Journal of Computational Design and Engineering, 2024, **11**, 85–105

DOI: 10.1093/jcde/qwae017
Advance access publication date: 16 February 2024
Research Article

Generative artificial intelligence and building design: early photorealistic render visualization of façades using local identity-trained models

Hayoung Jo[1], Jin-Kook Lee [1,*], Yong-Cheol Lee [2] and Seungyeon Choo[3]

[1]Department of Interior Architecture and Built Environment, Yonsei University, Seoul 03722, Republic of Korea
[2]Louisiana State University, Baton Rouge 70803, Louisiana
[3]Kyungpook National University, Daegu 41566, Republic of Korea
*Correspondence: leejinkook@yonsei.ac.kr

Abstract

This paper elucidates an approach that utilizes generative artificial intelligence (AI) to develop alternative architectural design options based on local identity. The advancement of AI technologies has increasingly piqued the interest of the architecture, engineering, construction, and facility management industry. Notably, the topic of "visualization" has gained prominence as a means for enhancing communication related to a project, especially in the early phases of design. This study aims to enhance the ease of obtaining design images during initial phases of design by drawing from multiple texts and images. It develops an additional training model to generate various design alternatives that resonate with the identity of the locale through the application of generative AI to the façade design of buildings. The identity of a locality in cities and regions is the capacity for the cities and regions to be identified and recognized as a specific area. Among the various visual elements of urban and regional landscapes, the front face of buildings may play a significant role in people's aesthetic perception and overall impression of the local environment. The research proposes an approach that transcends the conventional employment of three-dimensional modeling and rendering tools by readily deriving design alternatives that consider this local identity in commercial building remodeling. This approach allows for financial and temporal efficiency in the design communication phase of the initial architectural design process. The implementation and utilization of the proposed approach's supplementary training model in this study proceeds as follows: (i) image data are collected from the target area using open-source street-view resources and preprocessed for conversion to a trainable format; (ii) textual data are prepared for pairing with preprocessed image data; (iii) additional training and outcome testing are performed using varied text prompts and images; and (iv) the ability to generate building façade images that reflect the identity of the collected locale by using the additional trained model is determined, as evidenced by the findings of the proposed application method study. This enables the generation of design alternatives that integrate regional styles and diverse design requirements for buildings. The training model implemented in this study can be leveraged through weight adjustments and prompt engineering to generate a greater number of design reference images, among other diverse approaches.

Keywords: architectural design, design visualization, generative artificial intelligence, additional training model, diffusion model, image generation, façade design

1. Introduction

In the domain of architecture, engineering, construction, and facility management projects, design concepts and requirements have traditionally been conveyed through a combination of visual representations and written and verbal descriptions. These visual images, along with accompanying text explanations, are crucial to effectively communicating design specifications to stakeholders who are involved in the process (Koutamanis, 2000). With the evolution of computing and digital technologies, architectural visualization, facilitated by high-performance rendering techniques that create realistic visualizations akin to photography, continues to play a pivotal role in design communication. Recent attention has turned toward text-to-image model-based generative artificial intelligence (AI) techniques for visual representations, which are among the broad advances in computing technology (Oppenlaender *et al.*, 2023). This technology is expected to provide innovative methodologies for automating visualization in the architectural industry. The present study proposes a creative approach to design visualization by using this technology to generate images of buildings that embody domain knowledge in architecture.

In architectural visualization, the design of building façades may have significant symbolic representation. Derived from the French word meaning "front" or "face", the façade of a building plays a crucial role in unifying the visual and aesthetic aspects of an entire building, and it serves as an intermediary between external and internal conditions (De la Fuente Suarez, 2020; Prieto & Oldenhave, 2021). The building façade, in terms of its visual aspect, can become a backdrop to urban spaces and regional life and influence how people perceive and adapt to urban spaces and regional life. Therefore, the visual and aesthetic aspects of a building façade have great significance, even in the physical realm (Bala-

Received: September 26, 2023. **Revised:** February 9, 2024. **Accepted:** February 12, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of the Society for Computational Design and Engineering. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

subramanian *et al.*, 2022; Kaplan, 1992; Lang, 1987; Ulrich, 1983). The building façade serves various roles and meanings within a building, and it encompasses aspects of the structure and the environment. However, this study focuses specifically on the visual and aesthetic aspects of building façades in urban and regional spaces. Our goal is to concentrate on the design identity of building façades and to develop a methodology for generating design images that reflect this identity. The objective of this study is to propose an approach that promotes and integrates automation in architectural visualization within the context of urban and regional spaces. Although traditional methods such as architect-generated idea sketches or tools such as three-dimensional (3D) modeling and GPU (graphics processing units) rendering can effectively visualize existing structures, a more efficient design tool may be necessary for iterative modifications and work. Our approach and scope of this study is depicted in Fig. 1.

To develop a method for generating building façade images that incorporate local identity, the research follows these steps. First, initial iterations of image generation were performed using generative AI to assess the performance of the base image model; these iterations resulted in approximately 10 000 generated images. This underscored the necessity for a model that could integrate local identity and led to the second step, which was to carry out additional training. The training steps included (i) constructing a training dataset through data acquisition and preprocessing, (ii) optimizing the hyperparameters, and (iii) generating additional training models. Third, the built models were applied to various building façade visualization scenarios and demos to showcase the generation of visual images that reflected design creativity. Additionally, a visualization tool that incorporated these scenarios and approaches was developed to enhance the utility of the research and to make it accessible to both experts and non-experts, thereby supporting a wider audience.

2. Background

2.1. Architectural design visualization automation

Automated architectural design visualization involves using technology and software to automatically generate visual representations of architectural designs, which is a process that is crucial for previewing building appearances before actual construction (Abrishami *et al.*, 2021). In the past, the process of architectural design would begin with manual drawing. However, with the advancement of computer technology, various innovations and developments, such as computer-aided design (CAD) software and the introduction of building information modeling (BIM) concepts, have led to simplification and automation in the complex architectural design process. These advances have significantly enhanced the entire design process (Abrishami *et al.*, 2021; Weber *et al.*, 2022; Yu & Choi, 2023). This development is intertwined with innovations in software, computing, and visualization, which have fundamentally altered how architectural designs are conceived, communicated, and presented (Castelo-Branco *et al.*, 2022).

Architects in the pre-computer era relied on manual techniques such as sketches, models, and photomontages to depict building façades and visual perspectives (Lever, 1973). The advent of CAD software revolutionized architecture by allowing architects to digitally create 3D models and facilitating their design modifications (Gero, 1983). However, early CAD primarily focused on technical drawings and often neglected realism and aestheticism (Greenberg, 1974; Jonson, 2005; Oxman, 2006); Sutherland, 1964). Design constraints were mitigated through advances in 3D modeling and rendering that enabled architects to develop digital models and engage in geometric design (Lewis & Sequin, 1998). Nevertheless, rendering these models into lifelike images remained a complex task.

BIM emerged as a solution by integrating architectural design with data, which resulted in digital models that encompassed both visual and performance-related information (Kalay, 2004). BIM uses predefined rules and parameters to automate the design process, and its software applications employ parametric design for efficient workflows (H. Kim *et al.*, 2018; Park *et al.*, 2021). BIM integrates attribute information from 3D modeling and 2D drawings, which has fostered research in BIM-based automated design (Lee & Kim, 2014; Lee *et al.*, 2010, 2012; Sanguinetti *et al.*, 2012; Shin & Lee, 2019; Shin *et al.*, 2014).

Furthermore, technologies such as augmented reality (AR) and virtual reality (VR) have evolved to offer immersive design experiences (Whyte et al., 2000). They bridge the gap between the virtual and real worlds by introducing novel visualization techniques (Milgram & Kishino, 1994; Yigitbas *et al.*, 2023; Han *et al.*, 2021). In addition, AI technologies such as machine learning (ML) and deep learning (DL) have automated architectural visualization using methods such as generative adversarial networks (GANs, Goodfellow *et al.*, 2014) and variational autoencoders (VAEs, Kingma & Welling, 2022), which have streamlined the image-generation process and reduced complexity (Caetano *et al.*, 2020; Karan *et al.*, 2021; Sönmez, 2018).

In summary, architectural design visualization has evolved through a progression from manual to automated techniques, which has resulted in a diverse array of tools. Technological advances have led to the development of various methods that range from manual sketches to automated approaches using AR/VR and AI. As technology continues to become integrated, the efficiency of realistic visualization has improved significantly. The next section will delve into AI technologies and showcase their potential in architectural design.

2.2. ML, large language models, and image generation models

ML has undergone substantial advances and has assumed a pivotal role across diverse domains, including speech recognition, natural language processing, visual object recognition, and object detection (LeCun *et al.*, 2015; Song, Lee, *et al.*, 2020). Of particular note was the emergence of GANs in 2014, which marked a paradigm shift in image generation research. GANs have demonstrated an ability to produce lifelike images by using a competitive training process that involves a generator and a discriminator (Goodfellow *et al.*, 2014). With the maturation of GANs, scholars have embarked on the exploration of techniques for enhancing the stability of GAN training (Arjovsky *et al.*, 2017; Gulrajani *et al.*, 2017; Karras *et al.*, 2018; Miyato *et al.*, 2018; Radford, Metz, *et al.*, 2016). Notably, Google DeepMind revealed findings indicating that larger batch sizes lead to enhanced image quality, although larger batch sizes require robust computational resources (Brock *et al.*, 2019). As strategies for GAN training solidified, the focal point shifted toward the generation of images endowed with specific attributes and the manipulation of generated images (Goetschalckx et al., 2020; Karras *et al.*, 2019, 2020; Zheng & Yuan, 2021).

In the realm of natural language processing, the Transformer network architecture, which incorporates attention mechanisms, has yielded improvements in output quality while reducing training duration (Vaswani *et al.*, 2017). Transformers have extended

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

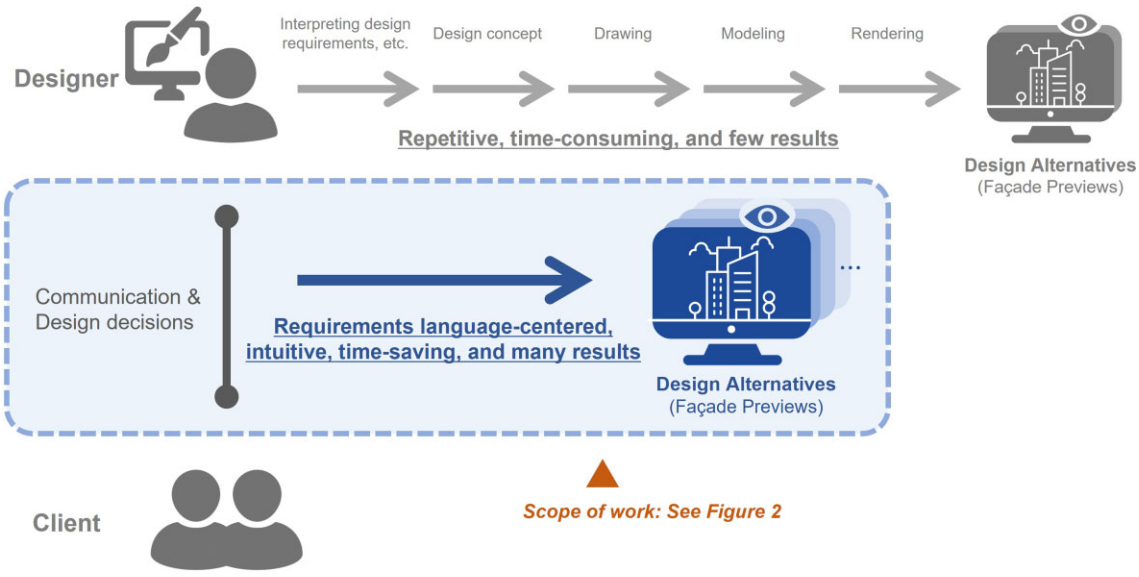


Figure 1: Overview of the paper (see Figs 2 and 3 for more specific process of this research).

their applicability to computer vision tasks that include image classification, image detection, and image retrieval (Z. Liu *et al.*, 2021; Radford, Kim, *et al.*, 2021). This transformative technology has engendered breakthroughs in natural language processing, as exemplified by models such as ChatGPT (OpenAI, 2022).

