

Article

Knowledge-Driven and Diffusion Model-Based Methods for Generating Historical Building Facades: A Case Study of Traditional Minnan Residences in China

Sirui Xu ¹ , Jiaxin Zhang ^{2,*}  and Yunqin Li ² 

¹ School of Architecture, Huaqiao University, Xiamen 361021, China; 2013131021@stu.hqu.edu.cn

² Architecture and Design College, Nanchang University, No. 999, Xuefu Avenue, Honggutan New District, Nanchang 330031, China; liyunqin@ncu.edu.cn

* Correspondence: jiaxin.arch@ncu.edu.cn

Abstract: The preservation of historical traditional architectural ensembles faces multifaceted challenges, and the need for facade renovation and updates has become increasingly prominent. In conventional architectural updating and renovation processes, assessing design schemes and the re-designing component are often time-consuming and labor-intensive. The knowledge-driven method utilizes a wide range of knowledge resources, such as historical documents, architectural drawings, and photographs, commonly used to guide and optimize the conservation, restoration, and management of architectural heritage. Recently, the emergence of artificial intelligence-generated content (AIGC) technologies has provided new solutions for creating architectural facades, introducing a new research paradigm to the renovation plans for historic districts with their variety of options and high efficiency. In this study, we propose a workflow combining Grasshopper with Stable Diffusion: starting with Grasshopper to generate concise line drawings, then using the ControlNet and low-rank adaptation (LoRA) models to produce images of traditional Minnan architectural facades, allowing designers to quickly preview and modify the facade designs during the renovation of traditional architectural clusters. Our research results demonstrate Stable Diffusion's precise understanding and execution ability concerning architectural facade elements, capable of generating regional traditional architectural facades that meet architects' requirements for style, size, and form based on existing images and prompt descriptions, revealing the immense potential for application in the renovation of traditional architectural groups and historic districts. It should be noted that the correlation between specific architectural images and proprietary term prompts still requires further addition due to the limitations of the database. Although the model generally performs well when trained on traditional Chinese ancient buildings, the accuracy and clarity of more complex decorative parts still need enhancement, necessitating further exploration of solutions for handling facade details in the future.

Keywords: architectural heritage; historical building facades; knowledge-driven method; diffusion model; low-rank adaptation model



Citation: Xu, S.; Zhang, J.; Li, Y. Knowledge-Driven and Diffusion Model-Based Methods for Generating Historical Building Facades: A Case Study of Traditional Minnan Residences in China. *Information* **2024**, *15*, 344. <https://doi.org/10.3390/info15060344>

Academic Editor: Tao Yin

Received: 7 May 2024

Revised: 4 June 2024

Accepted: 7 June 2024

Published: 11 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Conservation of architectural heritage complies with sustainable development objectives, increasingly incorporating new technologies and design approaches in such projects. Innovative applications of digital technologies enhance our understanding of heritage values and assist in their conservation, forecasting a future where digital tech is interwoven with the preservation of architectural historical heritage. In practice, the renewal of urban historic areas involves the refurbishment of building facades, a challenging task that requires architects to balance traditional architectural conservation theory and practice with graphical and textual considerations. As such, architects are required to meticulously design and plan to ensure that renovated historical areas continue to contribute cultural and historical value to both the city and the community.

Recent machine learning models have significantly improved in generating architectural images and have made transferring cultural heritage contents through digital media possible. Wang et al. (2022) present Bottleneck Concept Learner (BotCL), a concept-based framework for interpreting deep neural networks' behavior without explicit supervision. Tested on image classification tasks, BotCL enhances neural network interpretability [1]. Using deep learning and a dataset of traditional Chinese architecture from Jiangxi, China, a study implements a deep hashing retrieval method to provide high-accuracy recommendations, overcoming data scarcity [2]. Zhang et al. (2022) developed a framework to automatically generate synthetic datasets for building facade instance segmentation using city digital twins (CDTs). By rendering digital assets in a game engine, the framework produces synthetic street views and annotations [3]. Addressing the complexities of real-world facade parsing, a paper introduces the CFP dataset with high-resolution images and annotations. It proposes RTFP, using vision transformers and an efficient revision algorithm (LAFR) to improve segmentation results. The method shows superior performance on various benchmark datasets [4]. Zou et al. (2023) propose a machine learning-based method to quantify and evaluate architectural forms using a dataset labeled with typical features. The method identifies regional architectural styles in Hubei, China, showing geographic variations. It offers a quantitative tool for feature extraction and evaluation, useful in urban renewal processes [5]. Focusing on the digitalization of heritage in historic districts, a study develops a method combining typological plan analysis, Shape Grammar, and Grasshopper software (based on Rhino 7.0). Using Kulangsu's modern Western-style houses as a case study, the method classifies layout plans into prototypes and creates digital forms, aiding heritage culture databases and digital management [6]. These studies balance traditional architectural conservation theory and practice with new digital opportunities and AI applications.

Recent machine learning models have significantly improved in generating architectural images, yet the complexity and diversity of facade renovation projects continue to pose challenges in producing high-quality images. Challenges such as the rationality of image details, accuracy, and limited stylistic diversity necessitate further intervention from designers. The traditional architectural design process is typically a time-consuming and laborious procedure, including stages such as initial concept development, refinement, evaluation, and redesign. Especially in large-scale urban renewal projects, this process may involve a significant amount of repetitive work. The recent advent of artificial intelligence-generated content (AIGC) technologies has ushered in a new research paradigm for designing renewal plans for urban historic districts. AIGC, utilizing machine learning and deep learning algorithms, enables architects to efficiently optimize projects.

Compared to earlier technologies, advanced Stable Diffusion models offer more targeted capabilities, rapidly producing numerous high-quality images from specific prompts.

In the field of architectural history and heritage conservation, protecting and maintaining urban historic areas is of utmost importance. Exploring the integration of AIGC into new workflows for generating traditional architectural facades helps innovate methods for protecting and updating urban historical building groups, preserving the unique charm of urban architecture and culture, promoting cultural tourism and economic growth, and enhancing the city's attractiveness and competitiveness. And with the emergence of technology-assisted facade generation methods, evaluation methods for human-induced modifications of historical building facades are also established. A paper reviews human-induced impacts on UNESCO World Heritage Sites, emphasizing careful preservation of traditional architecture. Using the Amalfi Coast as a case study, it proposes evaluating anthropic alterations to historical facades to balance development and conservation [7].

This research aims to introduce a method utilizing Stable Diffusion for automatically generating traditional Chinese residential architectural styles, using Minnan residential facades in Southern Fujian villages as an example to provide architects with an inspirational and efficient workflow for historical building conservation and renovation. We apply the knowledge-driven method, integrating and analyzing historical documents, architectural

drawings, and historical photos, to better understand the historical value and original state of buildings in the Minnan region, providing a scientific basis for renovation and transformation. Stable Diffusion is a model designed for text-to-image generation tasks, capable of creating detailed images based on text descriptions. Architects can utilize prompt and counter-prompt words to source more inspiration in urban renewal projects, specifically for generating facades of traditional Minnan residences. We initially generate basic line drawings with Grasshopper, which serve as control factors for the ControlNet lineart to regulate the images. Concurrently, we train the related logistic regression and low-rank adaptation (LoRA) models to ensure that architects can adjust the style features of the facade images generated by the Stable Diffusion model, thereby enhancing the controllability of the generated image content. The innovative aspects of this paper are outlined as follows: (1) We utilize the diffusion model to study the facade styles of traditional Chinese residences in urban historical districts. Architects can generate facades of traditional Minnan residences using specific terminologies to gain insightful reference schemes; (2) We also provide architects with an open-source LoRA model containing a dataset of traditional Minnan residential facades; (3) We propose a qualitative evaluation method and a contrastive language–image pre-training (CLIP) score quantitative metric to measure the quality of the generated historical facade images.

The rest of this paper is organized as follows: Literature Review, Method and Materials, Experiments and Results, and Discussion and Conclusion. Each section, respectively, discusses the achievements and shortcomings in the research field, the sample collection and processing in this paper, the research methods combining Grasshopper with Stable Diffusion, the experimental process and results of generating Minnan architectural facades using the ControlNet model and LoRA model, and the significance of integrating Grasshopper and AIGC deep learning workflows for the facades of regional traditional buildings.

