing image and text data. For efficiency in image data collection, street-view functionality from portal sites API was employed, as described earlier. However, the distorted nature of 360-degree panorama images from street-view led to generating indistinct façade images, lowering image quality and accuracy. To address this, image preprocessing was conducted to correct distortions, resize images to a consistent size, and then pair them with text data to compile the dataset.

For model training, the LoRA (Low-Rank Adaptation of Large Language Models) approach was adopted to facilitate additional training of the Diffusion model (Hu, Shen, …& Chen, 2021). LoRA allows for rapid additional training of existing large-scale models within a short timeframe, without significant demands on GPU performance. Unlike other methods, LoRA generates relatively smaller additional training model files and offers the advantage of easily assessing style incorporation through adaptability changes in the model files. Thus, in this research, LoRA is employed to construct additional training models, optimizing hyperparameters to generate highly accurate images with minimal distortion. The optimization of hyperparameters, including adjustments to epochs, training batch size, and caption extensions, aims to enhance the accuracy and quality of the resulting images.

Fig. 2: Construction Process of the Additional Training Model

When conducting additional training using LoRA, model files with the extension ".safetensors" are generated. Inserting these generated model files into the model management folder of the Stable Diffusion Web-UI enables the models to function in the format of a text prompt, allowing the generation of desired images alongside the text data used for training. Furthermore, by adjusting the adaptability of the generated model files, a wide array of creative design images can be produced. Applying the additional training model file created using exterior images and text data of commercial buildings in the Seoul area, according to different weight values, results in images as shown in Table 3. When applying a weight of 0.1, images of buildings with views from different angles beyond the front facade are generated. As the weight approaches 1.0, images distinctly reflecting Seoul's facade design style are generated.

Table. 3: Test of Additional Training Models according to each weight

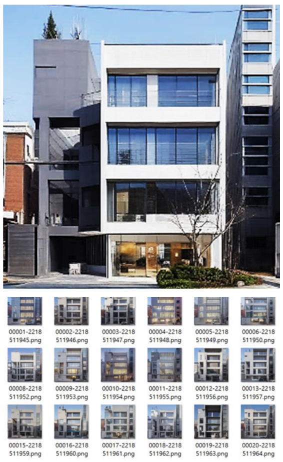
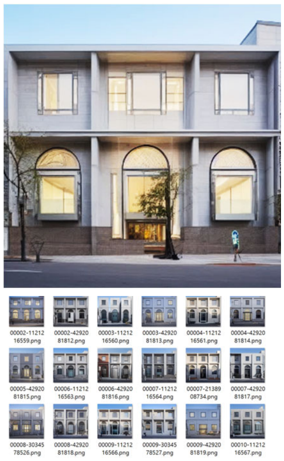
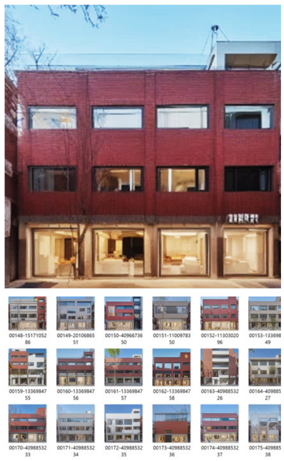
Weight	Generated Images
0.1	
0.5	

1.0	

4.2 Utilization Approaches of the Additional Training Model File

In this section, we demonstrate one example of an approach that can be applied in the early stages of architecture using the constructed additional-trained model files. We validated the images that could be generated by applying the model files using actual facade images of buildings in Seoul. When applying this method and providing detailed prompts, it was observed that images reflecting Seoul's facade design style could be generated.

Table. 4: Image generation from Each Input Image

		A	**B**	**C**
INPUT	Key Prompt	Building Façade reflects Seoul style		
	Detailed Prompt	Modern design style	An arched window	Red brick finish
	Utilized Model file	***Building Façade Design Style of Seoul.safetensors***		
	Images			
OUT-PUT	Generated Images	 …	 …	 …

5. CONCLUSION

In the initial design stages of existing buildings, facade design plans have traditionally relied on manual efforts by designers and architects, or methods involving 3D modeling tools and high-performance GPU renderers. These methods have necessitated repetitive tasks to facilitate communication with clients. This study discusses an approach that leverages the recent advancements in generative artificial intelligence, which is being actively applied in related fields, to generate facade design alternatives using image generation AI. Within the context of this research, we propose an approach that enables quick confirmation of building facade design plans reflecting regional facade identity in the early design stages and the generation of numerous alternatives.

According to the approach proposed in this study, it was confirmed that utilizing image generation AI can rapidly confirm building facade design plans, incorporating regional facade identity, and produce a multitude of alternatives. This approach was demonstrated through applying Seoul's facade design style using actual building images to showcase its effectiveness. Consequently, exceptional visualization images were generated.

Although there may be limitations in this study, particularly in constructing a fine-tuned model focused on Seoul, it holds significance in its potential to create and explore more diverse and domain-specific models using this methodology. This opens the door for further application-oriented research, leveraging more specific characteristics and domain knowledge to refine the approach.

REFERENCES

Lee, J.K., Lee, S., Kim, Y., & Kim, S. (2023). Augmented virtual reality and 360 spatial visualization for supporting user-engaged design, Journal of Computational Design and Engineering, Volume 10(3), Pages 1047–1059, https://doi.org/10.1093/jcde/qwad035.

Kim, J., & Lee, J.K. (2020) Stochastic Detection of Interior Design Styles Using a Deep-Learning Model for Reference Images. Appl. Sci. 10, 7299. https://doi.org/10.3390/app10207299.

Kim, Y., & Lee, J.K. (2022). PROCESSING OF 360 PANORAMIC IMAGES FOR ARCHITECTURAL INTERIOR IMAGE TRAINING ARCHIVE, *ConVR 2022 conference*.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. & Chen, W. (2021). Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840-6851.

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).

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

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.

Relph, E., Place and Placelessness, London (1976).

Rob Kier, Elements of Architecture, (1984).

OpenAI, "ChatGPT: Optimizing Language Models for Dialogue," 2023. [Online]. Available: https://openai.com/blog/chatgpt/.

https://www.midjourney.com/

https://beta.dreamstudio.ai/

https://playgroundai.com/

https://openart.ai/

ACKNOWLEDGEMENT

This work is supported in 2023 by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA163269).