buildings

MDPI

Article

Using Artificial Intelligence to Generate Master-Quality Architectural Designs from Text Descriptions

Junming Chen [1,†], Duolin Wang [1,†], Zichun Shao [1,†], Xu Zhang [2], Mengchao Ruan [1], Huiting Li [1] and Jiaqi Li [1,*]

[1] Faculty of Humanities and Arts, Macau University of Science and Technology, Macao 999078, China; jmchen@must.edu.mo (J.C.); dlwang@must.edu.mo (D.W.); 2220015871@student.must.edu.mo (Z.S.); 2220024237@student.must.edu.mo (M.R.); 2220002154@student.must.edu.mo (H.L.)

[2] College of Arts, Nanjing University of Information Science and Technology, Nanjing 210044, China; 20211231007@nuist.edu.cn

* Correspondence: jiqli@must.edu.mo

† These authors contributed equally to this work.

Abstract: The exceptional architecture designed by master architects is a shared treasure of humanity, which embodies their design skills and concepts not possessed by common architectural designers. To help ordinary designers improve the design quality, we propose a new artificial intelligence (AI) method for generative architectural design, which generates designs with specified styles and master architect quality through a diffusion model based on textual prompts of the design requirements. Compared to conventional methods dependent on heavy intellectual labor for innovative design and drawing, the proposed method substantially enhances the creativity and efficiency of the design process. It overcomes the problem of specified style difficulties in generating high-quality designs in traditional diffusion models. The research results indicated that: (1) the proposed method efficiently provides designers with diverse architectural designs; (2) new designs upon easily altered text prompts; (3) high scalability for designers to fine-tune it for applications in other design domains; and (4) an optimized architectural design workflow.

Keywords: architectural design; text to design; design process optimization; design quality; design style; diffusion model

Citation: Chen, J.; Wang, D.; Shao, Z.; Zhang, X.; Ruan, M.; Li, H.; Li, J. Using Artificial Intelligence to Generate Master-Quality Architectural Designs from Text Descriptions. *Buildings* **2023**, *13*, 2285. https://doi.org/10.3390/buildings13092285

Academic Editors: César Martín-Gómez, Ana Sofia Guimarães and Barbara Rangel

Received: 11 August 2023
Revised: 1 September 2023
Accepted: 6 September 2023
Published: 8 September 2023

1. Introduction

1.1. Background and Motivation

Often the icon of a city, excellent architecture can attract tourists and promote local economic development [1]. However, designing outstanding architecture via conventional design methods poses multiple challenges. For one thing, conventional design methods involve a significant amount of manual drawing and design modifications [2–4], resulting in low design efficiency [4]. For another thing, cultivating designers with superb skills and ideas usually proves difficult [4,5], hence low-quality and inefficient architectural designs [5,6]. Such issues in the construction industry warrant urgent solutions.

1.2. Problem Statement and Objectives

Artificial intelligence (AI) has been widely used in daily life [7–10]. Specifically, diffusion models can assist in addressing the low efficiency and quality in architectural design. Based on the machine learning concept, diffusion models are trained by learning knowledge from a vast amount of data [11,12] to generate diverse designs based on text prompts [13]. Nevertheless, the current mainstream diffusion models, such as Stable Diffusion [14], Midjourney [15], and DALL E2 [11], have limited applications in architectural design due to their inability to embed specific design style and form in the generated architectural designs (Figure 1). Considering that styles and shapes are crucial in architectural designs,

improvements to the diffusion models are needed to meet the specific requirements in generating particular architectural designs. This research aims to enhance the controllability and usability of diffusion models to generate master-quality architectural designs of specified design styles.

Figure 1. Mainstream diffusion models compared with the proposed method for generating architectural designs. Stable Diffusion [14] fails to generate an architectural design with a specific style, and the image is not aesthetically pleasing (left panel). The architectural design styles generated by Midjourney [15] (second from the left) and DALL E2 [11] (third from the left) are incorrect. None of these generated images met the design requirements. The proposed method (far right) generates architectural design in the correct design style. (Prompt: "An architectural photo in the Shu Wang style, photo, realistic, high definition").

1.3. Significance of Research

The efficiency and quality of architectural designs have been unsatisfactory [4,16–18]. In this context, this research proposes a method to generate architectural design based on text prompts. This method is mainly used in the conceptual scheme design stage to complete a rough design scheme quickly. It significantly improves design quality and efficiency [19–22]. For one thing, the proposed method can generate many designs from which designers can choose a satisfactory solution without manual drawing, thereby enhancing design efficiency [4,5]. For another thing, the method acquires design experience from the training data and generates a design beyond the learned data or designers [5,6]. Furthermore, introducing AI into architectural design can drive the industry toward intelligent transformation. The studied method employs AI to generate the design and eliminates the heavy reliance on intensive intellectual labor for innovative design and drawing of conventional methods, thus conforming to the future design trend.

1.4. Research Framework

The proposed method builds on an improved diffusion model to generate high-quality architectural designs upon textual prompts. The method involves data collection, loss function definition, diffusion model fine-tuning, and model utilization for design generation. To address the limited training data, we initially curated a new master builder dataset (i.e., MBD-8) with the assistance of professional designers [23,24]. Subsequently, we introduced a novel loss function to fine-tune the diffusion model, allowing it to generate architectural designs with specific stylistic features and improved quality. The fine-tuned diffusion model exhibits the capability to produce diverse architectural designs in batches based on textual prompts. Finally, we summarize a set of guiding words to effectively control the architectural design outcomes, which designers can utilize to exert greater control over the generation process and enhance the controllability of the generated designs. The research framework is presented in Figure 2.

