

Overview of Artificial Intelligence Painting Development and Some Related Model Application

Mingyang Liu^{1,*}

¹School of Computing and Data Science, Xiamen University Malaysia, Xiamen, Fujian, 361006, China

ABSTRACT. The application of Artificial Intelligence (AI) in painting has gained significant attention in recent years, promising to revolutionize the digital art world and enhance traditional painting processes. AI painting involves the use of computer algorithms and machine learning techniques to generate and manipulate digital images. This paper provides an overview of current research in AI painting, covering key findings, literature review, and analysis of the field's current status and future direction. Research has shown that AI algorithms can generate unique digital art forms and automate tasks in traditional painting processes, freeing up time and effort for artists. The field of AI painting holds immense potential and further research and development will likely lead to new and innovative applications. In addition, this technology has a long way to go on the road of commercialization.

1. INTRODUCTION

AI painting is a growing field that has the potential to revolutionize the creation and consumption of digital art. This technology is attracting increasing attention due to its potential to expand the creative horizons of artists and designers, and foster the development of innovative applications [1]. With AI painting, even individuals with limited technical skills can produce high-quality works of art in a matter of minutes, greatly lowering the barrier to entry in the field of digital art.

In recent years, research in the field of AI painting has made impressive strides, with a growing number of studies exploring innovative ways to use artificial intelligence to generate and manipulate images. For example, researchers have used advanced techniques like automatic learning and unsupervised learning to segment and analyze image details such as character posture and identity [2]. Additionally, the use of Deep Convolutional Generative Adversarial Networks (DCGAN) [3] has demonstrated the potential for unsupervised learning to outperform supervised learning in certain scenarios. Some advancements are achieved by improving convolution networks with the help of Generative Adversarial Networks [4]. And the development of diffusion model enables AI painting to draw high-resolution and realistic images [5] etc.

Despite these impressive achievements, there are still several challenges that need to be addressed in order to further advance the field of AI painting. Some of the key areas of focus include improving the quality and authenticity of images generated by AI, developing new algorithms that can produce a greater visual appeal and diverse styles, discuss the originality of artworks

produced by AI and its ethical issues as a unique personal expression, and exploring the potential of AI to create interactive and immersive digital art experiences.

2. RELEVANT MODELS

2.1. Generative Adversarial Networks

Generative Adversarial Networks (GAN) are a class of deep generative models that have gained significant attention in recent years due to their ability to generate high-quality, diverse and realistic images.

GAN were introduced by Goodfellow et al. (2014) [4], where two deep neural networks, generator and discriminator, are trained in an adversarial manner. The generator tries to generate synthetic samples that are indistinguishable from real ones, and the discriminator tries to correctly identify whether a sample is real or fake. Since then, GAN have been used in various computer vision and machine learning tasks, including image generation, style transfer, and representation learning.

The generator and discriminator are both deep neural networks, and the training process for GAN is accomplished through a minimax game, where the generator tries to minimize the error it makes in fooling the discriminator, and the discriminator tries to maximize its accuracy in distinguishing between real and fake samples. The game continues until a balance is reached, such that the generator is producing high-quality synthetic samples that are difficult to distinguish from real ones, and the discriminator is as accurate as possible in differentiating between the two.

