International Journal of Advanced Smart Convergence Vol.12 No.2 90-95 (2023)
http://dx.doi.org/10.7236/IJASC.2023.12.2.90

IJASC 23-2-10

Research on AI Painting Generation Technology Based on the [Stable Diffusion]

Chenghao Wang, Jeanhun Chung*

Doctor's Course, Dept. of Multimedia, Graduate School of Digital Image and Contents
Dongguk University, Korea
**Professor, Dept. of Multimedia, Graduate School of Digital Image and Contents*
Dongguk University, Korea

*how9951@gmail.com, * evengates@gmail.com*

Abstract

With the rapid development of deep learning and artificial intelligence, generative models have achieved remarkable success in the field of image generation. By combining the stable diffusion method with Web UI technology, a novel solution is provided for the application of AI painting generation. The application prospects of this technology are very broad and can be applied to multiple fields, such as digital art, concept design, game development, and more. Furthermore, the platform based on Web UI facilitates user operations, making the technology more easily applicable to practical scenarios. This paper introduces the basic principles of Stable Diffusion Web UI technology. This technique utilizes the stability of diffusion processes to improve the output quality of generative models. By gradually introducing noise during the generation process, the model can generate smoother and more coherent images. Additionally, the analysis of different model types and applications within Stable Diffusion Web UI provides creators with a more comprehensive understanding, offering valuable insights for fields such as artistic creation and design.

Keywords: *AI painting, Artificial Intelligence, Stable Diffusion, Checkpoint, LoAR,*

1. Introduction

1.1 Research Background

AI painting generation technology in the field of artificial intelligence art has become a popular research direction, combining elements of artificial intelligence and art to enable computers to have the ability to paint and create images through algorithms and data. However, the application of AI painting generation technology still faces many challenges, such as insufficient stability in the generated results and low controllability. These issues limit the scope of application of AI painting generation technology and hinder its further development. To address these issues, a stable diffusion-based AI painting generation technique has been developed. This technology effectively enhances the stability and controllability of generated images. Stable diffusion is an emerging mathematical approach that simulates the evolution process of an image by solving partial differential equations, allowing for control and adjustment of the image. Compared to traditional methods such as Convolutional Neural Networks (CNN), stable diffusion enables more precise control over details and

Manuscript Received: April. 10, 2023 / Revised: April. 15, 2023 / Accepted: April. 20, 2023
Corresponding Author: evengates@gmail.com (Jeanhun Chung)
Tel: +82-2-2260-3767, Fax: +82-2-2260-3766
Professor, Dept. of Multimedia, Gradute School of Digital Image and Contents, Dongguk University, Seoul, Korea

textures of the image, resulting in more realistic and high-quality image generation. To facilitate the use of this stable diffusion-based AI painting generation technology, the Stable Diffusion Web UI AI painting open-source software has been developed. This platform enables AI painting generation through simple web-based operations. Users can input parameters and set controls for the generation process through the Web UI interface, such as painting style, dimensions, sampling methods, and more. This allows them to obtain stable and high-quality artworks through the AI painting generation process.

In conclusion, AI painting generation technology is a novel research direction that enhances the quality and stability of generated images, offering new possibilities in digital art and design fields. The trained models can determine the style and content of AI paintings, and the combination of different model types can produce unexpected effects.

1.2 Research Objectives

Using artificial intelligence (AI) tools to create art has become a novel and effective method nowadays. By utilizing machine learning algorithms and neural networks, users can conceive complex and intricate artworks, showcasing a uniqueness and originality that surpasses traditional art creation.

The uniqueness and originality of the Stable Diffusion Web UI AI painting software rely on the application of models, which can be trained according to individual choices and requirements. The trained models can determine the style and content of AI paintings, and combining different types of models can produce unexpected effects. Therefore, a comprehensive and clear guide is needed to help users understand the different types and otherness of models in AI painting with Stable Diffusion Web UI. This study is based on the AI painting open-source software Stable Diffusion Web UI, analyzing its model types and exploring the technical characteristics within each model type, as well as investigating the differences in the generated images among different models.

2. The Operating Principle and Practical Applications of Stable Diffusion Web UI

2.1 Operating Principle

Stable Diffusion AI painting is an image generation method based on artificial intelligence technology. It utilizes a technique called "Stable Diffusion" to gradually generate complex and highly detailed images starting from simple ones. The diffusion model primarily consists of two processes: the forward and backward processes. Initially, AI starts with an image filled with random noise. Then, a series of diffusion processes are applied to this image. The process of diffusion is similar to dropping a droplet of ink into a glass of water. The ink spreads throughout the water, and after a few minutes, it becomes randomly distributed throughout the entire water. Subsequently, AI performs reverse diffusion, gradually restoring the details of the image. During this process, AI references a target image or style, allowing the generated image to gradually exhibit these characteristics. Ultimately, after multiple iterations of diffusion and reverse diffusion, AI generates an image with rich details and a specific style.