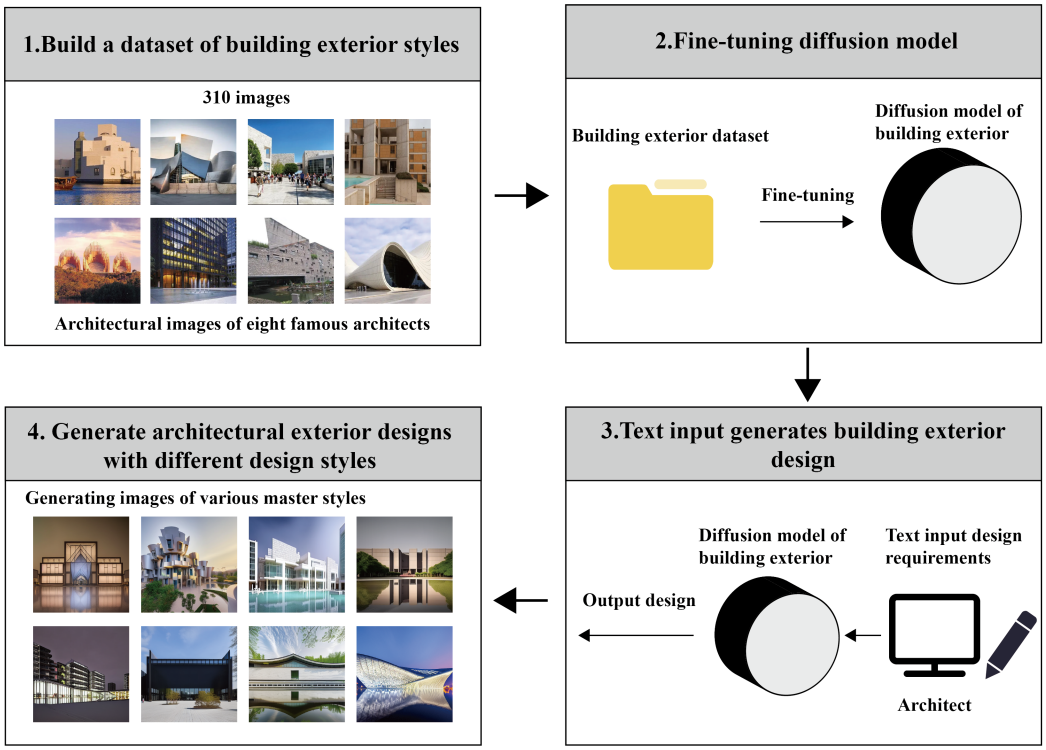


Figure 2. Research framework. In this research, we first collected a dataset of 310 images of architecture designed by eight renowned architects. Subsequently, we trained an improved diffusion model on the built dataset and the newly designed loss function. The enhanced diffusion model can directly generate master-quality architectural designs based on textual prompts. Additionally, controllable modifications can be achieved with the proposed guiding words.

1.5. Main Contribution

The proposed method leverages AI to enhance the conventional design process, thereby replacing the intellectual labor-intensive creative design and drawing work of the past. The findings indicate that the proposed method can generate diverse architectural designs in batches based on textual prompts, significantly improving design quality and efficiency. Figure 3 shows the effect of the proposed method on generating different master-quality architectural designs.

The main contributions of this research are as follows:

1. Proposing a method for generating architectural designs in batches based on textual prompts.
2. Enhancing the capabilities of the diffusion model to generate architectural designs with specified styles.
3. Optimizing the architectural design workflow.
4. Proving the superiority of the proposed method to other mainstream diffusion models for architectural design generation.

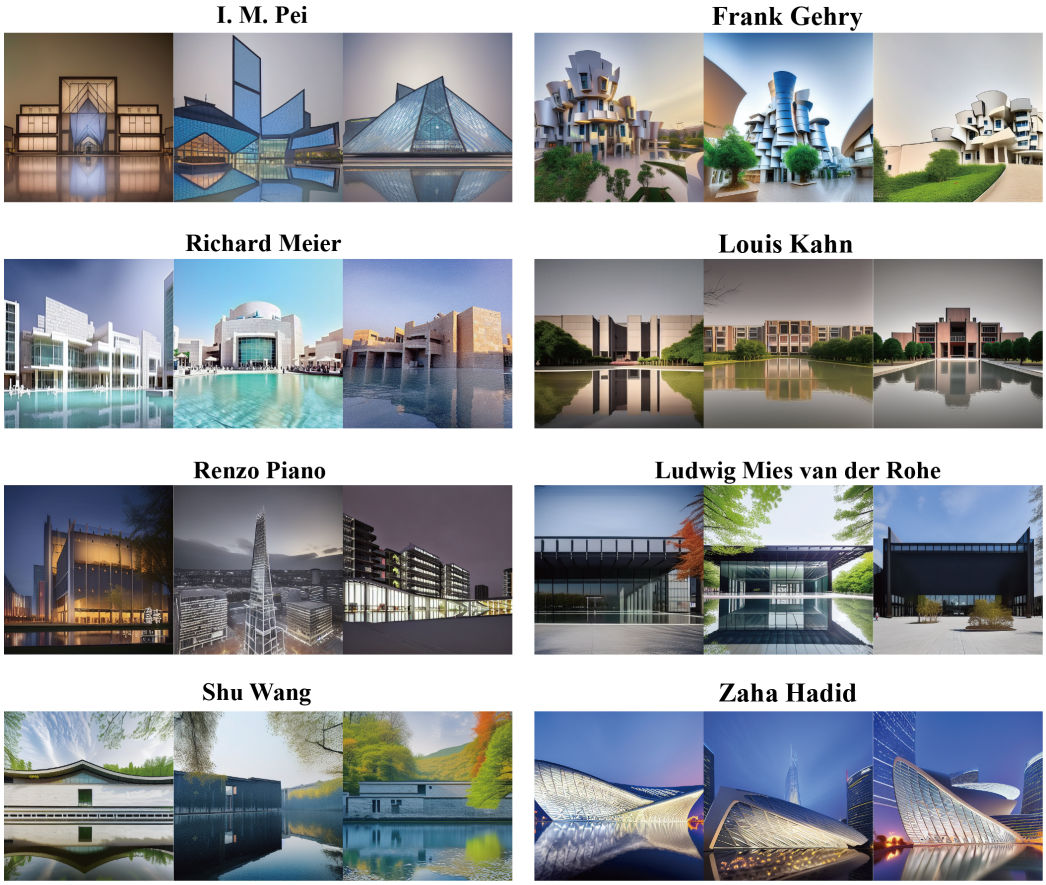


Figure 3. The improved diffusion model proposed in this research directly generates architectural designs with different styles according to text prompts. (Prompt: "An architectural photo in the style of (names of different architects), photo, realistic, high definition").

2. Literature Review

2.1. Architectural Design

Architectural designing relies on the professional skills and concepts of designers. Outstanding architectural designs play a crucial role in showcasing the image of a city [2–4]. Moreover, iconic landmark architecture in a city stimulates local employment and boosts the tourism industry [1,25].

Designers typically communicate architectural design proposals with clients through visual renderings. However, this conventional method has low efficiency and low quality. The inefficiency stems from the complexity of the conventional design process involving extensive manual drawing tasks [2,5], such as creating 2D drawings, building 3D models, applying material textures, and rendering visual effects [26]. This linear design process restricts client involvement in the decision-making until producing the final rendered images. If clients find the design not to meet their expectations upon viewing the final images, designers must redo the entire design, leading to repetitive modifications [2–4]. Consequently, the efficiency of this design practice needs improvement [26].

