HiLL PUBLISHING

Advances in Computer and Communications, 2023, 4(3), 163-166
https://www.hillpublisher.com/journals/acc/
ISSN Online: 2767-2875

Exploration and Improvement of the Stable Diffusion Model in the Field of Image Generation

Lei Liang

Faculty of Humanities and Arts, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, China.

How to cite this paper: Lei Liang. (2023) Exploration and Improvement of the Stable Diffusion Model in the Field of Image Generation. *Advances in Computer and Communication*, 4(3), 163-166.
DOI: 10.26855/acc.2023.06.011

Received: May 30, 2023
Accepted: June 28, 2023
Published: July 24, 2023

***Corresponding author:** Lei Liang, Faculty of Humanities and Arts, Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau, China.

Abstract

Image generation is an important research direction in the field of computer vision, which covers many tasks such as image synthesis, image conversion, image editing and other tasks. In recent years, the rapid development of deep learning technology has provided powerful tools for image generation, where the stable diffusion model has attracted much attention in the image generation field as a kind of image generation model. Stable diffusion model (Stable Diffusion Model) is a generative model based on the diffusion process. The basic principle is to gradually generate the target image by conducting the diffusion process on the noise image. Different from the traditional generation models such as generative adversarial networks (GANs) and variation auto-encoders (VAEs), the stable diffusion model can gradually control the details and quality of the image during the generation process, with good generation stability and sample quality. With the continuous exploration and application of stable diffusion model in the field of image generation, researchers have proposed many improvement methods, including the improvement of generation network structure, loss function and sample optimization method, to further improve the generation effect and generation speed of stable diffusion model. This paper aims to explore and summarize the application status and improvement methods of stable diffusion model in the field of image generation. Through the study and improvement of stable diffusion model, it can provide new methods and ideas to achieve more realistic, diversified and controllable image generation effects.

Keywords

Stable diffusion model, image generation, technology application

Introduction

Image generation is an important research direction in the field of computer vision, which has wide applications in the fields of virtual reality, games, design and art. Traditional image generation methods are mainly based on rules and models, but it is difficult to generate high-quality, realistic and diverse images. In recent years, the rapid development of deep learning technology has provided new solutions for image generation, especially the emergence of generative models such as generative adversarial networks (GANs) and variation auto-encoders (VAEs), which have achieved remarkable results. The stable diffusion model gradually generates the target image by conducting the diffusion process on the noisy image, which has the characteristics of generating stability and high sample quality. Different from the traditional generation model, the stable diffusion model can gradually control the details and quality of the image during the generation process, with good generation control and interpretability [1]. It is of great theoretical and practical significance to deeply study the application status and improvement method of stable diffusion model in the field of image generation for promoting the development of image generation technology, improving the generation effect and generation control.

1. Current application of the stable diffusion model in image generation tasks

1.1 Image synthesis

The stable diffusion model shows a good potential for application in image synthesis tasks. For example, images with realism and diversity can be gradually synthesized by performing the diffusion process on noisy images. The stable diffusion model can generate images with different styles, textures and structures by controlling the parameters and steps of the diffusion process, and the resulting results are relatively stable in detail and quality.

1.2 Image conversion

The stable diffusion model is also promising in image transformation tasks. For example, by introducing the information of the target image in the diffusion process, the conversion effect of the image style migration, color conversion, image repair and so on can be realized. The stable diffusion model can gradually control the conversion degree and effect of the image during the generation process, and the generating results have good smoothness and consistency [2].

1.3 Image editing

The stable diffusion model can also be applied to image editing tasks, such as image content editing, style editing, etc. Local or global editing of the image is possible by performing a diffusion process on a noisy image and introducing editing operations during the diffusion process. For example, fine editing of the image can be achieved by adjusting the parameters in the diffusion process to control the intensity and range of editing.

1.4 Generate the control system

Compared with the traditional generative model, the stable diffusion model shows advantages in generating control. By controlling the parameters and steps in the diffusion process, the style, quality and details of the generated image can be more finely controlled. This makes the stable diffusion model more controllable and interpretable in generation tasks, which helps to meet the requirements for generation effects in practical applications [3].

