Generating Interior Design from Text

Subjects: Construction & Building Technology

Contributor: Junming Chen , Zichun Shao , Bin Hu

Because interior design is subject to inefficiency, more creativity is imperative. Due to the development of artificial intelligence diffusion models, the utilization of text descriptions for the generation of creative designs has become a novel method for solving the aforementioned problem.

diffusion model   text generation design   interior design

1. Introduction

There is a huge demand for interior design worldwide, but existing design approaches or methodologies may not fully meet these needs [1][2][3]. One reason for this phenomenon is that the interior design process is complicated, and frequent changes lead to low design efficiency [1][3][4][5]. In addition, designers form fixed design methods to save time, resulting in a lack of innovation [1][6][7]. Therefore, it is important to improve the efficiency of interior design and address the lack of innovation.

With the introduction of the diffusion model [8][9], it is possible to solve the problems of low efficiency and a lack of creativity in interior design [10][11][12]. The advantage of the diffusion model is that it can learn prior knowledge from massive image and text description pairing information [13][14][15][16]. The trained diffusion model can generate high-quality and diverse images by inputting text descriptions in batches. Using the diffusion model for interior design can generate design schemes in batches for designers. This method can effectively improve the efficiency of design and creative generation [17][18][19][20].

Although diffusion models work well in most domains, they generate poor image quality in a few domains. Especially in the field of interior design, which requires highly professional skills, conventional diffusion models cannot produce high-quality interior designs. For example, the current mainstream diffusion models Midjourney [21], Dall E2 [13], and Stable Diffusion [22] cannot generate high-quality design images with specified decoration styles and space functions (**Figure 1**). The correct decoration style and space function are very important to interior design, and thus it is urgent to solve the above problems.

Figure 1. Images generated by mainstream diffusion models are compared with those generated by the researchers' method. Midjourney will produce a lot of redundant objects, and the image is not realistic (**left one**). The object generated by DALL E2 is incomplete and has the wrong space size (**second from the left**). The placement and spatial scale of things generated by Stable Diffusion are incorrect (**third from the left**). None of these images are up to the interior design requirements, and the proposed method (**far right**) improves the above problems. (prompt word: "Realistic, Chinese-style study room, with desks and cabinets").

In order to batch-generate designs with specific decoration styles and space functions, this study created a new interior decoration style dataset and retrained the diffusion model to make it suitable for interior design generation. Specifically, this study first collected a brand new Interior Decoration Style and Space Function (IDSSF-64) dataset from professional designers to solve the problem of a lack of training datasets for this task. IDSSF-64 includes the classification of decoration styles and space functions. Then, researchers proposed a new loss function, which adds style-aware reconstruction loss and style prior preservation loss to the conventional loss function. This function forces the diffusion model to learn the knowledge of decoration styles and space functions and retains the basic knowledge of the original model. The new model proposed in this study uses a new loss function and a new dataset to fine-tune training for interior design generation with specified decoration styles and space functions. The fine-tuning method does not need to retrain the whole model. It only requires a small number of images to fine-tune the model to obtain a better generation effect, thus significantly reducing the amount of training data and training time. The fine-tuned model can generate end-to-end interior designs in batches for designers to select, thereby improving design efficiency and creativity. The framework of this study is shown in **Figure 2**.
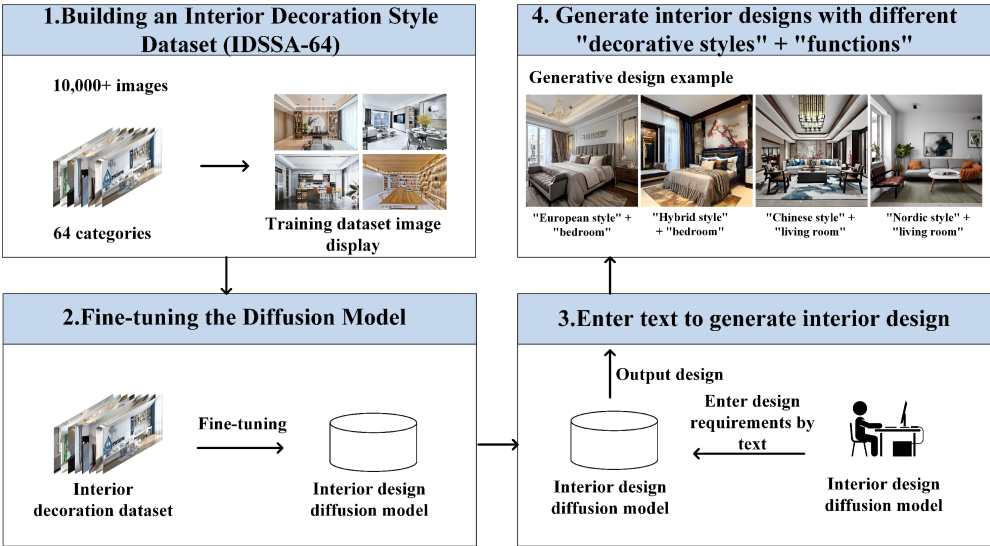
Figure 2. Study framework. This study first collects the interior decoration style dataset IDSSF-64 and then builds a diffusion model suitable for interior design through fine-tuning. Users can input a decoration style and space function into the fine-tuned diffusion model to directly obtain the design.

