The Generation and Analysis of Art Image based on Generative Adversarial Network

Jingfeng Li

Reading Academy, Nanjing University of Information Science & Technology, Nanjing, China

Keywords: GAN, CNN, Image Generation.

Abstract: Recent years have witnessed a significant surge in the attention garnered by Generative Adversarial Networks (GAN), owing to their remarkable capability to generate high-quality and realistic images. The objective of this research is to devise a model based on GAN that can effectively produce images with diverse and realistic attributes, ensuring a high level of quality. The proposed method involves training a GAN architecture consisting of a generator and a discriminator. In the training process, GAN engage in an adversarial game between the generator and discriminator models. The generator and discriminator of a GAN use a deep convolutional neural network (CNN) architecture to continuously improve performance. The generated images are efficiently transformed by a series of deconvolutional layers in the generator to incorporate random noise inputs. This study selects the dataset which include various artists about portraits. The proposed GAN model has been demonstrated to successfully generate high-quality images, as supported by the experimental results. The generated images exhibit diverse features and demonstrate the effectiveness of the GAN architecture in picking up the patterns in the portraits. In conclusion, this research highlights the potential of GANs in image generation.

1 INTRODUCTION

Artistic image generation is the automatic drawing of images through artificial intelligence techniques. It has vast applications in the fields of entertainment, advertising, design and education. This field is of great significance in bridging the gap between technology and art, democratizing art, and promoting interdisciplinary collaboration. In conclusion, artistic image generation has the potential to enhance visual experiences, stimulate creativity, and shape the future of the visual arts and creative industries.

In recent years, the field of generative adversarial network (GAN)-based image generation has witnessed remarkable advancements. Researchers have proposed a plethora of techniques and architectures aimed at elevating the quality and diversity of the generated images (Goodfellow et al 2014). Progressive GAN (PGAN) has achieved remarkable success in generating high resolution and realistic images by addressing the pattern collapse and training instability faced by GANs. PGAN proposed by Karras et al. employs a progressive training strategy. The images are generated from coarse to fine, resulting in more detailed and visually appealing outputs (Karras T et al 2017). This approach has been widely adopted and shown to outperform traditional GAN architectures. Another research direction is to explore the use of attentional mechanisms in GANs to improve the generation process. Self-attentive GAN (SAGAN) was proposed by Zhang et al. A self-attentive module was introduced to capture long-range dependencies. Additional various techniques and architectures are used to improve the quality of generated images (Zhang et al 2018). This technique has shown promising results in generating images with better global coherence and fine-grained details.

Furthermore, researchers have investigated the use of generative models based on variational autoencoders (VAEs) and GANs, known as VAE-GANs, to overcome limitations in traditional, GANs. Zhao et al. introduced the VAE-GAN framework, which combines the strengths of both models to generate high-quality images while maintaining a more stable training process (Zhao et al 2017). This approach has shown improvements in image quality and mode coverage. Additionally, techniques such as Wasserstein GANs (WGANs). WGANs, introduced by Arjovsky et al., utilize the Wasserstein distance metric to provide a more stable training process and improve the quality of generated images (Adler and

Li, J.
The Generation and Analysis of Art Image Based on Generative Adversarial Network.
DOI: 10.5220/0012799200003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 224-228
ISBN: 978-989-758-705-4

Lunz 2018). Spectral normalization, proposed by Miyato et al., helps stabilize the training of GANs by constraining the Lipschitz constant of the discriminator network (Miyato 2018). In conclusion, recent advancements in GAN-based image generation have explored techniques such as progressive training, VAE-GANs, and to improve the stability and quality of generated images, researchers have proposed various techniques and architectures. Over time, researchers have come up with many ways to improve and extend Gans. The Progressive GANs (PGans) are proposed by Tera et al in and Variation, which achieve higher quality, more stable, and more diverse image generation by gradually increasing the resolution of generators and discriminators (Karras et al 2017). In addition, their another research in 2019 StyleGAN is introduced to achieve better image generation quality and controllability by controlling the Style vector of the generator (Karras et al 2019). Also, Yunjey Choi et al in 2018 presents StarGAN, through the use of a generator and discriminator, a unified GAN method has been developed to enable the conversion of images between multiple domains, thereby achieving the capability of multi-domain image conversion (Choi et al 2018).

