CanFuUI: A Canvas-Centric Web User Interface for Iterative Image Generation with Diffusion Models and ControlNet

Qihan Hu[1], Zhenghui Xu[1], Peng Du[2], Hao Zeng[1(✉)], Tongqing Ma[1(✉)], Youbing Zhao[1,3], Hao Xie[1], Peng Zhang[4], Shuting Liu[4], Tongnian Zang[5], and Xuemei Wang[1]

[1] Communication University of Zhejiang, Hangzhou 310018, China
{hao.zeng,matongq}@cuz.edu.cn
[2] Uber Technologies Inc., 1725 3rd Street, San Francisco, CA 94158, USA
[3] University of Bedfordshire, Luton LU1 3JU, UK
[4] Jiangsu Dongyin Intelligent Engineering Technology Research Institute, Nanjing 211111, China
[5] Jiangsu CRRC Digital Technology Co. Ltd., Nanjing 210000, China

Abstract. Today, various AI generation tools are emerging in succession. And the majority of existing tools are predominantly model-centric in design, resulting in steep learning curves and high usability thresholds for users. Moreover, current user interfaces lack built-in image editing capabilities, forcing users to rely on external software even for basic image editing tasks. Considering that most image generation is an iterative process, this limitation significantly hampers user experience and creative potential. Instead, this paper proposes a novel canvas-centric design that seamlessly integrates editing functionalities into the UI called CanFuUI, streamlining secondary image processing. Users can crop, modify, and annotation of specific regions of generated images within the same canvas in CanFuUI. Furthermore, canvas content is utilized as preprocessed images, directly integrated into the ControlNet preprocessing procedure, reinforcing the customization capabilities of AI-generated outputs.

Keywords: WebUI · ComfyUI · Diffusion Models · Canvas-Centric UI

1 Introduction

The landscape of AI image generation tools has experienced an explosive expansion with the emergence of diffusion models [5,7,9–12] based open-source solutions. However, most existing AI generation tools primarily focus on model-centric design, offering high flexibility, but also presenting a steep learning curve for users. Additionally, the ability to perform secondary editing on generated

Q. Hu and Z. Xu—Contributed equally to this work.

Many thanks to Mr. Yihui Shen for his generous funding support.

F. Zhao and D. Miao (Eds.): AIGC 2023, CCIS 1946, pp. 128–138, 2024.
https://doi.org/10.1007/978-981-99-7587-7_11

images is often limited, hindering direct image manipulation. Prominent examples like Stable Diffusion WebUI [2] and ComfyUI [1], which are widely utilized, pose challenges to users who need to possess a deep understanding of checkpoint, LoRA [6], ControlNet [12] models, expertly manipulate parameters like cfg, Step, Sampling methods, and employ sophisticated prompt techniques to achieve the desired artistic result. Furthermore, as user expectations continue to increase, the utilization of AI to generate images that meet those expectations becomes increasingly challenging.

For novice users in the field of AI artistry, their limited knowledge may lead to simplistic prompts or basic parameter adjustments, resulting in unsatisfactory outcomes. Another major challenge in stable diffusion research is the intricate process of blending or replacing specific elements from multiple generated images. For instance, merging a character into a scene requires several cumbersome steps, such as converting both the character and scene into sketches, erasing unnecessary elements, integrating the character sketch into the scene sketch, and finally using ControlNet to generate the final image. These operations that involve multiple tools require a significant amount of time and effort.

To address these challenges, we propose a new web user interface named CanFuUI that streamlines operations, reduces the learning curve, and integrates settings, parameters, and prompt techniques to generate high-quality images into selectable options. Additionally, we present an innovative approach, canvas-centric editing, that allows users to crop, modify, and draw on specific regions of the generated images. All edits can be performed seamlessly within a single canvas, and the resultant images, as well as preprocessed images obtained using ControlNet, are saved in the gallery, allowing for easy drag-and-drop usage from the toolbar. These advances significantly simplify the workflow, saving substantial time and effort.

Our main contributions lie in the following aspects:

- We introduce the first canvas-centric design AI generation tool that integrates editing functionalities into the UI, streamlining the cumbersome process of secondary image processing.
- We propose using canvas contents as preprocessed images, directly utilized in ControlNet when activated.
- We pre-select and integrate model parameters into style options, reducing the complexity and lowering the barrier to user engagement.