The fine-tuned diffusion model generative design method proposed in this study has changed the design process, and interior design efficiency and creativity have improved. The model can generate a variety of indoor spaces and ensure that the generated content meets the design requirements. **Figure 3** demonstrates the interior design effects of different decoration styles and spaces generated by the model. The figure shows that the model understands the decoration styles and space functions. Each generated object appears in a suitable position, resulting in high-quality interior design.
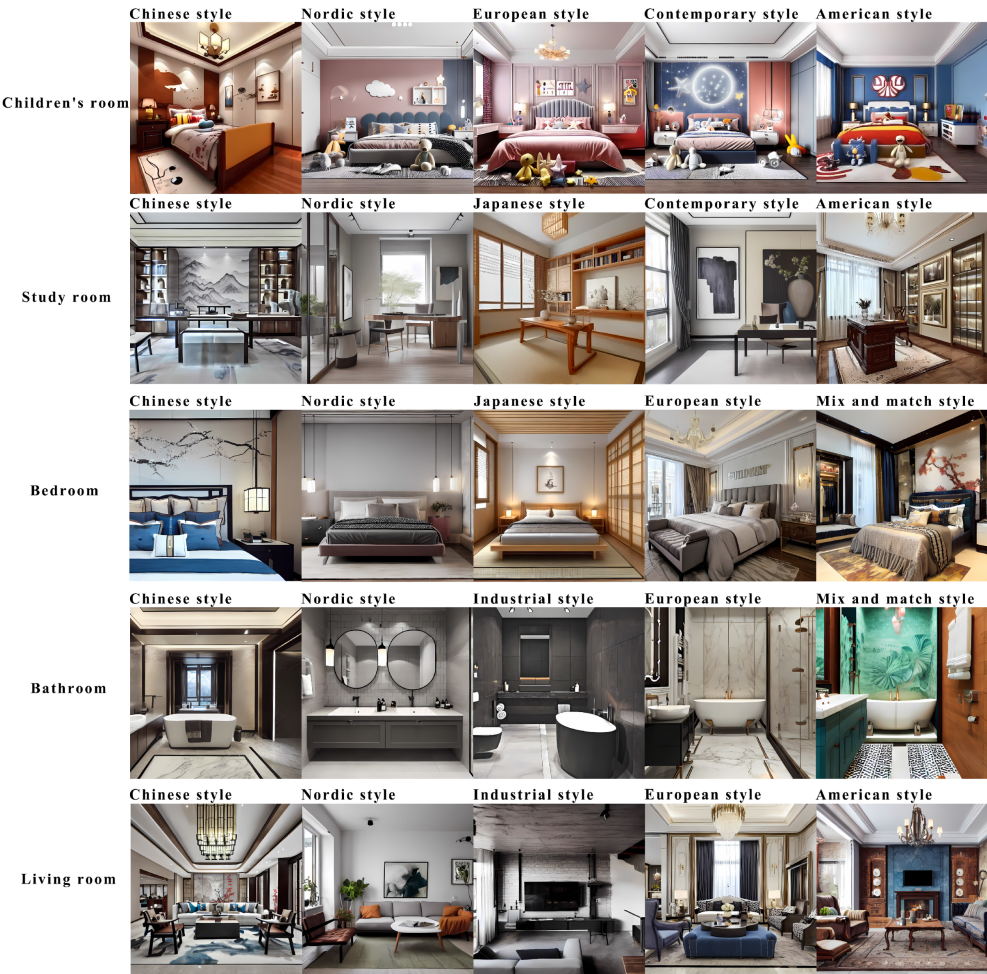

Figure 3. Interior design images generated by the diffusion model for different decoration styles and space functions.

2. Conventional Interior Design Process

Interior design usually means that designers use their art and engineering knowledge to design interior spaces with specific decorative styles for clients. Designers must choose appropriate design elements to shape the decoration

style, such as suitable tiles, furniture, colors, and patterns. A strong decoration style is key to making the design unique [3].