The main objective of this study is to propose a novel image generation model with improved GAN to enhance performance on various artist portrait datasets. Specifically, progressive training strategy allows generating images in a coarse-to-fine manner. This improves the performance of high resolution and realistic image generation. Second, attention mechanisms are incorporated into the GAN framework to improve the generation process. By utilizing the self-attention module to capture long-range dependencies. The generated images are enhanced, improving both their global consistency and fine-grained details. Meanwhile, this paper analyzes and compares the prediction performance of the proposed model with other state-of-the-art GAN architectures. In summary, this study offers valuable insights into the advancement of GAN-based image generation techniques that integrate attention mechanisms and regularization techniques.

2 METHODOLOGY

2.1 Dataset Description and Preprocessing

The dataset used in this study, named Art Portraits dataset, is from Kaggle (Dataset 2023). And the Art Portraits dataset consists of 4117 samples, with a batch size of 64 which reduced the image size to (64,64), presuming. So that it will be computationally less taxing on the GPU. The dataset contains that most of the images are portraits. A portrait is a painting representation of a person, The face is predominantly depicted portraits along with expressions and postures.

Before conducting the experiments, the Art Portraits dataset underwent preprocessing step---Normalization. For the data normalization, the data in the range is converted between 0 to 1. This helps in fast convergence and makes it easy for the computer to do calculations faster. Each of the three RGB channels in the image can take pixel values ranging from 0 to 256. Dividing it by 255 converts it to a range between 0 to 1. And there are some instances which shown by Fig. 1 (Dataset 2023).

Figure 1: Images from the Art Portraits dataset.

2.2 Proposed Approach

The field of computer vision and artificial intelligence has been revolutionized by the technology of GANs-based image generation. Comprising a generator and a discriminator as their core components, GANs belong to the category of deep learning models. These models are designed to generate and discriminate between data samples, enabling them to learn and produce realistic outputs in various domains. The basic core idea lies in the fact that the training generator of the GAN is used to generate highly realistic images. Additionally, these images are almost indistinguishable from real images. At the same time, after the discriminators have been trained iteratively, the discriminative model is able to distinguish between real and generated images. This adversarial fitting process motivates the two network models to continuously improve their performance, thus generating more convincing and high-quality information. The main module of GANs-based image generation involves a two-step process. From the beginning, the generator starts by taking random noise as input and employs it to generate an image. The generated image is subsequently forwarded to the discriminator for evaluation, along with real images from a training dataset. The input needs to be classified as real or synthetic image. Work is given to the discriminator. The objective of the training is for the generator to progressively enhance its ability to deceive the discriminator by producing images that are progressively more challenging to differentiate from real images. This adversarial training process leads to the improvement of both the generator and discriminator over time.

Fig. 2 illustrates the image generation process based on gass. First, the image material is imported with a batch size of 64, and this batch of data is standardized, and then the GAN model is established. Taking a random noise as input, the generator network transforms it into sample data by generating outputs. This process is repeated multiple times to produce diverse samples from the given seed noise input, a learning curve is drawn and evaluated based on the training data.

Figure 2: The pipeline of the model (Picture credit: Original).

2.2.1 Generator and Discriminator

In the GAN model, the Generator network assumes the crucial role of generating fresh images. It takes random noise as input and undergoes a transformative process to produce an image that closely resembles the desired target domain. The generator learns to map the input noise to the output image by leveraging deep neural networks, such as convolutional neural networks (CNN) or generative models like Variational Autoencoders (VAEs). The generator produces images that are close to real images. The inner workings of the generator effectively capture the underlying patterns and arrangements of the data set. The discriminator network, on the other hand, acts as a critic. Through learning, the discriminator is able to react quickly to images and effectively distinguish between images that are real or fictitious by the generator that fall within the target range. The discriminator is also implemented using deep neural networks, typically CNNs. And the discriminator, on the other hand, is trained to classify images as either real or fake. Its primary goal is to effectively identify the generated images and distinguish them from real images with high accuracy.