2 Related Work

Utilizing the Stable Diffusion model [9], remarkable image quality has been achieved that exceeds that of traditional GAN models [3,4,8]. However, due to the limited input conditions of the Stable Diffusion model, the ControlNet [12] model enables a wider range of input conditions, thus granting us greater control over the generated images.

Currently, the most popular user interfaces for Stable Diffusion are Stable Diffusion WebUI [2] and ComfyUI [1]. While the Stable Diffusion WebUI adopts

a functional model-centric design, it requires users to have a comprehensive understanding of multiple models, have the ability to adjust various parameters, and have experience crafting diverse and effective prompts. On the other hand, ComfyUI centers on a node-based design, requiring users to master the intricate workflow of AI-generated image production. Both WebUI and ComfyUI lead to suboptimal user experiences in terms of image creation, modification, and adjustments. In contrast, CanFuUI is built as canvas-centric, significantly enhancing user convenience for image editing, adjustments, and creative endeavors.

3 Canvas-Centric Web UI

In order to enhance user convenience, we have organized CanFuUI into four distinct modules: AI Artistry (a), Canvas (b), Toolbar (c) and Asset Library (d), as shown in Fig. 1.

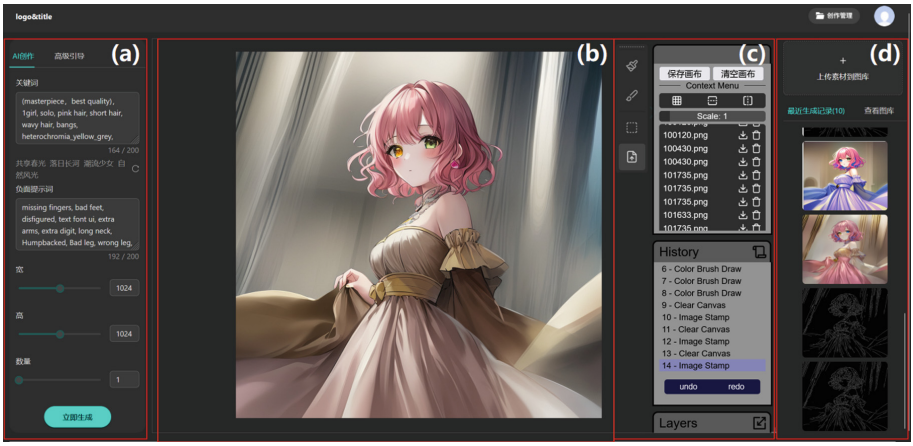


Fig. 1. Overview of CanFuUI design

3.1 AI Artistry

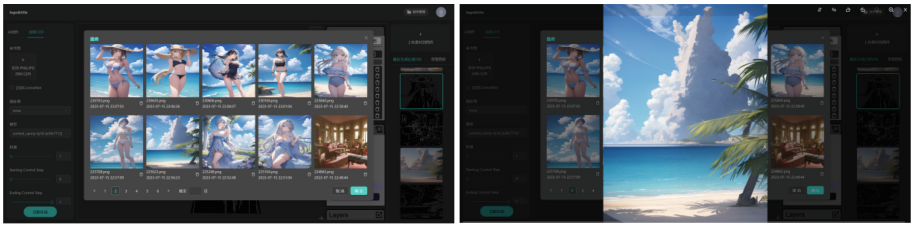
The extent of utilization of the AI artistry module generally depends on the user's proficiency in AI drawing techniques. For those users who seek artistic entertainment through AI drawing, an extensive degree of parameter adjustment and model selection might not be necessary. On the contrary, they may prefer fewer manual interventions and entrust more tasks to AI generation. As a result, we have simplified the creative interface, reducing the parameters options less frequently used while integrating various parameters, models, and prompt techniques into the style options shown in Fig. 1(a). This empowers users to effortlessly select their desired artistic styles for image generation. Moreover, for more demanding users, we have introduced ControlNet as an advanced guidance tool, facilitating further refinement of the generated images.

3.2 Canvas and Toolbar

The Canvas module is shown in Figs. 1(b) and (c), which offers a scalable canvas where AI-generated images are showcased, and these images can also be directly utilized as pre-processed images for ControlNet. Users have the flexibility to add their own generated images or import external images to the canvas, allowing for seamless manipulation using the diverse tools available in the toolbar. The toolbar primarily supports functions such as drawing, selection, cropping, and image insertion as demonstrated in Fig. 1(c). Additionally, the Canvas module incorporates support for layer operations, facilitating users to make secondary modifications to the generated images with ease.