Since 2020, methods grounded in diffusion process have risen to prominence within the landscape of DL-based image generation. These methodologies generate images progressively by iteratively updating pixel values (Ho *et al.*, 2020; Song, Meng, *et al.*, 2022). Concurrently, researchers have delved into AI models that facilitate the conversion of text to images, which signify significant advances within the domain of image generation (Ramesh *et al.*, 2021, 2022; Rombach *et al.*, 2022; Saharia *et al.*, 2022).

Although substantial research has been dedicated to DL-based image generation, its potential within the realm of architectural design visualization remains largely unexplored. This study introduces a groundbreaking proposition for architectural design visualization by harnessing the capabilities of image-generation AI models and recognizing their transformative influence within the landscape of image generation. This section aims to explore the application of these advanced ML techniques and to find innovative avenues to enhancing architectural design visualization via AI-facilitated image training models.

With the progress in large language models (LLMs) and the aforementioned image generation technology, generating architectural visualization images from textual input has become feasible. Referred to as text-to-image generation, this process produces exceedingly lifelike images and serves as a versatile tool for crafting an extensive array of architectural visualization content. As AI technology continues to advance, the role of text-to-image generation is anticipated to become pivotal in the realm of architecture. Consequently, the integration of image-generation AI augments the potential for creative exploration beyond conventional methodologies.

2.3. New opportunities: from mass geometry to render visualization

In the domain of architecture, the depiction of mass geometry has evolved into a typical technique of expression known as photorealistic rendering; this evolution owes to several technological advances described in the previous section and holds the potential to become a pivotal axis of architectural progress on a macroscopic scale, as it connects the conceptual and communicative aspects of architecture (Rafiq *et al.*, 2005).

First, photorealistic rendered images transcend mass geometry and enable architects to vividly convey their design intentions to clients. By presenting architectural concepts in a manner akin to reality, these images act as intermediaries between architectural drawings and the experiential aspect of architectural spaces (Lawrence, 1993). Such visualizations promote a shared understanding among stakeholders (Bragança *et al.*, 2014). Second, visualization allows not only architectural experts but also stakeholders, clients, and even the public to grasp architectural visions that transcend the realm of architectural terminology and technical intricacies (Ibrahim & Rahimian, 2010). Through visualized images such as photorealistic renderings, individuals can comprehend the characteristics, ambiance, of the envisioned architecture and its interactions with its surroundings, all of which enable informed decision-making. This shift from mass geometry to photorealistic rendered images makes complex architectural concepts more universally and broadly communicable and fosters easy communication.

In summary, incorporating visualization images such as photorealistic renderings beyond mass geometry into the architectural design process facilitates efficient communication during the early stages of architectural design. It also drives informed decision-making based on information and enhances creative design. Conventional architectural visualization relies on complex technical processes and requires GPUs and specialized hardware; however, the use of generative AI, as discussed in the preceding section, enables the effective acquisition of numerous detailed visualization images without the need for separate GPU renderers.

As mentioned in the introduction, the design of a building's façade can serve a visual and aesthetic reference within a specific area and street. Integrating the visualization process of building façade design efficiently at the early stages of architectural design, along with the use of photorealistic design representations, leads to numerous design alternatives and produces reference images related to design of building exterior. This, in turn, facilitates the

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

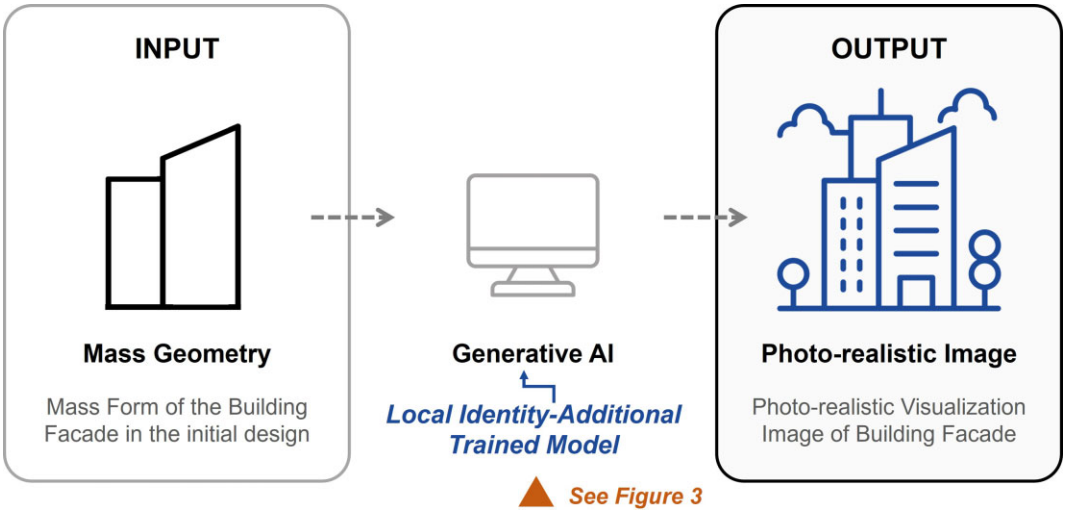


Figure 2: New opportunity of image generation from low level of detail geometry to photorealistic render images of façades (see Fig. 3 for more specific process of developing local identity-additional trained model).

creation of building façade designs that show creative diversity (Goldschmidt, 2004). Macroscopically, such a process contributes to enriching and enlivening landscape design elements, which include specific regions and streets (Lee & Park, 2023; Tandon & Sehgal, 2017).

The following section explores the application of generative AI to architecture, and it confirms the obtainable architectural images and details the specific utility of this approach. As outlined in the introduction, this study focuses on the design aspect of a building's façade within the domain of architectural elements. Specifically, this investigation aims to determine the feasibility of effectively generating architectural visualization images by emphasizing local identity as a pivotal design consideration during the design of the façade of a building. Therefore, as depicted in Fig. 2, this aims to project the domain knowledge of "local identity", particularly in the field of architecture and visualization of building façades, onto existing generative AI models. The objective of this study is to verify the applicability, practicality, and general applicability of AI technology in the fields of architecture and design.

3. LLM-Based Image Generation Approaches for the Design of Building Façade Alternatives

3.1. Cloud-based generation by generative AI platforms

AI-based image-generation platforms are now being made available to not only developers but also the general public through a convenient user-friendly interface. These platforms, as they are cloud-based, are accessible to anyone. Anyone who wants to create specific images can easily achieve the desired results using various platforms and prompts. Among these platforms, widely used ones provide excellent image-generation capabilities, including functionality such as text-to-image and image-to-image, along with additional adjustment features. Furthermore, applications that are developed using generative AI have been deployed as applications in the field of architecture (DreamStudio, n.d.; Midjourney, n.d.; OpenAI, 2022). When applications are installed in 3D programs in the architecture field, they allow the rendering of interior and exterior images of 3D models when given as input, producing them as output images.

Providing a detailed analysis of the user interface (UI) for all platforms is beyond the scope of this study. Instead, the study compares and test examples of the building exterior images that can be generated using these kinds of platforms to verify their capabilities.

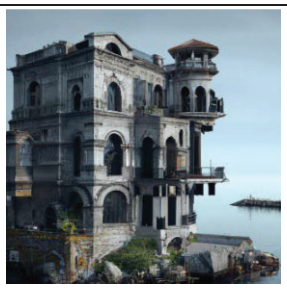
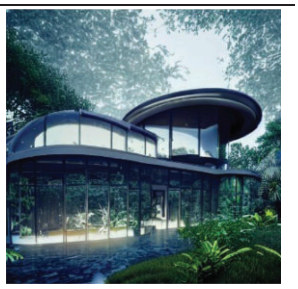
Among the text prompts that were used as input, "trigger words" that represented keywords included "realistic architecture" and "building façade image". Other prompts such as "high quality" and "high detailed image" were also used to enhance the quality and detail of the generated images. The term "trigger word" refers to the most crucial word in the input text (J. Kim *et al.*, 2023). The use of relatively simple and concise words as trigger words increased the freedom of the platforms in generating building exterior images. This facilitated precise testing and an understanding of the resulting images through comparative analysis, with an emphasis on examining the platforms' capabilities based on the chosen trigger words. Upon examining the image examples generated by each platform, the research observed that using generative AI-yielded remarkable and visually captivating architectural façade images. The image examples in Table 1 are all instances that were uploaded by users to social networking sites or AI image generation sharing platforms, and they primarily offer insights into the trends associated with user intentions for platform usage and the types of images the users intent to generate.

Images generated on the A and B platforms were primarily used to create fantastic and fairytale-like visuals and users often used these two platforms for tasks that involved generating creative and unrealistic images. In contrast, the C platform was observed to produce more realistic building exteriors than did the other two platforms. Therefore, although all three commonly used platforms generated building exterior images with extremely high quality and detail, the generation of realistic architectural images that one might encounter in real-life settings was predominantly observed on the C platform.

The diffusion model that is used on the C platform is a type of deep generative neural network composed of processes such as VAE, U-Net, and a selective text encoder. The study focuses on the application of this architecture to facilitate the generation of diverse, high-quality design images (Rombach *et al.*, 2022).

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Table 1: Example of the output images generated by the image generation platforms (Instagram, 2022; OpenArt, n.d.).

		A (OpenAI, n.d.)	B (Midjourney, n.d.)	C (Dreamstudio, n.d.)
INPUT	Case	Building facade		
	Trigger words	Realistic architecture, Building façade image		
OUTPUT	Generated Image			

When a diffusion model-based platform is used, as depicted in Table 1, a wide array of images can be generated. These images are generated from input prompts and showcase the model's ability to craft building design images that extend beyond conventional architectural façades and its creativity. This is efficiently achieved through the implementation of a Transformer-based text encoder into the diffusion model process. The model efficiently transforms the input text into embedding vectors for each token, which enable the image decoder of the image generator to create images based on the provided. By following this procedure, the research can craft design-oriented and creative images guided by text using the currently available platforms.

In this context, the current study employs a diffusion model to generate photorealistic images akin to actual building imagery, as these models are renowned for their reliability. The diffusion model harnessed by the mentioned platforms is employed in this study. The following section examines whether the diffusion model when used in conjunction with the platform can effectively capture local identity and produce architectural façade images aligned with this identity.

3.2. Local generation by base models without additional training

As mentioned in the introduction, the design of building façades serves as an intermediary between the external environment and internal spaces (Prieto & Oldenhave, 2021) and has potential to shape the backdrop of urban spaces and local living (Kotler *et al.*, 2008). These building façades encapsulate the identity of the region in which they are situated and enable individuals to experience a sense of place-based belonging through this identity (Relph, 1976). The local identity of a location can encompass various meanings and experiences, including the internal aspects of identity felt by people. Local identity, as a unique characteristic, is shaped by the historical narrative, cultural subtleties, and traditions inherent to a specific region or community. While this identity can be preserved, it may also change or fade due to modernization, globalization, standardization, and other factors. Therefore, the concept of local identity can be explained contextually; however, defining the identity that is specific to a specific area is subject to variability and subjectivity (Erickson & Roberts, 1997).

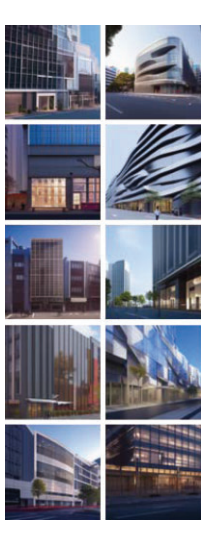
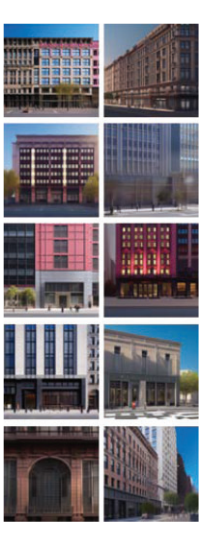
In this section, image-generation tests were conducted, focusing on the individual aspects of local identity among various characteristics, particularly the visual elements related to the design of building façades. The research initially examined façade images generated by the existing diffusion model. This study does not aim to define or explore the identity of a specific region but instead explores façade images that can be generated based on a variety of text prompts, including various regional names. Furthermore, unlike platforms that provided services based on the cloud in the previous section, image tests were conducted by installing locally (GitHub, 2023). As a result, a more diverse range of models was directly applied than in the previous section where tests were conducted based on the cloud. Additionally, this approach allowed for efficient and quick generation of a large number of images based on desired prompts.