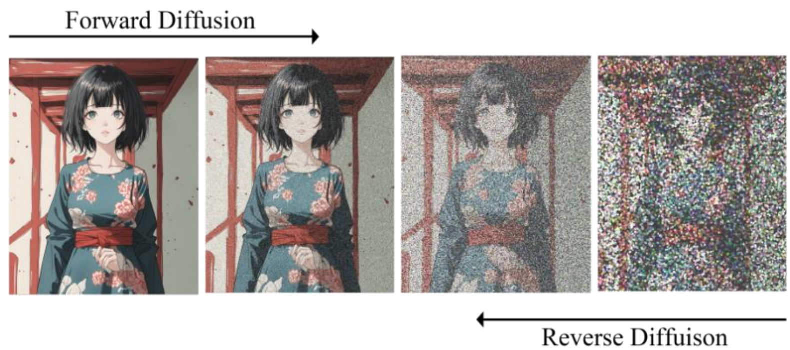


Figure 1. Stable Diffusion Process

However, the diffusion process operates in the image space, which can be computationally intensive, slow, and memory-consuming, especially when generating high-resolution images. Therefore, Stable Diffusion is a latent diffusion model that operates not in the high-dimensional image space but by first compressing the image into a latent space. Latent diffusion reduces memory and computational costs by applying the diffusion process on a lower-dimensional latent space instead of using the actual pixel space.

2.2 Practical Applications of AI Painting

Stable Diffusion Web UI is a browser interface based on the Gradio library, providing a user-friendly graphical interface for interacting with Stable Diffusion. With just a set of text descriptions, AI starts generating images. The main functionalities include text-to-image generation and image-to-image generation (such as local redraw and image up scaling). It is also possible to generate another image based on a specified image or generate images based on specified actions. Even with a wire frame image, AI can fill in the appropriate content according to your intentions. In addition to these functionalities, it also supports training custom models to enhance the quality of image generation.

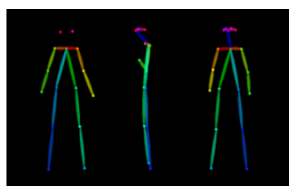
With the main basic generation functionalities of Stable Diffusion Web UI, one can quickly generate desired images. However, to generate specific styles, it is necessary to rely on models.

2.2.1 Text-to-Image. Text-to-image generation is an important basic creative function in Stable Diffusion Web UI, where AI can automatically convert user-inputted keywords (prompts) into images. For example, by inputting "Absurdres, best quality, 1girl, solo, standing, eye focus, black theme, (from side:0.5), looking at viewer, headphones, <lora:Color water_v4:1>" and clicking the generate button, the text-to-image generation feature will automatically generate an image that matches the description.

2.2.2 Image-to-Image. It can automatically transform uploaded images into another artistic style. By selecting a photo and entering keywords in the Prompt section, the image-to-image generation feature, based on the model, can generate the desired image by clicking the generate button.

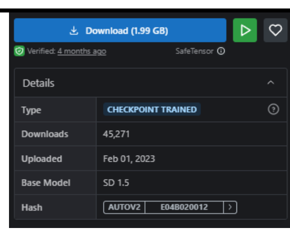
2.2.3 Control Net. It effectively addresses the challenge of accurately generating poses in AI painting. Purely relying on keyword-based control is insufficient to meet the need for precise control over details. The core capability of Control Net lies in enabling more controllable generation of the final image results by implementing specific settings.

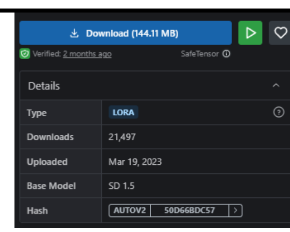
Table 1. Application of Stable Diffusion Web UI AI Painting

	Input	Output
Text to Image	Absurdres, best quality, 1girl, solo, standing, eye focus, black theme, (from side:0.5), looking at viewer, headphones, <lora:Color water_v4:1>	
Image to Image		
ControlNet		

3. AI Model Types

Table 2. Model Types

<table>
<tr><th>Feature</th><th>

Checkpoint Models</th><th>

LoRA Models</th></tr>
<tr><td>Model architecture</td><td>Diffusion model</td><td>Diffusion model</td></tr>
<tr><td>Training dataset</td><td>Massive dataset of image</td><td>Massive dataset of image</td></tr>
<tr><td>Model size</td><td>Bigger</td><td>Smaller</td></tr>
<tr><td>Image quality</td><td>High quality</td><td>High quality</td></tr>
<tr><td>Computation al cost</td><td>High computational cost</td><td>Low computational cost</td></tr>
</table>

Stable Diffusion is an open-source software, and to harness its powerful capabilities, it is essential to choose the right model.Currently, most models are trained from V1.4 or V1.5, and through the theoretical and practical operations conducted by AI in advance, it is discovered that the model in the Stable Diffusion Web UI is a pre-trained model. It is trained on a large dataset of images and can be used to generate images with various styles and quality. Additionally, there are differences in the effects of different models generated and the ways in which models are trained.