3.3 Asset Library

The Asset Library module maintains a record of the ten most recently generated images for user selection presented in Fig. 1(d). Previous generated images are seamlessly integrated into the gallery depicted in Fig. 2(a), allowing users to preview their past generations as in Fig. 2(b). Additionally, users can easily add these images to the canvas or directly export them. The Asset Library module also facilitates the upload function, allowing the import of external images into the gallery.

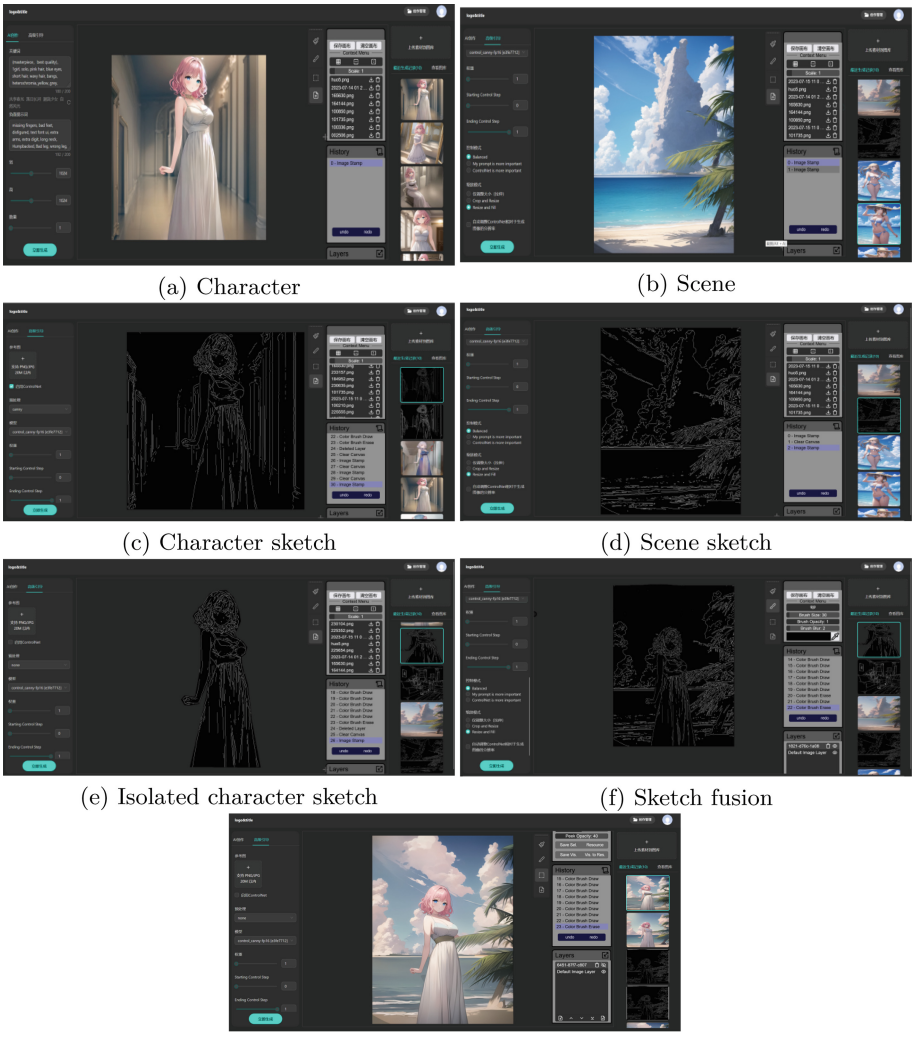


(a) Gallery (b) Preview

Fig. 2. Asset library

3.4 Use Case

In the AI Artistry module, users can effortlessly obtain desired images by entering prompt phrases, as illustrated in Fig. 3(a) and Fig. 3(b). For more advanced operations, such as incorporating a character from Fig. 3(a) into Fig. 3(b), users can utilize ControlNet's canny model within the advanced options to transform images in Fig. 3(a) and Fig. 3(b) into sketches, represented as Fig. 3(c) and Fig. 3(d). Then, using the brush and cropping tools in the toolbar, the character can be accurately isolated to obtain the image Fig. 3(e). Subsequently, the image shown in Fig. 3(e) can be seamlessly integrated with the scene sketch, producing the image presented in Fig. 3(f). Finally, employing AI, the image of the result of the fusion can be generated as in Fig. 3(g).