As shown in Table 2, the research aimed to generate and examine images by inputting trigger words, such as "building façade", "Korea style building façade", "Tokyo style building façade", and "Manhattan style building façade" in the text prompts. Upon generating tens of thousands of images through prompts containing various trigger words, this research observed that when images based on keywords associated with countries, cities, and regions were generated, distinct design commonalities or overarching thematic similarities emerged according to the specific prompts. Furthermore, the addition of location names to the prompts resulted in exterior building images that exhibited greater detail and specificity than in cases where the nation, city, and region names were not provided. Along similar lines, scaling down the input of nation, city, and region names yielded images of higher quality and finer detail.

The generated images exhibited a notable level of visual resemblance to images that had been sourced from portal site searches or actual photographs, particularly in the context of well-known nations, renowned cities, famous regions, and Western design styles. For instance, as evident from Table 2, in the case of Korean-style building façade images, traditional Chinese-style hanok architecture was frequently generated, whereas images of buildings in Western Gothic architectural style were also frequently produced. It is worth noting that the visual similarities were especially pronounced for widely recognized countries, famous cities, renowned regions, and prevalent Western design styles. This trend

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Table 2: Images generated by entering various scale-specific building façade design style prompts.

		Without a regional keyword	With regional keywords		
			National scale	City scale	Street scale
INPUT	Case	Building facade			
	Used Model	Stable Diffusion (v2.1)			
	Trigger words	building facade, storefront image	**Korea style** building facade, storefront image	**Tokyo style** building facade, storefront image	**Manhattan style** building facade, storefront image
OUTPUT	Output images	 ...	 ...	 ...	 ...

underscores the efficacy of the image-generation process in capturing and representing these distinct architectural and cultural attributes within the generated images.

3.3. Local generation by using additional training models

The applicability of AI models in the field of architecture was established from the image-generation tests conducted on the fundamental generative AI models mentioned in the preceding sections. In particular, our assessment focused on building façades, which enabled an exploration of their specific utility. Consequently, the research obtained insights into the diversity of images learned for each text value of the references and prompts for building façade designs.

Thus, this study aims to identify building façade designs that incorporate crucial local identities during the initial stages of design, which is achieved through additional training of the existing model using region-specific images. This study endeavors to generate a range of practical images that mirror the design identity of a region by training on 4 the façades of buildings located within that region. This facilitates the generation of images that resemble real-world scenarios, thereby prompting discussions regarding practical applications such as marketing, building design, and remodeling based on.

Moreover, this approach could be further extended to the currently available generative AI model-based applications and platforms, as mentioned in Section 3.1. This extension may lead to greater efficiency and automated processes during the early stages of design and communication. In contrast to existing applications, where generating realistic rendering images, such as photos, from mass-only 3D models during the early stages of design is a challenge and achieving high-quality rendering images that reflect the desired design requires intricate prompts and iterative adjustment processes, the proposed approach aims to explore creative design solutions by applying additional training with domain-specific knowledge to the field of architecture. This could result in more sophisticated and realistic rendering images during the initial stages of design.

In Section 4, a detailed description of the modules that constitute the supplementary training model developed in this research is provided, along with the presentation of the resultant images that were achieved with its implementation. An outline of the development process for the upcoming steps is presented in Fig. 3. Furthermore, unlike the models and training data used in recently deployed applications and platforms, all the datasets that are constructed in this study use openly available data that has been directly preprocessed. Our intention is to make this method accessible for future use by general users or architects.

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

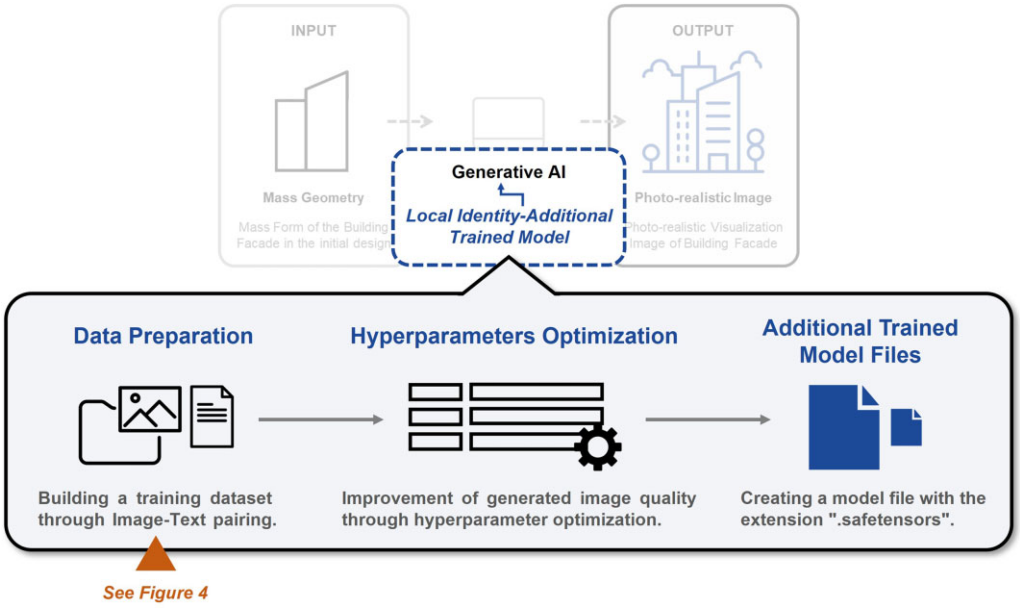


Figure 3: Approach to additional training to build the local identity-trained model (see Fig. 4 for more specific process about data preparation).

4. Additional Trained Model for Building Façade Design Alternatives

4.1. Data preparation and preprocessing

Our intent in this section is to elucidate the process of constructing a dataset of building façade images for additional training. The collection of building façade images for the proposed supplementary training model was facilitated by leveraging the street-view functionality of portal websites, thereby streamlining the process of image acquisition (J. Zhang *et al.*, 2022).

The images obtained through the Street-View API (Google, n.d.; Street View Download 360, n.d.) were captured using a 360-degree panoramic camera. Although these 360-degree panorama images are valuable data that are suitable for providing immersive visual experiences such as AR and VR (Lee *et al.*, 2023), they may not be suitable for use as data because of the distortion of individual buildings within the images. Therefore, to mitigate distortion in the output images, a correction process was applied to the 360-degree panorama images. Automatic segmentation was performed using a DL model implemented with TensorFlow and Python (Kim & Lee, 2022). This process took 360-degree panoramic images as input and produced segmented images having a resolution of 1280 × 720 pixels based on the field of view as output. For each 360-degree panoramic image, the segmentation resulted in 72 images from two angles, 120 and 90 degrees per image (Zhang & Hu, 2022).

Initially, the acquired 360-degree panorama image data were automatically segmented and among the segmented images, those containing intact building objects were selected and categorized. The categorized images were then uniformly adjusted to 512 × 512 pixels for ease of training. Subsequent operations involved aligning and orienting the building shapes within the images both vertically and horizontally, as well as adjusting the brightness levels. Such preprocessing tasks could enhance the usability of the training dataset and secure the potential for use in various tasks, as well as help improve the results and accuracy of the learning (J. Zhang *et al.*, 2022). Graphic adjustments were made using image editing software from "A" company (Adobe, n.d.) and brightness adjustments were uniformly processed using the data processing functionality of "R" company (Roboflow, n.d.). Automated tools for the 360-degree panoramic image segmentation and selection were developed based on advanced DL technology. This DL model, which was derived from the prominent YOLO series, was trained to detect building objects in images and videos (Redmon *et al.*, 2016).

After the image preprocessing phase in this study was complete, captions were assigned to each image to pair them with specified textual data. The text dataset was constructed using the Bootstrapping Language-Image Pre-training for unified vision-language understanding and generation model, which automatically generated text files. The constructed texts were then modified using a revision process (Li *et al.*, 2022). The captions consisted of keywords related to the region, such as "building façade image with local design style applied" and "design alternatives for building façade with local identity applied", along with design keywords such as "the design of building façades predominantly featuring red brick finishes" and "the design of building façades characterized by large glass windows and an open layout". The entire process for the training dataset is shown in Fig. 4.

The dataset for the additional training model was constructed by collecting and preprocessing image data that were paired with these keywords. In the next section, the study reviews the output images that can be generated using the supplementary training model based on this dataset.

4.2. Hyperparameter optimization

Optimizing hyperparameters during the constructing of additional training models improves the accuracy and quality of images generated by the additional training models. Hyperparameters are parameters that are set before the training process begins to control the behavior of the training algorithm and modify the performance of the model. This optimizes the process of building additional training models, which helps the models perform optimally for the specific tasks or problems they perform and improves their accuracy, generalization, and other desired performance metrics. Thus, by adjusting these hyperparameters, this research improves a model's ability to capture relevant patterns and

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

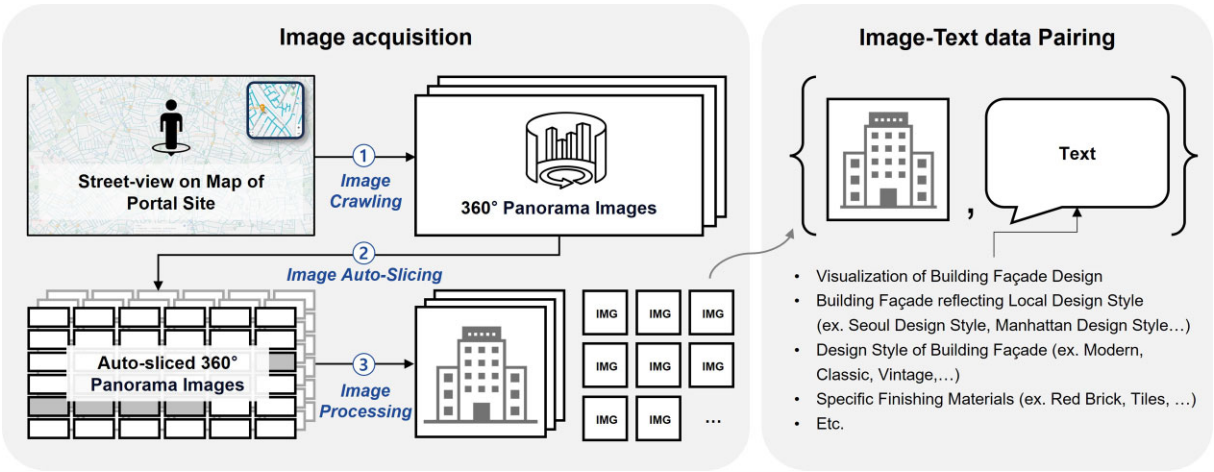


Figure 4: The process of constructing a training dataset.

makes predictions by aligning the behavior and characteristics of the model with new training data and training objectives (D. Liu *et al.*, 2023). In addition, hyperparameter optimization processes are generally used to advertise or underfit and efficient resources can be used to minimize computational resources such as memory, processing power, or training time (Feurer & Hutter, 2019).

Hyperparameters have various types of settings; however, this study focused on epochs, learning rates, learning batch sizes, and fine-tuning the images of additional training models to have high accuracy and high quality in relation to the desired style. This study employed the low-rank adaptation (LoRA) approach for LLMs to build an additional training model, with the intent to configure and train the parameters effectively. Instead of training all the weights of the pre-trained model, LoRA freezes all the weights and trains parameters efficiently by adding trainable rank decomposition matrices (Hu *et al.*, 2021). With the use of LoRA, additional training reduces the number of parameters and effectively minimizes memory usage (Valipour *et al.*, 2023; Zeng & Lee, 2023). This achieves a higher throughput with significantly fewer parameters. Therefore, to optimize model training, this research systematically diversified and deepened the training for each combination and selected the most efficient hyperparameter combination to evaluate the quality of the generated images.

4.3. Training and testing of additional training models

This section performs additional training and generates model files. Using the LoRA model mentioned in the previous section for additional training generates model files with a specific extension (e.g., "safetensors"). When these files are incorporated into the base diffusion model, they enable a more effective generation of desired images. During the additional training process, text data that are used in the training are input as prompts and prompt engineering techniques are employed to enhance the quality of the images that are generated. This process yields high-quality building façade images, as shown in Table 3. Trigger words such as "image of the building façade of the city, design style of Korea commercial building, and building façade design of Korea" were input and images were generated by providing a "positive prompt" and by focusing solely on enhancing image quality or reflecting details of the image itself. Conversely, a "negative prompt" was also generated to prevent distortion and maintain image quality (Liu & Chilton, 2022).

