

Generative AI-powered architectural exterior conceptual design based on the design intent

Mengnan Shi¹, JoonOh Seo^{1,*}, Seung Hyun Cha², Bo Xiao³ and Hung-Lin Chi¹

¹Department of Building and Real Estate, The Hong Kong Polytechnic University, Hung Hom, Kowloon, 999077, Hong Kong

²Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon, 34141, Republic of Korea

³Department of Civil, Environmental, and Geospatial Engineering, Michigan Technological University, Houghton, MI 49931, USA

*Correspondence: joonoh.seo@polyu.edu.hk

Abstract

In the architectural exterior design domain, design intent is usually expressed by textual design intent [e.g., client needs, architectural language (AL)] and non-verbal design intent (e.g., sketch). However, existing generative AI-based methods for automated architectural exterior conceptual design can only use the general image description as the prompt. Thus, despite its potential, existing generative image AI cannot produce appropriate design alternatives that meet various design requirements. Enabling automated architectural exterior conceptual design requires solving two problems: teaching the AI model to understand textual design intent and allowing generative AI to combine textual design intent with non-verbal design intent. The study aims to propose an automated architectural exterior conceptual design approach by incorporating domain-specific prompting strategies and sketch-to-image synthesis into fine-tuned generative image AI models. In the proposed approach, textual design intent annotations (including client needs and AL) are added to architectural images and general image description annotations. Web crawler and ChatGPT automatically extract design intent-related annotations from online sources for famous architectural works that are used as training images. The constructed dataset is then used to fine-tune a generative AI model [i.e., Stable Diffusion (SD)] via the Lora algorithm, teaching the AI model to understand textual design intent. Also, ControlNet is used to control the generation process of the SD model to enable the generative AI to reflect the design intent expressed by the sketches. The proposed approach is validated by comparing generated images from our approach with those from two existing models. The results show that the proposed method can successfully generate architectural exterior conceptual design images that fulfil the requirements based on the architectural design intent. The proposed approach is expected to streamline and facilitate time-consuming and demanding iterative processes during a conceptual design phase.

Keywords: architectural exterior conceptual design, design intent, generative AI, Stable Diffusion

Nomenclature

| | | | |
|-------------------------------------|--|--|---|
| x : | Input image | $R(\theta)$: | Regularization term |
| ε : | Noise vector | λ_{reg} : | Regularization coefficient |
| α : | Coefficient determining the noise level | c : | Control vector |
| $L_{adv}(\theta)$: | Adversarial loss | $f(attributes)$: | Encoding function that maps the sample's attributes or features to the control vector |
| $L_{stab}(\theta)$: | Stability loss | $G(z, c)$: | Generative model with control vector c and input noise z |
| λ : | Weighting coefficient that balances the two losses | \mathcal{L} : | Loss function for ControlNet training |
| $CNN_{pre-trained}$: | Pre-trained deep convolutional neural network | $x_{controlled}$: | Controlled sample generated by the model |
| $\theta_{large-scale}$: | Parameters of the pre-trained model on a large-scale dataset | $\theta_{previous}$: | Previous parameters of the model |
| θ_{frozen} : | Parameters of the frozen layers | θ : | Model parameters |
| $freeze(\theta_{1...k})$: | Function to freeze the parameters of the first k layers | $J(\theta; x, y)$: | Loss function |
| W_{new} : | Weights of the new fully connected layer | $\mathcal{L}(G(z, c; \theta), target)$: | Loss function that measures the difference between generated samples and target |
| b_{new} : | Biases of the new fully connected layer | $\theta_{1...k}$: | Parameters of the first k layers |
| $\theta_{updated}$: | Updated parameters of the model | $Output_{new}$: | Output of the new fully connected layer |
| η : | Learning rate | | |
| $\nabla_{\theta} J(\theta; x, y)$: | Gradient of the loss function J concerning the parameters θ | | |
| $J_{reg}(\theta; x, y)$: | Regularized loss function | | |

1. Introduction

Conceptual design is a creative process, usually in the early stages of architectural design. At this stage, the architect and client explore and develop essential ideas and concepts about the

Received: May 23, 2024. Revised: August 14, 2024. Accepted: August 14, 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

architectural project, including themes, design styles, and basic structures (Xia et al., 2008; Castro Pena et al., 2021). The conceptual design provides the basis for the project's overall direction and helps ensure that subsequent design phases move in the right direction (Pourzolfaghar et al., 2014). In particular, one of the crucial goals during the early conceptual design stage is to determine the visual appearance of a building, such as the building's shapes and external design, that meet both functional and aesthetic requirements (Castro Pena et al., 2021). Traditional architectural conceptual design is a communication process between clients and architects, where clients provide for their needs. Then, the architects propose various design solutions to meet clients' needs and design intent based on their professional knowledge through a series of meetings and discussions with the client. During this iterative process, various forms of visual representation are commonly used for the client and the architect to agree on the outcomes, ranging from sketches to modelling. However, conceptual design generation is time-consuming and mentally demanding, and the design quality is subject to human intervention (Qiu et al., 2002).

To support this complex design process during the architectural exterior conceptual design, several automated architectural methods based on artificial intelligence (AI) have been proposed, including those based on fractals (Joye, 2011), swarm intelligence (Sengupta & Mishra, 2014), and meta-cellular automata (Coates et al., 1996). For example, Anzalone & Clarke (2003) developed an architectural exterior conceptual design tool called CAAD using the principles of meta-cellular automata. The tool automatically generates architectural solutions with meta-cellular automata morphology that are evaluated and selected according to the user's needs and preferences. However, these approaches focus on design optimization given architectural design needs, and they have difficulty generating realistic architectural exterior conceptual design drawings. Generative image AI empowered by large-scale foundation models is recently gaining attention in architectural exterior conceptual design. Generative AI is at the forefront of current research in AI. Contrasting with traditional AI, which is typically focused on tasks like data analysis and prediction, the generative image AI [e.g., DALL-E (Li et al., 2023), Stable Diffusion (SD; Dehouche & Dehouche, 2023)] has demonstrated its ability to create detail-rich new content by understanding and replicating patterns learned from data in the areas of natural language processing and image generation, and has developed applications in areas such as drug discovery and music composition. Recently, studies have explored the feasibility of using generative AI for architectural design (Chen et al., 2023). For example, in the commercial sector, Mnml.ai (Mnml.ai, 2024, July 24) can generate realistic renderings of buildings based on user-input image description prompts.

Despite the success of these approaches, they still present challenges in understanding design intent and handling complex design requirements. In the domain of architectural design, design intent is expressed through specific client needs and architectural design language that reflects an architect's design concept (Krügelis, 2018; Song et al., 2020; Chen & Kitagawa, 2023). However, it is still questionable whether existing generative AI tools can understand these architectural contexts to create new designs, as they only use general image description texts to train existing building images. Thus, they could not understand the design intent implicitly represented in building design. Also, another challenge is that most generative image AI tools (i.e., text-to-image models) rely on prompts as a means of interaction between designers and AI models. Thus, reflecting the architect's design intent is impossible and is often difficult to formulate in words.

In response to the above challenges, this study aims to propose a framework for a generative AI-powered automated architectural exterior conceptual design approach by incorporating domain-specific prompting strategies and sketch-to-image synthesis into fine-tuned generative image AI models. The proposed framework is demonstrated using famous architectural works by well-known architects. First, this study creates domain-specific building image datasets from famous architectural works by textual design intent annotation (i.e., client needs and architectural design languages) to the traditional image descriptions annotation. To streamline the time-consuming procedure of creating datasets, this study uses web crawlers to collect images of famous architectural works and a ChatGPT interface to automatically extract corresponding textual annotations representing each architectural work's design intents from web-based articles. The constructed dataset is then used to fine-tune a generative AI model (i.e., SD) via the Low-Rank Adaptation (Lora) algorithm, teaching the generative AI to understand textual design intents from an architectural exterior conceptual design perspective. Also, to reflect the design intents represented in non-verbal forms such as sketches when creating new building design images, ControlNet is used to control the generation process of the SD model. The proposed framework is qualitatively validated by comparing generated design images from our approach with those from two existing models [i.e., Mnml.ai (Mnml.ai, 2024, July 24) and Architectural Schoolteacher (Technology, 2024, July 24)]. It is expected that the proposed approach could accelerate an iterative architectural exterior conceptual design process by quickly creating and visualizing various conceptual design alternatives based on both the client's needs and an architect's ideas and creative concepts.

2. Literature Review of AI-Powered Architectural Exterior Conceptual Design

Architectural exterior conceptual design methodology has evolved through stages of manual, automated, and intelligent development. The traditional manual design process is time-consuming and labour-intensive, and human factors affect the quality of the design (Pérez, 2017). Automation of architectural exterior conceptual design can help improve design efficiency and quality and is of great research significance (Dutta & Sarthak, 2011).

The continuous evolution and enhancement of computer modelling software have increasingly enabled studies to leverage these tools for the semi-automated conceptual design of architectural exteriors. Utilizing such software facilitates designers in swiftly generating and revising models of architectural appearances, thus expediting the conceptual design phase and yielding superior outcomes with ease. For instance, designers can employ software like 3ds Max (Baltus & Žebrauskas, 2019) and Revit (Mora et al., 2008) for modelling and image rendering, creating lifelike visualizations of architectural exteriors. Nonetheless, it is essential to note that these methods still require considerable manual work from architectural designers.

Conceptual design methods for architectural exteriors, grounded in AI, represent a burgeoning approach that incorporates AI techniques like fractal geometry, swarm intelligence, and deep learning for devising exterior design solutions. These advanced methods aid designers more effectively comprehend user needs and design limitations, enabling the creation of highly customized design solutions tailored to specific requirements. For example, Wen et al. (2010) developed an architectural exterior conceptual design tool called Fractal Architect using the

principles of fractal geometry. The tool automatically generates architecture solutions with fractal forms, which are evaluated and selected according to the user's needs and preferences. Sharafi *et al.* (2015) developed an ant colony algorithm-based architectural exterior conceptual design method for optimizing the energy consumption of an architecture. The method automatically generates various architectural scenarios evaluated and selected based on multiple design objectives, such as energy consumption and comfort. Rapone & Saro (2012) developed an architectural exterior conceptual design method using a particle swarm algorithm to optimize the design of the curtain wall façade of an office. Zhao (2021) combined parametric modelling, a building performance simulation engine and an optimization algorithm to propose an optimal design optimization method based on building design objectives. The method can generate optimal window and wall proportions in less than 2 s. This compares to about two weeks architects spend using traditional simulation engine-based methods. Yi & Kim (2022) proposed a multi-objective optimization method for architectural design based on swarm intelligence algorithms to meet multiple design requirements simultaneously. Coates *et al.* (1996) developed a conceptual design tool for architecture called Cellular Automata Designer using the principles of multi-state meta-cellular automata. The tool automatically generates architectural solutions with multi-state cellular automata morphology. It evaluates and selects them based on multiple design goals, such as the architecture's aesthetics, structure and function. Kakooee & Dillenburg (2024) exploit the potential of deep reinforcement learning algorithms to optimize the spatial design of buildings. Experiments show that the method can automatically explore a broader range of design options, thus facilitating the discovery of innovative solutions. Gan *et al.* (2024) proposed a building design method integrating Generative Adversarial Networks (GANs) and Multi-Objective Optimization algorithms, which can generate spatial design solutions for multiple buildings in less than 5 min. Chang *et al.* (2020) proposed a deep learning and Electroencephalography (EEG) signal-based approach for building design image preference recognition. The method can support selecting building appearance design solutions by considering user needs.