1.5 Application and expansion

In addition to the above-mentioned tasks of image synthesis, image conversion, and image editing, the stable diffusion model can also be applied to other image generation tasks, such as image generation completion, image generation interpolation, etc. At the same time, the stable diffusion model can also be combined with other generation models to form a mixed model to further improve the generation effect and generation control.

2. Improved method for the stable diffusion model

2.1 Diffusion path control

Diffusion path control is a method to improve the stable diffusion model, aiming to make the generation process more controllable and flexible by controlling the diffusion path and speed of the generated image. Here are some possible diffusion path control methods. First, the learnable diffusion step size. In the stable diffusion model, the generation process is performed by gradually spreading the noisy image into a real image. The morphology and details of the generated images can be controlled by introducing a learnable diffusion step, which controls the amplitude and direction of the diffusion at each step. For example, convolutional neural networks or other learnable modules can be used to generate the step size for each step of the diffusion, thus making the generation process more flexible and controllable. Second is the nonhomogeneous diffusion. In the traditional stable diffusion model, each diffusion step is uniform, namely, the same noise is added to the whole image [4]. More abundant and diverse features can be obtained in the generated images by introducing non-uniform diffusion processes, for example, adding more noise in some regions. For example, images with more rich and diverse features can be generated by introducing different diffusion steps or diffusion modes into different regions in the generated image. Third is the multi-scale diffusion. The diffusion process at different scales can be used to generate images with different details and diversity.

For example, multi-scale noisy images or multi-scale generative networks can be introduced to provide the resulting images with features at different scales by conducting diffusion processes at different scales. This can help to generate images more rich and diversified, thus improving the generation effect.

2.2 Generate the control mechanism

First is the conditional generation. By introducing condition information, such as an image label, a semantic mask, etc., as a control signal for generating the generated image, the form, content, or style of the generated image is controlled according to the condition information during the generation process. For example, the noise, color, texture of the generated image, etc., can be adjusted according to the conditional information to generate the image meeting the conditional requirements. Second, guide the image. A guide image can be introduced as a reference for the generated image, thus guiding the generation path of the image during the generation process. For example, the difference between the generated and guided images can be controlled during generation, thus achieving fine control of the morphology and content of the generated image. Third is style control. We can introduce style control mechanism and control the style of the image during generation. For example, the image with different styles can be adjusted by introducing the style vector of the generated image at each step of the generated image. Fourth is user interaction. The form, content or style of the generated image can be controlled through interaction with users, such as user input and user feedback. For example, the form or content of the generated image can be adjusted according to the user's input, thus enabling user-customized image generation. Fifth is dynamic control. The diffusion path or speed of the generated image can be adjusted according to the dynamic information in the generation process, such as the current state of the generated image and the generated speed, so as to realize the dynamic control of the generated image. For example, the generation parameters can be dynamically adjusted according to the quality or generation speed of the generated image, thus controlling the generation path of the image during the generation process [5]. Motion capture: also known as "motion capture", is a technology to record and process the actions of people or other objects. From the technical point of view, the essence of motion capture is to measure, track and record the motion track of objects in three-dimensional space. Each of the 17 physical inertial sensors includes a gyroscope, an accelerometer, and a magnetometer. It can sense the rotation around three axes in space, and calculate roll pitch and heading through complex algorithms. The communication equipment includes the data output by the sensor, and calculates the position of the limbs relative to the "main heart bone". At the same time, special algorithms are used to help calculate the position of the main heart bone relative to the ground. After these information are transferred to the computer, the computer will model the 3d scene by the information. "Vtuber" also uses this technology.

2.3 Model architecture improvement

First is the multi-scale diffusion. Multi-scale diffusion paths can be introduced to consider image features at different scales in the generation process, thus generating more abundant and detailed images. For example, a diffusion model can be applied at different levels or resolutions of the generated image, thereby generating images with different scale features. Second, the pyramid spreads out. Multiple diffusion models can be combined into a single pyramid structure, so as to generate images on different levels of pyramids at the same time, thus realizing multi-level image generation. For example, different generative models or parameters can be used at different levels of the pyramid to generate images with different details and diversity. Third is the attention mechanism. An attention mechanism can be introduced to weight different areas or features of the image during generation, thus enabling local control of the generated image. For example, the weights of different areas of the generated image can be generated by the self-attention mechanism, which can control the local features of the image during generation. Fourth is the structural improvement of the generator and the discriminator. The structure of the generator and discriminator can be improved to improve the quality and diversity of generated images. For example, deeper generator or discriminator networks can be introduced, or more complex generator or discriminator structures, such as recurrent neural network (RNN) and convolutional neural network (CNN), can be used to improve the representation ability and generation effect of the model. Fifth is the morphological diffusion. A morphological diffusion model can be introduced to control the morphology of the image during generation. The morphological diffusion model can control the morphological structure of the image by introducing morphological operations, such as corrosion, expansion, etc., to control the morphological features of the generated image.