2. Literature Review

2.1. *The Demand for Traditional Building Protection for the Generative Building Facade Design Method*

With urban development, traditional architectural ensembles are confronting severe preservation issues. W. Liang, Ahmad, and Mohidin (2023) conducted a systematic literature review of UNESCO and ICOMOS heritage conservation documents, revealing an evolution in the conservation focus from tangible to intangible attributes and advocating for sustainable conservation practices [8]. Traditional Chinese architecture is a vital component of architectural heritage, and residential structures constitute a significant portion of these traditional buildings. Traditional Chinese residences often exist in clusters and have distinct regional stylistic characteristics. Generative design has been widely explored in the conservation and renovation of architectural heritage, aiding architects in design decisions and offering certain applicability to architectural districts and communities. A paper investigates generative design's potential in architectural contexts, presenting a framework that supports exploration beyond traditional models. It highlights a case study of a residential block in Sweden, showcasing the framework's effectiveness in expanding design possibilities [9]. Generative design is always driven by evidence and knowledge of the heritage. One study applies evidence-based medical practices to the virtual restoration of architectural heritage. It proposes a scientific, data-driven framework for restoration, emphasizes rigorous, evidence-based decision-making, and provides a detailed case study of the Bagong House in Wuhan [10]. The facades of residential buildings hold significant value for the streetscape within the community and the overall architectural unity, prompting many studies to develop generative design methods to regenerate traditional districts or architectural group facades [11]. For example, Tang, Wang, and Shi (2019) utilized new data collection and processing technologies to gather information and establish databases, followed by cognitive surveys and morphological analysis to quantify and evaluate historical features, abstract traditional facade elements, quantitatively generate referential facades, and devise widely accepted conservation plans and guidelines [12]. The research

also identifies and evaluates the importance of semantic values in architectural heritage conservation. By analyzing a broad range of documents, it identifies 40 key semantic values, emphasizing the need for multi-level collaboration for holistic conservation [13]. Existing research indicates that the preservation and renewal of traditional Chinese architectural groups still have significant shortcomings, with an ongoing need for substantial efforts in conserving and updating these structures [14]. Generative architectural facade design methods offer insights for preserving the appearance of traditional architectural groups. Stuart Hall discusses the ideologies and cultural representations embedded in cultural products, highlighting how these products influence and shape social perceptions and identity formation. Emerging technologies, such as the internet, have transformed the modes of cultural production and dissemination, presenting new challenges and opportunities [15]. Generative design methods utilize algorithms and parametric design tools to reproduce the forms and details of traditional architecture. This approach preserves the visual characteristics while conveying the cultural and historical significance behind them. But in facade renovation projects, there exists a nuanced complexity and diversity, compounded by a myriad of challenges. A study assessment revealed that certain alternative facade solutions, compared to demolition, could enhance urban landscape integration and preferences [16]. An article re-evaluates façadism, highlighting its potential in adaptive reuse and contemporary architecture despite past criticism, and explores various forms and their implementation in modern practice [17]. In another study, the new construction within the urban fabric of the historical city and the surrounding urban context, which uses modern building systems and materials, and the architectural formation of the facades that are entirely incompatible with those used in the old urban fabric caused visual pollution in Ibb's historical city [18].

2.2. *The Need for Traditional Minnan Residences' Preservation in China*

"Minnan ancient cuo" refers to traditional residences in Southern Fujian located along the southeastern coast of China (as shown in Figure 1), representing a typical example of traditional Chinese residential architecture. "Minnan", which means Southern Fujian, describes the geographic position of this type of residence. In Minnan dialect, "cuo" means house, and "red brick cuo" refers to houses built with red bricks. The deep and unique cultural heritage of Minnan has cultivated the rich architectural expressions of red brick houses, predominantly featuring the three or five spans of the main chamber, often constructed with multi-tiered swallowtail and arched horseback ridges, complemented by geomantically favorable saddle walls (decorated with features called "shuichedu" and "guidai"), and red brick or white stone walls, all while maintaining a high level of craftsmanship amid diverse forms. Exploring the style transfer techniques for residences based on the Stable Diffusion model in this region is valuable for proposing generative renovation methods for traditional Chinese residential facades while preserving the historical and cultural characteristics of the buildings.

2.3. *The Application of GANs to Generate the Facade*

With the continual advancements in artificial intelligence, generative adversarial networks (GANs) have introduced new possibilities in the field of architectural design. This technology learns from a vast array of data samples to automatically generate images with specific styles, providing designers with a wealth of design inspiration and rapid project iteration capabilities. As a machine learning algorithm widely used in the generation of 2D images, GANs have demonstrated significant capabilities in image creation. The core concept of the algorithm involves two neural networks competing against each other to generate realistic images, with subsequent research evolving in various directions. GANs have been widely applied in the field of image generation. Through the adversarial process, they are able to generate high-quality, high-resolution images, showing great potential for traditional architectural facades and urban transformation. For instance, in generating traditional architectural facades, a study created Chinese traditional architectural facades

by labeling the data of samples and elements, high-lighting the potential and development possibilities of traditional facade design methods. A system is also proposed to aid architects by allowing intervention in the image generation process for facade design. It uses text-to-image retrieval to choose a base image, then employs adversarial generation networks to create diverse images from which users can select, fostering idea convergence and divergence [19]. Additionally, GANs can also play a role in urban redevelopment design. Previous works have developed learning tools for urban redevelopment based on GANs, successfully preserving the unique facade designs of urban districts [20,21]. Thus, GANs can, to a certain extent, enhance the efficiency of facade design in the early stages of design for architects. A paper surveys generative adversarial networks (GANs), detailing their data generation capabilities, various applications, and training challenges, and reviews proposed solutions for stable training [22]. A systematic review analyzes recent GAN models in image segmentation, identifying applications across multiple domains, and outlining challenges and future research directions based on 52 selected papers [23]. A study surveys advancements in GAN design and optimization to address training challenges, proposing a new taxonomy for solutions and discussing various GAN variants and their relationships, highlighting promising research directions [24].



Figure 1. Each part of a Minnan residence.

However, despite the great potential of GANs in architectural design, there are still challenges during the training process. These challenges include network architecture design, the choice of objective functions, and the appropriateness of optimization algorithms. Some studies have noted issues such as mode collapse and instability during the training process [25], and training a high-fidelity GAN model requires a large amount of sample data, yet existing training methods usually require substantial data to achieve good results [26]. Bachl and Ferreira (2019) critique current generative adversarial networks (GANs) and conditional GANs (CGANs) for their unsuitability for generating images of nonexistent buildings. It introduces a custom architecture proven through experiments to perform better in generating the architectural features of major cities [27]. Addition-

ally, straightforward pruning could risk losing critical information, and excessive sparsity might lead to model collapse [28]. Some studies also indicate certain difficulties during usage [29–31]. Therefore, training an effective GAN model requires not only a large amount of sample data and neural architecture engineering but also some skills [32].

2.4. Stable Diffusion

Compared to GANs, diffusion models started later in the field of image generation but have achieved more significant accomplishments in various aspects [33]. Ho, Jain, and Abbeel (2020) modified the diffusion model, proposing the denoising diffusion implicit models (DDIMs), which significantly sped up the sample generation process while maintaining high sample quality, being 10 to 50 times faster than the traditional denoising diffusion probabilistic models (DDPMs) [34]. This model generates clear data by learning the reverse diffusion process, thus further producing high-quality images [35]. This article proposes a text-to-image architecture, VQ-Diffusion, capable of generating more complex scenes and producing high-quality images in both conditional and unconditional settings [36]. Rombach et al. (2022) explored the potential of latent diffusion models (LDMs) for high-resolution image synthesis, emphasizing their efficiency and quality in generating detailed architectural visualizations [37]. These improvements further validate the effectiveness of diffusion models in semantic image synthesis, enhance their performance, and demonstrate improved visual quality in generated images. Although single-step model samples are not yet competitive with GANs, they are superior to previous likelihood-based single-step models. Future research may fully bridge the sampling speed gap between diffusion models and GANs without sacrificing accuracy [33]. Moreover, Kim and Ye (2021) further proposed the DiffusionCLIP model, which not only speeds up computation but also achieves nearly perfect inversion with fewer limitations in inversion and image processing compared to GANs [38]. The gap in computational speed between diffusion models and GANs is continuously narrowing while also diverging in terms of application scope. A paper explores the use of Stable Diffusion models for generating detailed and visually appealing building facades, integrating technology with cultural heritage in architecture. It compares the performance of diffusion models with GANs, showing superior results in detail and quality [39]. Therefore, the application of the diffusion model in the field of architecture has great potential.

Stable Diffusion is a text-to-image technology based on the diffusion model, developed by combining large-scale training data and advanced deep learning technologies, featuring more stable characteristics and greater adaptability. Stable Diffusion's most notable advantage is its ability to save time and effort [40]. In terms of learning, Stable Diffusion is known for its user-friendly interface, allowing even those who have never used artificial intelligence to easily generate high-quality images [40]. Users can quickly generate the desired images by simply providing a text description. In terms of application, Stable Diffusion not only saves time but also allows designers to quickly and conveniently experiment with different ideas and concepts. Additionally, Stable Diffusion demonstrates good stability and controllability in style transfer, providing designers with a broader creative space and encouraging its widespread use. A study that explores the use of text-to-image generators in the early stages of architectural design, involving architecture students, shows that these tools can foster creativity and idea discovery [41]. Stable Diffusion also offers substantial improvements over earlier versions by using a larger model architecture and novel conditioning schemes, competing well with leading image generators [42]. Therefore, Stable Diffusion can be more quickly adopted by designers and generate higher-quality images.

However, diffusion models have certain limitations in terms of generation stability and content consistency [43]. Specifically, when new concepts are introduced, diffusion models may forget previous text-to-image models, thus reducing their ability to generate high-quality images from the past. Recent studies indicate that LoRA models have the potential to enhance the accuracy of image generation in Stable Diffusion. This article proposes a content-consistent super-resolution method to enhance the training of diffusion

models (DMs), thereby increasing the stability of image generation. Smith et al. (2023) introduced the C-LoRA model to address past issues in continuous diffusion models [44]. Additionally, a study proposed the LCM-LoRA as an independent and efficient neural-network-based solver module for rapid inference with minimal steps on various fine-tuned Stable Diffusion models and LoRA models [45]. Yang et al. (2024) introduce Laplace-LoRA, a Bayesian method applied to low-rank adaptation (LoRA) of large language models, improving calibration and mitigating overconfidence by estimating uncertainty, especially in models fine-tuned on small datasets [46]. Therefore, the use of LoRA models can generate images with a stable style while significantly reducing memory usage and can fine-tune images to further meet designers' needs for architectural renovation.