Recently, the generative AI is gaining significant attention. For example, GANs and variants of transformer models have demonstrated significant capabilities in creating realistic and coherent output (Goodfellow *et al.*, 2014). DALL-E (Li *et al.*, 2023) shows the ability to create novel, high-quality images based on textual descriptions. In addition, generative models have been used in drug discovery, music composition, and video game design, demonstrating various applications in innovation and automated creation processes (Elgammal *et al.*, 2017). Generative AI models have been extended to create conceptual architectural designs. For example, in academia, Chen *et al.* (2023) proposed a method for generating architectural designs using AI that can batch-generate high-quality architectural designs based on prompts. This method can improve the efficiency and quality of architectural design and optimize the workflow of architectural design. Jo *et al.* (2024) applied generative AI to the design of building facades, capable of generating various design solutions that resonate with the character of local buildings. In addition, several architectural design assistance programs claim to be commercially available based on generative AI. This software can assist in the generation of conceptual design plans for buildings to a certain extent. For example, in the commercial sector, Veras (EvolveLAB, 2024, July 24) can render images based on user-input sketches by selecting architec-

tural styles; Architectural schoolteacher (Technology, 2024, July 24) can generate high-quality architectural exterior conceptual designs based on prompts by rendering input images in specified styles.

Generative AI shows potential in architectural exterior conceptual design. However, existing methods still present challenges in understanding both textualized architectural design intent [client needs and architectural language (AL)] and non-textualized design intent (sketches) simultaneously.

3. Methodology

This study proposes a generative AI-powered automated architectural exterior conceptual design approach that can understand and reflect architectural domain-specific inputs, including textual and non-textual design intents for developing architectural exterior conceptual design. The overall framework of the proposed method is shown in Fig. 1. The framework consists of (i) a collection of datasets for training the generative AI model, (ii) fine-tuning the generative AI model, and (iii) generating conceptual designs with a controlled fine-tuned generative AI model. Specifically, first, this study creates a domain-specific architectural image dataset from famous architectural works through textual design intent annotations (i.e., client needs and architectural design language) and traditional image description annotations. In particular, this study uses a web crawler to collect images of famous architectural works to simplify the time-consuming dataset creation process. It uses ChatGPT to automatically extract corresponding textual annotations representing the design intent of each architectural work from the web articles. Then, using the constructed dataset, the generative AI model (i.e., the SD model) is fine-tuned by the Lora algorithm to teach the generative AI to understand textual design intent from a perspective of architectural exterior conceptual design. In addition, ControlNet is used to control the generation process of the stabilized diffusion model to reflect the design intentions for shape and form required in architectural mass modelling, expressed in non-verbal forms such as sketches. The proposed method uses textual descriptions and sketches as inputs. It is not constrained by parameters (e.g., specific heights and numbers of floors, which are more critical in the mid-and late-stage of architectural design) to ensure that a diverse range of creative exterior design solutions can be provided to the architect in the early architectural design stage (conceptual design stage).

3.1 Defining architectural design intent for model inputs

Architectural design intent refers to the goals and concepts pursued in the architectural design process, which usually involves several aspects of the architecture's function, form, space, materials, structure, environment, and culture (Krügelis, 2018). The expression of architectural design intent can take many forms, including text and non-verbal (e.g., images; Chen & Kitagawa, 2023). The process usually begins with the client's design requirements during the conceptual design phase of an architecture, as shown in Fig. 2. The architect further uses this to clarify the design intent and communicates and confirms this with the client through sketches from the architecture client's presentation. The final conceptual design proposal will synthesize and respond to the results of all the above information and communication. Therefore, in the conceptual design phase, the client's requirements, the

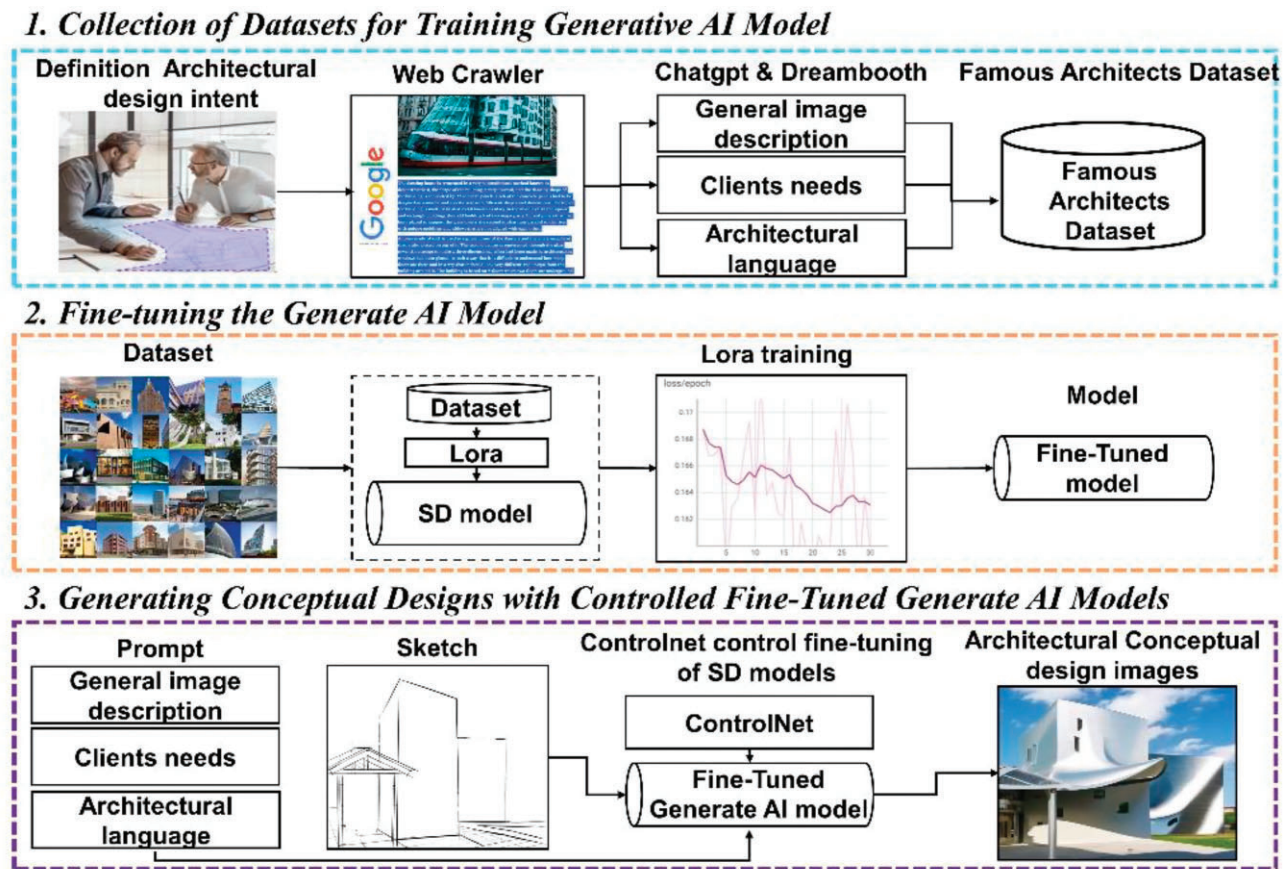


Figure 1: Research framework.

architectural design language and the architectural design sketches are the primary means of expressing the architectural design intent.

3.1.1 Architectural client needs

Architectural client needs are the specific requirements and expectations of the client for the architectural design, functionality and construction of a building project. These needs typically include building functionality, budget, schedule, style and aesthetics, sustainability and environmental protection, and codes and standards (Wikberg *et al.*, 2014). These requirements are central considerations in the architectural design and planning process, and they can influence the success of a project. During the conceptual design phase of an architectural project, the client needs to work with the architect and other stakeholders to clearly define these essential requirements and objectives (Thyssen *et al.*, 2010). Table 1 shows the main items of architectural client requirements and corresponding descriptions for 'Dancing House' as examples.

3.1.2 AL

AL is the set of terms and principles to describe and understand architectural design and expression. This includes elements and concepts from various aspects, such as form, function, materials, technology, and culture (Eilouti, 2018). AL can help people better understand and evaluate architectural works and is an essential tool for architects and designers to communicate design ideas and concepts. AL can communicate design intent, explain the creation of architectural forms and spaces, and how architecture responds

to cultural, social, and environmental needs. It is also important to note that different architects and designers may have unique AL that reflects their design philosophy and creative style. Table 2 shows the main items of AL and corresponding descriptions for 'Dancing House' as examples.

3.1.3 Architectural design sketches

Adopting appropriate and easy-to-understand visualization techniques can communicate the design intent more clearly. Architectural design sketches, as preliminary hand drawings or drawings that express the designer's creativity and ideas, play an essential role in architectural exterior conceptual design (Chen *et al.*, 2008). In terms of the form of the composition, it usually includes floor plans, elevations and sections. Regarding the level of compositional detail of sketches, they include rough and fine sketches. In addition, sketches are usually simple, quick, and free from strict proportions and specifications.

These sketches aid designers, clients, and other stakeholders in visualizing and comprehending the fundamental concepts and layouts of architectural appearance design. They are also helpful in designing the shape and form required for architectural mass modelling.

3.2 Automated training data collection by web crawlers, ChatGPT, and Dreambooth

The proposed data acquisition process is shown in Fig. 3. First, images of famous architectural works and related articles that describe these works are obtained from web pages through



Figure 2: Examples of architectural design intent for ‘Dancing House’ [Dancing House (Hartoonian, 2010), also known as Fred and Ginger, is a modern building located in Prague, Czech Republic].