(a) Character (b) Scene

(c) Character sketch (d) Scene sketch

(e) Isolated character sketch (f) Sketch fusion

(g) Fusion result image

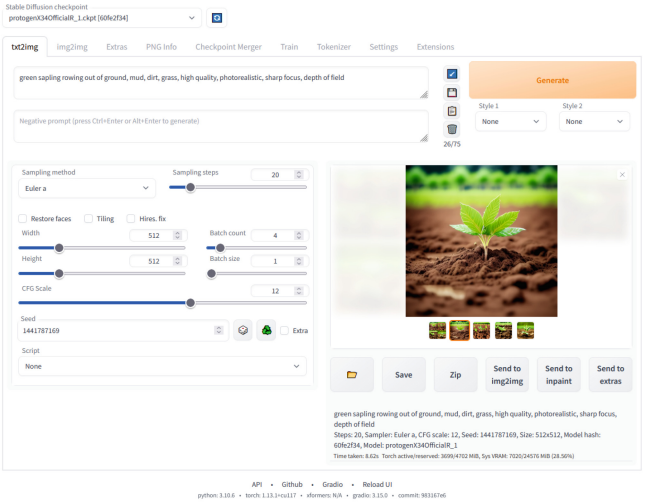
Fig. 3. Use case

4 Comparison with WebUI and ComfyUI

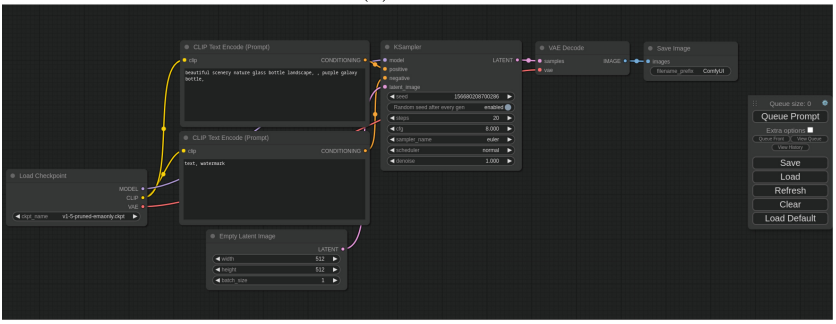
Compared to WebUI and ComfyUI presented in Fig. 4(a) and Fig. 4(b) respectively, the advantages of CanFuUI design approach are as follows:

(1) Our approach seamlessly integrates image generation and image editing into a unified interface, eliminating the need for external editors. In addition, the generated images are automatically recorded and displayed on one side, facilitating users in their creative iterations. (2) The presence of a resizable canvas streamlines image editing; this feature absents in both Stable Diffusion

WebUI and ComfyUI. (3) The community enables users to access image assets and facilitates communication, interaction, and feedback among users. (4) We divide image generation into two modules, AI Artistry and Advanced Guidance. AI Artistry involves fewer parameter adjustments, making it simple and user-friendly, while Advanced Guidance incorporates additional functionalities like ControlNet. (5) We provide a dedicated asset library, allowing users to upload and use image materials. Users can also upload the required image assets to the library through the community feature.



(a) WebUI



(b) ComfyUI

Fig. 4. WebUI and ComfyUI

5 Implementation

CanFuUI is developed using the React and UmiJS frameworks, along with the AntD UI component library. UmiJS is an enterprise-level React application

framework introduced by the Ant Group, designed to facilitate routing and lifecycle management in React applications. By utilizing an UmiJS-based framework, CanFuUI becomes easier to develop and extend. The AntD UI library, also developed by the Ant Group, offers modern UI components that adhere to the Ant Design guidelines for various frameworks such as React, Vue, Angular, and more. In CanFuUI, we use AntD 5.x UI components.