* Corresponding author: a1278336390@gmail.com

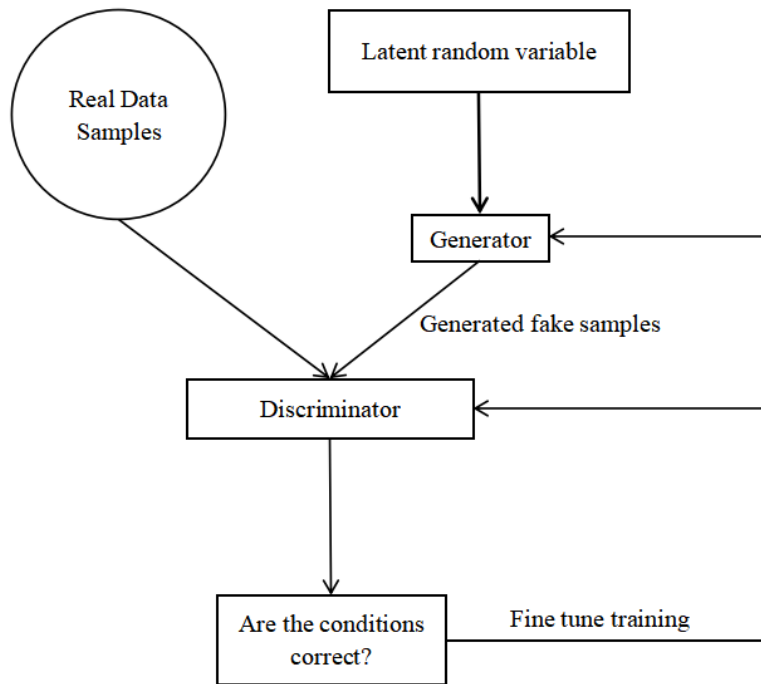


Fig. 1. Generative Adversarial Networks schematic diagram. [6]

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

The GAN mathematical formula [4] refer to "Equation (1)", where G=generator; D=discriminator; P_{data}(x)=distribution of real data; P(z)=distribution of generators; x=sample of P_{data}(x); Z=sample of P(z); D(x)=discriminator network; G(z)=generator network [6].

Overall, the principle of GAN is to use the adversarial relationship between the generator and discriminator to learn a generative model that can produce synthetic data samples that are similar to real ones. The process allows for unsupervised representation learning, as the generator and discriminator can learn to extract and use important features from the data to generate high-quality synthetic samples.

However, GAN also have several known disadvantages [7], including:

Instability: GAN can be difficult to train due to their unstable loss landscape and the possibility of mode collapse, where the generator produces limited variations of the training data.

Training data requirements: GAN require a large amount of training data to learn the underlying distribution of the data and generate high-quality samples.

Model collapse: GAN may produce images that lack diversity, with certain modes or features being ignored by the generator.

Evaluation: It can be challenging to evaluate the quality of GAN-generated samples, as there is no objective metric for measuring the similarity between the generated and real data.

Computational resources: GAN can be computationally expensive to train and require

significant computing resources, such as Graphics Processing Units or Tensor Processing Units.

GAN is not suitable for processing discrete data, such as text.

Certainly, with these shortcomings, there are also some optimization models. Here, four popular improved GAN models, their advantages, disadvantages and principles are briefly provided.

2.1.1. Four popular improved GAN models

Conditional GAN (CGAN): CGAN was proposed by Mirza and Osindero (2014) [8] to enable the generation of specific types of data, such as images of certain classes or with certain attributes. CGAN extends the original GAN architecture by conditioning both the generator and the discriminator on additional input, such as a class label or an attribute vector and transforms the unsupervised GAN into a semi-supervised or supervised model. This allows for more precise control over the generated samples and enables applications such as image-to-image translation. One disadvantage of CGAN is that it requires the availability of the conditioning information at both training and inference time, which may not always be feasible.

Deep Convolutional Generative Adversarial Networks (DCGAN) are a type of GAN proposed by Radford et al. (2015) [3] that use convolutional neural networks in both the generator and discriminator networks. DCGAN are capable of generating higher quality images than previous GAN models by learning hierarchical representations of the input images. The generator network takes random noise as input and generates images that are progressively refined through deconvolutional layers. The discriminator network, on

the other hand, learns to distinguish between real and fake images. DCGAN have been used for a variety of applications, including image synthesis, image editing, and style transfer. However, like Style-based GAN [2], DCGAN also require a significant amount of computing

resources and training time. This kind of model may cause the over-generation problem, which is another type of model collapse. So strange pictures may be generated, but the discriminator still treats them as real pictures.



Fig. 2. Strange pictures caused by over-generation. [9]

Wasserstein GAN (WGAN): WGAN was proposed by Arjovsky et al. (2017) [10] to address the instability of the original GAN training process. WGAN replaces the discriminator with a critic that outputs a scalar score, and enforces a Lipschitz constraint on the critic by clipping the gradients of the critic's weights. This leads to a more stable training process and a smoother loss landscape, which allows for better convergence and higher quality generated samples. One of the main disadvantages of the WGAN model is that it can be more difficult to train than traditional GAN. This is because the gradient clipping used to enforce the Lipschitz constraint can cause problems with vanishing or exploding gradients, which can make it harder to converge to a stable solution. Additionally, the WGAN loss function can be less intuitive to interpret than the standard GAN loss, which can make it harder to diagnose problems or make improvements to the model.