Table 1: specific items and examples of construction client needs.

| Items | Descriptive | Examples of ‘Dancing House’ |
|--|---|---|
| Functional needs | How the building should meet its users’ basic functional and activity needs. | As a mixed-use building containing office space, a restaurant and a gallery, its functional needs include the provision of suitable office environments, dining space, and art display space. |
| Aesthetic and style needs | The appearance, style, and artistic expression that the building should have. | The building, consisting of static and dynamic parts, became a cultural center, symbolizing the transition of Czechoslovakia from a communist regime to a parliamentary democracy. |
| Technological and innovation needs | Consider the innovation and application of building technology. | Many innovative building techniques and materials were used to realize its unique form and structure. |
| Sustainability and environmental needs | Building design should consider environmental impacts, including energy efficiency, material selection, and eco-friendliness. | While the ‘Dancing House’ does not explicitly emphasize sustainable design, new building designs often consider these factors. |

Table 2: Specific items of AL and examples.

| Items | Descriptive | Examples of ‘Dancing House’ |
|------------------------|---|---|
| Form and space | The basic shape, structure, and spatial layout of a building. | Shaped like two dancing men, deconstructionist architecture creates a dynamic and flowing space through curved and irregular forms. A vast twisted metal structure tops the building. |
| Materials and textures | Types, characteristics, and finishes of materials used for construction. | The façade uses mainly glass and concrete, creating a modern and industrial sense of materials. |
| Colour and tone | The effect of light and shadow created by a natural or artificial light source. | The window and façade design allow natural light to enter the interior space fully, creating a rich light and shadow effect. |
| Proportion and scale | Size, proportion, and relationship of architectural elements and spaces. | There are two main sections. The first is a glass tower, reduced in height by half and supported by curved columns; the second runs parallel to the river and is characterized by undulating forms and unaligned windows. |

Articles from Google Search

The Dancing House-A Unique Architectural Design

The Dancing House or also known as Fred and Ginger is in the city of Prague in the Czech Republic. It is the building for the second half of the 20th century. The building has been designed by two architects from different regions: Frank Gehry, from the United States and the other was an American Canadian architect Frank Gehry. The design of the building is so different and unique that one is not able to understand whether it is a sculpture or a building. The building was designed in 1985 and it took five years to complete the construction. The design of the building was so different and unique that one is not able to understand whether it is a sculpture or a building. The building was designed in 1985 and it took five years to complete the construction. The design of the building was so different and unique that one is not able to understand whether it is a sculpture or a building. The building was designed in 1985 and it took five years to complete the construction.

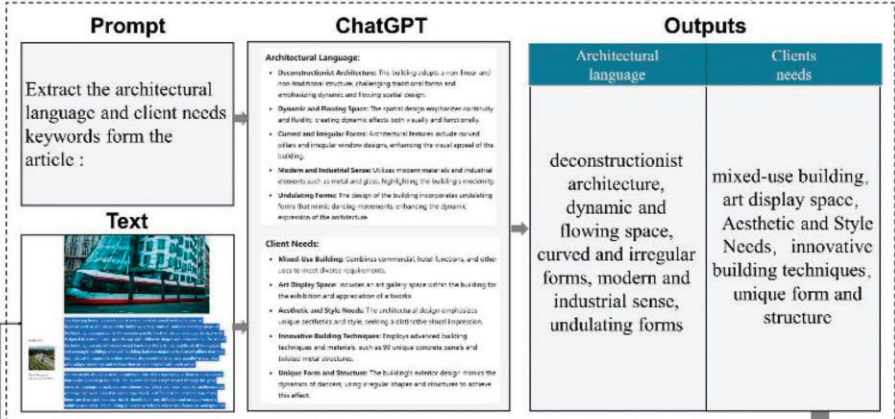
History:

The location of the dancing house carries great importance as it was built on the site where a house was destroyed by the US bombing in Prague during 1945. Around 15 years the damaged property was not given importance and then a famous architect presented the idea to the next owner of the land that what he had in mind. He was a neighbor to this land area as well and he came into some power in the later years. His influence and connections helped the architect give shape to his idea with all the support. This man was both known as both Fred. He was a well-known leader and later became the president of Czechoslovakia. Due to the history of the land the Dutch insurance company had agreed to sponsor the building required for the building. The sponsorship required naming up with a famous architect from another region, a few architects declined the offer as the land area was very small but the American Canadian architect took up the challenge.

Structure of the Dancing House:

The dancing house is considered as a very controversial method known as deconstruction, the shape of the building is very unusual and the dancing shape of

Generate client needs and architectural language tags



Generate general image distribution tags

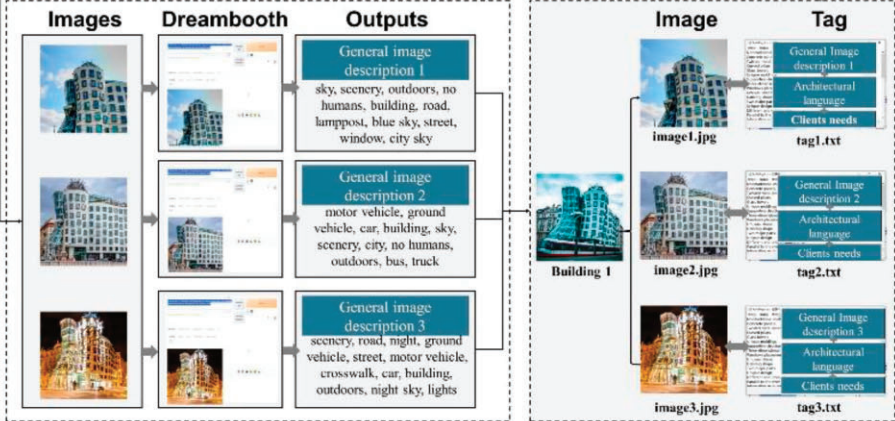


Figure 3: Data collection flow.

keyword searches. Second, the prompt (“Extract the AL and client needs keywords from the article”) and the web text are entered into ChatGPT to obtain the above-defined client needs and AL keywords. After that, the images are input into Dreambooth (a method for automatically extracting text descriptions from images; Ruiz et al., 2023) to obtain the regular image descriptions of the images. Finally, the regular image description is combined with extracted keywords representing client needs and architectural design language as annotations for

the images. To improve the efficiency of data collection, the above process uses a web crawler to automatically retrieve web pages and a ChatGPT interface to automate Q&A, thus realizing the automation of dataset construction. Retrieved images and corresponding annotations (general image descriptions, client needs and AL) were manually verified for model training. Given the uncertainty of ChatGPT’s responses each time, in order to ensure the quality of the dataset, the retrieved images and the corresponding annotations (general image descriptions,

client needs and AI) were manually checked for model training.

In general, existing methods for training an SD model must be paired with image-text data. Instead of using single paired data, the proposed approach utilizes each architectural work as a fundamental training unit. For example, multiple images of each building captured from various angles and against diverse backgrounds are compiled for each architectural work. As a result, these images from each architectural work collectively share the same text annotations, enabling them to reflect image variations according to diverse views and backgrounds.

3.3 Fine-tuning generative AI to learn textual design intent via Lora

In order to enable a generative AI to generate architectural images according to the textual design intent, SD (Ni et al., 2023), a deep learning model related to GANs (Hitaj et al., 2017) is adopted and further fine-tuned using the training data collected from the previous step. The SD model is among the most advanced generative AI models available today, is open-source, receives regular updates, and is user-friendly for researchers. The key idea of the SD model is to introduce stability enhancement mechanisms to improve the stability of GANs during training and the quality of generated samples. Traditional GANs can face problems such as mode collapse during training, resulting in a lack of diversity in the samples generated. SD helps to address this by introducing noise into the generation process and through a series of stability enhancement techniques.

Specifically, SD uses progressive noise injection to gradually increase the strength of the noise to help the generator better explore the sample space. The noise injection can be represented as

$$x' = x + \alpha \cdot \varepsilon \quad (1)$$

where x is the input image, ε is a noise vector, and α is a coefficient determining the noise level.

In addition, it employs a stability-enhancing loss function that helps improve the model's training stability. This loss function can be formulated as

$$L(\theta) = L_{adv}(\theta) + \lambda \cdot L_{stab}(\theta) \quad (2)$$

where $L_{adv}(\theta)$ is the adversarial loss, $L_{stab}(\theta)$ is the stability loss, and λ is a weighting coefficient that balances the two.

These improvements allow SD to generate more diverse, high-quality images that are more stable and controllable compared with traditional GANs. To further fine-tune the SD model, Lora (Hu et al., 2021) is applied to enhance domain knowledge. The principle of fine-tuning SD models involves the following steps.

- 1) Loading a pre-trained model: An SD model trained on a large-scale dataset is loaded. This model includes a deep convolutional neural network and a generative network (Encoder and decoder structure in Fig. 4b) in SD.

$$CNN_{pre-trained} = f(\theta_{large-scale}) \quad (3)$$

- 2) Freezing Layers: As the first few layers of the model include low-level feature extractors, these layers (Encoder block in Fig. 4b) are frozen. This is because these layers have already learned generalized features, while we are mainly concerned with fine-tuning the model to fit task-specific

high-level features.

$$\theta_{frozen} = freeze(\theta_{1...k}) \quad (4)$$

where $\theta_{1...k}$ represents the parameters of the first k layers.

- 3) Replacing the top layers: Next, we replace the output layers (output layer of the decoder block connection in Fig. 4b) of the model to fit our specific task. We can replace the output layer with a fully connected layer with an appropriate number of neurons for classification tasks.

$$Output_{new} = W_{new} \cdot x + b_{new} \quad (5)$$

New W_{new} and new b_{new} are the weights and biases of the new fully connected layer.

- 4) Fine-tuning Training: Now, we fine-tune the model using task-specific training data. The fine-tuning process connects the task-dependent loss function to the model's output and then updates the model's parameters through optimization algorithms such as backpropagation and gradient descent.

$$\theta_{updated} = \theta_{previous} - \eta \cdot \nabla_{\theta} J(\theta; x, y) \quad (6)$$

where η is the learning rate, $\nabla_{\theta} J(\theta; x, y)$ is the gradient of the loss function J with respect to the parameters θ , and (x, y) is the input and target output pairs for the training data.

- 5) Regularization and tuning the learning rate: To avoid overfitting, regularization techniques such as weight decay or dropout are usually applied. In addition, the learning rate must be adjusted to ensure that the model converges to the appropriate weights. Regularization can be introduced as an additional term in the loss function:

$$J_{reg}(\theta; x, y) = J(\theta; x, y) + \lambda_{reg} \cdot R(\theta) \quad (7)$$

where $R(\theta)$ represents the regularization term, and λ_{reg} is the regularization coefficient.