Text descriptions often do not accurately generate final results, nor can they understand complex texts. ControlNet is a neural network architecture designed to add spatial conditional controls to large pre-trained text-to-image diffusion models (such as Stable Diffusion) [47]. It can precisely control the generated images using conditions such as edges, depth, and human poses [48]. Uni-ControlNet allows the use of different local and global controls within the same model, further reducing the fine-tuning cost and model size and enhancing the controllability and composability of text-to-image conversion. ControlNet-XS shows advances in controlling pre-trained networks such as Stable Diffusion for text-to-image generation, focusing on efficient and effective adaptations for image-guided model control, and helping generate street scenes [49]. Therefore, we have incorporated the ControlNet neural network architecture into the model, using line art to control architectural outlines, thereby minimizing excessive modifications to traditional buildings and exploring the efficiency of models combining line art control conditions with ControlNet. By adding extra conditions, ControlNet enables better control over the generated images. Using ControlNet's lineart model allows for the exploration of different architectural facade styles and assessing the impact of different weights on the generation results. Using diffusion models to generate architectural facades offers more options for preserving the appearance of historic districts.

3. Method and Materials

3.1. Methodology

This study aims to propose a method of identifying and collecting historical architectural facade images from specific urban areas and further applying them to urban renewal and transformation to ensure the continuity of the city's historical context. The workflow, after experimentation and optimization, is shown in Figure 2. Initially, based on the field survey elevation drawings in books on Minnan residential architecture by Chinese publishers, simplified line drawing facades are generated using Grasshopper (based on Rhino 7.0), serving as the ControlNet in Stable Diffusion. ControlNet enhances Stable Diffusion by providing precise control over the image generation process. In this case, it guides the diffusion process with line drawing facades, which serve as additional constraints, improving the accuracy, consistency, and overall quality of the generated building facades. Then a series of traditional ancient residence facade images matching the expected style are photographed from traditional Minnan villages, and all images are processed. Subsequently, elements on each facade image are distinguished and labeled with text, forming a training dataset for the LoRA model. LoRA (low-rank adaptation) in Stable Diffusion is used to fine-tune the model efficiently and ensure that the images generated by Stable Diffusion exhibit the stylistic characteristics of the image dataset used to train the LoRA model. Lastly, based on the ControlNet and LoRA models, architects can automatically generate facade images of specific historical styles using the Stable Diffusion model based on text or reference images.

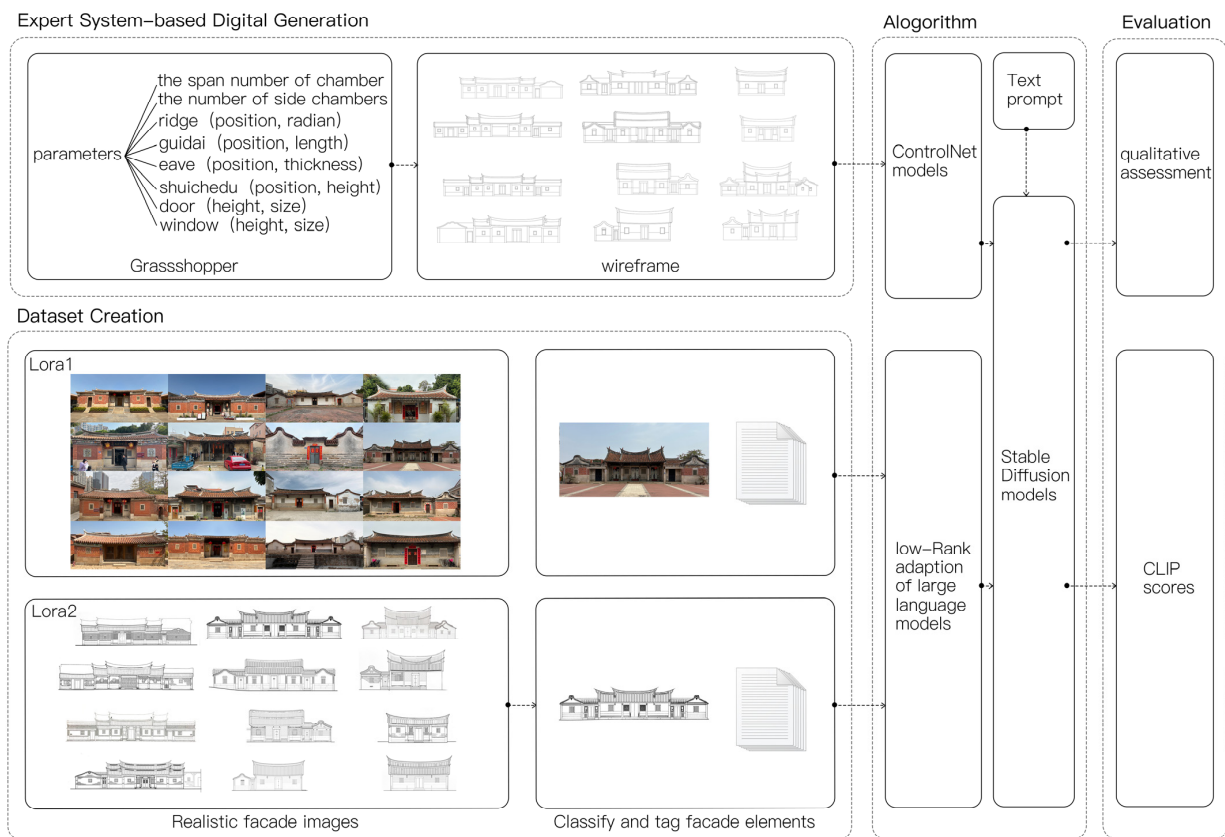


Figure 2. Methodology workflow.

3.2. Image Dataset and Dataset Processing

This study focuses on traditional Minnan residential groups, collecting existing photos of Minnan ancient houses that are distributed in Duishan Village, Houxi Village, and Lunshang Village in Xiamen's Jimei District; Xinan Village, Qinqiao Village, and Haicang Village in Xiamen's Haicang District; and Wulin Village and Wudian City in Quanzhou, showcasing traditional Minnan ancient house facades. A dataset of arcade facade images with paired text labels has been constructed, comprising 34 images of Minnan ancient house facades. Each image was preprocessed, and text labels were generated using Deepbooru, followed by manual modifications based on image characteristics. Figure 3 shows the dataset after preprocessing, which includes real photo correction, cropping to a rectangle, and uniform adjustment to a resolution of 768×384 pixels. Based on the generated text labels and classifications of feature elements of Minnan ancient house facades, we have applied LoRA model trigger labels and specific text labels to each image. Specific text labels include the external environment of Minnan ancient house facades, swallowtail ridges, and the number of the side chamber on either side. Figure 4 displays examples of Minnan ancient house facades and their corresponding text labels.

3.3. The Usage of Grasshopper to Form a Simplified Line Draft

The lineart model in ControlNet serves as a preprocessor for facade images, capable of controlling the form of the LoRA model. From an architectural standpoint, Minnan residential facades are highly typified, commonly featuring the main chamber with three or five spans and side chambers on both sides. Based on survey drafts of Minnan ancient house facades from the literature, simplified line drafts of Minnan ancient house facades covering basic elements are generated with the aid of Grasshopper, including features such as "guidai" (position, length), eaves (position, thickness), main ridges (position, angle), the number of side chambers (one on each side), "shuichedu" (position, height), the span number of the main chamber, doors (height, size), and windows (height, size), as shown

in Figure 5. Due to translation barriers, the terminology of Chinese architectural heritage has not been effectively aligned with that of world architectural heritage. Some Chinese architectural elements and structures lack specific professional terms in the global context. For instance, the term “hucuo” in traditional Minnan residences is directly translated into English as “protective chamber”, whereas it actually signifies a “side chamber”. Similarly, “shuichedu” and “guidai” are decorative elements on building facades, derived from the Minnan dialect, lacking corresponding terms in the shared terminology of world architectural heritage. Hence, in this manuscript, they are represented in Chinese phonetics. Grasshopper allows for the adjustability of each facade element’s dimensions, providing flexibility and better control over Stable Diffusion training.



Figure 3. Image collection.

The creation of the Minnan ancient house facade begins with the line of the front chamber, from which three branches generate three parts of the building (as shown in Figure 6). First is the main chamber of the facade; second is the corridor of the facade; and third is the main chamber of the facade. To generate these three parts, components such as expression, extend curve, divide curve, and line SDL are used to write the generation logic, then baked to form the line-drawn facade (as shown in Figure 7). Next, taking the main chamber as an example, the generation of this part of the facade is described. Firstly, the line of the eave of the facade is generated based on the height and thickness input by the number slider. Secondly, based on the line of the eave, the lines of the structure and ridge of the facade are generated according to the inputted height and span of the main

chamber and the position and radian of the ridge. The generation of the swallowtail ridges and saddle walls relies on the employment of the following curve formulas:

$$\begin{cases} z = x^2 \\ z = x^2 - yx + 1/2(y = nx, 1 \leq y \leq 2) \\ z = -3.5\cos(x^2 - yx) + 3(y = nx, 1 \leq y \leq 2) \end{cases} \tag{1}$$



Figure 4. Image tags.