The low quality of architectural designs is ascribed to the challenges of training an excellent designer and the long process of improving design ability. The lack of design skills in the designer renders it difficult to improve the design quality [4–6]. Yet, enhancing design capabilities is a gradual process, where the designer must continuously learn new design methods and explore different design styles [2,3,6,27–29]. Meanwhile, seeking the best design solutions under complex conditions also poses significant challenges for the designer [2,5].

All these factors ultimately result in inefficient and low-quality architectural designs [2,5]. Therefore, new technologies must be promptly introduced to the architecture industry to address these issues.

2.2. Diffusion Model

In recent years, diffusion models have rapidly evolved into the mainstream image-generation models [19–22,30,31], which could enable designers to acquire images swiftly, thus significantly improving architectural design efficiency and quality [13,32].

The conventional diffusion models comprise the forward and backward processes. In the forward process, noise is continuously added to the input image, transforming it into a noisy image. The backward process aims to recover the original image from the noisy image [33]. By learning the image denoising process, the diffusion models acquire the ability to generate images [32,34]. When the need arises to generate images with specific design elements, designers can incorporate text prompts into the denoising process of the diffusion model to generate consistent images and achieve controllability over the generated results [35–40]. The advantage of using text-guided diffusion models for image generation lies in their simplicity to control the generation of images [12,14,41–43].

Despite the excellent performance of diffusion models in most fields, their applications in architectural design still have room for improvement [35,44]. Specifically, the limitation arises from acquiring vast amounts of Internet data for training, which lack high-quality annotations with professional architectural terminology. As a result, the model fails to establish connections between architectural design and architectural language during the learning process, making it challenging to exert guidance in architectural design generation using professional design vocabulary [45–48]. Therefore, it is essential to collect high-quality architectural design images, annotate them with relevant information, and subsequently, fine-tune the model to adapt them to architectural design tasks.

2.3. Model Fine-Tuning

Diffusion models learn new knowledge and concepts through entire retraining or fine-tuning for new scenarios. Due to the massive cost of whole-model retraining, the need for large image datasets, and the long training time [11,15], model fine-tuning is currently the most-feasible.

There are four standard fine-tuning methods. The first is Textual Inversion [11,36,46,49], i.e., freezing a text-to-image model and only providing the most-suitable embedding vector to embed new knowledge. This method offers fast model training and minimal generated models, but ordinary image generation effects. The second is the Hypernetwork [47] method, i.e., inserting a separate small neural network into the middle layer of the original diffusion model to affect the output. The training speed of this method is relatively fast, but the image-generation effect is average. The third is LoRA [48], i.e., assigning weight to the attention cross-layer to allow the learning of new knowledge. This method can generate models averaging several hundred MB in size with better image-generation effects after medium training time. The fourth is the Dreambooth [45] method, i.e., fine-tuning the original diffusion model as a whole. Using this method, a prior-preservation loss is designed to train the diffusion model and enable it to generate images conforming to the prompt while preventing overfitting [50,51]. Rare vocabulary is recommended when naming new knowledge to avoid language drift due to similarities with the original model's vocabulary [50,51]. This method only requires 3 to 5 images on a specific topic with corresponding textual descriptions to fine-tune for a particular case and match the specific textual description with the characteristics of the input image. The fine-tuned model generates images based on specific topic words and general descriptors [31,46]. As the entire model is fine-tuned using the Dreambooth method, the yielded results are usually the best among these methods.

3. Methodology

This research proposes using an improved diffusion model to generate architectural designs. The method employed a novel architectural design dataset (MBD-8) and a new design style loss function to fine-tune the diffusion model, thus enabling it to generate high-quality architectural designs of different styles in batches. In this way, designers are provided with numerous high-quality design schemes in the design stage, which effectively solves the low efficiency and quality in architectural design.

3.1. Collect Building Datasets

The dataset was built by collecting images from architectural design websites. Specifically, 310 images of high-quality architectural designs by eight renowned architects were gathered. To facilitate design style pairing with the images, each image was annotated with the name of the corresponding architect to represent the specific architectural design style. An overview of MBD-8 is provided in Table 1, listing the number of designs of different architecture types designed by each master architect.

Table 1. Architecture type distribution of different architects in MBD-8.

	I. M. Pei	Frank Gehry	Richard Meier	Louis Kahn	Renzo Piano	Ludwig Mies van der Rohe	Shu Wang	Zaha Hadid
Expo Centers	23	18	17	14	14	11	36	29
Mansions	8	8	3	4	15	13	4	7
Houses	9	16	12	21	9	13	4	2
Total	40	42	32	39	38	37	44	38

3.2. Build the Training Loss Function

In this research, a new composite loss function was proposed, i.e., adding the architectural style to the conventional loss function (Equation (1)) as a loss to form Equation (2). The first part of Equation (2) is the regular loss, which ensures the model learns an architectural style. The second part is the prior knowledge loss, which prevents the diffusion model from forgetting old knowledge while learning new knowledge. Thus, the two parts of Equation (2) ensure that the diffusion model retains the original basic understanding while remembering the architectural style.

The basic diffusion model can be expressed as Equation (1):

$$\mathbb{E} = \frac{1}{N}\sum_{i=1}^{N} ||\hat{Y}_\theta(\alpha_t Y + \sigma_t \epsilon, \boldsymbol{g}) - Y||_2^2, \tag{1}$$

where $\mathbb{E}$ is the average loss, and the diffusion model reduces the loss through training, i.e., denoising the noisy image; $\hat{Y}_\theta$ is a pre-trained text-to-image diffusion model that receives a noisy image vector $\alpha_t Y + \sigma_t \epsilon$ and a text vector $\boldsymbol{g}$ as inputs and predicts a noise-free image; the training process uses a squared error loss to optimize the model and reduce the difference Y between the expected and actual images.

The fine-tuned diffusion is expressed as Equation (2):

$$\mathbb{E} = ||\hat{Y}_\theta(\alpha_t Y + \sigma_t \epsilon, \boldsymbol{g}) - Y||_2^2 + \lambda_w ||\hat{Y}_\theta(\alpha_{t'} Y_{pr} + \sigma_{t'} \epsilon', g_{pr}) - Y_{pr}||_2^2. \tag{2}$$

By adding the architectural style as a loss function to Equation (1), Equation (2) solves the problem that the conventional diffusion models cannot generate architecture designs of specified styles. The first term of Equation (2) is Equation (1), which measures the difference between the image generated after fine-tuning and the actual image used for training. The second term is a loss term that preserves the knowledge of the original model. Smaller differences between the images generated by the fine-tuned model and the original frozen diffusion model indicate better abilities to retain the knowledge of the original model. λ_w

is the weight of the two losses, which is automatically adjusted for the best generation effect. The new loss function incentivizes the model to retain the original diffusion model's knowledge while learning the architectural style so that the fine-tuned diffusion model can generate architectural designs of specified styles.