3.4 Controlling generative AI models with sketches by ControlNet

To enable the proposed method to match design intents expressed in sketches during the process of generating conceptual design images, ControlNet (Zhang et al., 2023) is added to the pipeline of the fine-tuned SD model. ControlNet is a technique for controlling GANs, which can be used in conjunction with SD or other GAN models to control certain aspects of the generated samples, such as the attributes of the samples, the content or the style. The following is the basic principle of ControlNet for controlling SD, as shown in Fig. 4.

- 1) Introducing a Control Vector: The network structure of ControlNet is shown in Fig. 4a as an encoder-decoder structure. One of the critical concepts of ControlNet is the introduction of a Control Vector, which contains information about the features or attributes of the samples we want to control. We can design the Control Vector as an encoding that represents different expressions. The formula for the Control Vector can be expressed as

$$c = f(attributes) \quad (8)$$

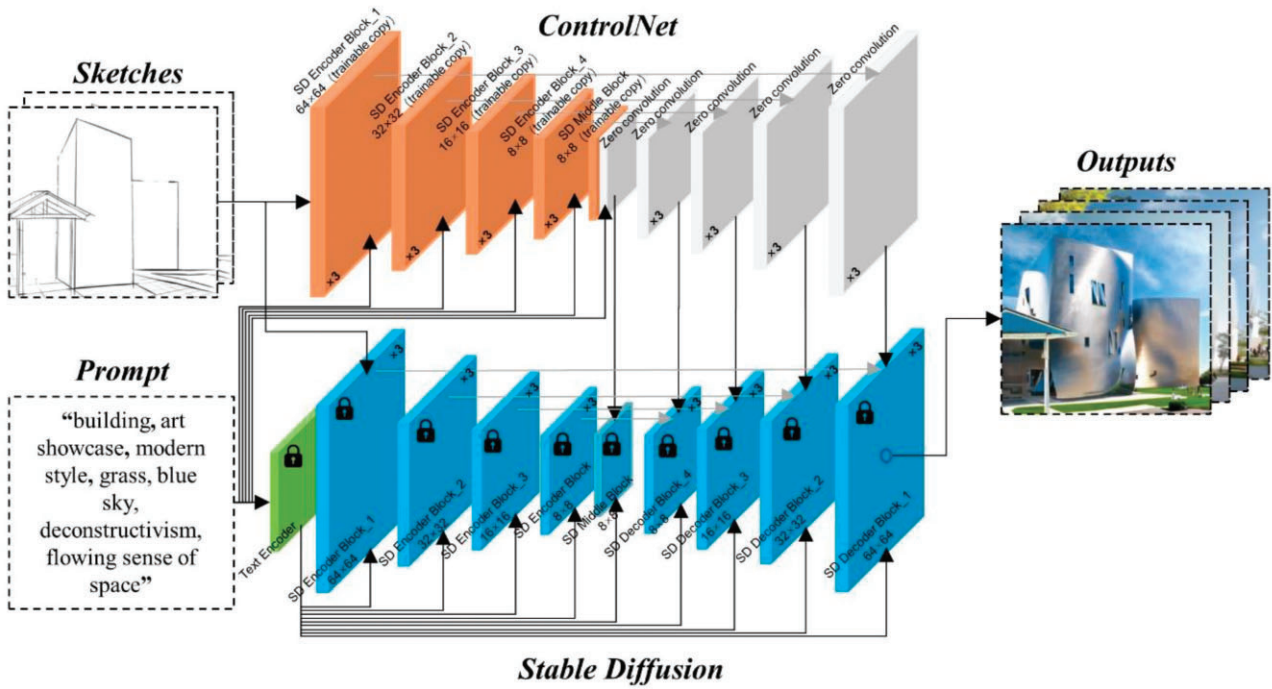


Figure 4: Schematic diagram of controlling generative AI models by ControlNet.

where c represents the Control Vector, and f is an encoding function that maps the sample's attributes or features to the Control Vector. The sample described is the sketch and prompt in Fig. 4.

- Combining with a generative model: ControlNet uses control vectors with a generative model, such as SD. This usually involves feeding the control vectors into some part of the generative model to influence the generation process. As shown in Fig. 4, the output information of the decoder block of ControlNet is input to the decoder block of SD. The formula can describe the process of combining the control vector with the generative model:

$$G(z, c) \quad (9)$$

where G represents the generative model, z is the input noise to the generative model, and c is the control vector.

- Training the ControlNet: In the training phase, we typically need to train the ControlNet to generate appropriate control vectors and integrate the control vectors with the generative model. The optimization of a loss function can represent the training process:

$$\min_{\theta} \mathcal{L}(G(z, c; \theta), \text{target}) \quad (10)$$

where θ represents the model parameters, and \mathcal{L} is a loss function that measures the difference between the generated samples and the target.

- Generating controlled samples: Once ControlNet is trained, it can create controlled samples. We can control the generative model by inputting different control vectors to generate samples with different features or attributes. This process can be represented by

$$x_{\text{controlled}} = G(z, c) \quad (11)$$

where $x_{\text{controlled}}$ represents the controlled sample (Images of buildings in Fig. 4) generated by the model.

3.5 Validation method

In the domain of computer vision, existing validation methods for 'txt-to-image' generative AI have primarily focused on appraising the quality of the generated images (Chen et al., 2023; Wang et al., 2023). However, relevant studies that applied 'txt-to-image' generative AI remain relatively limited in architectural design. Given that the assessment of architectural design is inherently subjective and encompasses aesthetic considerations, Chen et al. (2023) employed a questionnaire survey to subjectively evaluate images produced by generative AI by using several assessment criteria related to design quality (e.g., overall impression, design details, architectural integrity, consistency in architectural style etc.) that professional architects scored. However, these criteria would not be enough to assess the capability of the proposed method to reflect the design intent.

In this regard, this study validates the proposed approach by assessing the performance of the proposed model in terms of whether specific design intent is well reflected in generated architectural exterior conceptual design images. In particular, the outputs (i.e., generated design images) from the proposed method are compared with those from two existing generative image AI models [Mnml.ai (Mnml.ai, 2024, July 24) and Architectural schoolteacher (Technology, 2024, July 24)], when given same inputs (e.g., textural design intent as prompts and non-verbal design intent as sketches). For example, the assessment focuses on the extent to which each model effectively reflects (i) sketches, (ii) general image descriptions, (iii) client needs, and (iv) AL, in addition to (v) overall design quality. The assessment is based on a questionnaire survey conducted by invited professionals using these five criteria.



Figure 5: Dataset of famous architects.

4. Experimental Demonstration Using Famous Architectural Works

The experimental demonstration shows the overall procedures of the proposed method and qualitatively describes its effectiveness in realizing architectural exterior conceptual design based on design intent. Furthermore, to ascertain the benefits of the proposed method, sample images from the two existing methods mentioned above are presented along with those from the proposed method.

4.1 Datasets

For model demonstration, a dataset was created comprising 2021 images of 198 buildings designed by six famous architects (Frank Gehry, Louis Kahn, Ludwig Mies van der Rohe, Renzo Piano, Richard Meier, and Zaha Hadid), given that the works of renowned architects can be accessed online. Figure 5 shows examples of images from these buildings. The images were resized uniformly to meet the prerequisites of the training process (512×512 -pixel resolution). Annotations for these images include (i) general image descriptions used in generative image AI model training, (ii) keywords that define client needs of architectural works, and (iii) keywords related to AL that describes architects' design concepts. While general image descriptions were obtained using Dreambooth, specific keywords related to design intent (client needs and AL) were extracted from online articles using ChatGPT. Subsequently, the author meticulously reviewed the extracted textual annotations and had them confirmed by two experienced architectural experts (with more than 3 yr of architectural design experience), considering that conventional image descriptions are intuitive. The architects primarily inspected the user requirements, and AL implied behind the images to eliminate any weird annotations. Figure 6 presents an analysis of the frequency of labelled

annotations, encompassing two types of AL, client needs, and various general image descriptions.

4.2 Generative AI model fine-tuning

The SD model deployed in this research is built on a Pytorch framework. The SD model has been extensively pre-trained on a large dataset [the LAION-5B dataset (Schuhmann *et al.*, 2022) used for SD training, containing 5.8 billion 'image-text' pairs collected from the internet], which provides it with the fundamental ability to generate images from text. To enhance its capabilities in architectural design, it needs to be fine-tuned. The fine-tuning was conducted on a Windows 10 platform with 64 GB of RAM and a 12 GB video memory GPU. We set the epoch limit to 30, with a batch size of one, optimizing our fine-tuning with the Lora algorithm. To prevent overfitting, dynamic learning rates were utilized. The specific learning rates for the text encoder and the U-Net within the SD model are detailed in Fig. 7a and b, respectively. Figure 7c illustrates the loss curves throughout the fine-tuning phase, indicating a decreasing trend in the loss function, signalling convergence.

4.3 Inputs for experimental demonstration

To generate new architectural exterior conceptual design images from the fine-tuned model, four input prompts (as textual design intent) and two sketches (as non-verbal design intent) are chosen, as shown in Table 3 and Fig. 8, producing eight combinations as model inputs. The four prompts are all combinations of image descriptions, user requirements, and AL. Sketch 1 in Fig. 8a is a high-rise building, while Sketch 2 in Fig. 8b is a building consisting of three parts.