These images, which were considered to be part of the testing phase, exhibited a notably higher level of accuracy and quality than did the Korean-style façade images generated earlier and shown in Table 2. The additional training model files in Table 3 showcase examples of images that embody the façade design identity of Korea. In this study, similar additional training model files were tested and trained on a regional and national basis. The subsequent section constructs an additional training model for the design of building façades with local identity, as has been proposed in this study. In addition, the study intends to demonstrate its utility and value by applying it for specific scenarios.

5. Demonstration

5.1. Application scenarios for commercial building façades in Korea

This section validates the research approach by using a wider scope for the building façade image-generation models constructed in the previous section. The scope of our validation includes neighborhood facilities in Seoul, which is the capital city of South Korea. Specifically, the study constructs additional training models that emphasize the local identity of the building façades of these neighborhood facilities. Subsequently, this study proposes the use of software and web applications based on these models. The details of the scope and the targets used to demonstrate this study are identified in Fig. 5.

Neighborhood facilities refer to designated structures that provide the necessities for daily life in residential areas. They encompass a broad range of buildings, including commercial facilities, educational institutions, cultural venues, and medical facilities (J. Jacobs, 1961; Park & Yang, 1997). The buildings of such community facilities are located along major roads in commercial or residential areas, and they form vibrant urban spaces (A. B. Jacobs, 1993). As a space that is inhabited by various stakeholders, such as residents, merchants and visitors, these structures possess sociocultural and commercial value (Alexander, 1977; Galster, 2001). Therefore, this study has chosen buildings of a multifunctional nature that are characterized by diverse commercial values as the focus of our investigation, with a particular emphasis on the design and aesthetic elements of their façades.

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Table 3: Example of an image generated by applying an additional training model file.

Generated Image	Used Prompt
	***Trigger words:*** **Image of the building facade of city, Design Style of Korea Commercial Building, Building Façade design of Korea** *Prompt for Image Quality:* (Positive) photorealistic, hyper detailed, good quality, realistic shadow, photorealistic building, v-ray rendering image, high detail of buildings, nice exterior of building, image like rendered based on twin motion into 8K pixel size, architectural image, high dynamic range (Negative) bad quality, normal quality, bad detail, blurry, foggy, bad quality shadow, drawing, sketch, poor quality, ugly, pixelated, low resolution, saturated, high contrast, over sharpened, split image, distortion

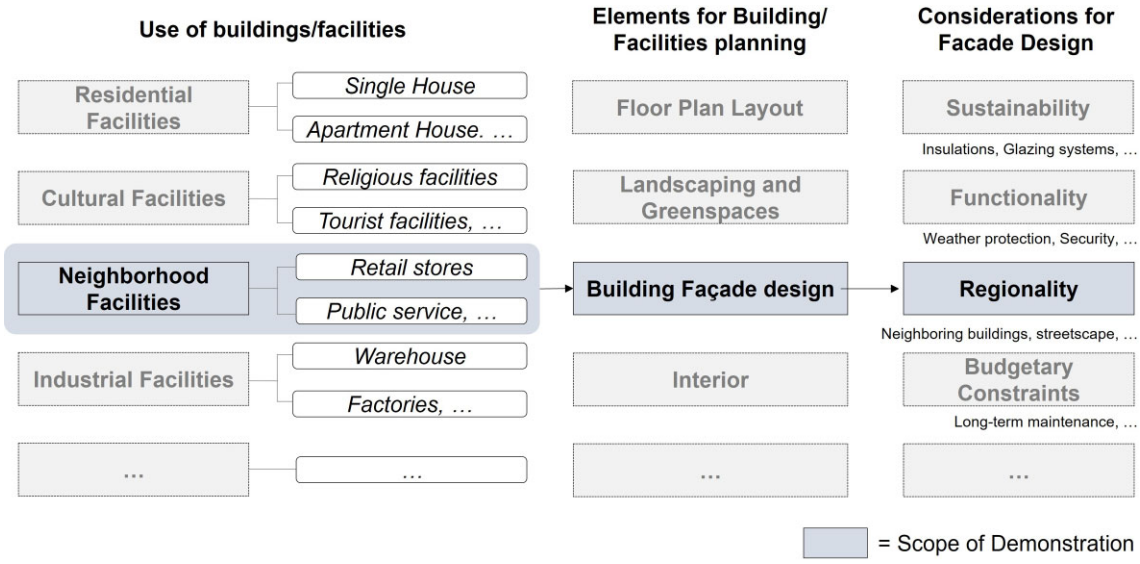


Figure 5: Conceptual structure for explaining the scope used in the additional training models to be conducted in this demonstration.

These application cases demonstrate the service that building façades provide as a connection between interior spaces and the urban environment. Consequently, they showcase the various design alternatives that could be generated. The proposed additional training models take input in the form of images and text and produce the desired design replacement images as photorealistic output images. The research aims to validate the utility and value of the modules within the additional training models proposed in this study and to implement specific application scenarios.

5.2. Additional training for the target design styles of building façades

This section divided the steps into three main stages according to the additional training process described in Section 4: (i) preparing the dataset by pairing street-view images collected using the API of the portal site with related textual data, (ii) conducting training using the Stable Diffusion's Web UI based on optimized hyperparameters, and (iii) generating the newly trained model and verifying the resulting images.

Furthermore, the additional training models are summarized in Fig. 6. In the data collection phase, as described in Section 4.1, this research obtained 360-degree panoramic images using the street-view feature provided by a portal website. The efficient extraction of building façades was achieved using image slicing and DL-based techniques, as outlined in the dataset construction step. The images collected during this process represented various conditions, such as weather, seasons, and time, enabling the generation of façade images that were more realistic and diverse. Images that were distorted because of building shape distortion or obstructed views, such as images containing trees or vehicles, were excluded. All other images underwent preprocessing to prepare the dataset. This preprocessing included resizing to the most effective size of 512 × 512 pixels for training in LoRA. In addition, corrections were made for any distorted building shapes. Subsequently, pairing with textual data allowed us to construct a training dataset, as illustrated in Fig. 7. To further train the model, the objective was to generate new and creative images from the training dataset for additional training according to user requirements. Therefore, there was no separate step for the creation or splitting of the testing dataset; the training was conducted directly. There-

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

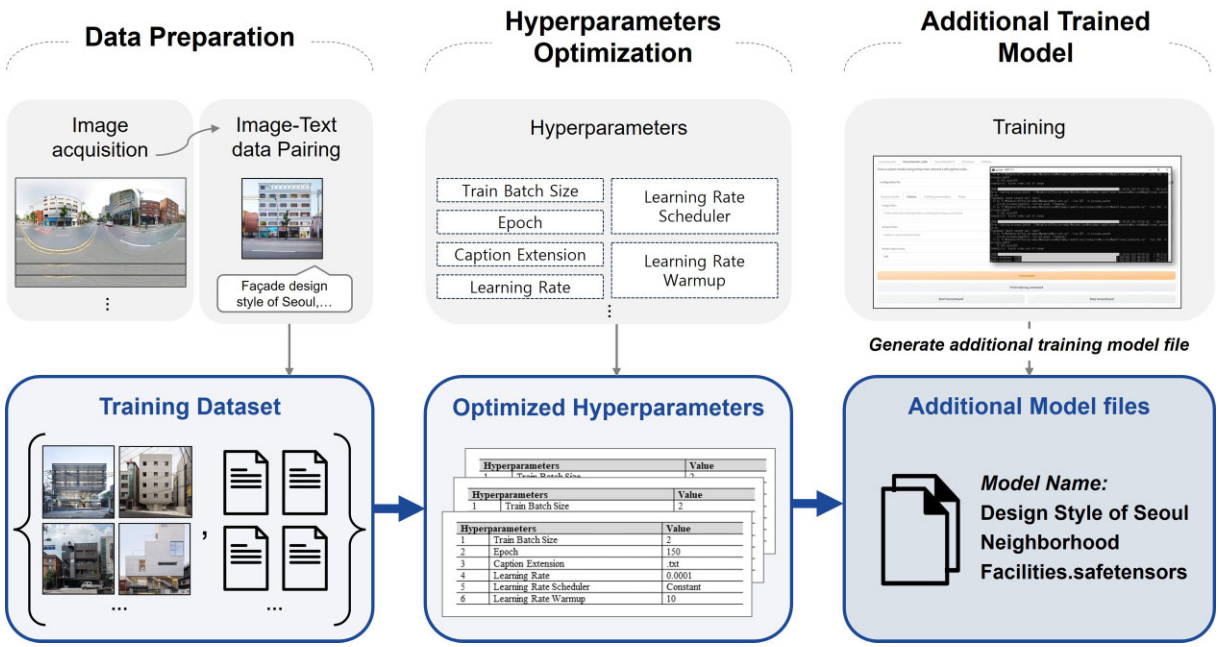


Figure 6: Process of additional training steps in this demonstration.

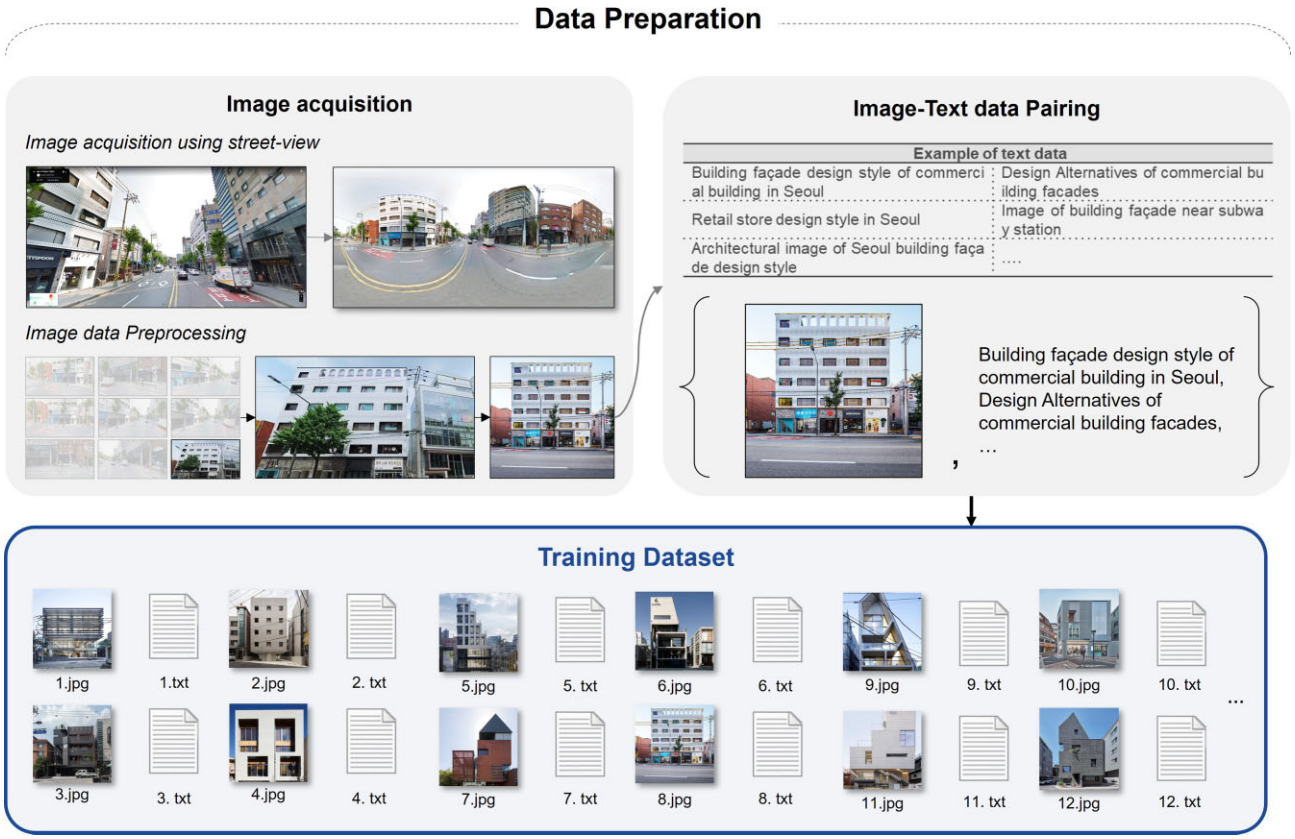


Figure 7: Data preparation step of demonstration.

fore, for the training data, approximately 2000 image datasets and approximately 2000 text datasets were used. During the training process, an A6000 GPU was employed with a VRAM of 47.5 GB on a PC with a CPU specification of 275 GB. The training duration was approximately 120 min and resulted in the generation of additional training model files.