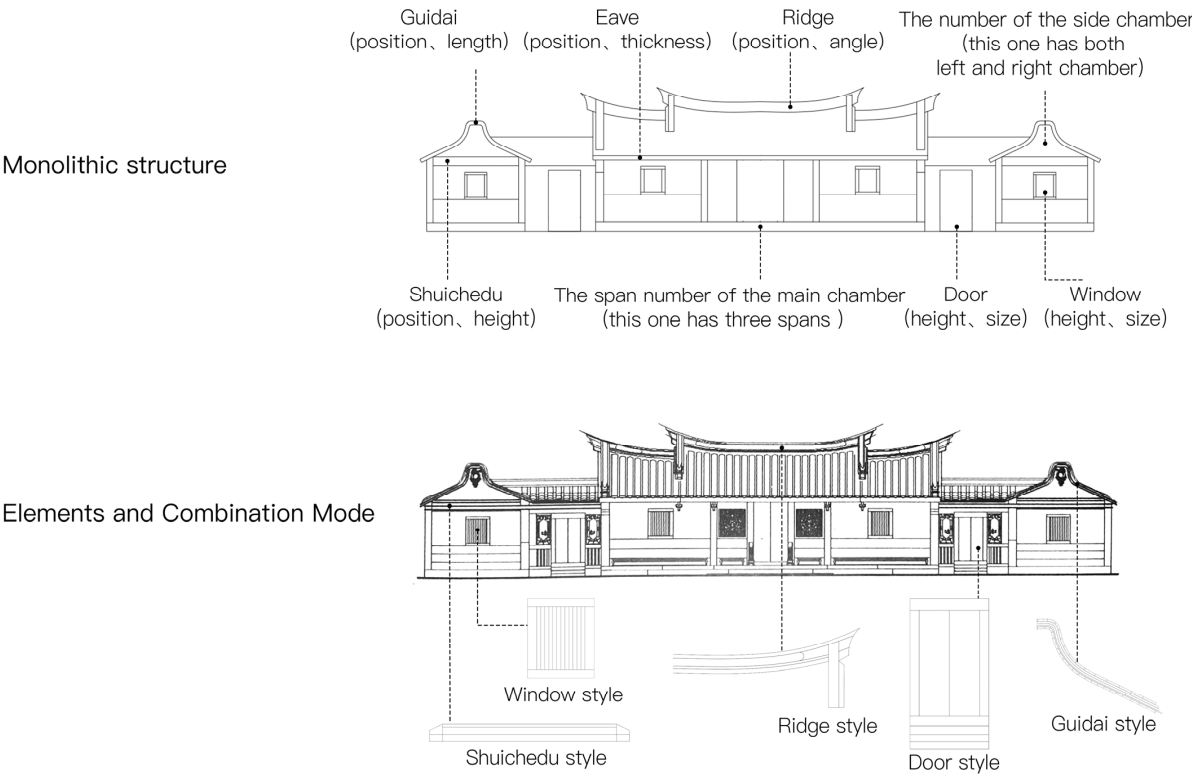


Figure 5. Structure of elements of the facade of Minnan residences.

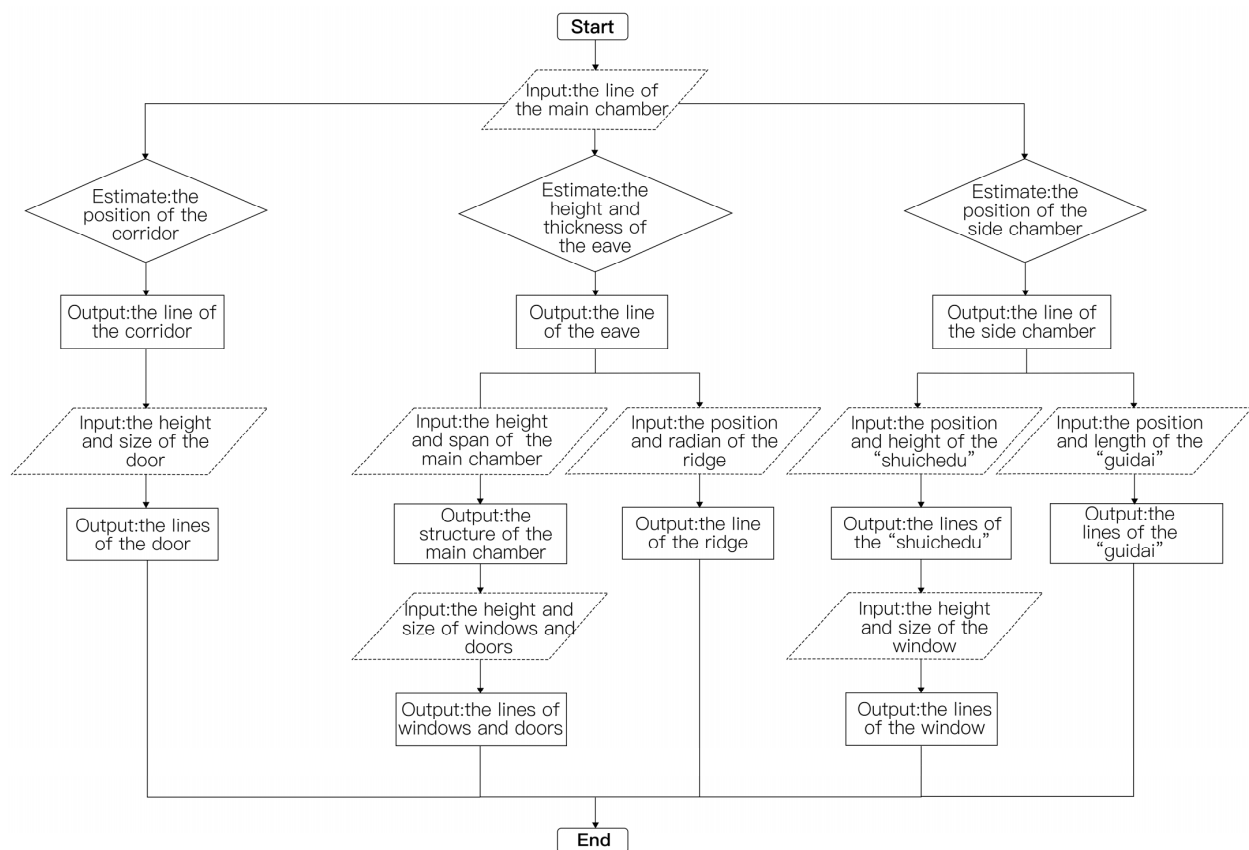


Figure 6. The flow diagram of the architecture facade generation.

The base curve is divided by equidistant points on it, and the points are moved to new places in the z coordinate direction, according to the curve formulas. The variable ‘n’ represents the number of points taken on the curve, while ‘x’ and ‘y’ help control the domain and step of the curves. Lastly, based on the lines of the structure of the main chamber, the lines of windows and doors are generated according to their height and size.

3.4. The Usage of Stable Diffusion to Preprocess Line Drafts

This study primarily utilizes Stable Diffusion as a tool for generating facades in the style of urban historical districts. Stable Diffusion is a latent diffusion model capable of generating new images using prompts and redrawing existing images. To ensure the accuracy, diversity, and efficiency of generated images and to more effectively and finely guide the renovation of traditional architectural areas with historical features, this study employs the LoRA and ControlNet models to assist in generating images of traditional Minnan residential facades. The LoRA model is a pre-trained model used for fine-tuning larger models, which can be trained more quickly with smaller datasets. ControlNet is a neural network architecture that enhances pre-trained image diffusion models based on image prompts and control conditions provided by users. The LoRA model, assisted by the canny model in ControlNet, effectively provides conditions for training Stable Diffusion, better controlling the edges and forms of the architecture during training. Figure 8 shows the results of processing images using the lineart model in ControlNet. Therefore, using the LoRA model trained on a custom Minnan ancient house facade image dataset and the ControlNet model with simple line drawings generated by Grasshopper as conditional inputs, Stable Diffusion can accurately and faithfully generate Minnan ancient house facades designed by architects using prompts.