4.4 Examples of generated images

Using the inputs defined above, sample architectural exterior conceptual design images are generated from the proposed and two

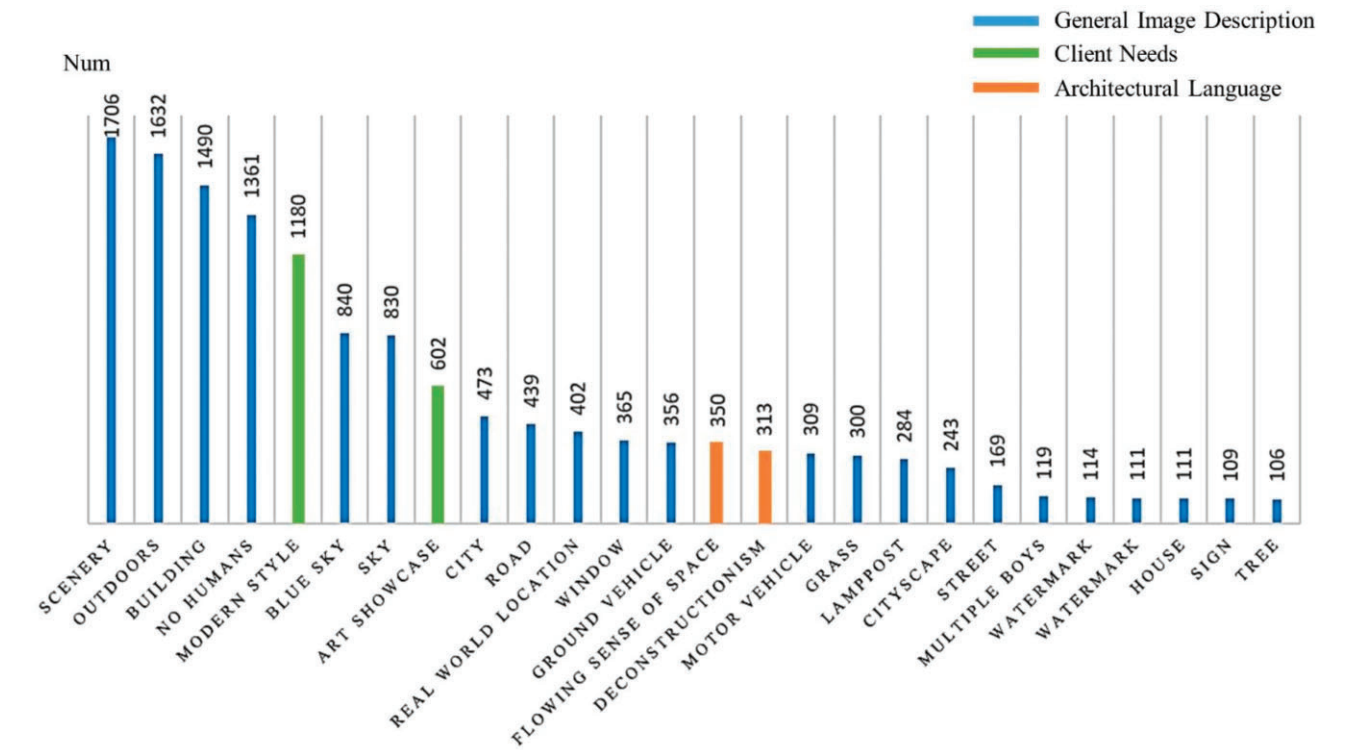


Figure 6: Distribution of image annotations in the dataset.

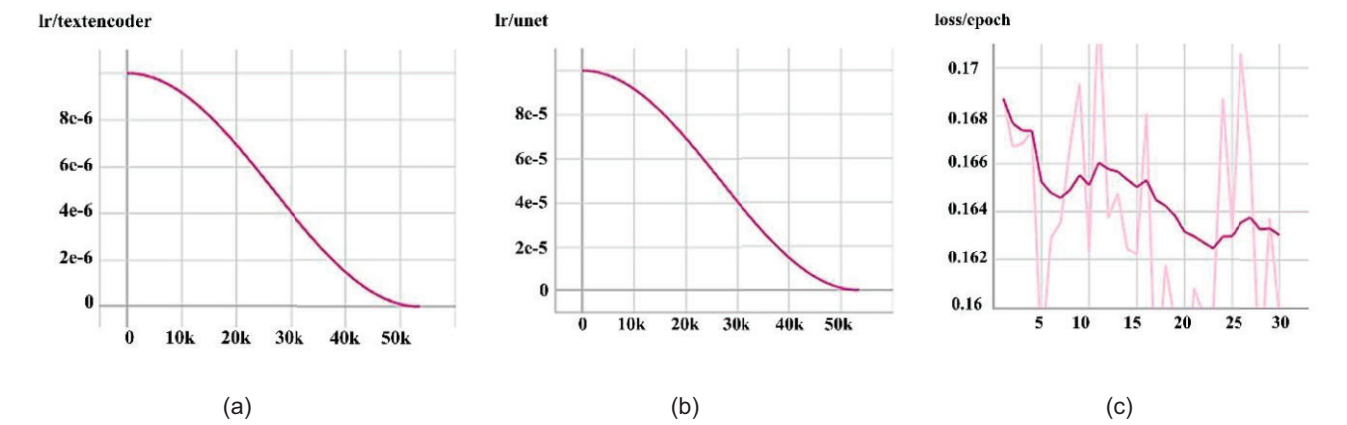


Figure 7: Fine-tuning learning rate and loss: (a) plot of variation in text encoder learning rate, (b) graph of change in U-Net learning rate, and (c) loss curves.

Table 3: Prompts.

| ID | Prompt |
|----|---|
| 1 | "Design a building located beside a road, reflecting a modern style. The building should embrace deconstructivism." |
| 2 | "Design a building with the function of an art showcase, featuring a flowing sense of space. There should be no humans in the generated images." |
| 3 | "Design a modern style building. The building should embody a flowing sense of space. There should be some trees beside the building in the generated images." |
| 4 | "Design a building with the function of an art showcase, embodying a modern style. The architecture should be designed deconstructively, featuring a flowing sense of space. Generate images of the building on grass under blue sky conditions." |

'Modern style' means minimalist design, functional spaces, use of new materials like steel and glass, and clean lines; 'art showcase' refers to a building designed as a work of art itself or a space optimized for displaying art, emphasizing creative aesthetics and functional design; 'deconstructivism' refer to a style characterized by fragmentation, non-linear shapes, and the appearance of controlled chaos, challenging traditional design conventions; 'flowing sense of space' describes design with interconnected areas, creating a continuous, dynamic spatial experience.

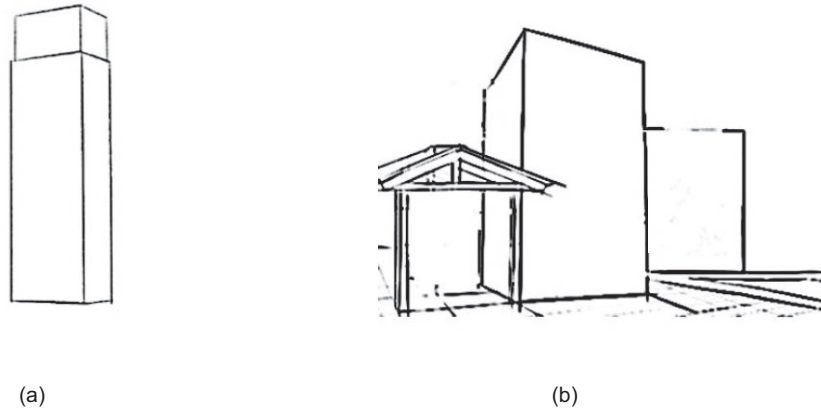


Figure 8: Sketches: (a) Sketch 1, and (b) Sketch 2.

existing methods 9Method 1: Architectural schoolteacher (Technology, 2024, July 24), Method 2: Mnml.ai (Mnml.ai, 2024, July 24)] as shown in Fig. 9. The proposed method successfully generated images that align with the requirements set by sketches and textual descriptions. For instance, Sketch 1's design intent of a two-part structure (with a shorter top and longer bottom) is depicted in the generated images, regardless of prompt variations. Similarly, Sketch 2's three-part structure, including a pavilion and two buildings, is effectively realized. In comparison, Method 1 and Method 2 generate images consistent with the sketches. Regarding general description adherence, all methods satisfactorily incorporate the 'building' element. The proposed method notably succeeds in incorporating additional elements like roads (Prompt1), no humans (Prompt2), trees (Prompt3), and grass with a blue sky (Prompt4), outperforming Method 1 and Method 2. Furthermore, in meeting client needs, all three methods produce images with a 'modern style' and 'art showcase' functionality. However, the proposed method's outputs are appealing and fitting for an art showcase landmark. Regarding architectural design language, the proposed method effectively generates images in a 'deconstructivism' style (Prompt1 and Prompt4) and with a 'flowing sense of space' (Prompt2 and Prompt3), a feat not fully achieved by Method 1 and Method 2.

5. Validation and Results

This study designed a questionnaire survey based on the evaluation criteria outlined in Section 3.5 to quantitatively assess the strengths and weaknesses of our proposed method, as illustrated in Fig. 10. Specifically, we devised eight questionnaires to align with the eight combinations of prompts and sketches. Each questionnaire initially presented images generated by the three methods. It is important to note that to ensure accurate assessment, we randomized the order of the three methods in each questionnaire to form three groups. Respondents were instructed to assess the degree to which the sketches, general image descriptions, client needs, and AL corresponded to the generated images through visual observation. The response options included 'Not Matched-1', 'Poorly Matched-2', 'Matched-3', 'Well Matched-4', and 'Very Well Matched-5'. We manually extracted keywords from the input prompts after categorization to enhance accurate judgment, as detailed in Table 4. Furthermore, participants were requested to evaluate the overall design quality using five options: 'Poor

Quality-1', 'Low Quality-2', 'Moderate Quality-3', 'High Quality-4', and 'Exceptional Quality-5'. To ensure comprehensive and expert evaluation, we enlisted professionals in the architectural field (certified architects currently employed at architectural design firms) to complete these questionnaires through the crowdsourcing website Mturk (Amazon, 2024, July 24). We received a total of 50 questionnaires. After excluding responses from individuals with less than one year of architectural experience, we ended up with 39 valid questionnaires; among them, 84.62% have been employed for more than three years, and 61.54% have been employed for more than 5 yr.

Figure 11 displays the distribution of questionnaire scores for the three methods across five categories. The illustration clearly shows that the proposed method garners higher top-tier scores in general image description, client needs, and AL. The scores regarding design quality are relatively consistent across all methods. Notably, scores of 3 and above indicate successful alignment between images and their corresponding sketches and descriptions. In this context, Table 5 tabulates the matching rates of the three methods in these various criteria. The data reveal that the proposed method excels in general image description, client needs, and architectural design language, registering a matching rate exceeding 80%. Although its performance in sketch matching is somewhat lower, the overall average matching rate for the proposed method is 80.69%. In comparison, the two existing methods fall short of this benchmark, with method 1 achieving 75.64% and method 2 reaching 73.00%.