2.2.2 GAN

The GAN model's significance lies in its ability to generate new and original images that resemble real artworks. It has opened up new possibilities for artistic expression, creative design, and data synthesis. GAN is a generative model for deep learning characterized by two adversarial networks, generative and discriminative, to learn the distribution of data and generate new samples.The structure of a GAN consists of two main components: a generator and a discriminator. The generator is responsible for generating false samples from random noise and trying to deceive the discriminator. The discriminator, on the other hand, is a binary classifier that distinguishes between real samples and false samples generated by the generator. The generator receives a random noise vector as input, maps it to the data sample space through a series of transformations, and generates spurious samples. The discriminator receives the true samples and the false samples generated by the generator and tries to distinguish the classes. The goal of the discriminator is to maximize the ability to correctly classify the true and false samples. The generator's goal is to minimize the discriminator's ability to discriminate the generated false samples, even if the discriminator is unable to distinguish false samples from true samples. By iteratively training the generator and the discriminator, the two networks work against each other and gradually improve their performance. The generator and discriminator can be modeled using a deep neural network, such as a multilayer perceptron (MLP) or CNN. The input to the generator network is a random noise vector and its output is a generated sample. The input to the discriminator network is real samples or generated samples and its output is the classification result for the input samples. Through the adversarial training process, GAN is able to optimize the dynamic balance between the generator and the discriminator to generate more realistic samples. It has a wide range of applications in areas such as generating images, language modeling, and audio synthesis. By training the GAN model on a dataset of existing artworks, it can learn the underlying patterns, styles, and textures present in the training data and generate new artworks that capture the essence of the art style. In the implementation flow of this experiment, the GAN model is trained using a two-step process. Initially, the generator network receives random noise as input and utilizes it to generate an image. Subsequently, the discriminator can discriminate between the generated images and the real images in the training database. The discriminator is then used to evaluate the screened situation and provide feedback to the generator.

Based on the discriminative feedback, the generator is automatically updated to produce increasingly realistic images that can fool the discriminator. The iterative process persists until the generator achieves the ability to produce high-quality images that closely resemble authentic artworks. Fig. 3 shows the general flow of GAN.

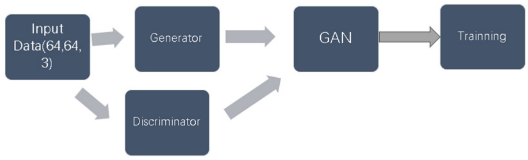

Figure 3: The processing of GAN (Picture credit: Original).

2.2.3 Loss Function

The loss function comprises of two components: Generator and discriminator loss. Generation loss quantifies the extent to which the generator spoofs the discriminative model, while discrimination loss measures the ability to discriminate between real and occurring samples, as follows:

$$\text{Generator} = -\log(\mathrm{K}(\mathrm{N}(\mathrm{z})) \tag{1}$$

where discriminator is describe as K while N represents the generator, and z represents the input noise.

$$Discriminator\ Loss = -\log\big(K(x)\big) - log\left(1 - K\big(N(z)\big)\right) \tag{2}$$

where K is the symbol of the discriminator, x represents real samples, N represents the generator, and z represents the input noise.

2.3 Implementation Details

The implementation of the GAN-based image generation system involves several important aspects, including the system's background, data augmentation techniques, and hyperparameters. The system is constructed based on the GAN architecture, comprising two fundamental components: Generator Discriminator Network. The generator network model is responsible for generating additional inputs, on the other hand, the discriminator network model evaluates the veracity of the resulting image. The system leverages the adversarial training process to improve the generator's ability to create realistic and visually appealing artworks. While to enhance the diversity and quality of the generated images. Data augmentation techniques are utilized, which involve the application of various transformations, such as rotation, scaling, and flipping, to augment the training dataset. Data augmentation helps the model generalize better and produce more varied and realistic artworks. And the system's performance heavily relies on the selection of hyperparameters. The factors that influence the training process of GANs include the learning rate, batch size, number of training epochs, and network architecture. The learning rate determines the step size in the optimization process. The batch size, on one hand, determines the quantity of samples processed in each training iteration. On the other hand, the number of training epochs signifies the frequency at which the complete dataset is traversed through the model during training. The network architecture refers to the specific design and configuration of the generator and discriminator networks.

3 RESULTS AND DISCUSSION

This chapter aims to analyze the learning curve after the training of the current model and discuss the image quality created by the GAN model.