3.3. Fine-Tuning the Master Builder Model

We used MBD-8 to fine-tune the diffusion model. Specifically, these graphic data labeled with different architectural master design styles were input into the diffusion model with a novel composite loss function for training. Considering the involvement of architectural style loss items in the loss function, the model can learn various architectural design styles during training and bind them to our preset prompts. After that, the designer can generate the specified architectural design style by importing the preset prompts.

3.4. Generating Designs Using Fine-Tuned Models

Designers can easily use the fine-tuned diffusion model to obtain high-quality architectural designs by entering architectural terms and guiding words. The proposed method replaces the conventional design processes of drawing, modeling, and rendering with a generative design approach. With the proposed method, generating a design takes only 10 s on a computer with a graphics card with 12 GB VRAM. Thus, the architectural design workflow is optimized, and the design efficiency and quality are improved. A comparison between the conventional design method and the proposed method is presented in Figure 4.

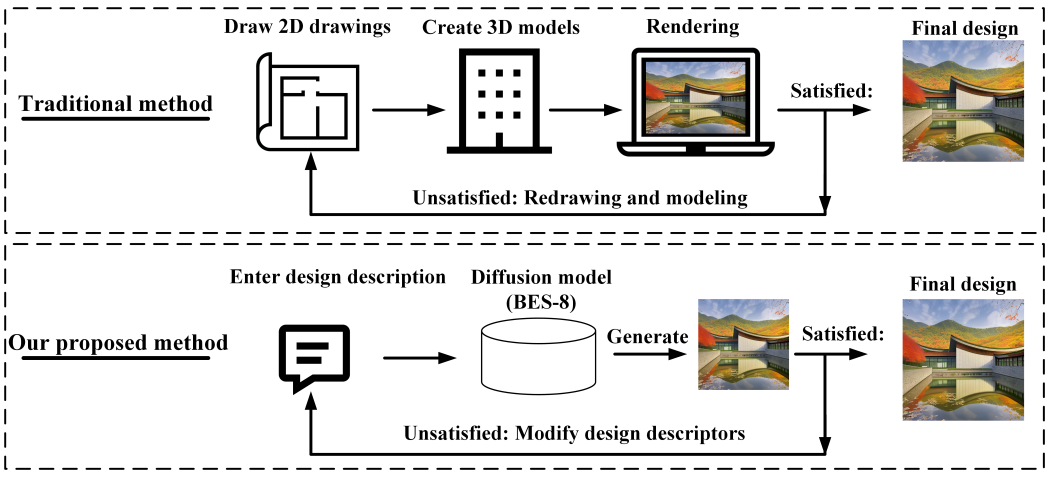


Figure 4. Conventional design methods compared with the proposed method. The conventional design processes include preparing 2D drawings, building 3D models, and producing renderings. The proposed method directly generates design renderings based merely on text prompts. If modifications are required, conventional methods must redo the entire design, while the proposed method only needs a modified prompt to regenerate the design.

3.5. Evaluation Metrics

Evaluating generative architectural design images is challenging. Conventional automated image-evaluation methods only allow image aesthetic or composition assessment [52,53], and the design content assessment is still lacking. Evaluating architectural design content quality requires reasonable evaluation indicators and expert manual scoring. For this purpose, establishing diverse evaluation metrics is necessary. In collaboration with senior architects, this study developed a set of evaluation indexes applicable to generative architectural design, covering the whole and details of generative images and enabling a comprehensive assessment of the design quality. Eight evaluation items were determined, including "overall impression", "architectural details", "architectural integrity", "lighting relationship", "architectural realism", "composition", "architectural background",

and "consistent architectural style". The evaluation used a two-category scoring method, and professional designers scored the images: one point for those that meet the scoring requirements and zero for the others.

Elaborations on the content and importance of each evaluation index were given as follows. We believe that "overall impression" and "consistent architectural style" are the most-critical indicators. Specifically, the former refers to the designer's general perception of the generated image. When the image is considered generally beautiful and has no obvious errors, it is typically feasible; the latter evaluates whether the input text prompts generate a building with the corresponding design style. Consistency of architectural style is essential for generative design. In addition, the "architectural details" can judge whether the design details of buildings in an image are clear and reasonable. "Architectural integrity" refers to the evaluation of the building's integrity. "Lighting relationship" examines the correctness of lights, shadows, and colors in generative images. "Architectural realism" indicates the realism degree of designed buildings. "Composition" can provide insights into the rationality of the building location. "Architectural background" is defined as the authenticity of the generated architectural background.

4. Experiments and Results

4.1. Implementation Details

The fine-tuned diffusion model was implemented based on Pytorch. The experiment platform was a Windows 10 system running on 64 GB of RAM and 12 GB of VRAM. The number of iterations was 30,000 steps, and each training time was four hours. During preprocessing, the input images were automatically cropped to a 512×512 resolution. Mirror flipping was adopted as the data-enhancement method. The learning rate was 2×10^{-6}; the batch size was four; XFormers and FP16 were used to accelerate the calculations.

4.2. Visual Qualitative Assessment

Visual evaluation is essential for assessing the quality of the generated images. The images generated by the proposed method were visually compared with those generated by three mainstream diffusion models, i.e., Stable Diffusion, Midjourney, and DALL E2. The same guiding words were used for the different models to generate images to facilitate the comparison. The generated images are presented in Figure 5.