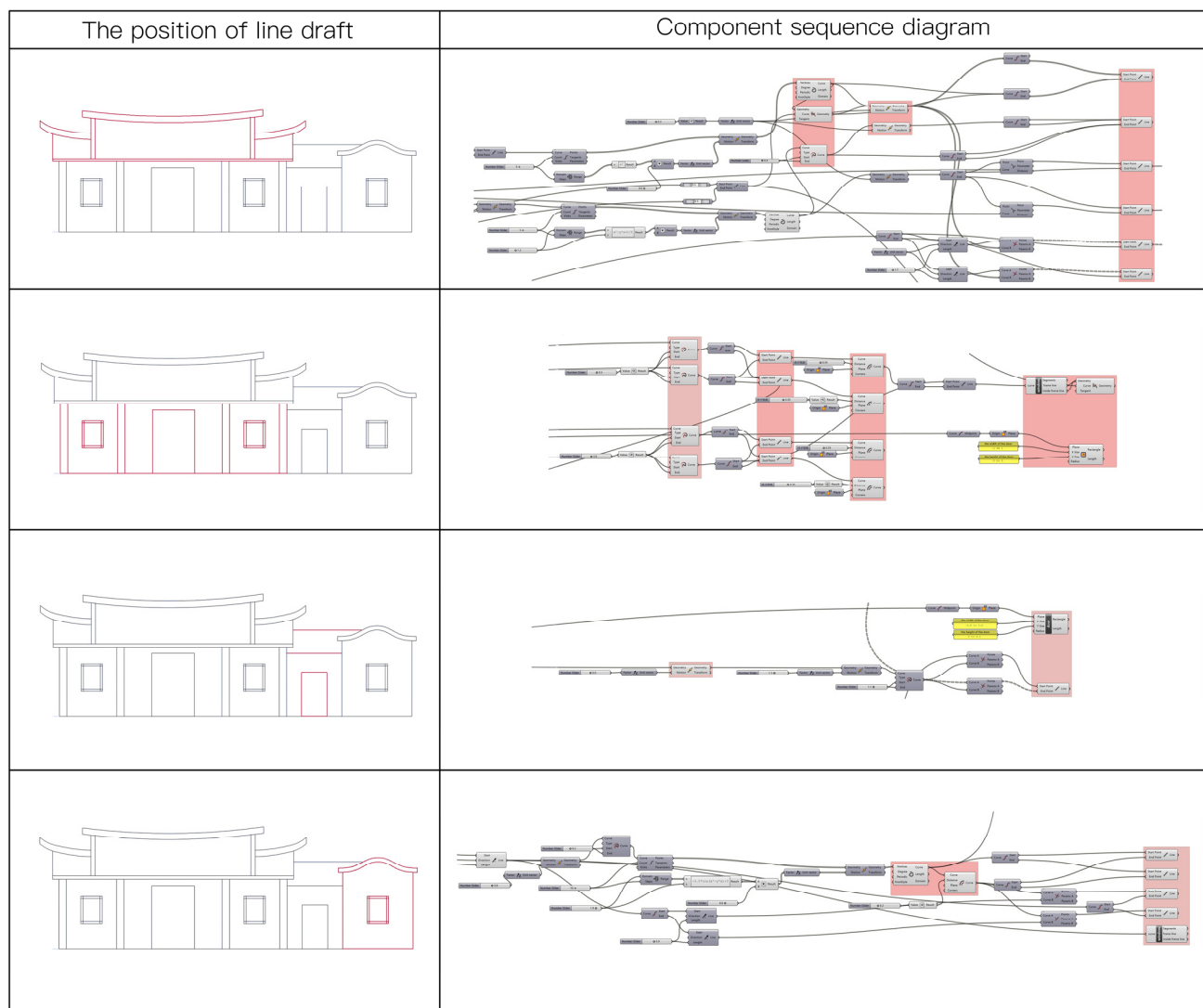


Figure 7. Component sequence diagrams for different positions of line draft (the component in the red frame is the output part).

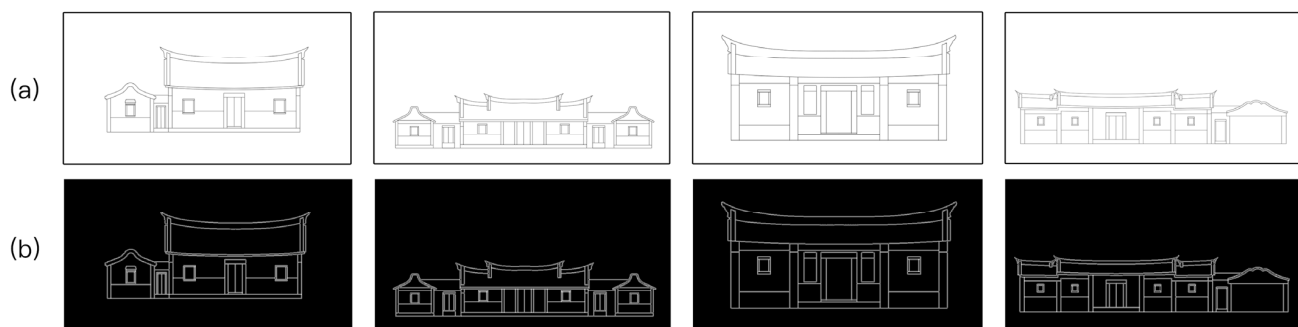


Figure 8. ControlNet canny model processing: (a) the line draft generated by Grasshopper; (b) the line draft processed by the ControlNet canny model.

4. Experiments and Results

4.1. Arcade Facade LoRA Model

We based our work on a dataset of traditional Minnan ancient house facade images and a dataset of detailed line drawing images of Minnan ancient houses, using the ChilloutMix_Chilloutmix-Ni-pruned-fp32-fix.safetensors base model, to train an image

LoRA model named “gucuolimian” and a line drawing LoRA model named “minnangu-cuo” on an NVIDIA GeForce RTX 3060 graphics card. Figure 9 shows the imaging effects of Stable Diffusion under the influence of these two LoRA models.

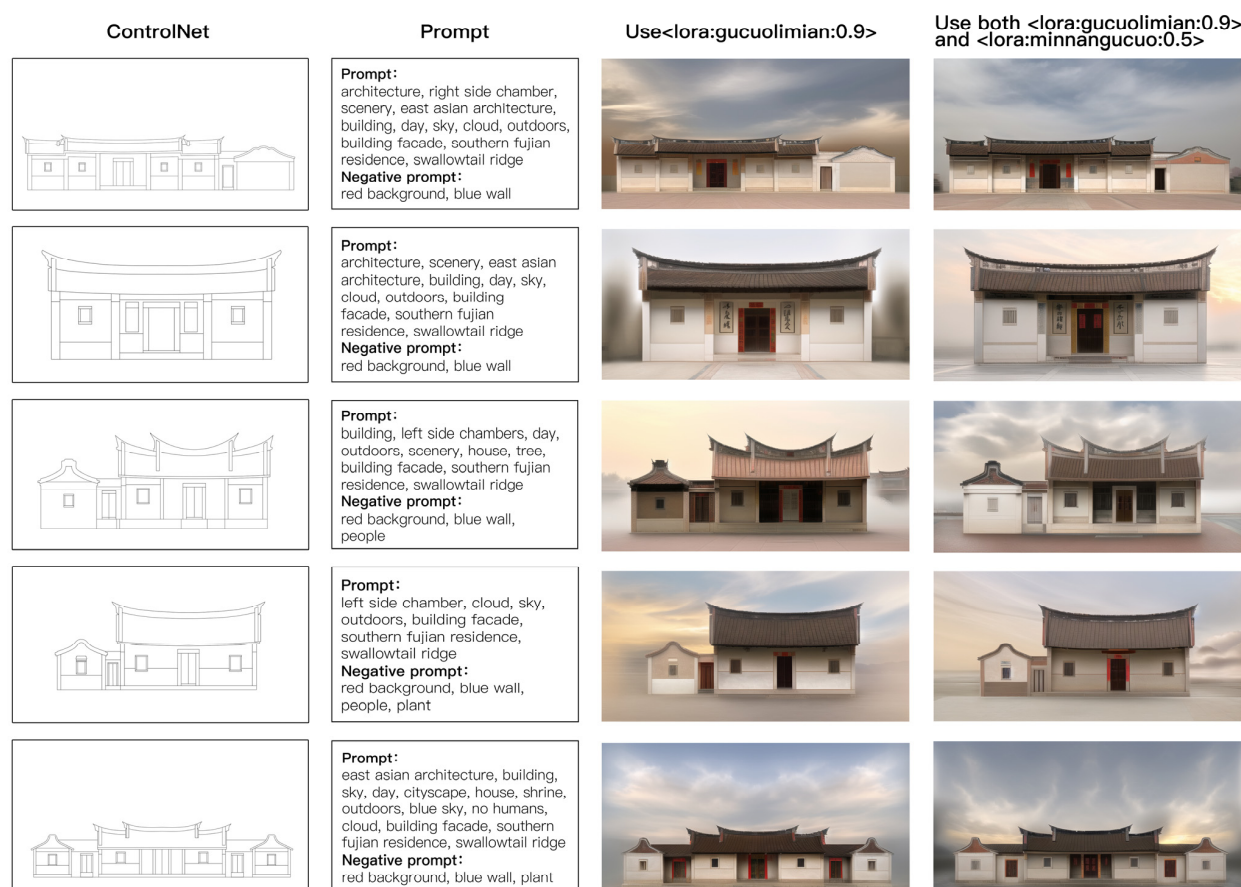


Figure 9. Training results of 2LoRA models.

Based on different simple line drawings and corresponding prompts, we tested weights between 0.7 and 1 and found that a weight of 0.9 produced the best facade generation results. Initially, we attempted to solely control the Stable Diffusion model with the “gucuolimian” LoRA model (whose dataset consists of real images of traditional Minnan ancient houses) at a weight of 0.9. The generated facade generally met the characteristics of Minnan ancient houses, but the details in the roof tiles and decorative bands called “guidai” were disordered and did not accurately reflect reality. After applying the “minnangucuo” model (whose dataset consists of detailed line drawings of traditional Minnan ancient houses) at a weight of 0.5, the accuracy of the roof tiles and decorative bands significantly improved, making the overall image more consistent with the traditional Minnan residential facade appearance.