2.4 Data augmentation and training strategies

First is data enhancement. The training data of the model can be enriched by enhancing the diversified data of the input image, so as to improve the generalization ability of the model and the diversity of the generated results. For example, more diverse training samples can be generated through image transformation operations such as rotation, translation, zoom, and flip, so that the model can better process different image styles and contents when generating images. Second is incremental training. Incremental training strategies can be adopted to gradually introduce new training data and styles, so that the model can gradually adapt to different types of image generation tasks. For example, you can conduct initial training with a simple set of training data, and then gradually introduce more complex and diverse training data, so that the model can better learn richer and more complex image generation patterns. Third is the unsupervised / self-supervised training. Unsupervised or self-supervised training strategies can be adopted, thus using label-free data to train the model. For example, unsupervised training can be carried out by automatically generating the adversarial loss in the adversarial network (GAN), or self-supervised training can be carried out through self-reconstruction loss, so that the image generated by the model itself can be used as the target for model training. Fourth is conditional generation. Thus, a conditional generation strategy can be introduced to consider additional conditional information, such as category labels, and attribute labels, so as to achieve more refined control over the generated image. For example, conditional information can be fused with a diffusion model to control specific properties or styles of the image during generation, thus generating a more personalized image. Fifth is reinforcement learning. Reinforcement learning strategies can be introduced, so that the generative model can self-adjust and optimize according to the feedback of the generated results. For example, reinforcement learning algorithms can be used to optimize the generation strategy of the generated model by introducing the evaluation index of the generated results as the reward signal, so as to improve the quality and diversity of the generated results.

Epilogue

The stable diffusion model has wide application potential in the field of image generation, and has achieved some results in different tasks and scenarios. The performance of the stable diffusion model and the quality of the generated results can be further improved by improving the model architecture, controlling the diffusion path, introducing the generative control mechanism, data enhancement and optimization of training strategies. These improvement methods can be used individually or in combination, selected and adjusted according to specific needs and problems. However, there are still some challenges in the field of image generation, such as the diversity and authenticity of the generated results, the efficiency of training and inference of the model, and the controllability of the generation process. Future studies can further explore and improve these problems, and combine advanced technologies such as deep learning, reinforcement learning, self-supervised learning, etc., to further improve the application performance and practical application value of the stable diffusion model in the field of image generation. Overall, the stable diffusion model, as an emerging generative model method, still has a huge research space and application prospect in the field of image generation.

References

[1] Li Zonglin, Zhang Shengping, Liu Yang, Zhang Zhaoxin, Zhang Weigang, Huang Qingming. Text-driven face image generation and editing based on a multi-level residual mapper [J / OL]. Journal of Software: 1-15 [2023-04-20]. DOI:10.13328/j.cnki.jos.006767.

[2] Lai Lina, Mi Yu, Zhou Longlong, Rao Jiyong, Xu Tianyang, Song Xiaoning. Summary of Generative adversarial networks and text image generation methods [J / OL]. Computer Engineering and application: 1-23 [2023-04-20]. http://kns.cnki.net/kcms/detail/11.2127.TP.20230314.1549.022.html.

[3] Yang Hongyu, Yang Fan. Adversarial sample detection method based on image denoising and image generation [J / OL]. Journal of Hunan University (Natural Science Edition): 1-10 [2023-04-20]. http://kns.cnki.net/kcms/detail/43.1061.n.20230308.1449.002.html.

[4] Jiang Nianyun. High technology and its industrial development mechanism are used to look at it from the history of computer development [J]. Technology think-tank, 2023(03):12-16.DOI:10.19881/j.cnki.1006-3676.2023.03.02.

[5] Li Chun. Computer Development based on Intelligent Information Processing [J]. China New Communications, 2023, 25 (04): 31-33.