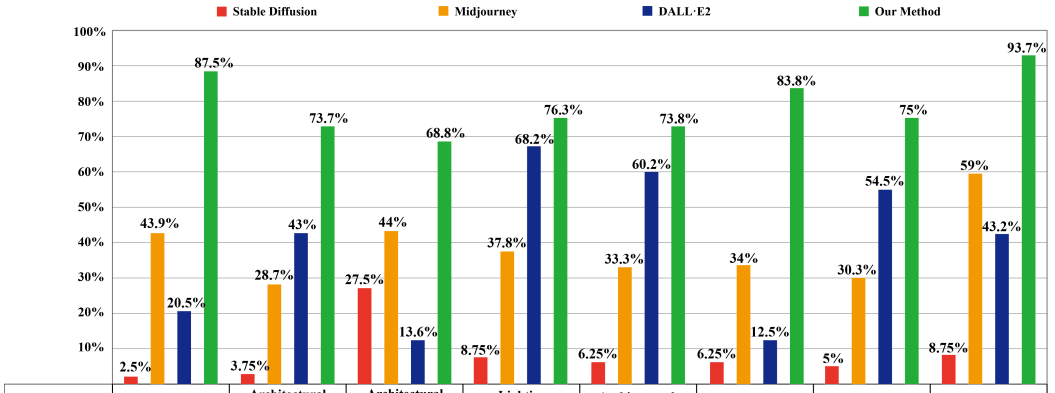
Figure 5 shows that Stable Diffusion failed to generate aesthetically pleasing designs with specified styles. In comparison, Midjourney performed significantly better. Midjourney understood the design styles of a few renowned architects and could generate corresponding designs. The images generated by DALL E2 were more realistic, but mostly lacked specific architectural styles and exhibited deficient architectural completeness. Moreover, a common drawback was observed in all three mainstream models. Specifically, these models lack knowledge of non-Western design styles, likely due to the limited architecture designed by non-Western architects in the training data. Consequently, these models may produce biased results when generating designs in the styles of architects from other countries. For example, when generating architectural designs in the style of Chinese architect "Shu Wang", all three models failed to accurately reproduce a consistent style. In contrast, the fine-tuned model demonstrated the capability to understand various architectural design styles, generating designs with complete compositions, creativity, and design details, while maintaining a high level of realism.

4.3. Quantitative Evaluation

The architectural designs generated by the proposed method were quantitatively compared with those generated by Midjourney, DALL E2, and Stable Diffusion. Each model generated 10 images in each of the eight different styles, resulting in 320 architectural design images. Then, the ten invited professional designers quantitatively evaluated the images based on eight indicators, namely, "overall impression", "architectural details", "architectural integrity", "lighting relationship", "architectural realism", "composition",

"realistic architectural background", and "consistent architectural style". One point was awarded if the image met the requirements for each evaluation indicator, and no point was awarded otherwise. Finally, the average score of each indicator was calculated and converted into a percentage to obtain the quantitative score for each model. The scores of the different diffusion models are shown in Figure 6.

Figure 5. Comparison between the proposed method and mainstream methods in generative architectural design. Each method generated architectural designs in six different styles, a total of 24 images. (Prompt: "An architectural photo in the style of (well-known architect's name), photo").



	Overall impression	Architectural details	Architectural integrity	Lighting relationship	Architectural realism	Composition	Realistic architectural background	Consistent architectural style
Stable Diffusion	2.5%	3.75%	27.5%	8.75%	6.25%	6.25%	5%	8.75%
Midjourney	43.9%	28.7%	44%	37.8%	33.3%	34%	30.3%	59%
DALL·E2	20.5%	43%	13.6%	68.2%	60.2%	12.5%	54.5%	43.2%
Our Method	87.5%	73.7%	68.8%	76.3%	73.8%	83.8%	75%	93.7%

Figure 6. Quantitative evaluation of the different models when generating architectural designs.

As shown in Figure 6, the four models showed considerable differences in their architectural designs. Among them, Stable Diffusion had the lowest score, with scores of each evaluation indicator below 10%, showing a clear gap with the other three models. It was almost impossible for the Stable Diffusion model to generate architectural designs

meeting the design requirements. The DALL E2 and Midjourney models had their merits and demerits. Compared with the DALL E2 model, the Midjourney model performed better in terms of "consistent architectural style", "overall impression", "architectural integrity", and "composition", with scores being 15.8%, 23.4%, 30.4%, and 21.5% higher than those of DALL E2, respectively. The images generated by DALL E2 scored 14.3%, 26.9%, 24.2%, and 30.4% higher than those of Midjourney in terms of the "architectural details", "architectural realism", "realistic architectural background", and "lighting relationship". DALL E2 had more advantages than Midjourney when generating realistic images.

Compared with the other models, the proposed model took the lead in all eight scoring metrics, especially the "consistent architectural style", reaching 93.7%. In addition, the proposed model successfully overcame the problem of the other three models, i.e., the difficulties in generating designs with styles of foreign architects, exhibiting a massive advantage in generating diverse architectural design styles. Considered comprehensively, the designs generated by the proposed model proved superior to those of the other models and showed higher usability.

4.4. Special Guiding Words

Appropriate guiding words play a major role in controlling the generated architectural designs. Therefore, this research summarized six categories of applicable guiding words with a clear impact on architectural design, namely "building type", "illumination", "angle of view", "environment", "architectural style", and "time". Then, cue words that could effectively control these elements were experimented on. By testing different guiding words, a summary of the vocabulary capable of precisely controlling the image generation was obtained and listed in Table 2.

Table 2. Six types of guiding words with obvious effects on architectural design generation.

Building Type	Illumination	Angle of View	Environment	Architectural Style	Time
Museum	Day light	Top view	Residential area	Modernism	Morning
Residential	Night light	Side view	Business district	Bauhaus	Noon
Hotel	Natural light	Aerial view	City center	Functionalism	Evening
Apartment	Cinematic light	Front view	Plaza	Neofuturist	Night
Commercial build	Sun light	Low-angle view	Park	Minimalism	Spring
Office building	Moon light		Countryside	Postmodern	Summer
Exhibition hall			Villa area	Gravel path	Autumn
Concert hall				Garden	Winter
				Canal	

For example, "building type" guiding words were used to adjust the type and scale of the building; "illumination" guiding words were used to adjust the lighting conditions of the generated image and change the atmosphere of the building; "angle of view" guiding words were used to adjust the angle of view of the image and change the observing angle and perspective of the building; "environment" guiding words were used to adjust the surrounding environment of the building; "architectural style" guiding words were used to control the style of the generated architectural design; "time" guiding words could control the observation period and season of the building and change its atmosphere.

Figure 7 demonstrates the effects of the guiding words above in controlling image generation. Figure 7A shows images generated by combining "Architectural style" and "Building type" guiding words. Figure 7B demonstrates the combination of "Illumination" and "Architectural style" guiding words. Figure 7C shows the combination of "Angle of view" and "Architectural style" guiding words. Figure 7D shows the combination of "Environment" and "Architectural style" guiding words. The combination of these guiding words achieved effective control over the generated results.