4.2. Assessment System

4.2.1. Qualitative Evaluation

We conducted a qualitative assessment of the generated Minnan residential facades using the analytic hierarchy process (AHP). The AHP is a hierarchical weighting decision analysis method. Its basic principle involves breaking down the decision-making problem based on its nature and the ultimate goals into several different factors. These factors are then organized based on their interrelations and hierarchical positions to form a multi-level structural model. Weights are then applied to determine the priority order between higher and lower levels, thus simplifying the decision-making process. Based on the objective of generating accurate and complete Minnan residential architectural facades, we established

five criteria for evaluating the quality of the facade images: visual authenticity, cultural accuracy, creativity and beauty, technology quality, and overall picture coordination. These criteria were further subdivided into smaller categories under each main category to assess the architectural facade styles of traditional Chinese Minnan building groups.

According to the AHP, the quality of Minnan residential facade images generated by Stable Diffusion can be measured using the following dimensions (as shown in Table 1):

Table 1. Qualitative evaluation system of the generated traditional Minnan residential facade.

Criteria	Weight	Further Criteria	Maximum Limit
V isual authenticity	30	T he authenticity of details	13
		C olor accuracy	12
		L ighting effects	5
C ultural accuracy	30	A rchitectural style	12
		D ecorative elements	10
		A rchitectural layout	8
C reativity and beauty	10	I nnovative elements	5
		A esthetic design	5
T echnology quality	15	C larity	5
		N oise level	5
		G eneration errors	5
O verall picture coordination	15	E lement coordination	4
		E nvironmental integration	8
		V isual focus	3

The first is visual authenticity, which is the similarity of the generated images to the real-world facades of Minnan residences. This includes the authenticity of details, color accuracy, and lighting effects. Detail authenticity refers to whether the details of architectural elements like tiles, swallowtail ridges, and stone carvings are clear and realistic. Color accuracy indicates whether the colors match those of real Minnan residential facades, and natural light refers to whether the effects of natural and artificial light on the facade are realistic.

The second is cultural accuracy, which refers to the accuracy with which the images depict the characteristics of Minnan residences (such as style and decorative elements). Through the expert evaluation and analytic hierarchy process, the factors are evaluated through a survey of Minnan architectural research experts and construction experts in China. Based on experts' suggestions, these dimensions include architectural style, decorative elements, and architectural layout. Architectural style refers to whether the building's characteristics, such as saddle walls and swallowtail ridges, are accurately reflected. Decorative elements refer to whether doors, windows, and carvings (including "guidai" and "shuichedu") reflect the characteristics of Minnan culture. Architectural layout refers to whether the overall layout of the building facade conforms to the traditional layout pattern of Minnan residences, with the main chamber in the middle and side chambers on both sides.

Third is creativity and beauty, which refers to the performance of the image in innovation and aesthetic design, including innovative elements and aesthetic design.

Innovative elements refer to the integration of novel design elements while maintaining the traditional style of Minnan residences. Aesthetic design refers to whether the overall visual effect is harmonious and beautiful and whether it possesses artistic expressiveness.

Fourth is technology quality, which pertains to the technical execution quality of the generated images, including clarity, noise level, and generation errors. Clarity refers to whether the image is clear and the details are easily discernible. Noise level refers to the presence of unnecessary noise or artifacts in the image. Generation errors refer to checking for significant generation mistakes, such as inconsistencies in the architectural structure.

Fifth is overall picture coordination, which evaluates the coordination and consistency among different parts of the image, including element coordination, environmental integration, and visual focus. Element coordination refers to whether different architectural elements are consistent in style and color. Environmental integration refers to whether the architecture naturally blends into its environmental background. Visual focus refers to whether the image guides the observer's attention to the focal points when viewed.

We collected qualitative metrics for line art1, line art2, line art3, line art4, and line art5 and created a radar chart. (as shown in Figure 10). It is observed that for visual authenticity and cultural accuracy, the facade quality generated by superimposing the “minnangucuo” LoRA model (trained from line drawings) with a weight of 0.5 on the “gucuolimian” LoRA model (trained from images) with a weight of 0.9 (marked as 2LoRA in the diagram) is generally higher than that generated by the “gucuolimian” LoRA model alone (marked as 1LoRA in the diagram). For the indicators of creativity and beauty, technology quality, and overall picture coordination, influenced by the current level of Stable Diffusion technology, there is not much difference in the quality of Minnan architectural facades generated under the conditions of 2LoRA and 1LoRA. It is evident that in terms of displaying the similarity between generated images and the real-world facades of Minnan residences, as well as the accuracy in depicting characteristics of Minnan residences (such as style and decorative elements), the combined generation approach of two LoRAs (one is trained from images and the other is trained from line drawings) has significant advantages.

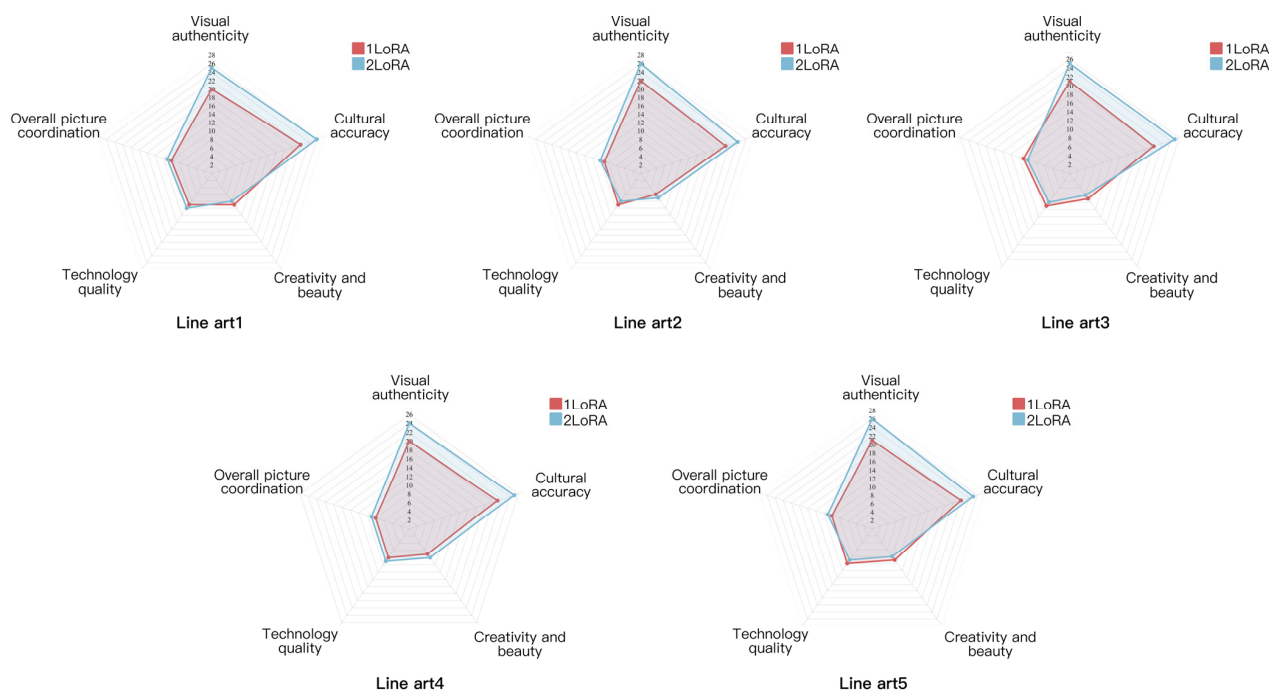


Figure 10. Radar images of qualitative evaluation.

4.2.2. CLIP Score

CLIP scoring is an evaluation method used to measure the relevance between image content and textual descriptions. Developed by OpenAI, CLIP is a multimodal model that can understand the relationships between images and text. The core of this method lies in its ability to pre-train on a vast amount of image–text pairs, mapping images and text into the same vector space to measure their similarity.

CLIP scores range from 0 to 1, with higher CLIP scores indicating greater consistency between the model-generated content and its description. In this study, for different line drawings, we tested facade images generated by the “gucuolimian” LoRA model at a weight of 0.9, as well as those produced by layering the “gucuolimian” LoRA model at a weight of 0.9 with the “minnangucuo” LoRA model at a weight of 0.5, using CLIP to

calculate the semantic similarity of 10 generated images to Minnan residential facades, resulting in 100 ($5 \times 10 \times 2$) CLIP scores. Figure 11 shows the box plot of the actual CLIP scores.