The quality of an additional training model is greatly affected by the optimization of its hyperparameters, such as number of epochs, learning rate, learning rate scheduler, and learning rate warm-up. These hyperparameters directly affect a model's ability to adapt and learn when it incorporates new data. The number of epochs determines the number of training iterations, whereas

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Table 4: Overview of hyperparameters optimization.

	Hyperparameters	Description	Value
1	Train batch size	The number of training examples or samples that are processed together in each iteration or batch during the training process.	2
2	Epoch	A complete iteration through the entire training dataset during the training process of a model.	150
3	Caption extension	The process of enhancing or expanding the captions associated with images or visual content. It involves adding more descriptive or informative text to provide a more detailed or comprehensive description of the content.	.txt
4	Learning rate	Second parameters that determine the step size or speed at which the parameters of the model are updated during training.	0.0001
5	Learning rate scheduler	A technique used for ML to dynamically adjust the learning rate during training, which modifies the learning rate according to predefined rules or conditions instead of using a fixed learning rate throughout the training.	Constant
6	Learning rate warm-up	A technique for gradually increasing the learning rate from a small value to a desired value or a scheduled value in the early stages of training.	10

the learning rate controls the step size for the weight adjustment and strikes a balance between convergence speed and stability. The learning rate scheduler modifies the learning rate over time, which aids in efficient training. In addition, the learning rate warm-up gradually increases the learning rate at the beginning of training to ensure stability. Properly configuring these hyperparameters is crucial to achieving a high-quality additional training model within the LoRA framework.

Therefore, the study conducted various tests for additional training using different hyperparameter types and settings to generate target building façades. Based on the results of the conducted tests, the hyperparameters for further training were set as follows: the batch size was set to 2, the number of epochs to 150, the learning rate to 0.0001, and the other parameters were set as specified in Table 4. In this demonstration, the optimization criterion for hyperparameters was set to a loss value below 0.02. The final values of the hyperparameters, which were determined through heuristic optimization, are presented in Table 4.

By using the model files generated through additional training, the research was able to obtain the output images shown in Table 5. As evident from Table 5, inputting textual descriptions, including location names, resulted in the generation of European-style building images instead of façades specific to Korean neighborhood facilities when the existing model was used. Even when the term "Seoul" was input, it was difficult to discern images tailored to Korean-style neighborhood facilities. However, when the additional training model files implemented in this demonstration were used, it became possible to generate images that represented the architectural façade design style of the facilities in Seoul's neighborhood. These generated images exhibited a level of accuracy and quality that is generally difficult to distinguish in real images and incorporate creativity trained from the dataset. The following section describes the method for evaluating the results of the additional training model constructed in this study through quantitative measurements based on the generated images, as shown in Table 5.

5.3. Quantitative measurement and evaluation of generated images and developed model

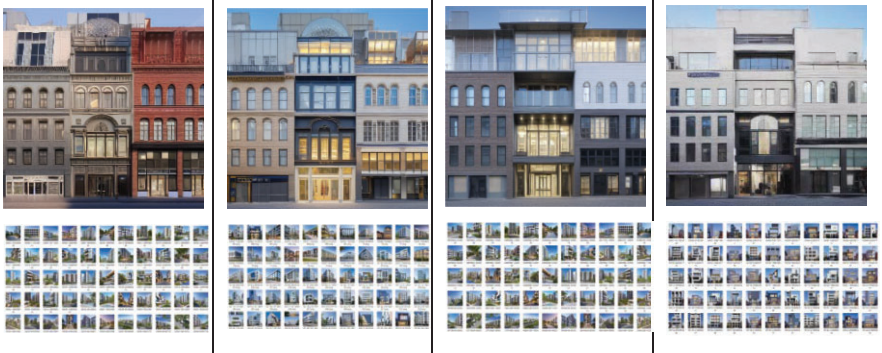
This section establishes a quantitative measurement method for evaluating the images generated by the model that received additional training based on generative AI. The study then conducts a brief evaluation. The results for the model trained on generative AI may have a quantitative distance from the accuracy measurements of traditional ML and DL models. Efforts to increase model training accuracy and validation accuracy, as seen in tasks such as the image recognition of cats or dogs (Parkhi *et al.*, 2012) and flower type classification (Nilsback & Zisserman, 2008), may not be as relevant, especially given the qualitative architectural domain knowledge of local identity that this study focuses on in generating alternative designs for the exterior appearance of buildings during the initial design phase. Instead, the generated results may be more closely related to human preferences and tendencies.

Therefore, to quantitatively evaluate the results of the generative AI-based additional training model, this study constructed an available AI training model. Presenting numerical values and stochastically evaluating this model was considered to be a convincing approach because it aligns with a focus on human preference and the tendency of the results. This motivated us to build and use an evaluation training model for this purpose.

Before evaluating other images, the research conducted a brief, quantitative evaluation of the generated images in Table 5. This quick evaluation encompassed three main criteria, "fidelity of domain issue", "image quality", and "similarity" to the applied region. The "fidelity of domain issue" criterion assessed whether the generated images aligned with the façades of the buildings. "Image quality" evaluated whether the generated images met the criteria for high-quality visuals. "Similarity to the applied region" was quantified as the QE score to determine how effectively the regional design style of the applied additional training model files was reflected. This evaluation was conducted using a simplified 5-point Likert scale and employed a DL-based image recognition algorithm based on a Convolutional neural network (CNN) based pre-trained inception model (LeCun *et al.*, 2015; Xia *et al.*, 2017). The DL-based image recognition model was trained on randomly collected and preprocessed images of building façades, including those representing the architectural styles of neighborhood facilities in Seoul. The constructed image dataset, which comprised approximately 700 images, was gathered from portal sites and social media platforms (Kim & Lee, 2020; Krizhevsky *et al.*, 2017; Szegedy *et al.*, 2016). The dataset was later split into a training dataset (85%) and a testing dataset (15%). The model was trained until it achieved a training accuracy and validation accuracy close to 1. Each generated image was tested using the trained image recognition model and its similarity to the design of building façades

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Table 5: Generating façade design of neighborhood facilities in Seoul based on implemented additional training model.

			A	B	C	D
INPUT	Case	Building Type	Neighborhood Facilities			
		Region	Seoul			
	Prompt	Trigger words	Design Style of Seoul Neighborhood Facilities, Design of Urban Neighborhood Facilities, Design of Seoul Neighborhood Living Facilities, Front View of Neighborhood Facilities Building in Seoul			
	ATM	Model name	Base Model only	Design Style of Seoul Neighborhood Facilities.safetensors		
		Weight	0	0.3	0.7	1
OUTPUT	Generated Images					
Quick Evaluation			Fidelity of domain issue: Good Image quality: Good QE Score: Very Low	Fidelity of domain issue: Good Image quality: Good QE Score: Low	Fidelity of domain issue: Good Image quality: Good QE Score: High	Fidelity of domain issue: Good Image quality: Good QE Score: Very High

QE Score: The QE Score is an evaluative metric used to quantitatively assess whether the desired design identity of the target region is reflected in the generated images, with detailed information available in Fig. 8.

QE Score: The QE Score is an evaluative metric used to quantitatively assess whether the desired design identity of the target region is reflected in the generated images, with detailed information available in Fig. 8.

images was rapidly measured based on the resulting image recognition rates. To quantify the QE score, the training and validation accuracy of the DL-based image recognition model constructed for this purpose, along with the established scale based on them, can be observed in Fig. 8.

The study tested the generated images shown in Table 5 using this trained image recognition model. The results showed that for Case A, the similarity to the design style of Seoul's local facilities was approximately 14%, which was rated "very low". For Case B, it was 27%, which was rated "low". Case C achieved 68%, which was rated "high", and Case D achieved 86%, which was rated "very high". This indicates that when images are generated with adjustments to the application rate of additional training model files, the resulting images for each application rate appropriately reflects the local design style of the building façade. Therefore, this research has confirmed that the additional training model files constructed in this demonstration are effective for their intended purpose. The following section further demonstrates the utility of the constructed additional training model files by applying them to different cases.

5.4. Generating building façade design images based on additional trained models

This section confirms the generation of creative building façade designs that reflect design requirements by applying various additional training model files constructed in this research. As an illustrative example, Table 6 presents examples of images generated based on prompts containing trigger words. In the cases of A, B, and C in Table 6, the target building type was set as 'local facilities', and the selected geographical area was Seoul.

For Case A, the objective was to generate building façade images that reflect the design style of Seoul's local facilities, with a particular focus on placing large windows on the ground floor to capture people's attention. In Case C, the goal was to generate building exterior images for the Seongsu-dong area of Seoul, incorporating design elements such as the use of red brick finishes to reflect the design style of local facilities in Seongsu-dong. In Case C, the aim was to generate building façade images that reflect the design style of Seoul's local facilities, and some of them were designed to align with the style of Seongsu-dong's building façades, aiming to generate creative design alternatives that meet various design requirements.

The primary intention of this demonstration was not to define the identity or design style of a specific region. However, as evident from the table above, Case A yielded various design alternatives that clearly reflected the building façade style of local facilities in Seoul. The image quality was also exceptional, with minimal distortion of building forms when compared with images of newly constructed neighborhood facilities. Similarly, in Case C, the design style of local facilities in the Seongsu-dong area of Seoul was distinctly reflected in the generated images. Particularly notewor-

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

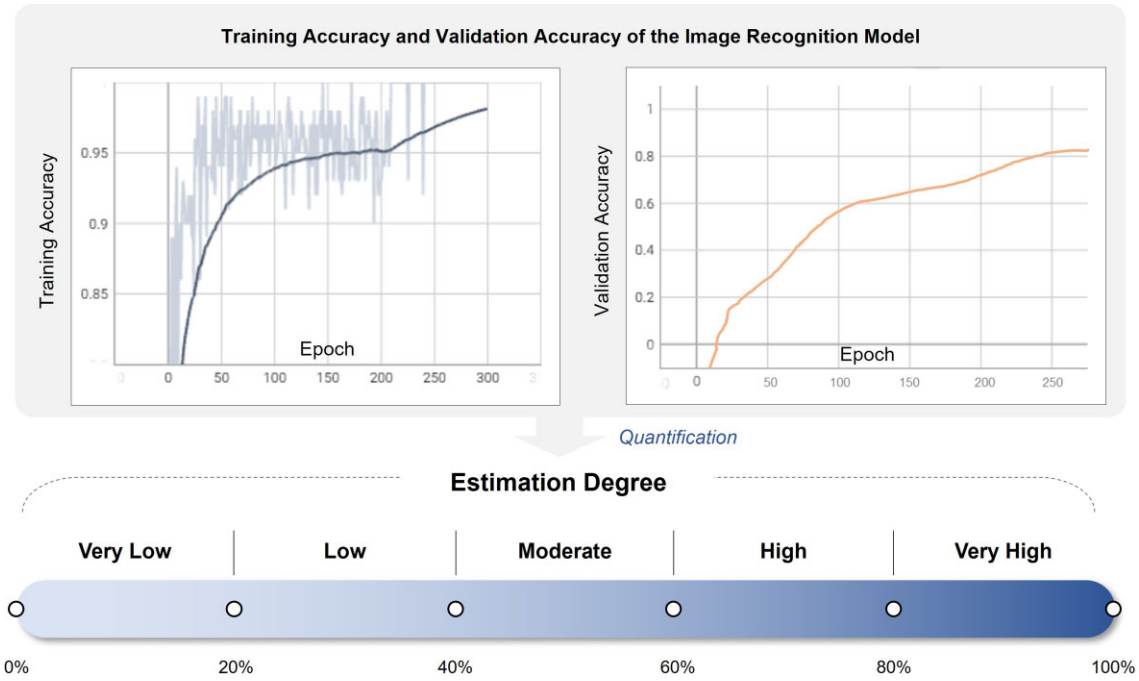


Figure 8: Quantitative evaluation scale based on epoch-wise training accuracy and validation accuracy of the trained image recognition model, utilized for quick assessment of generated images using a 5-point Likert scale.

thy were the design images featuring building façades made of red bricks, which can be considered a unique design identity of Seongsu-dong. This can be predicted based on the learning data derived from Seongsu-dong's building façades, along with various research papers, policy documents, and related literature (Cho & Chi, 2020; Kim & Lee, 2016; Nahm & Jang, 2016).