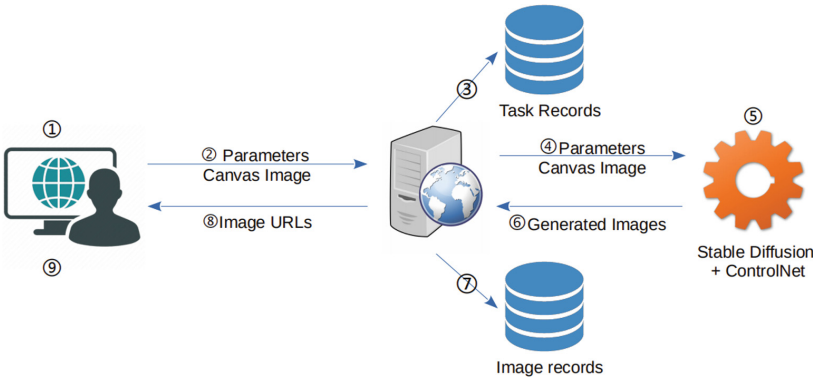


Fig. 5. The workflow of CanFuUI frontend and backend

The core editing function of CanFuUI is implemented using the Canvas APIs. Both sketches and image displays are rendered on an HTML canvas. The implementation consists of two main parts. The first part is the canvas-based view layer, responsible for rendering images, sketches, selections, and other visual elements. The second part includes various editing tools such as the eraser, pen, selector, and layer and history management. During page initialization, a canvas instance is created. Each tool selection triggers specific operations on the canvas instance. For example, when the pen tool is selected, a 2D context is created using the canvas's 'getContext('2d')' operation. The 2D context listens to mouse events and records the mouse trajectories to enable drawing functionality.

The interaction between the frontend and the stable diffusion backend is established through Restful HTTP requests. Prompts and parameters for Stable Diffusion and ControlNet are provided as JSON parameters within these requests. The Stable Diffusion engine generates images based on the specified parameters. The backend then saves these images and returns records containing the image URLs back to the frontend. The general workflow of CanFuUI is shown in Fig. 5.

6 Evaluation

A cohort of 32 first-year students majoring in Digital Media Technology from the College of Media Engineering, Communication University of Zhejiang, was

selected as test users for the survey. After a brief tutorial on how to use three distinct UIs (WebUI, ComfyUI and CanFuUI), we arranged 32 test users to individually employ these three UIs to generate the artwork shown in Fig. 3, and the minimum, average and maximum time consumption among the 32 test users was recorded in Table 1. Based on the data provided in Table 1, it is evident that the creative efficiency is higher when using CanFuUI compared to the other two UIs.

Table 1. Time consumption of test users employing three UIs

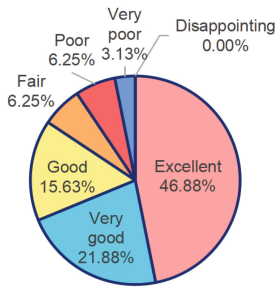
UI	Minimum time	Average time	Maximum time
WebUI	19 m 47 s	22 min 32 s	26 m 05 s
ComfyUI	21 m 14 s	26 m 12 s	30 m 24 s
CanFuUI	15 m 10 s	18 m 23 s	21 m 02 s

Furthermore, based on the questionnaire conducted with the 32 test users, 46.88% of the participants (15 test users) consider the CanFuUI design to be excellent and approximately 84.39% of the participants (27 test users) expressed positive feedback on the overall usability and user experience of the CanFuUI design, as shown in the pie chart in Fig. 6(a). These test users found the CanFuUI design elegant and intuitive, allowing for rapid parameter customization and providing abundant functionality and operational options for users.

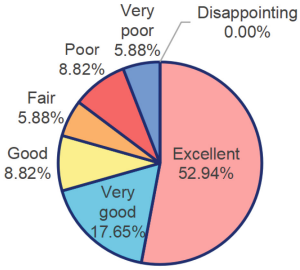
Out of the total 32 test users, approximately 52.94% (18 test users) rated the CanFuUI design as excellent in integrated image generation and editing. And about 84.39% (27 test users) give positive feedback as shown in Fig. 6(b). The result indicates that the test users are highly satisfied with the quality of the generated images, thus expressing that CanFuUI effectively meets their creative requirements.

Approximately 71.88% of the participants (23 test users) found the layer manipulation feature in the CanFuUI design to be useful, as shown in Fig. 6(c). CanFuUI design incorporates the functionality of layer operations, allowing users to make secondary modifications and adjustments to the generated images, thereby enhancing their ability to customize the images.