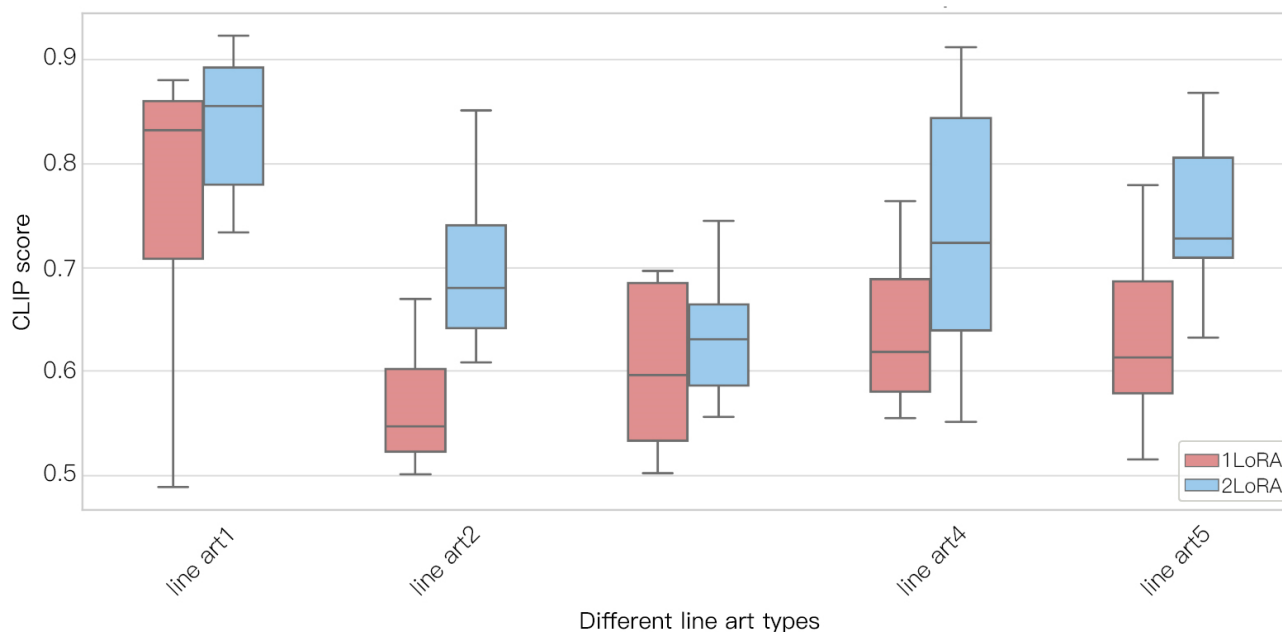


Figure 11. CLIP scores across pictures generated by different LoRAs.

In both experiments, we used the same seed to further analyze model performance under different weights to achieve more stable and predictable results. We compared the CLIP scores of the “gucuolimian” LoRA model trained from images at a weight of 0.9 (labeled as 1LoRA in the figure) with those of the same model superimposed with the “minnangucuo” LoRA model trained from line drawings at a weight of 0.5 (labeled as 2LoRA in the figure) and found that using two LoRAs (light blue) consistently yielded higher median values in various line arts compared to using a single LoRA (light red), indicating that the combined action of the image-based LoRA model and line drawing-based LoRA model can achieve higher measurement values or performance indicators. Images generated under the condition of superimposing the “minnangucuo” LoRA model trained from line drawings at a weight of 0.5 on the “gucuolimian” LoRA model trained from images at a weight of 0.9 were more consistent with textual descriptions of Minnan residential facades. This indicates that when generating Minnan architectural facades, the combined effect of the LoRA models trained from images and the LoRA models trained from line drawings can provide higher-quality generated results and align the semantics of images and texts effectively.

5. Discussion and Conclusions

In this study, based on the knowledge- and data-driven method, we propose a workflow that can generate simplified line drawings through Grasshopper with varying parameters as needed and accurately produce the desired images of traditional Minnan residential facades using these line drawings, ControlNet, and LoRA models, demonstrating Stable Diffusion’s precise understanding and execution of architectural facade elements. Based on Stable Diffusion, this study confirms that simplified Minnan line drafts associated with the LoRA model trained from images and the LoRA models trained from line drawings have a good imaging effect and improve the quality of the generated results. Drawing on historical insights, we employ our technical expertise to develop more accurate and appropriate methods for the restoration and alteration of buildings. The results of this study show great potential for knowledge-driven and diffusion model-based methods for generating

historical building facades in the renovation of traditional architectural complexes and districts, capable of generating regional traditional architectural facades that meet architects' requirements for facade style, size, and form based on existing images and descriptive prompts. Our research contributions can be summarized in four areas: (1) Our proposed workflow can successfully and efficiently generate high-fidelity facade images that conform to the overall historical architectural style of the study area, providing architects with inspiration and options in the early design stages of urban renewal and renovation. (2) Our approach adapts to the inherent constraints of the historical building renovation process, allowing architects to effectively control the generated results based on current photos, hand-drawn sketches, and verbal descriptions. (3) Our proposed workflow assists architects in virtually reconstructing and repairing historical buildings, using the method of generating images through Stable Diffusion to conduct virtual reconstruction and repair experiments. By simulating different restoration plans, it predicts the appearance and potential impacts of buildings post-restoration. Such experiments can help conservationists better understand the potential impacts of different restoration plans, enabling them to make more informed decisions. (4) In the future, our method may be applied to digital reconstruction and preservation, using Stable Diffusion to digitize historical buildings, including complex structures and details. This can help preserve records of historical buildings, enabling repairs and reconstruction when necessary.

Although we acknowledge these benefits, there are limitations that merit additional investigation. Firstly, one limitation is that the facade image labels processed with Deepbooru differ from reality. A possible explanation for this might be the issues associated with dataset bias and model limitations that Deepbooru may face. It requires manual input of new labels into the AIGC database that it cannot recognize, which affects the efficiency and accuracy of the training model. Secondly, during the process of manually inputting new labels, there are doubts regarding the accuracy of the terms used for labeling objects. For instance, when labeling the side room in traditional Minnan architecture, this study has been unable to demonstrate whether the English translation "side chamber" or the specialized term "hucuo" from the field of traditional Minnan residences in China will yield better results, and the differing impacts of these terms on the training model remain to be verified. Thirdly, while the overall effect of training traditional Chinese architecture is good, the complexity of some decorative components makes the details of the generated facades prone to misalignment and blurring, resulting in worse outcomes compared to more simplistic architectural facades. It is therefore likely that there are more suitable solutions to generate clearer and more accurate images. Fourthly, when this method is applied in the early design stages of urban renewal and renovation, the generated image results need to be professionally analyzed and evaluated before being implemented. If not, the issue of potential risks in automatically generating images that may give the wrong preconditions on which to base a restoration should be addressed, calling for the establishment of more comprehensive evaluation standards and systems for historical facade renovations that correspond with AI technology. Fifthly, the enhancement of human oversight in the utilization of AI-generated content is imperative. In the assessment system, informed people and trained specialists evaluate and achieve its plan towards sustainable and resilient renovation. Moreover, given that the primary users of most Minnan residences are local inhabitants, it is necessary to incorporate the resident perspective into the facade renovation evaluation system. This will enhance the community's proactive engagement in the preservation of Minnan residences, thereby achieving a sustainable, bottom-up approach to the comprehensive protection of these traditional residences. Sixthly, the criteria related to "creativity and beauty" lack clarity regarding the evaluation of this category. Generative design can integrate elements of traditional architecture into new projects, ensuring harmony with existing traditional architectural groups. This method allows for the flexible incorporation of traditional features into new designs, preserving cultural identity while meeting modern functional requirements. However, there are currently no clear metrics for assessing innovation in the reconfiguration of historical elements. It remains to

be determined whether this innovation pertains to the architectural function corresponding to the facade or to the form and appearance.

Author Contributions: Conceptualization, S.X. and J.Z.; methodology, S.X., J.Z. and Y.L.; data curation, S.X.; writing-original draft preparation, S.X. and J.Z.; writing-review and editing, J.Z. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available on request.