Style-based GAN (StyleGAN): StyleGAN was proposed by Karras et al. (2019) [2] to address the limitation of previous GAN models in generating high-resolution and diverse images. StyleGAN introduces a novel generator architecture that separates the generation of high-level attributes from the low-level details of the images, allowing for more control over the generated images. StyleGAN also uses adaptive instance normalization and progressive growing techniques to improve the quality and diversity of the generated images. One disadvantage of StyleGAN is that it requires a large amount of computing resources and training time to generate high-quality images.

2.2. Diffusion Model

Diffusion models are a family of generative models that have been widely used in various applications such as computer vision, sequence modeling, audio, artificial intelligence, etc [11]. The core principle of the diffusion model is to iteratively update a random noise vector until it approximates a target distribution [12]. This technique has proven to be effective in generating high-quality

samples that are visually appealing and exhibit realistic patterns [13].

The diffusion model is a type of autoregressive model that generates samples by updating a random noise vector, following a pre-defined set of diffusion steps. These steps are designed to gradually increase the complexity of the noise distribution by adding a small amount of noise to each element of the vector at each step. In this way, the diffusion process iteratively transforms the noise distribution into the target distribution. The key idea of the diffusion model is to sample a set of noise vectors, apply the diffusion steps to each of them, and use the resulting vectors to generate the samples [12]. The algorithm used in the diffusion model is called the Langevin dynamics [14], which is a well-studied algorithm in the field of statistical mechanics. The Langevin dynamics ensures that the updates to the noise vectors are smooth and gradually bring the distribution closer to the target distribution.

2.2.1. Some advantages and disadvantages of diffusion model

The diffusion model has several advantages over other generative models, such as GAN and Variational auto-encoder (VAE). VAE are another popular method for AI painting [15]. Unlike GAN, VAE are generative models that use a continuous latent space to generate new images. The key idea behind VAE is to map an image to a lower dimensional representation, called a latent code, and then use the latent code to generate new images. The generation of new images is achieved by sampling the latent space and mapping the samples back to the original image space. One of the main advantages of the diffusion model is that it does not require any adversarial training or complicated loss functions, which are commonly used in GAN [10]. This makes the training process more stable and less prone to mode collapse, a common issue in GAN. Another advantage of the diffusion model is that it does not require any additional parameters or network structures to generate diverse samples, which is a

common issue in VAE [15]. The diffusion model also allows for flexible conditioning, which enables the generation of samples that are conditioned on some input, such as text or images [16]. Finally, the diffusion model produces high-quality samples that exhibit realistic patterns and textures, which makes it suitable for a wide range of applications such as image generation [13].

Despite its advantages, the diffusion model also has some limitations. One of the main challenges of the diffusion model is that it can be computationally expensive, especially when dealing with high-dimensional data such as images or videos. This is because the diffusion model requires a large number of steps to generate high-quality samples, which can increase the computation time significantly [11]. Another limitation of the diffusion model is that it can be difficult to train, especially for large datasets. This is because the Langevin dynamics used in the diffusion model requires careful tuning of the learning rate and other hyperparameters [13]. Finally, the diffusion model may not be suitable for all types of data distributions. One such situation is when the data distribution is highly structured or complex. The diffusion model assumes that the spread of information or particles will be proportional to the density of the medium through which they are traveling. However, in highly structured or complex distributions, this assumption may not hold true.

To solve the problem of slow sampling and excessive calculation consumption, the existing improved algorithms are divided into four categories: Training Schedule, Training-Free Sampling, Mixed-Modeling, and Score-Diffusion Unification. For Data Structure Diversion, there are three types of methods: Continuous Space, Discrete Space, and Constrained Space. There are two methods for Likelihood Optimization, namely, Improved ELBO and Variable Gap Optimization. Finally, Dimension Reduction can be realized through Mixed-Modeling [11].

3. FUTURE PROSPECTS AND PROBLEMS

In the field of art, AI painting provides artists with new technologies and tools for creating art. An important aspect of AI painting is the interaction between the user and the AI system. In some cases, the user may provide high-level constraints or guidance to the AI system, such as a desired color palette or brush style [17]. This technology enables artists to explore new possibilities and improve their creativity. In the future, AI painting can also be used to repair old and damaged artworks.