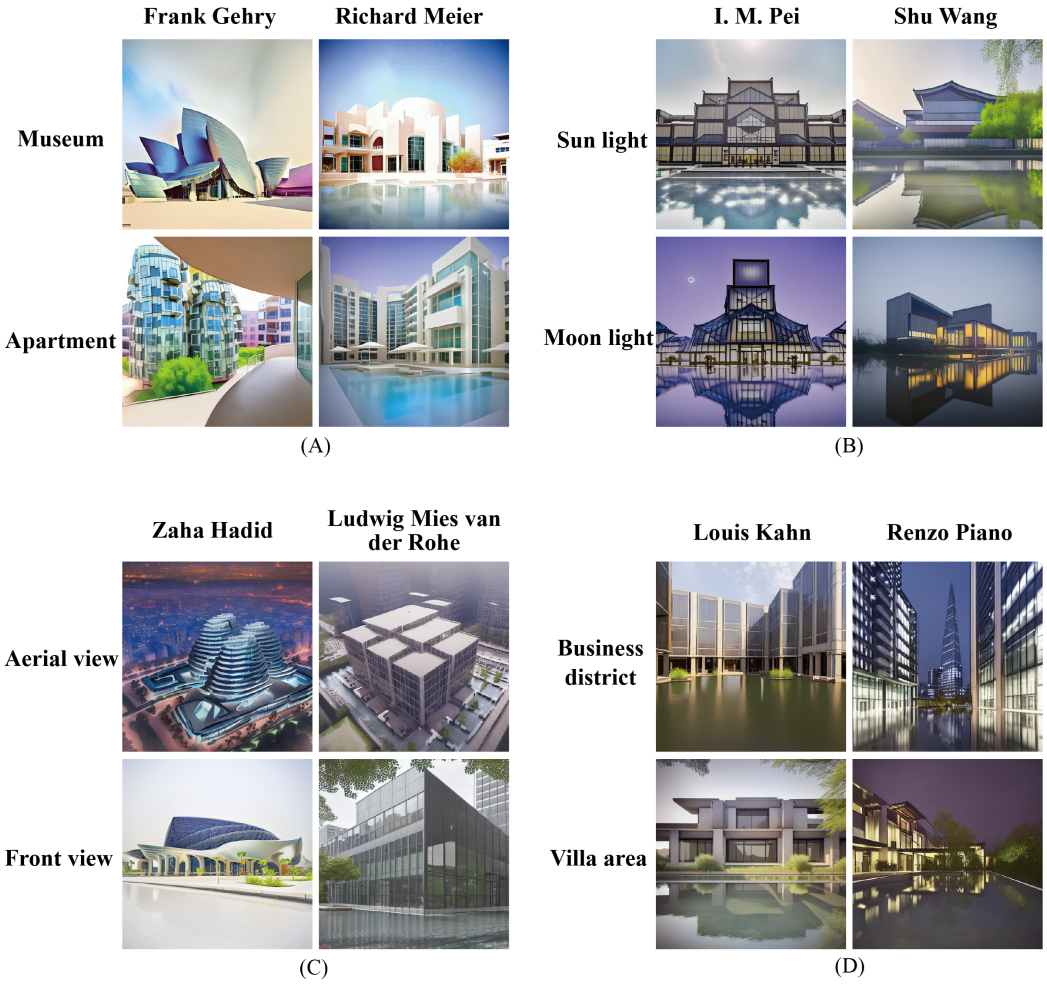


Figure 7. The subfigures (**A**–**D**) show the impact of different combinations of control vocabulary on the generation results. The guiding words summarized in this research achieved more-precise control over the architectural design generation results.

The effectiveness of the proposed method was proven, and the guiding words can help designers generate designs conforming to the requirements, thereby improving design efficiency and quality.

4.5. Generate Design Details Showcase

Figure 8 shows the architectural design generated by the proposed method, which is in the style of the famous Chinese architect "Shu Wang". The generated architectural design met the style requirements and was full of details. Specifically, a sloping roof design was adopted, and the materials included bricks, stones, and tiles, all essential manifestations of the "Shu Wang" style. In addition, the generated image had a good composition, fine light and shadow effects, and an excellent sense of realism. Therefore, the architectural design generated by the proposed method reached a usable level, capable of improving design efficiency and quality.

Figure 8. Architectural design in the "Shu Wang" style generated by the proposed method. (Prompt: "An architectural photo in the Shu Wang style, photo, realistic, high definition, museum, gravel path, garden, canal, sunlight, white clouds, autumn").

5. Discussion

This research demonstrated the effectiveness of the proposed approach through qualitative and quantitative analyses. In terms of qualitative analysis, the visual comparison with other methods proved that the proposed method can generate architectural designs of specified styles, an ability not achieved by other mainstream diffusion models. In terms of quantitative analysis, the quantitative data proved the superiority of the proposed method in all evaluation indicators. In particular, the proposed method has the advantage of generating specified architectural styles and usable designs. Taking the consistent architectural style indicator as an example, the proposed method performed well and surpassed Midjourney, DALL E2, and Stable Diffusion by 34.7%, 50.5%, and 84.95%, respectively, which fully demonstrated its effectiveness.

Potential ethical and bias risks exist in using AI to generate designs. For example, training AI relies on considerable data scraped from the Internet, and the clarification of data copyright is generally required. Legislating data copyright to define available data is necessary for protecting user privacy. Another example is that model training may be biased in the learned knowledge due to a lack of diversity in the collected datasets. The model training needs to consider cultural diversity. Furthermore, the rapid development of generative image technology has blurred the boundary between real and generated

images. Above all, managing or forcing unified annotated generated images is important; otherwise, misunderstandings may originate.

6. Conclusions

We propose a new AI approach for generative architectural design to help designers improve design quality. This method first constructs a master builder dataset (i.e., MBD-8) to solve the problem of insufficient training data and then proposes a new loss function. Then, the MBD-8 dataset and the loss function were employed to fine-tune the diffusion model. The fine-tuned model enables batch production of master-quality designs of a given architectural style. In the present work, guiding words that control the architectural design were summarized to ensure better controllability. The experiments showed that the proposed method can efficiently generate architectural designs in batches, improving design efficiency and quality.

This research has limitations. In application, the method primarily generates and modifies design concept solutions quickly. Despite its advantages in improving the efficiency and quality of scheme design, the designer still needs to remodel the generated renderings into a 3D model in the construction stage. For example, only the design styles of eight master architects were collected for the training data, which proved that the diffusion model could learn these styles to generate high-quality designs, while other architectural design styles were not collected nor tested. In addition, this research lacked the consideration of comprehensive quantitative evaluation indicators of architectural design. Furthermore, exploring more-effective guiding words for architectural generation is a future research direction.