Designers usually use decorative renderings to determine the final design with clients, but this approach is inefficient. The reason for this inefficiency is that the conventional interior design workflow is linear, and designers spend a lot of time drawing design images and cannot communicate with customers in real time, resulting in many revisions. The conventional design process is shown in **Figure 4**. Specifically, interior design usually requires designers to find intentions to discuss with customers and decide on the decoration style. Then, the designer starts to produce two-dimensional (2D) drawings and build corresponding three-dimensional (3D) models. Then, material mapping is assigned to the 3D model, lighting is arranged for the space, and renderings are obtained. Finally, the customer determines whether the design is suitable by observing the renderings [5]. The linear workflow requires designers to design step by step. Once the customer is unsatisfied with a particular node, the designer must redo the entire design, leading to low design efficiency.
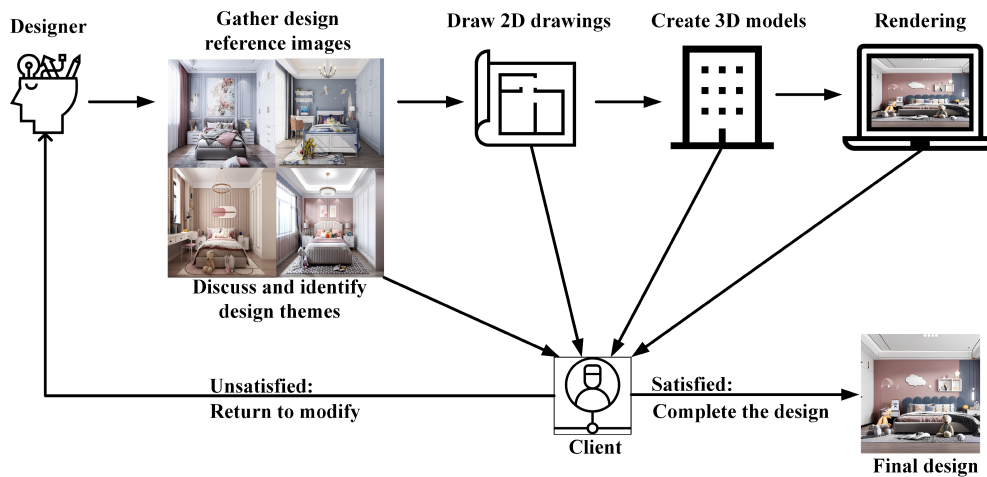
Figure 4. Conventional interior design process. Designers need to complete the design through a linear design process. If the customer is unsatisfied with the design during the process, then the designer must redo the entire design process.

At the same time, the cumbersome interior design workflow also suppresses creative design. On one hand, designers will form a fixed design method in the pursuit of efficiency so that they can quickly produce creative designs. On the other hand, even if the designer has a lot of creative inspirations, it takes a lot of labor to transform them into renderings, and they can only draw some of these within a limited time. Helping designers quickly obtain diversified interior design renderings is the key to solving the problems of low design efficiency and insufficient creativity.

Existing design automation mainly focuses on a particular process in the design, but fewer studies focus on end-to-end design [4][5]. This study achieves end-to-end generation of interior design by building a text-to-image diffusion model, thereby improving design efficiency and addressing the lack of creativity.

3. Text-to-Image Diffusion Model

The earlier diffusion model was proposed in 2015 [8] and has been continuously optimized and improved since then [23]. The improved model has become a new mainstream generative model due to its excellent productive image effects [23][24][25]. The diffusion model mainly includes forward and reverse processes, and the forward process continuously adds noise to the original image. The reverse process iteratively denoises purely random noise to restore the image. Diffusion models learn the denoising process to gain the ability to generate images [8][23]. Generating an image of a specified category or with specific features requires adding text guidance. Text-to-image-based diffusion models enable controlled image generation using text as a guiding condition [10][11][25][26][27]. An advantage of the text-guided diffusion model is that it can create images that match the meanings of the prompt words.

There are two ways to learn new knowledge in the diffusion model: to retrain the entire model and to fine-tune the model to make the model suitable for new scenarios. Considering the high cost of retraining the whole model, fine-tuning the model is more feasible. There are four commonly used methods for fine-tuning models. The first one is textual inversion [13][25][28][29]. The core idea of textual inversion to embed new knowledge into a model is to freeze the text-to-image model and only give the most suitable embedding vector for new knowledge. This approach does not require model changes and is similar to finding new representations in the model to represent new keywords. The second is the hypernetwork [30]. A hypernetwork is a separate small neural network. The model is inserted into the middle layer of the original diffusion model to affect the output. The third is LoRA [31]. LoRA adds its weight to the attention cross layer as fine-tuning. The fourth is DreamBooth [32], which expands the text-image dictionary of the target model and establishes a new type of association between text identifiers and images while using rare words to name new knowledge and train to avoid language drift [33][34]. At the same time, researchers designed a prior preservation loss function to solve the overfitting problem. This loss function prompts the diffusion model to produce different examples of the same category as the subject. This method only needs 3–5 images and corresponding text descriptions to complete the fine-tuning of a specific theme and match the detailed text descriptions with the characteristics of the input image. The fine-tuned model can generate images with the trained topic words and descriptors. DreamBooth usually works best among these methods because it fine-tunes the entire model.