In the case of Case B, the design style of Seoul's building façades was reflected, with the inclusion of the Seongsu-dong model in the image generation process. As a result, the images generated in Case C featured designs with red brick finishes on the front of local facilities in Seoul and included a series of square-shaped windows on the ground floor, along with open commercial spaces. The next section proposes tools to maximize the utility of these additional training models and provide examples of photo-realistic images that can be generated by applying these tools.

5.5. AIBIM-Design: implementation of AI rendering tool for photorealistic renders of façade, from mass geometry

This section describes additional scenarios that could be used during the early design stage of the architectural process. During the initial design phase for a building, various design alternatives may be required in the design of the façade and when only a basic mass form exists on the site where the design is needed, a variety of visually appealing design images may be needed for the communication with the client. In such cases, there is a need for high-quality images that resemble real-life situations, and these have traditionally involved the use of GPU-based rendering tools or software such as Photoshop. The study proposes approaches that can achieve efficiency in terms of cost and time during this process.

Figure 9 depicts the basic interface screens of the developed approach, which is the AIBIM-Design tool. As a prototype, it has been developed based on the Korean language, and captions have been provided for key functionality in the images. The tool is implemented to open the land type of the ".KBdgn" extension to load land and buildings or to allow users to set up land and draw buildings themselves. In addition, various types of building information such as land area, total floor area, building area, floor area, building area, and landscape area were obtained through a simple interface chart provided by the tool based on the land and buildings designed using the tool. The shape of the land building designed with this tool is generated with a simple mass geometry and to generate a building exterior perspective view of the user-designed mass, one generates a photorealistic design alternative in just a few seconds by applying the local design identity-training model constructed in this study.

The approach to generating photorealistic images based on mass geometry is implemented using the ControlNet algorithm (L. Zhang *et al.*, 2023). When the ControlNet algorithm is applied and the diffusion model is used, a canny edge detector is employed to detect the outline of the image, which initiates the initial preprocessing of the input image. After the outline is detected, the generation of photorealistic images is conducted based on the outline. In the AIBIM-Design tool implemented in this study, users can directly determine the priority for outline detection and model settings through the interface. For example, as shown in Fig. 9, setting "⑧ ControlNet Mode: Set ControlNet as Priority" emphasizes the lines of the outline more prominently according to ControlNet, which results in an image that closely matches the mass shape of the input image. Conversely, setting Model File Application as a higher priority prioritizes the application rate of the additional training model over the accuracy and detection of the outline, which leads to the generation of images.

The proposed examples can be applied as a new AI-based design approach during the early stages of architectural design decision-making and design communication. Based on the application approach of this study, designers can quickly generate images such as real-life images to facilitate easy communication

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Table 6: Generating façade design of neighborhood facilities based on implemented additional training model.

			A	B	C
INPUT	Case	Building Type	Neighborhood Facilities		
		Region	Seoul		
	Prompt	Trigger words	Design Style of Seoul Neighborhood Facilities, Front View of Neighborhood Facilities Building in Seoul	Design Style of Seoul Neighborhood Facilities, especially reflecting seoungsu-dong design style	Neighborhood facility of seongsu, Storefront design style of seongsu, Facade design style of seongsu-dong
		Design words	Large glass windows on the ground floor	Large glass windows on the ground floor, Red brick finish, a regular arrangement of the front window of a building	Red brick finish, a regular arrangement of the front window of a building
	ATM	Used File name	Design Style of Seoul Neighborhood Facilities.safetensors	Design Style of Seoul Neighborhood Facilities.safetensors / Design Style of Seoungsu-dong Neighborhood Facilities.safetensors	Design Style of Seoungsu-dong Neighborhood Facilities.safetensors
OUTPUT	Generated Images		  …	  …	  …
Quick Evaluation			Fidelity of domain issue: Good Image quality: Good QE Score: Very High	Fidelity of domain issue: Good Image quality: Good QE Score: Very High	Fidelity of domain issue: Good Image quality: Good QE Score: Very High

with clients. Therefore, if this generative model is used during the initial design phase of the architectural design process, designers not only gain a tool for rapidly generating real-life image data but also have the ability to adjust detailed designs.

Furthermore, through the additional training-based generative model developed in this study, it can serve as a design approach that strengthens the connection between the interior and exterior of buildings by generating architectural façade design images based on local identity.

This tool can be used to generate the examples shown in Table 7, along with the simpler interface mentioned earlier. In this interface, users can freely visualize the form of mass as an initial architectural model and apply the additional training models implemented in this study to generate lifelike images. Examples of this can be found in Fig. 10. The images in Fig. 10 were all generated using the AI renderer feature of the AIBIM-Design tool proposed in this study. It can produce high-quality images in just a few seconds and generate numerous design alternatives from a single initial mass design.

6. Available Resources and Limitations of Work

The proposed approach was demonstrated with a focus on Seoul, South Korea. However, a limitation of this research arises from the fact that it did not provide examples integrating the identities of various other regions. Furthermore, it may be a limitation that the concept of local identity deals with somewhat complex notions, and, as a result, it cannot be conclusively demonstrated that the generated images reflecting the regional design style model developed in this study for Seoul, South Korea, ex-

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

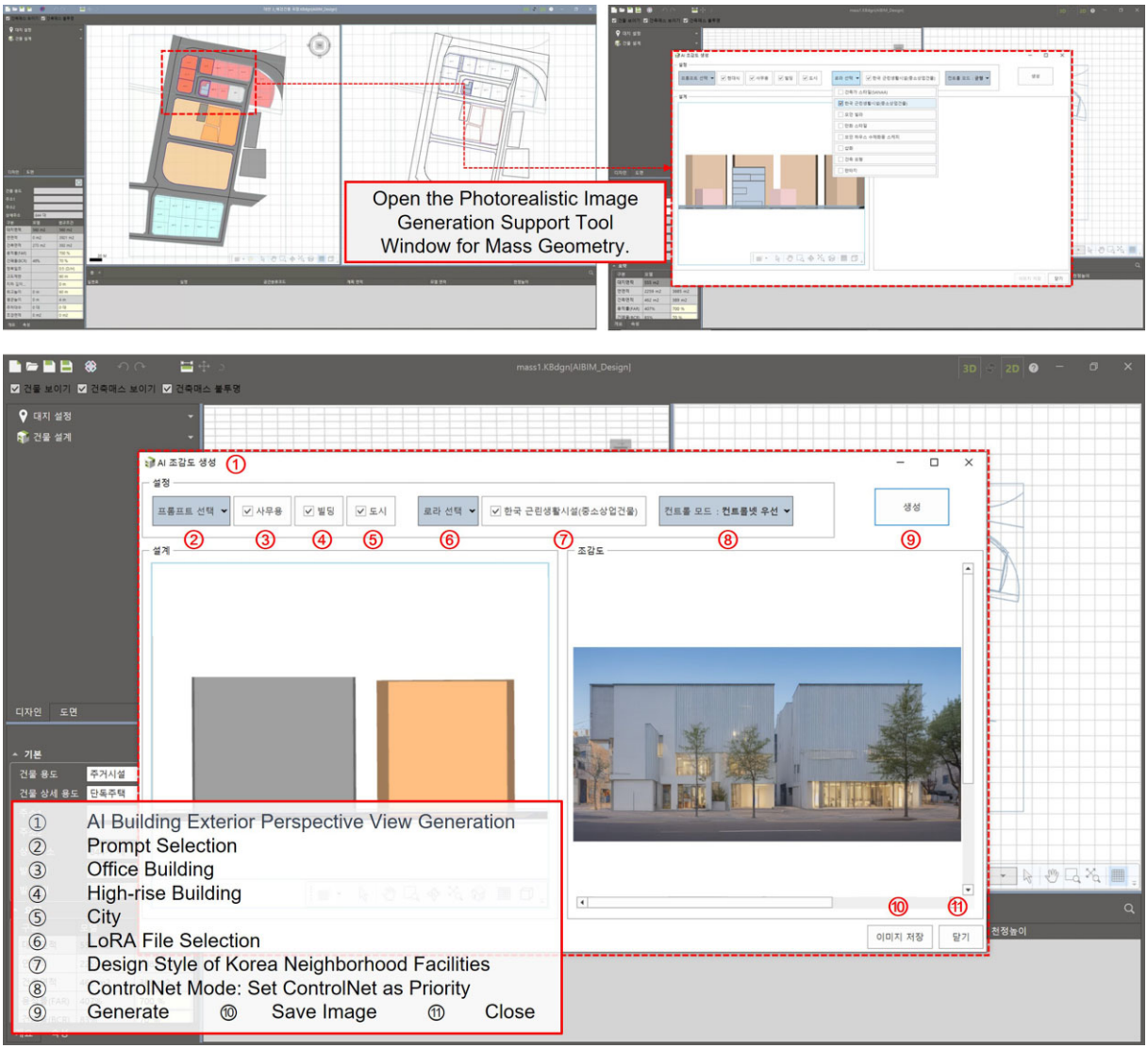


Figure 9: Using AIBIM-Design to generate a photorealistic image from mass geometry.

hibit 100% accuracy. Nevertheless, by acquiring and preprocessing image data of diverse architectural structures from various regions in Seoul through the street-view portal, the foundation for the developed approach was established. Furthermore, this study proposed a module to secure data from more countries and regions through this approach, allowing for the generation of additional training model files that reflect the façades' identities from more regions. This, in turn, enables the generation of highly creative design alternatives by integrating various design identities, extending beyond the effective application of local identity within various design considerations for building façades demonstrated in the showcase.

In the demonstration of this study, the effective application of local identity within various design considerations for building exteriors was primarily emphasized. Additionally, the current focus has been on discussing the generation of photorealistic images from the architectural form of mass geometry, highlighting the applicability and utility in real-world scenarios. However, there may be limitations to the applicability of this research in practical use, considering factors such as its concentration on generating photorealistic images from the architectural form of mass geometry. Nevertheless, it is expected that in the future, the process and approach presented in this study could be expanded to achieve integrated decision automation, encompassing other architectural considerations (e.g., energy efficiency, economic feasibility in terms of cost, etc.).

As one application and implementation example, in this study, a web page has been developed where users can directly download and utilize high-quality image examples generated through additional training model files and the application of model files. An example of the webpage is shown in Fig. 11. While the results of the proposed method in this study are generated as image files, there may be limitations in directly applying or integrating them into 3D modeling. However, it should be noted that all generated images are of exceptionally high quality and represent photo-realistic characteristics, thereby rapidly generating creative design alternatives. Therefore, it is expected that this could lead to generating or combining more data in the future and proposing approaches for utilizing reference images and models.

7. Conclusions

This paper proposed a new approach and tool for generating photorealistic building façade images. The use of photorealistic im-

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Table 7: Image generation from each mass image of buildings.

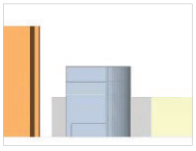
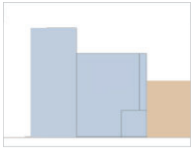
			A	B	C	D
INPUT	Case	Building Type	Neighborhood Facilities			
		Region	Seoul			
	Prompt	Trigger words	Design Style of Seoul Neighborhood Facilities, Design of Urban Neighborhood Facilities, Design of Seoul Neighborhood Living Facilities, Front View of Neighborhood Facilities Building in Seoul			
		Design words	Large glass windows on the ground floor	Overall Louver Design of Exterior Wall Finishing Material	Design of balcony and rooftop spaces with the glass railing	Piloti Structure, Exterior Wall Design Composition of Concrete Block
	ATM	Used File name	Design Style of Seoul Neighborhood Facilities.safetensors			
	Images					
OUTPUT	Generated Images		  ...	  ...	  ...	  ...
Quick Evaluation			Fidelity of domain issue: Good Image quality: Good QE Score: Very High	Fidelity of domain issue: Good Image quality: Good QE Score: Very High	Fidelity of domain issue: Good Image quality: Good QE Score: Very High	Fidelity of domain issue: Good Image quality: Good QE Score: Very High



Figure 10: Examples of building façade images generated by "AI façade renderer" in the AIBIM-Design tool.