Acknowledgments: We would like to thank the editors and anonymous reviewers for their constructive suggestions and comments, which helped improve this paper's quality.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, B.; Li, L.; Nakashima, Y.; Nagahara, H. Learning Bottleneck Concepts in Image Classification. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 10962–10971.
2. Ma, K.; Wang, B.; Li, Y.; Zhang, J. Image Retrieval for Local Architectural Heritage Recommendation Based on Deep Hashing. *Buildings* **2022**, *12*, 809. [\[CrossRef\]](#)
3. Zhang, J.; Fukuda, T.; Yabuki, N. Automatic Generation of Synthetic Datasets from a City Digital Twin for Use in the Instance Segmentation of Building Facades. *J. Comput. Des. Eng.* **2022**, *9*, 1737–1755. [\[CrossRef\]](#)
4. Wang, B.; Zhang, J.; Zhang, R.; Li, Y.; Li, L.; Nakashima, Y. Improving Facade Parsing with Vision Transformers and Line Integration. *Adv. Eng. Inform.* **2024**, *60*, 102463. [\[CrossRef\]](#)
5. Zou, H.; Ge, J.; Liu, R.; He, L. Feature Recognition of Regional Architecture Forms Based on Machine Learning: A Case Study of Architecture Heritage in Hubei Province, China. *Sustainability* **2023**, *15*, 3504. [\[CrossRef\]](#)
6. Zhang, K.; Zhang, N.; Quan, F.; Li, Y.; Wang, S. Digital Form Generation of Heritages in Historical District Based on Plan Typology and Shape Grammar: Case Study on Kulangsu Islet. *Buildings* **2023**, *13*, 229. [\[CrossRef\]](#)
7. Cucco, P. Heritage impact assessment in UNESCO WHS. An approach for evaluating human-induced alterations in traditional building's facades. In *Transition: Challenges and Opportunities for the Built Heritage, Proceedings of the Conference Colloqui.AT.e 2023, Bari, Italy, 14–17 June 2023*; EdicomEdizioni: Milan, Italy, 2023; pp. 177–192.
8. Liang, W.; Ahmad, Y.; Mohidin, H.H.B. The Development of the Concept of Architectural Heritage Conservation and Its Inspiration. *Built Herit.* **2023**, *7*, 21. [\[CrossRef\]](#)
9. Mukkavaara, J.; Sandberg, M. Architectural Design Exploration Using Generative Design: Framework Development and Case Study of a Residential Block. *Buildings* **2020**, *10*, 201. [\[CrossRef\]](#)
10. Zhang, Z.; Zou, Y.; Xiao, W. Exploration of a Virtual Restoration Practice Route for Architectural Heritage Based on Evidence-Based Design: A Case Study of the Bagong House. *Herit. Sci.* **2023**, *11*, 35. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Zhang, J.; Fukuda, T.; Yabuki, N.; Li, Y. Synthesizing Style-Similar Residential Facade from Semantic Labeling According to the User-Provided Example. In *HUMAN-CENTRIC, Proceedings of the 28th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Ahmedabad, India, 18 March 2023*; Association for Computer-Aided Architectural Design Research in Asia (CAADRIA): Hong Kong, China, 2023; Volume 1, pp. 139–148.
12. Tang, P.; Wang, X.; Shi, X. Generative Design Method of the Facade of Traditional Architecture and Settlement Based on Knowledge Discovery and Digital Generation: A Case Study of Gunanjie Street in China. *Int. J. Archit. Herit.* **2019**, *13*, 679–690. [\[CrossRef\]](#)
13. Taher Tolou Del, M.S.; Saleh Sedghpour, B.; Kamali Tabrizi, S. The Semantic Conservation of Architectural Heritage: The Missing Values. *Herit. Sci.* **2020**, *8*, 70. [\[CrossRef\]](#)
14. Kuang, Z.; Zhang, J.; Huang, Y.; Li, Y. Advancing Urban Renewal: An Automated Approach to Generating Historical Arcade Facades with Stable Diffusion Models. In Proceedings of the Habits of the Anthropocene, 43rd ACADIA Conference, University of Colorado, Denver, CO, USA, 26–28 October 2023; Volume II, pp. 616–625.
15. Hall, S. Cultural Studies and Its Theoretical Legacies. In *Cultural Studies*; Routledge: New York, NY, USA, 1991; ISBN 978-0-203-69914-0.
16. Serra, J.; Iñarra, S.; Torres, A.; Llopis, J. Analysis of Facade Solutions as an Alternative to Demolition for Architectures with Visual Impact in Historical Urban Scenes. *J. Cult. Herit.* **2021**, *52*, 84–92. [\[CrossRef\]](#)
17. Plevioets, B. Juxtaposing inside and Outside: Façadism as a Strategy for Building Adaptation. *J. Archit.* **2021**, *26*, 541–558. [\[CrossRef\]](#)

18. Alwah, A.A.Q.; Li, W.; Alwah, M.A.Q.; Drmouh, A.A.K.; Shahrah, S.; Tran, D.T.; Xi, L.B. Difficulty and Complexity in Dealing with Visual Pollution in Historical Cities: The Historical City of Ibb, Yemen as a Case Study. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *601*, 012045. [\[CrossRef\]](#)
19. Haji, S.; Yamaji, K.; Takagi, T.; Takahashi, S.; Hayase, Y.; Ebihara, Y.; Ito, H.; Sakai, Y.; Furukawa, T. Façade Design Support System with Control of Image Generation Using GAN. *IIAI Lett. Inform. Interdiscip. Res.* **2023**, *3*, LIIR068. [\[CrossRef\]](#)
20. Sun, C.; Zhou, Y.; Han, Y. Automatic Generation of Architecture Facade for Historical Urban Renovation Using Generative Adversarial Network. *Build. Environ.* **2022**, *212*, 108781. [\[CrossRef\]](#)
21. Ali, A.K.; Lee, O.J. Facade Style Mixing Using Artificial Intelligence for Urban Infill. *Architecture* **2023**, *3*, 258–269. [\[CrossRef\]](#)
22. Jabbar, A.; Li, X.; Omar, B. A Survey on Generative Adversarial Networks: Variants, Applications, and Training. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 157. [\[CrossRef\]](#)
23. Aggarwal, A.; Mittal, M.; Battineni, G. Generative Adversarial Network: An Overview of Theory and Applications. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100004. [\[CrossRef\]](#)
24. Saxena, D.; Cao, J. Generative Adversarial Networks (GANs). *ACM Comput. Surv. (CSUR)* **2021**, *54*, 63. [\[CrossRef\]](#)
25. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved Training of Wasserstein Gans. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
26. Saxena, D.; Cao, J.; Xu, J.; Kulshrestha, T. Re-GAN: Data-Efficient GANs Training via Architectural Reconfiguration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 16230–16240.
27. Bachl, M.; Ferreira, D.C. City-GAN: Learning Architectural Styles Using a Custom Conditional GAN Architecture. *arXiv* **2020**, arXiv:1907.05280.
28. Saxena, D.; Cao, J.; Xu, J.; Kulshrestha, T. RG-GAN: Dynamic Regenerative Pruning for Data-Efficient Generative Adversarial Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 4704–4712.
29. Liang, K.J.; Li, C.; Wang, G.; Carin, L. Generative Adversarial Network Training Is a Continual Learning Problem. *arXiv* **2018**, arXiv:1811.11083.
30. Wei, X.; Gong, B.; Liu, Z.; Lu, W.; Wang, L. Improving the Improved Training of Wasserstein GANs: A Consistency Term and Its Dual Effect. *arXiv* **2018**, arXiv:1803.01541.
31. Soviany, P.; Ardei, C.; Ionescu, R.T.; Leordeanu, M. Image Difficulty Curriculum for Generative Adversarial Networks (CuGAN). In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 3463–3472.
32. Kurach, K.; Lucic, M.; Zhai, X.; Michalski, M.; Gelly, S. The Gan Landscape: Losses, Architectures, Regularization, and Normalization. In Proceedings of the International Conference on Learning Representations (ICLR 2019), New Orleans, LA, USA, 6–9 May 2019.
33. Nichol, A.; Dhariwal, P.; Ramesh, A.; Shyam, P.; Mishkin, P.; McGrew, B.; Sutskever, I.; Chen, M. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. *arXiv* **2022**, arXiv:2112.10741.
34. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 6840–6851.
35. Gu, S.; Chen, D.; Bao, J.; Wen, F.; Zhang, B.; Chen, D.; Yuan, L.; Guo, B. Vector Quantized Diffusion Model for Text-to-Image Synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10696–10706.
36. Wang, W.; Bao, J.; Zhou, W.; Chen, D.; Chen, D.; Yuan, L.; Li, H. Semantic Image Synthesis via Diffusion Models. *arXiv* **2022**, arXiv:2207.00050.
37. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695.
38. Kim, G.; Ye, J.C. DiffusionCLIP: Text-Guided Image Manipulation Using Diffusion Models. *arXiv* **2021**, arXiv:2110.02711.
39. Lyu, Z.; Li, Z.; Wu, Z. Research on Image-to-Image Generation and Optimization Methods Based on Diffusion Model Compared with Traditional Methods: Taking Façade as the Optimization Object. In Proceedings of the Phygital Intelligence; Yan, C., Chai, H., Sun, T., Yuan, P.F., Eds.; Springer Nature: Singapore, 2024; pp. 35–50.
40. Yıldırım, E. Text-to-Image Artificial Intelligence in a Basic Design Studio: Spatialization from Novel. In Proceedings of the 4th International Scientific Research and Innovation Congress, Rome, Italy, 3–5 February 2022.
41. Paananen, V.; Oppenlaender, J.; Visuri, A. Using Text-to-Image Generation for Architectural Design Ideation. *Int. J. Archit. Comput.* **2023**, 14780771231222783. [\[CrossRef\]](#)
42. Podell, D.; English, Z.; Lacey, K.; Blattmann, A.; Dockhorn, T.; Müller, J.; Penna, J.; Rombach, R. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. *arXiv* **2023**, arXiv:2307.01952.
43. Sun, L.; Wu, R.; Zhang, Z.; Yong, H.; Zhang, L. Improving the Stability of Diffusion Models for Content Consistent Super-Resolution. *arXiv* **2023**, arXiv:2401.00877.

44. Smith, J.S.; Hsu, Y.-C.; Zhang, L.; Hua, T.; Kira, Z.; Shen, Y.; Jin, H. Continual Diffusion: Continual Customization of Text-to-Image Diffusion with C-LoRA. *arXiv* **2023**, arXiv:2304.06027.
45. Luo, S.; Tan, Y.; Patil, S.; Gu, D.; von Platen, P.; Passos, A.; Huang, L.; Li, J.; Zhao, H. LCM-LoRA: A Universal Stable-Diffusion Acceleration Module. *arXiv* **2023**, arXiv:2311.05556.
46. Yang, A.X.; Robeyns, M.; Wang, X.; Aitchison, L. Bayesian Low-Rank Adaptation for Large Language Models. *arXiv* **2024**, arXiv:2308.13111.
47. Zhang, L.; Rao, A.; Agrawala, M. Adding Conditional Control to Text-to-Image Diffusion Models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 3836–3847.
48. Zhao, S.; Chen, D.; Chen, Y.-C.; Bao, J.; Hao, S.; Yuan, L.; Wong, K.-Y.K. Uni-ControlNet: All-in-One Control to Text-to-Image Diffusion Models. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 11127–11150.
49. Zavadski, D.; Feiden, J.-F.; Rother, C. ControlNet-XS: Designing an Efficient and Effective Architecture for Controlling Text-to-Image Diffusion Models. *arXiv* **2023**, arXiv:2312.06573.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.