References

1. Wang, Y.; Liang, C.; Huai, N.; Chen, J.; Zhang, C. A Survey of Personalized Interior Design. In Computer Graphics Forum; Wiley Online Library: Hoboken, NJ, USA, 2023.

2. Ashour, M.; Mahdiyar, A.; Haron, S.H. A Comprehensive Review of Deterrents to the Practice of Sustainable Interior Architecture and Design. Sustainability 2021, 13, 10403.

3. Bao, Z.; Laovisutthichai, V.; Tan, T.; Wang, Q.; Lu, W. Design for manufacture and assembly (DfMA) enablers for offsite interior design and construction. Build. Res. Inf. 2022, 50, 325–338.

4. Karan, E.; Asgari, S.; Rashidi, A. A markov decision process workflow for automating interior design. KSCE J. Civ. Eng. 2021, 25, 3199–3212.

5. Park, B.H.; Hyun, K.H. Analysis of pairings of colors and materials of furnishings in interior design with a data-driven framework. J. Comput. Des. Eng. 2022, 9, 2419–2438.

6. Sinha, M.; Fukey, L.N. Sustainable Interior Designing in the 21st Century—A Review. ECS Trans. 2022, 107, 6801.

7. 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.

8. 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, 6–11 July 2015; Bach, F., Blei, D., Eds.; PMLR: New York, NY, USA, 2015; Volume 37, pp. 2256–2265.

9. Croitoru, F.A.; Hondru, V.; Ionescu, R.T.; Shah, M. Diffusion models in vision: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 2023, 1–20.

10. 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.

11. Ho, J.; Salimans, T. Classifier-Free Diffusion Guidance. arXiv 2021, arXiv:2207.12598.

12. Nichol, A.Q.; Dhariwal, P. Improved denoising diffusion probabilistic models. In Proceedings of the International Conference on Machine Learning, Virtual Event, 18–24 July 2021; Meila, M., Zhang, T., Eds.; PMLR: New York, NY, USA, 2021; Volume 139, pp. 8162–8171.

13. Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical text-conditional image generation with clip latents. arXiv 2022, arXiv:2204.06125.

14. 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 Advances in Neural Information Processing Systems; Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2022; Volume 35, pp. 36479–36494.

15. Song, Y.; Ermon, S. Improved techniques for training score-based generative 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. 12438–12448.

16. Song, Y.; Ermon, S. Generative modeling by estimating gradients of the data distribution. In Advances in Neural Information Processing Systems; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32.

17. 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.

18. 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, PMLR, Baltimore, ML, USA, 17–23 July 2022; pp. 16784–16804.

19. 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, 18–22 June 2023; pp. 6007–6017.

20. 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.

21. Borji, A. Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. arXiv 2022, arXiv:2204.06125.

22. 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.

23. Song, J.; Meng, C.; Ermon, S. Denoising Diffusion Implicit Models. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.

24. Jolicoeur-Martineau, A.; Piché-Taillefer, R.; Mitliagkas, I.; des Combes, R.T. Adversarial score matching and improved sampling for image generation. In Proceedings of the International Conference on Learning Representations, Virtual Event, 3–7 May 2021.

25. Dhariwal, P.; Nichol, A. Diffusion models beat gans on image synthesis. In Advances in Neural Information Processing Systems; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 8780–8794.

26. 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 Advances in Neural

Information Processing Systems; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 19822–19835.

27. 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.

28. 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.

29. 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, BC, Canada, 11–17 October 2021; pp. 14347–14356.

30. 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), Addis Ababa, Ethiopia, 26–30 April 2020.

31. 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 Event, 25–29 April 2022.

32. 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, 18–22 June 2023; pp. 22500–22510.

33. 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; Association for Computational Linguistics: Hong Kong, China, 2019; pp. 4385–4395.

34. 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 Event, 13–18 July 2020; Daumé, H., III, Singh, A., Eds.; PMLR: New York, NY, USA, 2020; Volume 119, pp. 6437–6447.