Future work will involve the following aspects:

1. Building more-comprehensive architectural style datasets and expanding the architectural styles generated by the diffusion model;
2. Establishing automatic evaluation indicators and algorithms suitable for architectural design;
3. Establishing a more-comprehensive text-to-image guiding word dictionary for better control over the text-to-image results;
4. Exploring the feasibility of directly using the diffusion model to generate trimodal models or videos.

Author Contributions: Conceptualization, J.C. and Z.S.; data curation, J.C. and M.R.; formal analysis, J.C.; funding acquisition, J.L.; investigation, J.C.; methodology, J.C.; project administration, J.L.; resources, J.L.; software, J.C.; supervision, J.L.; validation, J.C. and J.L.; visualization, J.C. and M.R.; original draft writing, J.C., Z.S., M.R. and H.L.; review and editing, J.C., D.W., Z.S. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Foundation of China Key Project of Art Science "Research on Chinese Animation Creation and a Theoretical Innovation under the Construction of National Cultural Image" (Grant No. 20AC003), the project FRG-23-021-FA and granted by the Research Fund of Macao University of Science and Technology (FRG-MUST), and the Digital MediaArt, Key Laboratory of Sichuan Province, Sichuan Conservatory of Music "Research on emotional paradigm of virtual idol fans" (Grant No. 22DMAKL05).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The labeled dataset used to support the findings of this research is available from the authors upon request.

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

References

1. Liu, S.; Li, Z.; Teng, Y.; Dai, L. A dynamic simulation study on the sustainability of prefabricated buildings. *Sustain. Cities Soc.* **2022**, *77*, 103551. [CrossRef]
2. Luo, L.z.; Mao, C.; Shen, L.y.; Li, Z.d. Risk factors affecting practitioners' attitudes toward the implementation of an industrialized building system: A case study from China. *Eng. Constr. Archit. Manag.* **2015**, *22*, 622–643. [CrossRef]
3. Gao, H.; Koch, C.; Wu, Y. Building information modelling based building energy modelling: A review. *Appl. Energy* **2019**, *238*, 320–343. [CrossRef]
4. Delgado, J.M.D.; Oyedele, L.; Ajayi, A.; Akanbi, L.; Akinade, O.; Bilal, M.; Owolabi, H. Robotics and automated systems in construction: Understanding industry-specific challenges for adoption. *J. Build. Eng.* **2019**, *26*, 100868. [CrossRef]
5. Zikirov, M.; Qosimova, S.F.; Qosimov, L. Direction of modern design activities. *Asian J. Multidimens. Res.* **2021**, *10*, 11–18. [CrossRef]
6. Idi, D.B.; Khaidzir, K.A.B.M. Concept of creativity and innovation in architectural design process. *Int. J. Innov. Manag. Technol.* **2015**, *6*, 16. [CrossRef]
7. Bagherzadeh, F.; Shafighfard, T. Ensemble Machine Learning approach for evaluating the material characterization of carbon nanotube-reinforced cementitious composites. *Case Stud. Constr. Mater.* **2022**, *17*, e01537. . [CrossRef]
8. Shi, Y. Literal translation extraction and free translation change design of Leizhou ancient residential buildings based on artificial intelligence and Internet of Things. *Sustain. Energy Technol. Assess.* **2023**, *56*, 103092. [CrossRef]
9. Chen, J.; Shao, Z.; Zhu, H.; Chen, Y.; Li, Y.; Zeng, Z.; Yang, Y.; Wu, J.; Hu, B. Sustainable interior design: A new approach to intelligent design and automated manufacturing based on Grasshopper. *Comput. Ind. Eng.* **2023**, 109509. [CrossRef]
10. Chen, J.; Shao, Z.; Hu, B. Generating Interior Design from Text: A New Diffusion Model-Based Method for Efficient Creative Design. *Buildings* **2023**, *13*, 1861. [CrossRef]
11. Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical text-conditional image generation with clip latents. *arXiv* **2022**, arXiv:2204.06125.
12. Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.L.; Ghasemipour, K.; Gontijo Lopes, R.; Karagol Ayan, B.; Salimans, T.; et al. Photorealistic text-to-image diffusion models with deep language understanding. In *Proceedings of the Advances in Neural Information Processing Systems*; Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A., Eds.; Curran Associates, Inc.: New York, NY, USA, 2022; Volume 35, pp. 36479–36494. [CrossRef]
13. Nichol, A.Q.; Dhariwal, P. Improved denoising diffusion probabilistic models. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 18–24 July 2021; Meila, M., Zhang, T., Eds.; PMLR: New York, NY, USA, 2021; Volume 139, pp. 8162–8171. [CrossRef]
14. 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. [CrossRef]
15. Borji, A. Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. *arXiv* **2022**, arXiv:2210.00586.
16. Chen, C.; Tang, L. BIM-based integrated management workflow design for schedule and cost planning of building fabric maintenance. *Autom. Constr.* **2019**, *107*, 102944. [CrossRef]
17. Bonci, A.; Carbonari, A.; Cucchiarelli, A.; Messi, L.; Pirani, M.; Vaccarini, M. A cyber-physical system approach for building efficiency monitoring. *Autom. Constr.* **2019**, *102*, 68–85. [CrossRef]
18. Barreca, A. Architectural Quality and the housing market: Values of the late twentieth century built heritage. *Sustainability* **2022**, *14*, 2565. [CrossRef]
19. 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. [CrossRef]
20. Nichol, A.Q.; 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. In Proceedings of the International Conference on Machine Learning, Baltimore, MA, USA, 17–23 July 2022; pp. 16784–16804. [CrossRef]
21. Kawar, B.; Zada, S.; Lang, O.; Tov, O.; Chang, H.; Dekel, T.; Mosseri, I.; Irani, M. Imagic: Text-based real image editing with diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 6007–6017. [CrossRef]
22. Avrahami, O.; Lischinski, D.; Fried, O. Blended diffusion for text-driven editing of natural images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 18208–18218. [CrossRef]
23. Li, Y.; Zhang, R.; Lu, J.C.; Shechtman, E. Few-shot Image Generation with Elastic Weight Consolidation. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 15885–15896. [CrossRef]
24. Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J.W.; Wallach, H.; Iii, H.D.; Crawford, K. Datasheets for datasets. *Commun. Acm* **2021**, *64*, 86–92. [CrossRef]
25. Ivashko, Y.; Kuzmenko, T.; Li, S.; Chang, P. The influence of the natural environment on the transformation of architectural style. *Landsc. Archit. Sci. J. Latv. Univ. Agric.* **2020**, *15*, 101–108. [CrossRef]