Moreover, we evaluated the proposed method alongside two other methods regarding specific scores. First, the results of descriptive statistics, including mean and standard deviation, are shown in Table 6. Second, analysis of variance (ANOVA; St & Wold, 1989) is a statistical analysis method mainly used to compare whether the difference in means between two or more samples or groups is statistically significant. A two-by-two comparison can visualize the difference between the two methods more intuitively. Therefore, we also used IBM SPSS Statistics 25 software to conduct an ANOVA analysis of the scores of the three methods in the questionnaire data in the five aspects of two-by-two comparisons between different methods, applying Bonferroni correction (Napierala, 2012) to mitigate the risk of type I errors, and the results are shown in Table 7. Specifically, Method 2 is significantly better than Method 1 regarding matching sketches, with a mean difference of 0.375 ($P < 0.001$). Method 2 is significantly better than the proposed method, with a mean difference

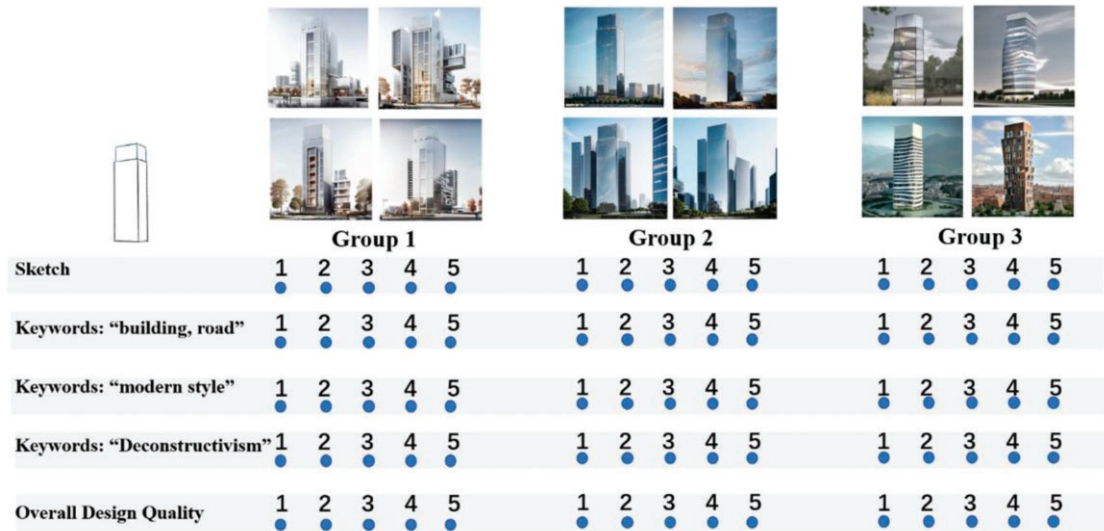


Figure 9: Results of three methods.

of 0.808 ($P < 0.001$). Regarding matching general image description, the proposed method is significantly better than method 1, with a mean difference of 0.455 ($P < 0.001$). The proposed method is significantly better than method 2, with a mean difference of 0.391 ($P < 0.001$). The difference between Method 1 and Method

2 was insignificant ($P = 0.496$). The proposed method is significantly better than Method 1 in matching client needs with a mean difference of 0.385 ($P < 0.001$). The proposed method is significantly better than method 2, with a mean difference of 0.484 ($P < 0.001$). The difference between Method 1 and Method

Please rate three groups of building images based on whether they match well with 1) the given sketch, and 2) the given keywords. Also, please evaluate the overall design quality.



| | Group 1 | | | | | Group 2 | | | | | Group 3 | | | | |
|------------------------------|---------|---|---|---|---|---------|---|---|---|---|---------|---|---|---|---|
| Sketch | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Keywords: "building, road" | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Keywords: "modern style" | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Keywords: "Deconstructivism" | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| Overall Design Quality | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |

Figure 10: Questionnaire.

Table 4: Keywords of the prompts.

| Prompt | General image description | Client needs | AL |
|--------|-----------------------------|------------------------------|--|
| 1 | 'Building, road' | 'Modern style' | 'Deconstructivism' |
| 2 | 'Building, no humans' | 'Art showcase' | 'Flowing sense of space' |
| 3 | 'Building, tree' | 'Modern style' | 'Flowing sense of space' |
| 4 | 'Building, grass, blue sky' | 'Art showcase, modern style' | 'Deconstructivism, flowing sense of space' |

2 was insignificant ($P = 0.232$). Regarding matching the AL, the proposed method is significantly better than Method 1, with a mean difference of 0.494 ($P < 0.001$). The proposed method is significantly better than method 2, with a mean difference of 0.753 ($P < 0.001$). Method 1 was significantly better than Method 2, with a mean difference of 0.260 ($P = 0.005$). Regarding design quality, the differences between the three methods were insignificant, with p-values greater than 0.05 for two-by-two comparisons. Furthermore, we applied the paired t-test (Shi et al., 2024) using IBM SPSS Statistics 25 software to enhance the robustness of the comparison, with the results presented in Table 8. On the metrics of Sketch, Image description, Client needs, and AL, the proposed method significantly outperformed the other two methods; in terms of Design quality, the differences between the three were insignificant.

In summary, the proposed method is significantly better ($P < 0.0167$) in terms of matching general image description, client needs and AL without significant differences ($P > 0.0167$) in design quality with the two existing methods.

6. Discussion

This study proposed an architectural exterior conceptual design method that generates images that align with the design intent. We designed a survey to assess the proposed method's effectiveness and advantages and compared it with two existing approaches. The results indicate that the proposed method can gen-

erate architectural images by design intent; the overall average matching rate for the proposed method is 80.69%. Furthermore, it is significantly better ($P < 0.0167$) than existing methods in matching general image descriptions, client needs, and AL.

The success of the proposed methodology stems from several aspects in terms of understanding textualized architectural design requirements. The proposed method constructs an architectural design dataset. It adds client needs and AL to the dataset on top of regular image descriptions, thus creating a new dataset containing architectural domain knowledge. Based on this, we employ the Lora technique to migrate the domain knowledge into generative AI. As a result, when we use prompts containing textualized design intent as input, the generative AI can understand and generate matching images. The ways of designing domain prompts and boosting generative AI with domain knowledge are all affiliated with prompt engineering (Oppenlaender, 2022). Therefore, this research highlights the importance of Prompt Engineering in augmenting the capabilities of generative AI in architectural design.

In terms of understanding sketches, the success of the proposed method stems from controlling the image generation process of the generative AI with ControlNet. Given that the proposed method is mainly used in the mass modelling of the conceptual design process, which is more concerned with the shape and form of the building. Therefore, as shown in Fig. 12a, the adopted rough sketch showing the pavilion and the two buildings, as well as their shape and spatial relationship, when inputted into prompt4, successfully generates an image of the building that satisfies both

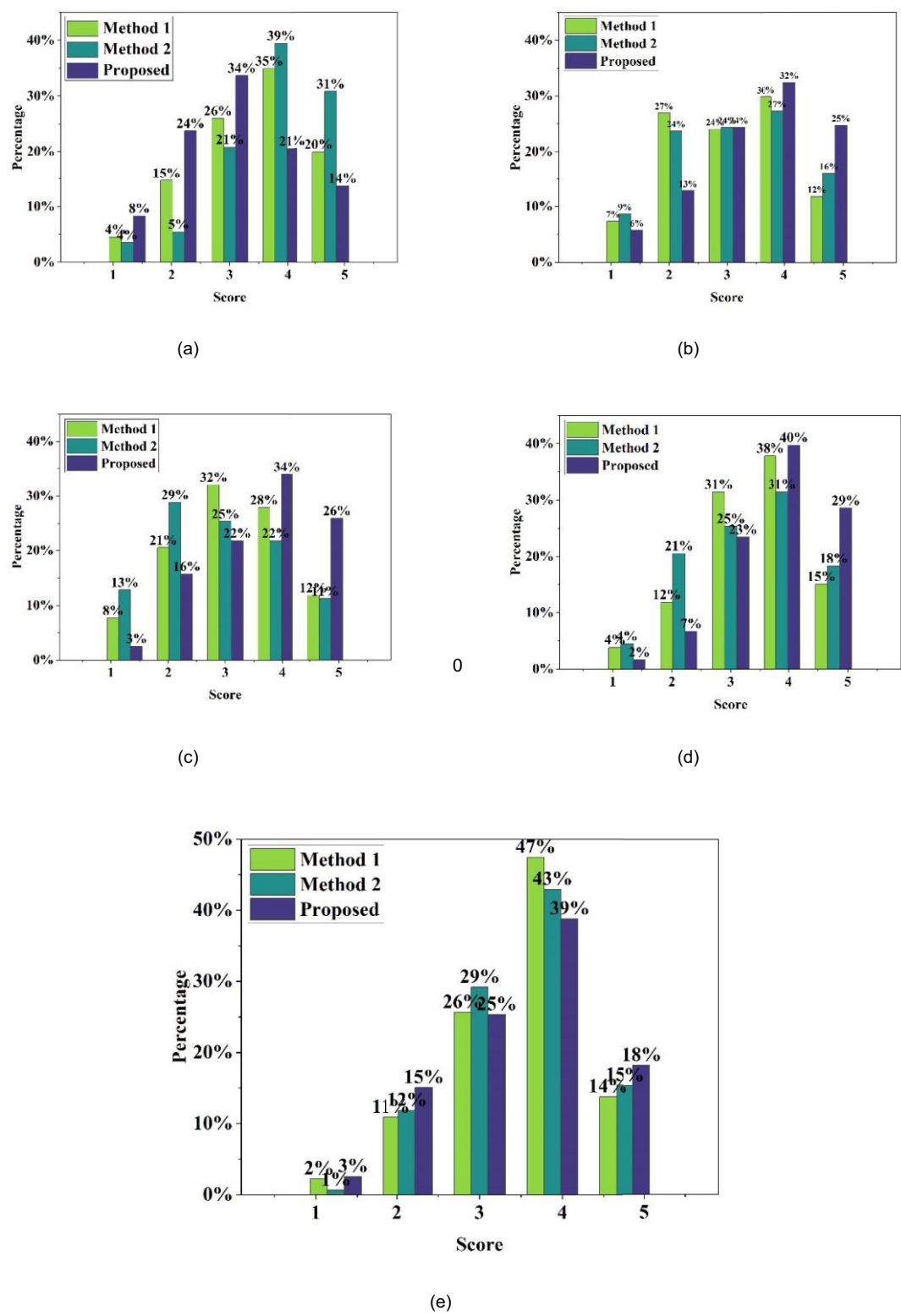


Figure 11: Histogram of multiple indicators of scoring ratio: (a) sketch-matching score, (b) general image description matching score, (c) client needs matching score, (d) AL matching score, and (d) AL matching score.