In the design industry, AI painting can be used to generate graphics and logos. AI systems can learn from previous designs and generate new designs according to user specifications, thus reducing the time and effort required to create designs. AI painting can also be used in the fashion industry to generate new designs and predict future fashion trends. For example, in the future, fashion brand LV may use AI algorithm to design the latest brand trademark, predict fashion trends and optimize the inventory management system.

AI painting can be used to create special effects, backgrounds and characters in movies and video games in the entertainment industry. After several iterations in the future, the AI system may be able to generate realistic 3D models, animations and textures to enhance the visual appeal of these products. In video games, AI painting may be used to create dynamic landscapes and environments to respond to user input.

In the future, the potential application of AI painting is not limited to the art, design and entertainment industries. It can also be used to generate 3D medical images in healthcare, create 3D building models in buildings, and so on.

While the use of AI in creating art may offer exciting new possibilities for creative expression, it also raises significant questions about originality and authenticity in artistic practice, as well as ethical issues related to ownership and authorship.

One of the primary challenges in assessing the originality of AI-generated art is the question of who or what is actually producing the work. While the software and algorithms used to generate the artwork may have been created by human programmers, the specific output generated by the AI may be considered wholly original and not necessarily attributable to any individual creator. This can create a complicated legal and ethical landscape for issues such as copyright and intellectual property, as well as challenges in determining who should receive recognition or compensation for the artwork.

Another aspect to consider is the role of personal expression in artistic creation. Traditionally, art has been considered a unique expression of an individual artist's perspective and creativity, reflecting their experiences, emotions, and values [18]. While AI-generated art can certainly produce striking and visually compelling images, it may be argued that the absence of personal intent or subjective experience in the AI undermines the notion of it as a unique personal expression. This could impact the value and meaning of the artwork for viewers, collectors, and critics, and raise questions about the role of technology in shaping the cultural significance of art. In addition, the development of artificial intelligence may cause large-scale unemployment, technological monopoly and other problems in human society.

As the field of AI-generated art continues to evolve, it is important to consider these issues thoughtfully. Then we can come to engage in critical discussions around the benefits and potential risks of this innovative form of artistic practice.

4. CONCLUSION

In conclusion, AI painting is a rapidly growing field that has the potential to revolutionize the way we create and appreciate art. Through the use of deep learning algorithm and neural network, AI painting has an incomparable material base and efficiency. It can use different color, style, composition and other basic elements to render hundreds of different pictures with the same element style at one time. However, it is also possible to have problems with the authenticity and value

of artworks generated by AI, as well as the impact of AI on the role of human artists. In addition, there are also ethical problems in the use of data and algorithms to create and disseminate art. In the follow-up practice, these issues need to be discussed and explored.

In this paper, I mainly summarize the principles, advantages and disadvantages of the relevant development of several model algorithms related to AI painting. Then it looks forward to the future of AI painting and summarizes several possible problems.

REFERENCES

1. Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A Neural Algorithm of Artistic Style. arXiv preprint arXiv:1508.06576.
2. Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4401-4410).
3. Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.
4. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11), 139-144.
5. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684-10695).
6. CSDN AIStudio. (2022). Machine Learning Notebook - Generation of an Antagonistic Network (GAN) Overview and Getting Started Example. <https://blog.csdn.net/ZwaterZ/article/details/124840602>.
7. CSDN TensorFlow. (2019). Principle, advantages and disadvantages of GAN. <https://blog.csdn.net/chary8088/article/details/89681869>.
8. Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
9. CSDN GAN. (2022). Original GAN (DCGAN) defect. https://blog.csdn.net/weixin_41883450/article/details/127119414.
10. Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.
11. Cao, H., Tan, C., Gao, Z., Chen, G., Heng, P. A., & Li, S. Z. (2022). A survey on generative diffusion model. arXiv preprint arXiv:2209.02646.
12. Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840-6851.
13. Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems*, 34, 8780-8794.
14. Roberts, G. O., & Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 341-363.
15. Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
16. Kawar, B., Zada, S., Lang, O., Tov, O., Chang, H., Dekel, T., ... & Irani, M. (2022). Imagic: Text-based real image editing with diffusion models. arXiv preprint arXiv:2210.09276.
17. Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).
18. The teaching and research team of the art communication direction of the Central Academy of Fine Arts. (2022). Art. <https://baike.baidu.com/item/%E8%89%BA%E6%9C%AF/12004323?fr=aladdin>.