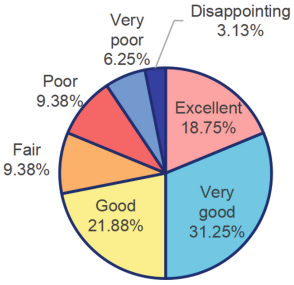
As presented in Fig. 6(d), approximately 40.63% of the participants (13 test users) perceived the asset library module and the gallery feature as being excellent. They expressed that the asset library's recording and management capabilities significantly facilitated their utilization and selection of the generated images. The survey result indicates that the advantages provided by CanFuUI have enhanced the creative experience of the user, the operational convenience and the creative flexibility, allowing them to effortlessly generate, edit, and customize high-quality images.



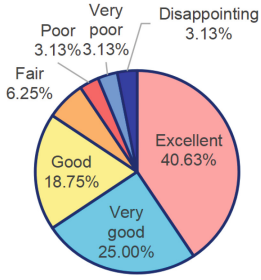
(a) Overall usability and user experience



(b) Integrated image generation and editing



(c) Layer manipulation



(d) Asset library module and gallery

Fig. 6. User evaluation demonstration

7 Conclusion

CanFuUI design simplifies user workflows and provides basic image editing tools and layer manipulation functionality, together with a convenient asset library. This helps users generate, edit, and create images with ease, leading to an enhanced user experience and satisfaction. Despite the numerous advantages of CanFuUI design, there are still some limitations and drawbacks that require further improvement and optimization as follows:

1. Currently, the range of parameter options in CanFuUI design is relatively limited, thereby restricting user control. Users seeking more fine-grained control over the generation process may wish to specify additional custom parameter choices.
2. CanFuUI design currently lacks real-time adjustments and interactive operations during image generation. Some users may prefer the ability to observe results during the generation process and make immediate adjustments to achieve more satisfactory results.
3. Although we offer certain model selection options, the current range is relatively narrow. A broader selection of diverse models would provide richer styles and effects, catering to a wider variety of user creative demands.

4. For new users who are just beginning to explore AI drawing, additional user tutorials and guidance may be required to help them better understand and master the functionalities and workflows of CanFuUI.

Our future endeavors involve offering customizable tools and editing options, incorporating interactive guidance and visual feedback, introducing intelligent prompts and suggestions, fostering community sharing and collaboration features, expanding output formats and resolutions, and optimizing the user interface's usability and responsiveness. We aim to provide users with a robust AI-based social platform for artistic creation.

Acknowledgements. We would like to thank the Zhejiang Provincial Blended First Class Online and Offline Course "Three-dimensional Character Design" (No. Z202Y22513), the Ministry of Education's Industry School Cooperation Collaborative Education Project "Research on PTA-Based Programming Training and Evaluation Model " (No. 202101151011) as well as the 17th batch Educational Reform Projects of Communication University of Zhejiang: "Cultivation and Practice of Computational Thinking in the Age of AI" for the generous funding support of the work referred to in this paper.

References

1. Stable diffusion ComfyUI. https://github.com/comfyanonymous/ComfyUI. Accessed 07 June 2023
2. Stable diffusion WebUI. https://github.com/db0/stable-diffusion-webui. Accessed 07 June 2023
3. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems, vol. 34, pp. 8780–8794. Curran Associates, Inc. (2021)
4. Goodfellow, I., et al.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) Advances in Neural Information Processing Systems, vol. 27. Curran Associates, Inc. (2014)
5. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851. Curran Associates, Inc. (2020)
6. Hu, E.J., et al.: LoRA: low-rank adaptation of large language models. In: The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, 25–29 April 2022. OpenReview.net (2022). https://openreview.net/forum?id=nZeVKeeFYf9
7. Jo, J., Lee, S., Hwang, S.J.: Score-based generative modeling of graphs via the system of stochastic differential equations. In: Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., Sabato, S. (eds.) Proceedings of the 39th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 162, pp. 10362–10383. PMLR (2022)
8. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8107–8116 (2020). https://doi.org/10.1109/CVPR42600.2020.00813

9. 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, pp. 10684–10695 (2021)
10. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, Lille, France, vol. 37, pp. 2256–2265. PMLR (2015)
11. Song, Y., Durkan, C., Murray, I., Ermon, S.: Maximum likelihood training of score-based diffusion models. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems, vol. 34, pp. 1415–1428. Curran Associates, Inc. (2021)
12. Zhang, L., Agrawala, M.: Adding conditional control to text-to-image diffusion models (2023)