3.1 Checkpoint

Checkpoint is currently a popular and common model, as it includes everything needed for generating images and does not require additional files. It is derived from the basic training of the Stable Diffusion model and can be considered as a secondary development. It is a diffusion model that generates images by gradually adding details and textures to low-quality images. The quality of the generated images also depends on the size and complexity of the images. Due to the computational resources required by the generative model, the Checkpoint model establishes key points to save the processed parts at critical positions. In case of any issues, the progress is not lost entirely, making it convenient to resume computations in the future. Therefore, the Checkpoint model has a large file size.

Advantages. Checkpoint models can save a significant amount of time and resources by resuming training from the point where it was interrupted during the training process. They also improve the performance of the model on new tasks by transferring knowledge from a model that has already been trained on similar tasks. Additionally, comparing the results of training different models on different datasets allows for performance comparison.

3.2 LoRA

LoRA model (Low-Rank Adaptation of Large Language Models), commonly referred to as fine-tuned models, is used to cater to a specific style or specified character attributes. In cases where the data similarity is very high, using fine-tuned models can save a significant amount of training time and resources, resulting in the desired outcomes. The LoRA model is a small model constructed through training on a customized set of a few images. It can be combined with larger checkpoint models and has the ability to influence the results generated by the checkpoint models. It can quickly provide creators with new creative inspiration, broaden their design ideas and directions, thereby better achieving their design goals.

Advantages. Pre-trained models can be shared and used to build many small LoRA modules for different tasks. One central model can serve multiple downstream tasks, saving parameter storage. Inference stage does not introduce additional computational overhead, and it is orthogonal to other parameter-efficient fine-tuning methods, allowing for effective combinations. Training on characters is relatively stable, yielding good results. Overall, the LoRA model saves training time, improves accuracy, and speeds up the creative process.

4. Conclusion

With the rapid development of deep learning and artificial intelligence, generative models have made significant achievements in the field of image generation. This study is based on the AI painting generation technique of Stable Diffusion Web UI. It explores the technical characteristics of different models and the differences in their application in AI painting. Through the research, it has been found that the training process of the Checkpoint Model requires high computational power and the model itself has a large size. However, the training process offers greater flexibility and freedom. Compared to the Checkpoint Model, the LoRA model is more user-friendly for most users, as it has lower computational requirements and a smaller file size.

It is better suited for ordinary users.

Based on the understanding of the differences between different models, it is important to apply models based on the specific circumstances during training and application. Combining different models can lead to unexpected effects and generate more creative and imaginative artwork. These different types of models represent different approaches to solving image generation problems. The diffusion models are based on the idea that natural images can be generated by gradually adding noise to a blank canvas. On the other hand, neural style transfer models are based on the idea that the style of one image can be transferred to another image to generate new images. Each of these methods has its own advantages and disadvantages. The Stable Diffusion Web UI allows users to combine them and create a variety of different effects.

References

[1] Maerten, Anne-Sofie, and Derya Soydaner. "From paintbrush to pixel: A review of deep neural networks in AI-generated art." *arXiv preprint arXiv:2302.10913* (2023).
DOI: https://doi.org/10.48550/arXiv.2302.10913

[2] Hutson, James, and Morgan Harper-Nichols. "Generative AI and Algorithmic Art: Disrupting the Framing of Meaning and Rethinking the Subject-Object Dilemma." *Global Journal of Computer Science and Technology: D*, 23.1 (2023).

[3] Deckers, Niklas, et al. "The Infinite Index: Information Retrieval on Generative Text-To-Image Models." *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*. 2023.
DOI: https://doi.org/10.1145/3576840.3578327

[4] Stable Diffusion Art, How does Stable Diffusion work?
https://stable-diffusion-art.com/how-stable-diffusion-work/#Diffusion_model

[5] Medium, Generate High-Quality Image Using Stable Diffusion Web UI
https://betterprogramming.pub/generate-high-quality-image-using-stable-diffusion-webui-de96d6947d85

[6] Hu, Edward J., et al. "Lora: Low-rank adaptation of large language models." *arXiv preprint arXiv:2106.09685* (2021).
DOI: https://doi.org/10.48550/arXiv.2106.09685

[7] Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.

[8] Mazzone, Marian, and Ahmed Elgammal. "Art, creativity, and the potential of artificial intelligence." *Arts*. Vol. 8. No. 1. MDPI, 2019.
DOI: https://doi.org/10.3390/arts8010026

[9] Zhang, Lvmin, and Maneesh Agrawala. "Adding conditional control to text-to-image diffusion models." *arXiv preprint arXiv:2302.05543* (2023).
DOI: https://doi.org/10.48550/arXiv.2302.05543

[10] Lee, Seongmin, et al. "Diffusion Explainer: Visual Explanation for Text-to-image Stable Diffusion." *arXiv preprint arXiv:2305.03509* (2023).
DOI: https://doi.org/10.48550/arXiv.2305.03509