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

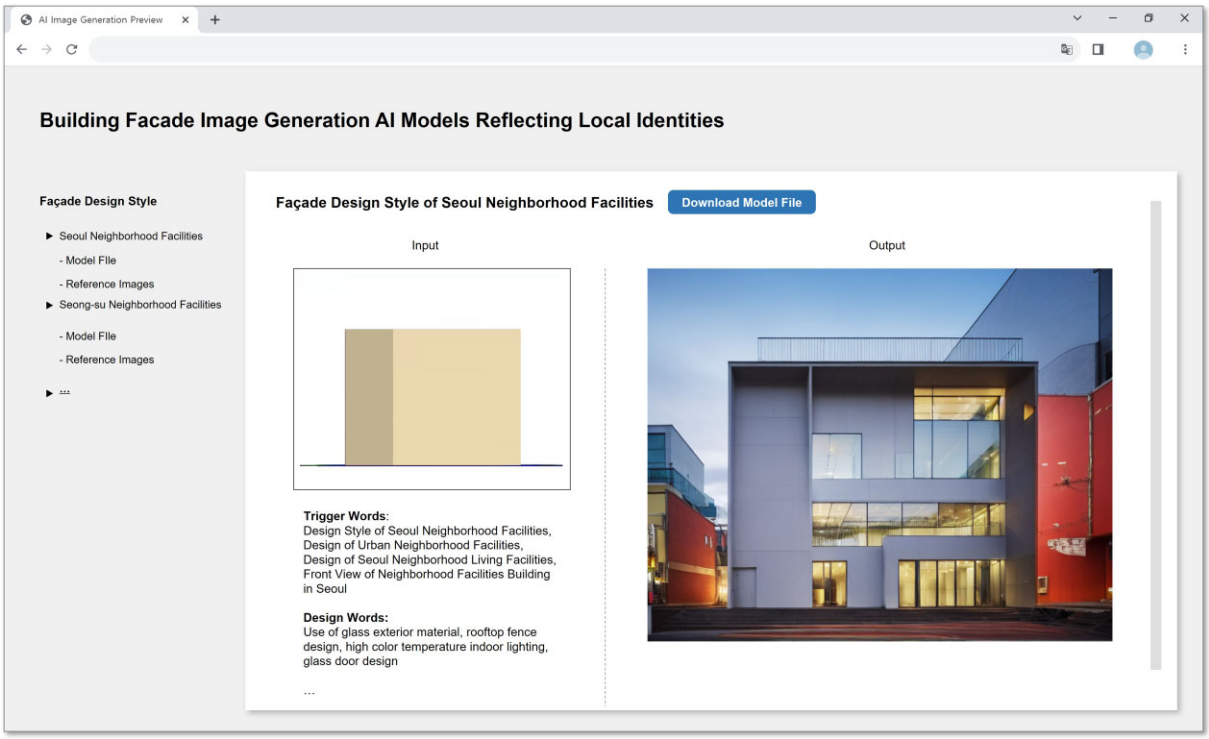


Figure 11: The interface of a webpage where the built available models from this demonstration can be downloaded.

ages in architectural visualization spans the entire lifecycle of a project, from the initial planning stages of a complex and multifaceted design process to the final construction phase. Such design visualization can effectively communicate designs to stakeholders and it can be used to test begins and facilitate design improvements. Consequently, the development of computational technologies and application tools for obtaining photorealistic images continues; however, existing tools often demand considerable time and resources. The approach proposed in this study leverages generative AI and is more efficient and scalable and generates a broader range of design alternatives than conventional methods.

The feasibility of this approach is demonstrated by using a specific design context, in which the design is used to generate façade design alternatives for specific types of building projects. This study confirmed through intensive testing whether an architectural image containing domain knowledge could be generated based on the image-generation model that existed previously. As a result, this research confirmed that a building façade image generated based on a prompt containing a region name could produce a relatively specific detail and high-quality image than an image of a prompt that does not include a region name. However, the existing model does not train design identities in various regions and model training was conducted to determine whether it could generate high-quality façade design images representing various creative processes by generating images reflecting more detailed of the local identities. Model training was conducted based on a dataset paired with text data by acquiring and preprocessing a large amount of image data through street-views on portal sites by region and various hyperparameter values were adjusted to conduct optimal model training. In this study, the training model was verified and tested for neighborhood facilities in Seoul within the scope of the demonstration to see if it could be effectively changed from mass geometry to a photorealistic image. Through this verification, this study implemented the AIBIM-Design tool as a generative AI-based visualization tool that could be easily accessible by architectural experts and also by the public. Several additional training model files that were built during the course of this study were input into the AIBIM-Design tool to confirm that real-life images could be generated in a short time by representing as many design identities as desired in existing mass-type models on land.

Except for the time allocated to development and further training, this research generated the actual building façade images using the models described in the paper. The total number of façade images generated was approximately 2000. The time and hardware resources consumed during this generation process were approximately one machine every 2 days. Most of the generated images exhibited meaningful and effective results, as reported in the project web resource. Among the numerous possibilities for design visualization, this study demonstrated this concept under very limited conditions and situations. However, the potential has been demonstrated and work is underway to construct additional training models. Future development directions include the following: horizontally, building additional training models for different building façade types, building categories, and visualization methods; vertically, extending beyond building façades to incorporate completely different kinds of visualizations, including multimode representations.

Acknowledgments

This work is supported in 2024 by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA163269). This work was supported by the National Research

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Foundation of Korea (NRF) grant funded by the Korea government (MIST) (No. NRF-2022R1A2C1093310).

Conflict of interest statement

None declared.

Data availability

The first or corresponding author (PI) can provide most of the data for training, and/or models that were used in this study upon a reasonable request, as well as the links in references and the technical resource section. Additionally, the paper includes references to the archives and links provided by the PI.

References

Abrishami, S., Goulding, J., & Rahimian, F. (2021). Generative BIM workspace for AEC conceptual design automation: Prototype development. *Engineering, Construction and Architectural Management*, **28**, 482–509. https://doi.org/10.1108/ECAM-04-2020-0256.

Adobe. (n.d.). Photo and image editing software for photographers. https://www.adobe.com/creativecloud/photography.html.

Alexander, C. (1977). *A pattern language*. Oxford University Press.

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning* (Vol. **70**, pp. 214–223). PMLR.

Balasubramanian, S., Irulappan, C., & Kitchley, J. L. (2022). Aesthetics of urban commercial streets from the perspective of cognitive memory and user behavior in Urban Environments. *Frontiers of Architectural Research*, **11**, 949–962. https://doi.org/10.1016/j.foar.2022.03.003.

Bragança, L., Vieira, S. M., & Andrade, J. B. (2014). Early stage design decisions: The way to achieve sustainable buildings at lower costs. *The Scientific World Journal*, **2014**, 1–8. https://doi.org/10.1155/2014/365364.

Brock, A., Donahue, J., & Simonyan, K. (2019). Large scale GAN training for high fidelity natural image synthesis. preprint (arXiv:1809.11096). https://arxiv.org/abs/1809.11096.

Caetano, I., Santos, L., & Leitão, A. (2020). Computational design in architecture: Defining parametric, generative, and algorithmic design. *Frontiers of Architectural Research*, **9**, 287–300. https://doi.org/10.1016/j.foar.2019.12.008.

Castelo-Branco, R., Caetano, I., & Leitão, A. (2022). Digital representation methods: The case of algorithmic design. *Frontiers of Architectural Research*, **11**, 527–541. https://doi.org/10.1016/j.foar.2021.12.008.

Cho, H., & Chi, S. (2020). Juxtaposed cities: The coexistence of gentrification and anti-gentrification policies—The case of Seongsu-dong, Seongdong-gu, Seoul. *Journal of the Korean Geographical Society*, **55**, 363–377.

de la Fuente Suárez, L. A. (2020). Subjective experience and visual attention to a historic building: A real-world eye-tracking study. *Frontiers of Architectural Research*, **9**, 774–804. https://doi.org/10.1016/j.foar.2020.07.006.

DreamStudio. (n.d.). https://beta.dreamstudio.ai/.

Erickson, B., & Roberts, M. (1997). Marketing local identity. *Journal of Urban Design*, **2**, 35–59. https://doi.org/10.1080/13574809708724395.

Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In *Automated machine learning* (pp. 3–33). Springer. https://doi.org/10.1007/978-3-030-05318-5_1.

Galster, G. (2001). On the nature of neighbourhood. *Urban Studies*, **38**, 2111–2124. https://doi.org/10.1080/00420980120087072.

Gero, J. S. (1983). Computer-aided architectural design: Past, present and future. *Architectural Science Review*, **26**, 2–5. https://doi.org/10.1080/00038628.1983.9697249.

GitHub. (2023). AUTOMATIC1111/stable-diffusion-webui: Stable diffusion web UI. https://github.com/AUTOMATIC1111/stable-diffusion-webui.

Goetschalckx, L., Andonian, A., Oliva, A., & Isola, P. (2020). GANalyze: Toward visual definitions of cognitive image properties. *Journal of Vision*, **20**, 297–297. https://doi.org/10.1167/jov.20.11.297.

Goldschmidt, G. (2004). Design representation: Private process, public image. In *Design representation* (pp. 203–217). Springer. https://doi.org/10.1007/978-1-85233-863-3_9.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (Vol. **27**, pp. 2672–2680). Neural Information Processing Systems Foundation.

Google. (n.d.). Street-View API: Google maps. https://maps.google.com/.

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

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of Wasserstein GANs. In *Advances in neural information processing systems* (Vol. **30**, pp. 5767–5777). Neural Information Processing Systems Foundation.

Han, M., Baek, K., Lee, K-T., Ko, S.,, & Kim, J-H. (2021). A Study on Supporting Design Decision Making on Office Building Remodeling Projects by Introducing Mixed Reality. *Korean Journal of Construction Engineering and Management* **22**, 3–12. https://doi.org/10.6106/KJCEM.2021.22.1.003.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In *Advances in neural information processing systems* (Vol. **33**, pp. 6840–6851).Neural Information Processing Systems Foundation.

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. preprint (arXiv:2106.09685). https://arxiv.org/abs/2106.09685.

Ibrahim, R., & Rahimian, F. P. (2010). Comparison of CAD and manual sketching tools for teaching architectural design. *Automation in Construction*, **19**, 978–987. https://doi.org/10.1016/j.autcon.2010.09.003.

Instagram. (2022). https://www.instagram.com/monomosite.

Instagram. (2022). https://www.instagram.com/surrealistly/?hl=en.

Jacobs, A. B. (1993). *Great streets*. MIT Press.

Jacobs, J. (1961). *The death and life of great American cities*. Random House.

Jonson, B. (2005). Design ideation: The conceptual sketch in the digital age. *Design Studies*, **26**, 613–624. https://doi.org/10.1016/j.destud.2005.03.001.

Kalay, Y. E. (2004). *Architecture's new media: Principles, theories, and methods of computer-aided design*. MIT Press.

Kaplan, S. (1992). Environmental preference in a knowledge-seeking, knowledge-using organism. In *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 581–600). Oxford Academic. https://doi.org/10.1093/oso/9780195060232.003.0017.

Karan, E., Asgari, S., & Rashidi, A. (2021). A markov decision process workflow for automating interior design. *KSCE Journal of Civil Engineering*, **25**, 3199–3212. https://doi.org/10.1007/s12205-021-1272-6.

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of GANs for improved quality, stability, and variation. preprint (arXiv:1710.10196). http://arxiv.org/abs/1710.10196.

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4396–4405). IEEE. https://doi.org/10.1109/cvpr.2019.00453.

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 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8107–8116). IEEE. https://doi.org/10.1109/cvpr42600.2020.00813.

Kim, H., Lee, J. K., Shin, J., & Choi, J. (2018). Visual language approach to representing KBimCode-based Korea building code sentences for automated rule checking. *Journal of Computational Design and Engineering*, **6**, 143–148. https://doi.org/10.1016/j.jcde.2018.08.002.