As shown in the Fig. 4, the discriminator loss fluctuates significantly compared with the generator loss, reaching a peak value of 1.3 between epoch0-25 and then slowly stabilizing at 0.8, while it is claimed that its loss reaches a minimum value close to 0.5 between epoch0-25 and then becomes gentle and close to 0.7. Through continuous iterative training, the generator and discriminator compete with each other, and finally reach a balance point, so that the generator can generate high-quality samples, and the discriminator can accurately distinguish between the real images and the generated images.

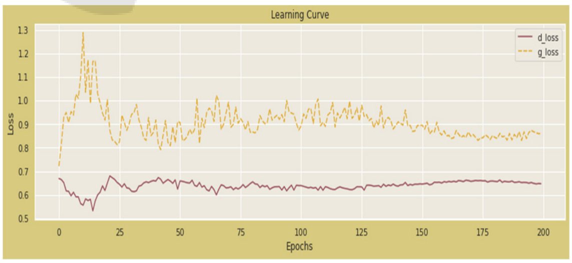

Figure 4: The learning curve (Picture credit: Original).

As shown in the following Fig. 5, the image results show the outline of the portrait, the GAN picked up the patterns in the portraits. But the character details and color rendering are not of high quality. This means that the model still needs to be further improved to meet the color rendering and the drawing of the five features of the characters.

Figure 5: The visualization of the result (Picture credit: Original).

4 CONCLUSION

This paper focuses on GANs-based image generation. A novel method that makes the art images is proposed to analyze and generate high-quality images. The proposed method involved a detailed process in which a generator and discriminator were trained in an adversarial manner to generate realistic images. A series of experiments are conducted to assess the capability of the proposed method. The experimental results show that Art using GAN achieves significant results in both image quality and diversity. The generated images exhibited high fidelity to the target distribution and showcased a wide range of variations. However, future work can be explored from the following aspects. First, Gans require more data when dealing with large databases, so you can consider increasing the amount of data. Second, there are many inconsistencies in the data, which is quite complicated for Gans to learn, so the results can be improved by cleaning up the data styles that have consistency. Overall, the proposed Art by GAN achieves promising results in GAN-based image generation. Continued future work will focus on exploring and enhancing this approach, thereby advancing image generation techniques and creating new opportunities for diverse applications in the fields of computer vision and artificial intelligence.

REFERENCES

I. Goodfellow J. Pouget-Abadie M. Mirza B. Xu D. Warde-Farley S. Ozair, A. Courville Y. Bengio "Generative Adversarial Networks," in Proceedings of the 27th International Conference on Neural Information Processing Systems, vol. 2014, pp. 2672–2680.

T. Karras T. Aila S. Laine J. Lehtinen "Progressive Growing of GANs for Improved Quality, Stability, and Variation." arXiv:2017, unpublished.

H. Zhang I. Goodfellow D. Metaxas A. Odena, "Self-Attention Generative Adversarial Networks," in Proceedings of the 32nd Conference on Neural Information Processing Systems, vol. 2018, pp.3-5

R. Zhao M. Mathieu Y. LeCun, "Energy-based Generative Adversarial Network," in Proceedings of the 5th International Conference on Learning Representations, arViv. 2017. Unpublished

J. Adler, S. Lunz, "Banach wasserstein gan", Advances in neural information processing systems, arXiv: 2017. Unpublished

T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, "Spectral Normalization for Generative Adversarial Networks," in Proceedings of the 6th International Conference on Learning Representations, vol. 2018, pp.2-4

T. Karras et al, "Progressive Growing of GANs for Improved Quality, Stability, and Variation". Karras T, Aila T, Laine S, et al. "Progressive growing of gans for improved quality, stability, and variation", arXiv:2017, unpublished.

T. Karras, S. Laine, T. Aila, "A style-based generator architecture for generative adversarial networks", Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, vol. 2019, pp. 4401-4410

Y. Choi, M. Choi, M. Kim, et al, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," Proceedings of the IEEE conference on computer vision and pattern recognition, 2018 pp. 8789-8797

Dataset, "Art Portraits," kaggle, 2023.9.5 , https://www.kaggle.com/datasets/karnikakapoor/art-portraits