26. Rezaei, F.; Bulle, C.; Lesage, P. Integrating building information modeling and life cycle assessment in the early and detailed building design stages. *Build. Environ.* **2019**, *153*, 158–167. [CrossRef]
27. Moghtadernejad, S.; Mirza, M.S.; Chouinard, L.E. Facade design stages: Issues and considerations. *J. Archit. Eng.* **2019**, *25*, 04018033. [CrossRef]
28. Yang, W.; Jeon, J.Y. Design strategies and elements of building envelope for urban acoustic environment. *Build. Environ.* **2020**, *182*, 107121. [CrossRef]
29. Eberhardt, L.C.M.; Birkved, M.; Birgisdottir, H. Building design and construction strategies for a circular economy. *Archit. Eng. Des. Manag.* **2022**, *18*, 93–113. [CrossRef]
30. Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Shao, Y.; Zhang, W.; Cui, B.; Yang, M.H. Diffusion models: A comprehensive survey of methods and applications. *arXiv* **2022**, arXiv:2209.00796.
31. Van Le, T.; Phung, H.; Nguyen, T.H.; Dao, Q.; Tran, N.; Tran, A. Anti-DreamBooth: Protecting users from personalized text-to-image synthesis. *arXiv* **2023**, arXiv:2303.15433.
32. Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; Bach, F., Blei, D., Eds.; PMLR: Cambridge, MA, USA, 2015; Volume 37, pp. 2256–2265. [CrossRef]
33. Croitoru, F.A.; Hondru, V.; Ionescu, R.T.; Shah, M. Diffusion models in vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10850–10869. [CrossRef] [PubMed]
34. Song, J.; Meng, C.; Ermon, S. Denoising Diffusion Implicit Models. In Proceedings of the International Conference on Learning Representations, Virtual Event, 26 April–1 May 2020. [CrossRef]
35. Liu, X.; Park, D.H.; Azadi, S.; Zhang, G.; Chopikyan, A.; Hu, Y.; Shi, H.; Rohrbach, A.; Darrell, T. More control for free! image synthesis with semantic diffusion guidance. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 289–299. [CrossRef]
36. Dhariwal, P.; Nichol, A. Diffusion models beat gans on image synthesis. In Proceedings of the Advances in Neural Information Processing Systems, San Francisco, CA, USA, 30 November–3 December 1992; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: San Francisco, CA, USA, 2021; Volume 34, pp. 8780–8794. [CrossRef]
37. Ho, J.; Salimans, T. Classifier-Free Diffusion Guidance. In Proceedings of the NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications, Cambrige, MA, USA, 14 December 2021. [CrossRef]
38. Ding, M.; Yang, Z.; Hong, W.; Zheng, W.; Zhou, C.; Yin, D.; Lin, J.; Zou, X.; Shao, Z.; Yang, H.; et al. Cogview: Mastering text-to-image generation via transformers. In *Proceedings of the Advances in Neural Information Processing Systems*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: New York, NY, USA, 2021; Volume 34, pp. 19822–19835. [CrossRef]
39. Gafni, O.; Polyak, A.; Ashual, O.; Sheynin, S.; Parikh, D.; Taigman, Y. Make-a-scene: Scene-based text-to-image generation with human priors. In Proceedings of the Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, 23–27 October 2022; Proceedings, Part XV; Springer: Cham, Switzerland, 2022; pp. 89–106. [CrossRef]
40. Yu, J.; Xu, Y.; Koh, J.Y.; Luong, T.; Baid, G.; Wang, Z.; Vasudevan, V.; Ku, A.; Yang, Y.; Ayan, B.K.; et al. Scaling Autoregressive Models for Content-Rich Text-to-Image Generation. *Trans. Mach. Learn. Res.* **2022**, *2*, 5. [CrossRef]
41. Cheng, S.I.; Chen, Y.J.; Chiu, W.C.; Tseng, H.Y.; Lee, H.Y. Adaptively-Realistic Image Generation from Stroke and Sketch with Diffusion Model. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 4054–4062. [CrossRef]
42. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. Acm* **2020**, *63*, 139–144. [CrossRef]
43. Ding, M.; Zheng, W.; Hong, W.; Tang, J. CogView2: Faster and Better Text-to-Image Generation via Hierarchical Transformers. In Proceedings of the Advances in Neural Information Processing Systems, Lyon, France, 25–29 April 2022. [CrossRef]
44. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning transferable visual models from natural language supervision. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 8748–8763. [CrossRef]
45. Ruiz, N.; Li, Y.; Jampani, V.; Pritch, Y.; Rubinstein, M.; Aberman, K. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 22500–22510. [CrossRef]
46. Gal, R.; Alaluf, Y.; Atzmon, Y.; Patashnik, O.; Bermano, A.H.; Chechik, G.; Cohen-Or, D. An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv* **2022**, arXiv:2208.01618.
47. Von Oswald, J.; Henning, C.; Grewe, B.F.; Sacramento, J. Continual learning with hypernetworks. In Proceedings of the 8th International Conference on Learning Representations (ICLR 2020), Virtual, 26–30 April 2020. [CrossRef]
48. Hu, E.J.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. In Proceedings of the International Conference on Learning Representations, Virtual, 25–29 April 2022. [CrossRef]
49. Choi, J.; Kim, S.; Jeong, Y.; Gwon, Y.; Yoon, S. ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 6–9 July 2021; pp. 14347–14356. [CrossRef]

50. Lee, J.; Cho, K.; Kiela, D. Countering Language Drift via Visual Grounding. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 4385–4395. [CrossRef]
51. Lu, Y.; Singhal, S.; Strub, F.; Courville, A.; Pietquin, O. Countering language drift with seeded iterated learning. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; Daume, H.D., Singh, A., Eds.; PMLR: Cambridge, MA, USA; Volume 119, pp. 6437–6447. [CrossRef]
52. Wang, W.; Wang, X.; Xue, C. Aesthetics Evaluation Method of Chinese Characters based on Region Segmentation and Pixel Calculation. *Intell. Hum. Syst. Integr. (Ihsi 2023): Integr. People Intell. Syst.* **2023**, *69*, 561–568. [CrossRef]
53. Wang, L.; Xue, C. A Simple and Automatic Typesetting Method Based on BM Value of Interface Aesthetics and Genetic Algorithm. In Proceedings of the Advances in Usability, User Experience, Wearable and Assistive Technology: Proceedings of the AHFE 2021 Virtual Conferences on Usability and User Experience, Human Factors and Wearable Technologies, Human Factors in Virtual Environments and Game Design, and Human Factors and Assistive Technology, Virtual, 25–29 July 2021; Springer: New York, NY, USA, 2021; pp. 931–938. [CrossRef]