Kim, J., Kim, G., Son, J., & Lim, H. (2023). Prompt language learner with trigger generation for dialogue relation extraction. *Applied Sciences*, **13**, 12414. https://doi.org/10.3390/app132212414.

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

Kim, S., & Lee, H. (2016). A case study on the gentrification process and characteristics of Seongsu-dong. *The Journal of Culture Contents*, **7**, 81–105.

Kim, Y., & Lee, J. K. (2022). Processing of 360 panoramic images for architectural interior image training archive. In *Proceedings of the 22nd International Conference on Construction Applications of Virtual Reality (CONVR 2022)* (pp. 886–891).

Kingma, D. P., & Welling, M. (2022). Auto-encoding variational Bayes. preprint (arXiv:1312.6114v11). https://arxiv.org/abs/1312.6114.

Kotler, P., Haider., D. H., & Rein., I. J. (2008). *Marketing places: Attracting investment, industry, and tourism to cities, states, and nations*. Free Press.

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

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, **60**, 84–90. https://doi.org/10.1145/3065386.

Lang, J. (1987). *Creating architectural theory: The role of the behavioral sciences in environmental design*. Van Nostrand Reinhold Co.

Lawrence, R. J. (1993). Architectural design tools: Simulation, communication and negotiation. *Design Studies*, **14**, 299–313. https://doi.org/10.1016/0142-694X(93)80026-9.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, **521**, 436–444. https://doi.org/10.1038/nature14539.

Lee, J. K., Eastman, C. M., Lee, J., Kannala, M., & Jeong, Y. S. (2010). Computing walking distances within buildings using the universal circulation network. *Environment and Planning B: Planning and Design*, **37**, 628–645. https://doi.org/10.1068/b35124.

Lee, J. K., & Kim, M. J. (2014). BIM-enabled conceptual modelling and representation of building circulation. *International Journal of Advanced Robotic Systems*, **11**, 127. https://doi.org/10.5772/58440.

Lee, J. K., Lee, J., Jeong, Y. S., Sheward, H., Sanguinetti, P., Abdelmohsen, S., & Eastman, C. M. (2012). Development of space database for automated building design review systems. *Automation in Construction*, **24**, 203–212. https://doi.org/10.1016/j.autcon.2012.03.002.

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

Lee, J., & Park, S. (2023). Current design guidelines' streetscape improvement for visual perception and walkability: A case study of Sejong City, Republic of Korea. *Frontiers of Architectural Research*, **12**, 423–443. https://doi.org/10.1016/j.foar.2022.11.002.

Lever, P. D. (1973). A photomontage system for site planning. *Computer-Aided Design*, **5**, 103–104. https://doi.org/10.1016/0010-4485(73)90005-5.

Lewis, R., & Séquin, C. (1998). Generation of 3D building models from 2D architectural plans. *Computer-Aided Design*. **30**, 765–779. https://doi.org/10.1016/S0010-4485(98)00031-1.

Li, J., Li, D., Xiong, C., & Hoi, S. (2022). BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *Proceedings of the 39th International Conference on Machine Learning* (Vol. **162**, pp. 12888–12900). PMLR.

Liu, D., Ouyang, H., Li, S., Zhang, C., & Zhan, Z.-H. (2023). Hyperparameters optimization of convolutional neural network based on local autonomous competition harmony search algorithm. *Journal of Computational Design and Engineering*, **10**, 1280–1297. https://doi.org/10.1093/jcde/qwad050.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 9992–10002). IEEE. https://doi.org/10.1109/iccv48922.2021.00986.

Liu, V., & Chilton, L. (2022). Design Guidelines for Prompt Engineering Text-to-Image Generative Models. *In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*,1–23. https://doi.org/10.1145/3491102.3501825.

Midjourney. (n.d.). https://www.midjourney.com/.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, **77**, 1321–1329.

Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. preprint (arXiv:1802.05957). https://arxiv.org/abs/1802.05957.

Nahm, K., & Jang, W. (2016). Small is beautiful: Elements of the regional industrial ecosystems of Seongsu Hand-made Shoe Industries, Seoul. Korea. *The Geographical Journal of Korea*, **50**, 197–210.

Nilsback, M.-E., & Zisserman, A. (2008). Automated flower classification over a large number of classes. In *Proceedings of the 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing* (pp. 722–729). IEEE. https://doi.org/10.1109/icvgip.2008.47.

OpenAI. (2022). Introducing ChatGPT. ChatGPT: Optimizing language models for dialogue. https://openai.com/blog/chatgpt/.

OpenArt. (n.d.). AI image generator that elevates and realizes your creative vision. https://openart.ai/discovery.

Oppenlaender, J., Silvennoinen, J., Paananen, V., & Visuri, A. (2023). Perceptions and realities of text-to-image generation. In *Proceedings of the 26th International Academic Mindtrek Conference* (pp. 279–288). https://doi.org/10.1145/3616961.3616978.

Oxman, R. (2006). Theory and design in the first digital age. *Design Studies*, **27**, 229–265. https://doi.org/10.1016/j.destud.2005.11.002.

Park, J., & Yang, W. (1997). Some distributional characteristics and planning strategies for the improvement of neighborhood-convenience facilities in urban residential areas. *Journal of the Architectural Institute of Korea*, **17**, 223–227.

Park, J., Chung, S., Lee, S., Song, J., & Kwon, S. (2021). Research on the Construction of an Automation Model for Maintenance Managers Based on Smart Devices. *Korean Journal of Construction Engineering and Management* **22**, 72–80. https://doi.org/10.6106/KJCEM.2021.22.1.072.

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Parkhi, O. M., Vedaldi, A., Zisserman, A., & Jawahar, C. V. (2012). Cats and dogs. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3498–3350). IEEE.

Prieto, A., & Oldenhave, M. (2021). What makes a façade beautiful? Architects' perspectives on the main aspects that inform aesthetic preferences in façade design. *Journal of Facade Design and Engineering*, **9**, 21–46. https://doi.org/10.7480/jfde.2021.2.5540.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning* (Vol. **139**, pp. 8748–8763).

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. preprint (arXiv:1511.06434). http://arxiv.org/abs/1511.06434.

Rafiq, Y., Packham, I., & Beck, M. (2005). Importance of visualization in decision making process in engineering design. In *Computing in civil engineering* (pp. 1–12). American Society of Civil Engineers (ASCE). https://doi.org/10.1061/40794(179)63.

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

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., & Sutskever, I. (2021). Zero-shot text-to-image generation. In *Proceedings of the 37th International Conference on Machine Learning* (Vol. **139**, pp. 8821–8831).

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779–788). https://doi.org/10.1109/cvpr.2016.91.

Relph, E. C. (1976). *Place and placelessness*. Pion.

Roboflow. (n.d.). Give your software the power to see objects in images and video. https://roboflow.com/.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10674–10685). https://doi.org/10.1109/cvpr52688.2022.01042.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Gontijo-Lopes, R., Salimans, T., Ho, J., Fleet, D., & Norouzi, M. (2022). Photorealistic text-to-image diffusion models with deep language understanding. In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022)*.

Sanguinetti, P., Abdelmohsen, S., Lee, J., Lee, J., Sheward, H., & Eastman, C. (2012). General system architecture for BIM: An integrated approach for design and analysis. *Advanced Engineering Informatics*, **26**, 317–333. https://doi.org/10.1016/j.aei.2011.12.001.

Shin, J., & Lee, J. K. (2019). Indoor walkability index: BIM-enabled approach to quantifying building circulation. *Automation in Construction*, **106**, 102845. https://doi.org/10.1016/j.autcon.2019.102845.

Shin, M., Lee, S. I., Lee, H., & Lee, J. K. (2014). Sensing the room: An integrated implementation process to visualize indoor temperature data on floor plans. *Visualization in Engineering*, **2**, 1–10. https://doi.org/10.1186/s40327-014-0010-2.

Song, J., Lee, J. K., Choi, J., & Kim, I. (2020). Deep learning-based extraction of predicate-argument structure (PAS) in building design rule sentences. *Journal of Computational Design and Engineering*, **7**, 563–576. https://doi.org/10.1093/jcde/qwaa046.

Song, J., Meng, C., & Ermon, S. (2022). Denoising diffusion implicit models. preprint (arXiv:2010.02502). https://arxiv.org/abs/2010.02502.

Sönmez, N. O. (2018). A review of the use of examples for automating architectural design tasks. *Computer-Aided Design*, **96**, 13–30. https://doi.org/10.1016/j.cad.2017.10.005.

Street View Download 360. (n.d.). https://svd360.istreetview.com/.

Sutherland, I. E. (1964). Sketch pad a man-machine graphical communication system. In *Proceedings of the SHARE Design Automation Workshop* (pp. 6.329–6.346). https://doi.org/10.1145/800265.810742.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2818–2826). https://doi.org/10.1109/cvpr.2016.308.

Tandon, M., & Sehgal, V. (2017). Traditional Indian religious streets: A spatial study of the streets of Mathura. *Frontiers of Architectural Research*, **6**, 469–479. https://doi.org/10.1016/j.foar.2017.10.001.

Ulrich, R. S. (1983). Aesthetic and affective response to natural environment. In *Behavior and the natural environment. Human behavior and environment* (Vol. **6**, pp. 85–125). Springer. https://doi.org/10.1007/978-1-4613-3539-9_4.

Valipour, M., Rezagholizadeh, M., Kobyzev, I., & Ghodsi, A. (2023). DyLoRA: Parameter efficient tuning of pre-trained models using dynamic search-free low-rank adaptation. preprint (arXiv:2210.07558). https://arxiv.org/abs/2210.07558.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (Vol. **30**, pp. 5998–6008). Neural Information Processing Systems Foundation. https://papers.nips.cc/paper/7181-attention-is-all-you-need.

Weber, R. E., Mueller, C., & Reinhart, C. (2022). Automated floorplan generation in architectural design: A review of methods and applications. *Automation in Construction*, **140**, 104385. https://doi.org/10.1016/j.autcon.2022.104385.

Whyte, J., Bouchlaghem, N., Thorpe, A., & McCaffer, R. (2000). From CAD to virtual reality: Modelling approaches, data exchange and interactive 3D building design tools. *Automation in Construction*, **10**, 43–55. https://doi.org/10.1016/S0926-5805(99)00012-6.

Xia, X., Xu, C., & Nan, B. (2017). Inception-v3 for flower classification. In *Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC)* (pp. 783–787). https://doi.org/10.1109/ICIVC.2017.7984661.

Yigitbas, E., Nowosad, A., & Engels, G. (2023). Supporting construction and architectural visualization through BIM and AR/VR: A systematic literature review. In *Human-computer interaction – INTERACT 2023. INTERACT 2023. Lecture notes in computer science* (pp. 145–166). Springer. https://doi.org/10.1007/978-3-031-42283-6_8.

Yu, E., & Choi, J. (2023). Development of building information modeling-based automation assessment process for universal design of public buildings. *Journal of Computational Design and Engineering*, **10**, 641–654. https://doi.org/10.1093/jcde/qwad018.

Zeng, Y., & Lee, K. (2023). The expressive power of low-rank adaptation. preprint (arXiv:2310.17513). https://arxiv.org/abs/2310.17513.

Zhang, J., Fukuda, T., & Yabuki, N. (2022). Automatic generation of synthetic datasets from a city digital twin for use in the instance segmentation of building facades. *Journal of Computational Design and Engineering*, **9**, 1737–1755. https://doi.org/10.1093/jcde/qwac086.

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Zhang, J., & Hu, A. (2022). Analyzing green view index and green view index best path using Google street view and deep learning. *Journal of Computational Design and Engineering*, **9**, 2010–2023. https://doi.org/10.1093/jcde/qwac102.

Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 3813–3824). https://doi.org/10.1109/iccv51070.2023.00355.

Zheng, H., & Yuan, P. F. (2021). A generative architectural and urban design method through artificial neural networks. *Building and Environment*, **205**, 108178. https://doi.org/10.1016/j.buildenv.2021.108178.

Downloaded from https://academic.oup.com/jcde/article/11/2/85/7609688 by guest on 18 October 2024

Received: September 26, 2023. **Revised:** February 9, 2024. **Accepted:** February 12, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of the Society for Computational Design and Engineering. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com