Table 5: Match rates of the three methods in different aspects.

| Methods | Sketch | General image description | Client needs | AL | Overall mean |
|----------|--------|---------------------------|--------------|--------|--------------|
| Method 1 | 80.77% | 65.71% | 84.29% | 71.79% | 75.64% |
| Method 2 | 91.03% | 67.63% | 75.00% | 58.33% | 73.00% |
| Proposed | 67.95% | 81.41% | 91.67% | 81.73% | 80.69% |

Table 6: Results of descriptive statistics for three methods (mean \pm standard deviation).

| Methods | Sketch | General image description | Client needs | AL | Design quality |
|----------|------------------|---------------------------|------------------|------------------|------------------|
| Method 1 | 3.51 \pm 1.102 | 3.12 \pm 1.152 | 3.48 \pm 1.011 | 3.16 \pm 1.116 | 3.60 \pm 0.934 |
| Method 2 | 3.88 \pm 1.020 | 3.18 \pm 1.212 | 3.38 \pm 1.134 | 2.90 \pm 1.209 | 3.61 \pm 0.908 |
| Proposed | 3.08 \pm 1.151 | 3.57 \pm 1.160 | 3.87 \pm 0.958 | 3.65 \pm 1.104 | 3.55 \pm 1.035 |

Table 7: Results of pairwise comparisons in five aspects using ANOVA for three methods.

| Different aspects | (I) Method | (J) Method | Mean difference (I–J) | Standard error | P value | 95% confidence interval | |
|-------------------|------------|------------|-----------------------|----------------|---------|-------------------------|-------------|
| | | | | | | Lower limit | Upper limit |
| Sketch | Method 1 | Method 2 | -0.375* | 0.087 | < 0.001 | -0.55 | -0.20 |
| | Method 1 | Proposed | 0.433* | 0.087 | < 0.001 | 0.26 | 0.60 |
| | Method 2 | Proposed | 0.808* | 0.087 | < 0.001 | 0.64 | 0.98 |
| Image description | Method 1 | Method 2 | -0.064 | 0.094 | 0.496 | -0.25 | 0.12 |
| | Method 1 | Proposed | -0.455* | 0.094 | < 0.001 | -0.64 | -0.27 |
| | Method 2 | Proposed | -0.391* | 0.094 | < 0.001 | -0.58 | -0.21 |
| Client needs | Method 1 | Method 2 | 0.099 | 0.083 | 0.232 | -0.06 | 0.26 |
| | Method 1 | Proposed | -0.385* | 0.083 | < 0.001 | -0.55 | -0.22 |
| | Method 2 | Proposed | -0.484* | 0.083 | < 0.001 | -0.65 | -0.32 |
| AL | Method 1 | Method 2 | 0.260* | 0.092 | 0.005 | 0.08 | 0.44 |
| | Method 1 | Proposed | -0.494* | 0.092 | < 0.001 | -0.67 | -0.31 |
| | Method 2 | Proposed | -0.753* | 0.092 | < 0.001 | -0.93 | -0.57 |
| Design quality | Method 1 | Method 2 | -0.010 | 0.077 | 0.901 | -0.16 | 0.14 |
| | Method 1 | Proposed | 0.045 | 0.077 | 0.560 | -0.11 | 0.20 |
| | Method 2 | Proposed | 0.054 | 0.077 | 0.479 | -0.10 | 0.21 |

*Applying Bonferroni correction to reduce the risk of Type I errors. Since each metric was compared three times, the significance level is $0.05/3 = 0.0167$. The name of the significantly better method in a two-by-two comparison is bolded.

Table 8: Results of pairwise comparisons in five aspects using paired t-tests for three methods.

| Different aspects | (I) Method | (J) Method | Paired differences (I–J) | t-value | Sig. (two-tailed) |
|-------------------|------------|------------|--------------------------|---------|-------------------|
| Sketch | Method 1 | Method 2 | 0.027* | -13.660 | < 0.001 |
| | Method 1 | Proposed | 0.028* | 15.401 | < 0.001 |
| | Method 2 | Proposed | 0.025* | 31.834 | < 0.001 |
| Image description | Method 1 | Method 2 | 0.017* | -3.863 | < 0.001 |
| | Method 1 | Proposed | 0.028* | -16.118 | < 0.001 |
| | Method 2 | Proposed | 0.027* | -14.131 | < 0.001 |
| Client needs | Method 1 | Method 2 | 0.022* | 4.471 | < 0.001 |
| | Method 1 | Proposed | 0.028* | -13.942 | < 0.001 |
| | Method 2 | Proposed | 0.028* | -17.079 | < 0.001 |
| AL | Method 1 | Method 2 | 0.025* | 10.443 | < 0.001 |
| | Method 1 | Proposed | 0.028* | -17.411 | < 0.001 |
| | Method 2 | Proposed | 0.026* | -28.453 | < 0.001 |
| Design quality | Method 1 | Method 2 | 0.060 | -0.159 | 0.873 |
| | Method 1 | Proposed | 0.065 | 0.686 | 0.493 |
| | Method 2 | Proposed | 0.067 | 0.804 | 0.422 |

*Applying Bonferroni correction to reduce the risk of Type I errors. Since each metric was compared three times, the significance level is $0.05/3 = 0.0167$. The name of the significantly better method in a two-by-two comparison is bolded.

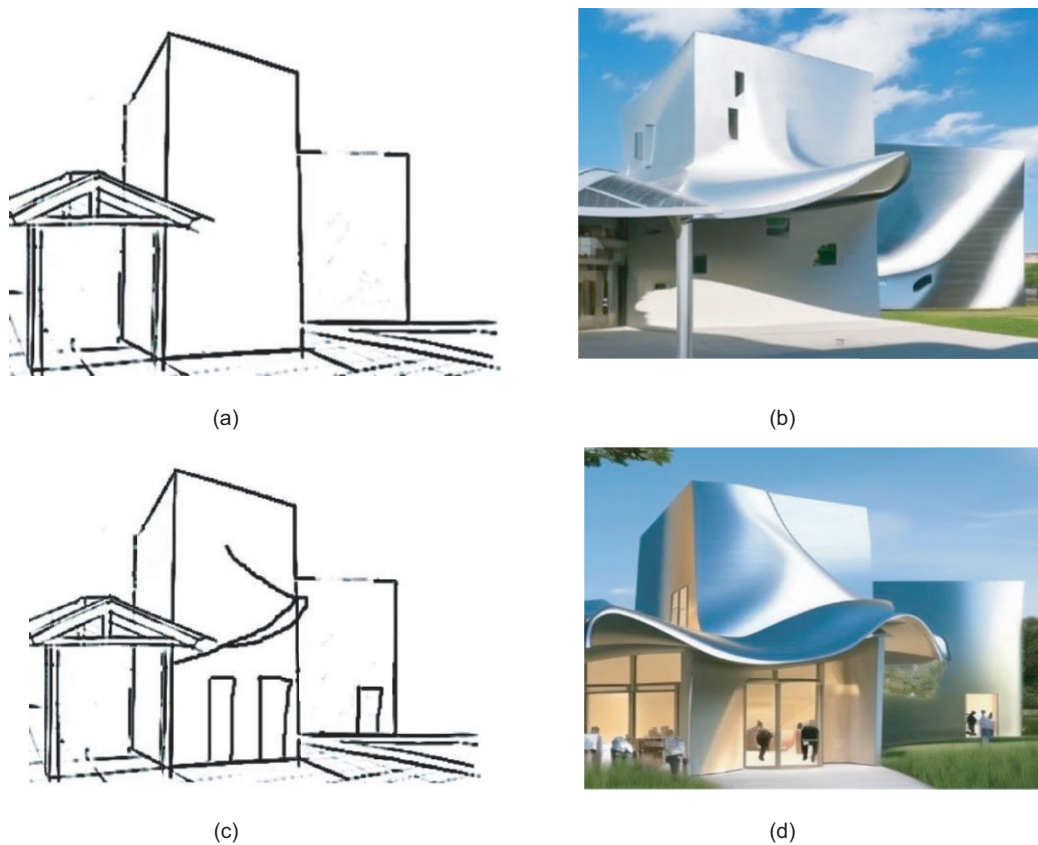


Figure 12: Comparison results of rough and more detailed sketches: (a) rough sketch, (b) generated image by rough sketch, (c) more detailed sketch, and (d) generated image by more detailed sketch.

the textualized architectural design language and the sketches of the buildings, as shown in Fig. 12b. The architectural exterior conceptual design process is an iterative process where the design of a building is continuously refined. Therefore, inspired by the generated image, the architect absorbed the design of the flow-sensitive eaves and further added the need for a door design, resulting in a refined sketch, as shown in Fig. 12c. When the same was entered in prompt 4, the architectural design matching the refined sketch was successfully generated. This indicates that, on the one hand, the proposed method can meet the design requirements of sketches with different levels of refinement; on the other hand, the proposed method can be used in a continuous iterative process in the conceptual design phase. Architects can cooperate with generative AI to generate conceptual designs that meet the design intent and are creative.

Despite the proposed method's success in this study's experiments, it still faces limitations. (i) The dataset's limited sources, comprising works of only six famous architects, restrict architectural style diversity. Capturing works from more architects is acknowledged as necessary to enrich the dataset and support diverse architectural designs. (ii) The types of text annotation in the dataset are constrained, relying on a restricted number of images for specific keywords. Future improvements involve acquiring more architectural images and adding richer textual annotations to address broader architectural design needs. (iii) Given that textual descriptions corresponding to architectural images are often incomplete or even missing. Considering how to construct a dataset under conditions of insufficient availability and comprehensiveness of textual de-

scriptions can help to increase the generality of the proposed method.

The use of generative AI in architectural exterior conceptual design has far-reaching implications. Generative AI can generate photorealistic images comparable to professional architects. This technology not only improves the efficiency and speed of design plan generation and shortens the initial drafting time for architects but also provides innovative design options and enhances creativity. In the future, developing a practical web interface or app based on the proposed method will benefit architects and accelerate the design process.

7. Conclusions

This study proposes an approach to automate the architectural exterior conceptual design based on architectural design intent. The contributions of this study are as follows: teaching generative AI to learn to understand textual design intent and allowing generative AI to combine textual and non-textual design intent. For this purpose, we constructed an architectural image dataset and added general image descriptions, client needs and AL. The SD model is fine-tuned using Lora to enable the generative AI to understand the textualized design intent. In addition, we used ControlNet to control the SD generation process to generate architectural conceptual images that conform to the sketches simultaneously.

We have designed comparative experiments and verified the effectiveness of the proposed method using a questionnaire. The results indicate that the proposed method can generate

architectural images by design intent; the overall average matching rate for the proposed method is 80.69%. It is significantly better ($P < 0.0167$) than the existing methods in terms of understanding general image descriptions, client needs and architectural design language (Mnml.ai and Architectural schoolteacher). This study demonstrates the potential of prompt engineering in enhancing the performance of generative AI in architectural design.

Further prompt engineering efforts are essential to support a broader range of complex and diverse architectural design intents. These include expanding the dataset to encompass a greater variety of data, which will aid the method in generating more diverse architectural concept sketches. Enriching the dataset with a broader array of keyword tags, including client needs and AI, will further enhance the generative AI's understanding of complex and nuanced architectural design intentions.

Conflict of interest statement

None declared.

Author Contributions

M.S.: Methodology, software, and writing (original draft). H.-L.: Writing (review and editing). J.S.: Conceptualization and supervision. S.H.C.: Writing (review and editing) and validation. B.X.: Writing (review and editing).

Data availability

Data will be available upon request.

