

Artificial Intelligence for Master-Quality Architectural Designs Generation

Subjects: **Architecture And Design**

Contributor: Junming Chen , Duolin Wang , Zichun Shao , Xu Zhang , Mengchao Ruan , Huiting Li , Jiaqi Li

The outstanding buildings designed by master architects are the common wealth of mankind. They reflect their design skills and concepts and are not possessed by ordinary architectural designers. Compared with traditional methods that rely on a lot of mental labor for innovative design and drawing, artificial intelligence (AI) methods have greatly improved the creativity and efficiency of the design process. It overcomes the difficulty in specifying styles for generating high-quality designs in traditional diffusion models.

architectural design

text to design

design process optimization

design quality

design style

diffusion model

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][3][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][8][9][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**).



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”).

2. 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][3][4]. Moreover, iconic landmark architecture in a city stimulates local employment and boosts the tourism industry [1][16].

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 [17]. 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][3][4]. Consequently, the efficiency of this design practice needs improvement [17].

The reason for the low quality of architectural design is that it is difficult to train excellent designers and the process of improving design capabilities is long. Designers' lack of design skills makes it difficult to improve design quality [4][5][6]. However, the improvement of design capabilities is a gradual process , and designers must constantly learn new design methods and explore different design styles [2][3][6][18][19][20]. At the same time, seeking the best design solution under complex conditions also brings huge challenges to designers [2][5].

All these factors ultimately lead to inefficient and low-quality architectural designs [2][5]. Therefore, new technologies must be introduced into the construction industry in a timely manner to solve these problems.

3. Diffusion Model

In recent years, the diffusion model has rapidly developed into a mainstream image generation model [21][22][23][24][25][26], which allows designers to quickly obtain images, thereby significantly improving the efficiency and quality of architectural design [13][27].

The traditional diffusion model includes forward process and backward process. During the forward process, noise is continuously added to the input image, transforming it into a noisy image. The purpose of the backward process is to restore the original image from the noisy image [28]. By learning the image denoising process, the diffusion model acquires the ability to generate images [27][29]. When there is a need to generate images with specific design elements, designers can incorporate text cues into the denoising process of the diffusion model to generate consistent images and achieve controllability over the generated results [30][31][32][33][34][35]. The advantage of using text-guided diffusion models for image generation is that they allow simple control of image generation [12][14][36][37][38].

Although diffusion models perform well in most fields, there is still room for improvement in their application in architectural design [30][39]. Specifically, the limitation comes from obtaining large amounts of Internet data for training, which lacks 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, which makes it challenging to use professional design vocabulary to guide architectural design generation [40][41][42][43]. Therefore, it is necessary to collect high-quality architectural design images, annotate them with relevant information, and then fine-tune the model to adapt it to the architectural design task.

4. Model Fine-Tuning

Diffusion models learn new knowledge and concepts through complete retraining or fine-tuning for new scenarios. Due to the huge cost of retraining the entire model, 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 methods of fine-tuning. The first is text inversion [11][31][41][44], which freezes the text-to-image model and provides only the most suitable embedding vectors to embed new knowledge. This method provides fast model training and minimal generative models, but the image generation effect is mediocre. The second is the Hypernetwork [42] method, which inserts a separate small neural network in the middle layer of the original diffusion model to affect the output. This method is faster to train, but the image generation effect is average. The third one is LoRA [43], which assigns weights to cross-layer attention to allow learning new knowledge. This method can generate models with an average size of hundreds of MB after moderate training time, and the image generation effect is good. The fourth is the Dreambooth [40] method, which is an overall fine-tuning of the original diffusion model. Using this method, a prior-preserving loss was designed to train the diffusion model, enabling it to generate images consistent with the cues while preventing overfitting [45][46]. It is recommended to use rare vocabulary when naming new knowledge to avoid language drift due to similarity with the original model vocabulary [45][46]. This method requires only 3 to 5 images of a specific subject and corresponding text descriptions, can be fine-tuned for specific situations, and matches specific text descriptions to

the characteristics of the input images. Fine-tuned models generate images based on specific topic terms and general descriptors [\[26\]](#)[\[41\]](#). Since the entire model is fine-tuned using the Dreambooth method, the results produced are usually the best of these methods.

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.
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.
3. Gao, H.; Koch, C.; Wu, Y. Building information modelling based building energy modelling: A review. *Appl. Energy* 2019, 238, 320–343.
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.
5. Zikirov, M.; Qosimova, S.F.; Qosimov, L. Direction of modern design activities. *Asian J. Multidimens. Res.* 2021, 10, 11–18.
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.
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.
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.
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.
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.
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.
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.
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.
15. Borji, A. Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. *arXiv* 2022, arXiv:2210.00586.
16. 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.
17. 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.
18. Moghtadernejad, S.; Mirza, M.S.; Chouinard, L.E. Facade design stages: Issues and considerations. *J. Archit. Eng.* 2019, 25, 04018033.
19. Yang, W.; Jeon, J.Y. Design strategies and elements of building envelope for urban acoustic environment. *Build. Environ.* 2020, 182, 107121.
20. 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.
21. 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.
22. 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.
23. Kavar, 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.

24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. Ho, J.; Salimans, T. Classifier-Free Diffusion Guidance. In Proceedings of the NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications, Cambridge, MA, USA, 14 December 2021.
33. 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.
34. 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.
35. 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.
 36. 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.
 37. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. Acn* 2020, 63, 139–144.
 38. 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*.
 39. 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.
 40. 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.
 41. 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.
 42. 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*.
 43. 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*.
 44. 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.
 45. 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.

46. 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.

Retrieved from <https://encyclopedia.pub/entry/history/show/111564>