References

- Amazon. (2024). Amazon Mechanical Turk (Mturk). <https://www.mturk.com> Accessed 24 July 2024.
- Anzalone, P., & Clarke, C. (2003). Architectural applications of complex adaptive systems. In *Proceedings of the 2003 Annual Conference of the Association for Computer Aided Design in Architecture*. Ball State University.
- Baltus, V., & Žebrauskas, T. (2019). Parametric design concept in architectural studies. *Architecture and Urban Planning*, **15**, 96–100. <https://doi.org/10.2478/aup-2019-0013>.
- Castro Pena, M. L., Carballal, A., Rodríguez-Fernández, N., Santos, I., & Romero, J. (2021). Artificial intelligence applied to conceptual design. A review of its use in architecture. *Automation in Construction*, **124**, 103550. <https://doi.org/10.1016/j.autcon.2021.103550>.
- Chang, S., Dong, W., & Jun, H. (2020). Use of electroencephalogram and long short-term memory networks to recognize design preferences of users toward architectural design alternatives. *Journal of Computational Design and Engineering*, **7**, 551–562. <https://doi.org/10.1093/jcde/qwaa045>.
- Chen, J., Wang, D., Shao, Z., Zhang, X., Ruan, M., Li, H., & Li, J. (2023). Using artificial intelligence to generate master-quality architectural designs from text descriptions. *Buildings*, **13**, 2285. <https://doi.org/10.3390/buildings13092285>.
- Chen, X., Kang, S. B., Xu, Y.-Q., Dorsey, J., & Shum, H.-Y. (2008). Sketching reality: realistic interpretation of architectural designs. *ACM Transactions on Graphics (TOG)*, **27**, 1–15. <https://doi.org/10.1145/1356682.1356684>.
- Chen, Y., & Kitagawa, K. (2023). Locally based architectural construction strategies in rural China: textual analysis of architects' Design thinking. *Sustainability*, **15**, 10573. <https://doi.org/10.3390/su151310573>.
- Coates, P., Healy, N., Lamb, C., & Voon, W. (1996). The use of Cellular Automata to explore bottom up architectonic rules. In *Proceedings of the Eurographics UK Chapter 14th Annual Conference*. Eurographics Association UK.
- Dehouche, N., & Dehouche, K. (2023). What's in a text-to-image prompt? The potential of stable diffusion in visual arts education. *Heliyon*, **9**, e16757. <https://doi.org/10.1016/j.heliyon.2023.e16757>.
- Dutta, K., & Sarthak, S. (2011). Architectural space planning using evolutionary computing approaches: a review. *Artificial Intelligence Review*, **36**, 311–321. <https://doi.org/10.1007/s10462-011-9217-y>.
- Eilouti, B. (2018). Concept evolution in architectural design: an octonary framework. *Frontiers of Architectural Research*, **7**, 180–196. <https://doi.org/10.1016/j.foar.2018.01.003>.
- Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). Can: Creative adversarial networks, generating “art” by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068*. <https://doi.org/10.48550/arXiv.1706.07068>.
- EvolveLAB. (2024). Veras. <https://www.evolveLAB.io/veras> Accessed 24 July 2024.
- Gan, W., Zhao, Z., Wang, Y., Zou, Y., Zhou, S., & Wu, Z. (2024). UGDAN: a new urban design inspiration approach driven by using Generative Adversarial Networks. *Journal of Computational Design and Engineering*, **11**, 305–324. <https://doi.org/10.1093/jcde/qwae014>.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, **27**, 1–9. <https://doi.org/10.48550/arXiv.1406.2661>.
- Hartoonian, G. (2010). Frank Gehry: roofing, wrapping, and wrapping the roof. *The Journal of Architecture*, **7**, 1–31. <https://doi.org/10.1080/13602360110114759>.
- Hitaj, B., Ateniese, G., & Perez-Cruz, F. (2017). Deep models under the GAN: Information leakage from collaborative deep learning. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*.
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). Lora: low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*. **10**, 1–26. <https://doi.org/10.48550/arXiv.2106.09685>.
- Jo, H., Lee, J.-K., Lee, Y.-C., & Choo, S. (2024). Generative artificial intelligence and building design: early photorealistic render visualization of façades using local identity-trained models. *Journal of Computational Design and Engineering*, **11**, 85–105. <https://doi.org/10.1093/jcde/qwae017>.
- Joye, Y. (2011). A review of the presence and use of fractal geometry in architectural design. *Environment and Planning B: Planning and Design*, **38**, 814–828. <https://doi.org/10.1068/b36032>.
- Kakooee, R., & Dillenburger, B. (2024). Reimagining space layout design through deep reinforcement learning. *Journal of Computational Design and Engineering*, **11**, 43–55. <https://doi.org/10.1093/jcde/qwae025>.
- Krūgelis, L. (2018). 3D printing technology as a method for discovering new creative opportunities for architecture and design. *Land-scape Archit. Art*, **13**, 87–94. <https://doi.org/10.22616/j.landarchart.2018.13.10>.
- Li, H., Gu, J., Koner, R., Sharifzadeh, S., & Tresp, V. (2023). Do DALL-E and Flamingo understand each other. *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- Mnml.ai. (2024). Architecture AI Design Assistant(mnml.ai). <https://mnml.ai/explore> Accessed 24 July 2024.
- Mora, R., Bédard, C., & Rivard, H. (2008). A geometric modelling framework for conceptual structural design from early digital architectural thinking.

- tural models. *Advanced Engineering Informatics*, **22**, 254–270. <https://doi.org/10.1016/j.aei.2007.03.003>.
- Napierala, M. A. (2024). What is the Bonferroni correction? In *Yaos Now* (pp. 40–41). https://docs.ufpr.br/~giolo/LivroADC/Material/S3_Bonferroni%20Correction.pdf Accessed 24 July 2014.
- Ni, Z., Wei, L., Li, J., Tang, S., Zhuang, Y., & Tian, Q. (2023). Degeneration-tuning: Using scrambled grid shield unwanted concepts from stable diffusion. *Proceedings of the 31st ACM International Conference on Multimedia*.
- Oppenlaender, J. (2022). Prompt engineering for text-based generative art. *arXiv preprint arXiv:2204.13988*, **6**, 1–18. <https://doi.org/10.48550/arXiv.2204.13988>.
- Pérez, R. I. P. (2024). *Blurring the Boundaries between Real and Artificial in Architecture and Urban Design through the Use of Artificial Intelligence*. Universidade da Coruña. <http://hdl.handle.net/2183/19688> Accessed 24 July 2024.
- Pourzolfaghar, Z., Ibrahim, R., Abdullah, R., & Adam, N. M. (2014). A technique to capture multi-disciplinary tacit knowledge during the conceptual design phase of a building project. *Journal of Information & Knowledge Management*, **13**, 1450013. <https://doi.org/10.1142/S0219649214500130>.
- Qiu, S., Fok, S., Chen, C., & Xu, S. (2002). Conceptual design using evolution strategy. *The International Journal of Advanced Manufacturing Technology*, **20**, 683–691. <https://doi.org/10.1007/s001700200207>.
- Rapone, G., & Saro, O. (2012). Optimisation of curtain wall façades for office buildings by means of PSO algorithm. *Energy and Buildings*, **45**, 189–196. <https://doi.org/10.1016/j.enbuild.2011.11.003>.
- Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE.
- Schuhmann, C., Beaumont, R., Vencu, R., Gordon, C., Wightman, R., Cherti, M., Coombes, T., Katta, A., Mullis, C., Wortsman, M., Schramowski, P., Kundurthy, S., Crowson, K., Schmidt, L., Kaczmarczyk, R., & Jitsev, J. (2022). Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, **35**, 25278–25294. <https://doi.org/10.48550/arXiv.2210.08402>.
- Sengupta, A., & Mishra, V. K. (2014). Integrated particle swarm optimization (i-PSO): An adaptive design space exploration framework for power-performance tradeoff in architectural synthesis. In *Proceedings of the Fifteenth International Symposium on Quality Electronic Design*. IEEE.
- Sharafi, P., Teh, L. H., & Hadi, M. N. (2015). Conceptual design optimization of rectilinear building frames: A knapsack problem approach. *Engineering Optimization*, **47**, 1303–1323. <https://doi.org/10.1080/0305215X.2014.963068>.
- Shi, Z., Jin, N., Chen, D., & Ai, D. (2024). A comparison study of semantic segmentation networks for crack detection in construction materials. *Construction and Building Materials*, **414**, 134950. <https://doi.org/10.1016/j.conbuildmat.2024.134950>.
- 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>.
- St, L., & Wold, S. (1989). Analysis of variance (ANOVA). *Chemometrics and Intelligent Laboratory Systems*, **6**, 259–272. [https://doi.org/10.1016/0169-7439\(89\)80095-4](https://doi.org/10.1016/0169-7439(89)80095-4).
- Technology, S. C. N. (2024). *Architectural Schoolteacher*. <https://jianzhuxuezhang.com> Accessed 24 July 2024.
- Thyssen, M. H., Emmitt, S., Bonke, S., & Kirk-Christoffersen, A. (2010). Facilitating client value creation in the conceptual design phase of construction projects: A workshop approach. *Architectural Engineering and Design Management*, **6**, 18–30. <https://doi.org/10.3763/aedm.2008.0095>.
- Wang, B., Zhu, Y., Chen, L., Liu, J., Sun, L., & Childs, P. (2023). A study of the evaluation metrics for generative images containing combinational creativity. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **37**, e11. <https://doi.org/10.1017/S0890060423000069>.
- Wen, W., Hong, L., & Xueqiang, M. (2010). Application of fractals in architectural shape design. In *Proceedings of the 2010 IEEE 2nd Symposium on Web Society*. IEEE.
- Wikberg, F., Olofsson, T., & Ekholm, A. (2014). Design configuration with architectural objects: Linking customer requirements with system capabilities in industrialized house-building platforms. *Construction Management and Economics*, **32**, 196–207. <https://doi.org/10.1080/01446193.2013.864780>.
- Xia, C., Zhu, Y., & Lin, B. (2008). Building simulation as assistance in the conceptual design. *Building Simulation*, **1**, 46–52. <https://doi.org/10.1007/s12273-008-8107-y>.
- Yi, H., & Kim, I. (2022). Differential evolutionary cuckoo-search-integrated tabu-adaptive pattern search (DECS-TAPS): A novel multi-hybrid variant of swarm intelligence and evolutionary algorithm in architectural design optimization and automation. *Journal of Computational Design and Engineering*, **9**, 2103–2133. <https://doi.org/10.1093/jcde/qwac100>.
- Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE.
- Zhao, S. (2021). Using artificial neural network and WebGL to algorithmically optimize window wall ratios of high-rise office buildings. *Journal of Computational Design and Engineering*, **8**, 638–653. <https://doi.org/10.1093/jcde/qwab005>.