Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/92/20241743

GAN-based image generation

Yunze Zhao

College of Imformation Science and Engineering, University of Jinan, Jinan, Shandong, China

arthurzhao0820@gmail.com

Abstract. Generative Adversarial Networks (GANs), as a deep learning model, have made significant progress in the field of image generation and style migration. This study aims to methodically investigate GAN-based image generation methods. First, this paper outlines the basic principles of GAN and its application in image generation, focusing on analyzing the structure and performance of representative models such as DCGAN, ProGAN and StyleGAN. This paper summarizes the improvement methods such as WGAN and LSGAN, and evaluates their efficacy in improving the stability of the model and the quality of the generated images in view of the problems of pattern collapse and instability faced by GANs in the training process. Finally, this paper discusses the limitations of current techniques and possible future directions, and suggests research prospects in the field of multimodal fusion and 3D image generation. The research in this paper provides a theoretical framework and useful suggestions for enhancing the application of GAN methods in image generation.

Keywords: Generative Adversarial Network, Image Generation, Face Generation, Artistic Image Generation, Image Style Migration.

1. Introduction

Generative Adversarial Network has rapidly become an important technology in the field of deep learning since it was proposed [1] . The unique architecture of GAN consists of a generator and a discriminator, and through the adversarial training of the two, the generator is gradually able to generate high-quality images that are similar to the real data distribution. This breakthrough brings novel solutions to tasks like image production and enhances the field of computer vision.

GAN can generate realistic, high-resolution images from random noise. The advent of models such as Progressive Growth GAN (ProGAN) has significantly improved the quality and diversity of generated images [2] and Deep Convolutional GAN (DCGAN) [3] .

Many studies on various models and their applications have been started in the fields of GAN image production and style migration research. DCGAN improves image quality by using deep convolutional neural networks [2]. ProGAN dramatically improves image generating stability and resolution by producing pictures layer by layer [3].

The purpose of this research is to conduct a thorough analysis of GAN-based image generation techniques and look at their performance and problems in real applications.

DOI: 10.54254/2755-2721/92/20241743

2. Basic Theory and Techniques of GAN

2.1. The basic structure of GAN

2.1.1. Principles. The Generative Adversarial Networks is a novel method used for unsupervised learning which can be defined by training a pair of networks in a competitive manner [4]. These two networks are the generative networks and the discriminative network, which are antagonistic. Let's make an analogy here: the discriminative network is like a appraiser and the generative network is like a counterfeiter.

Counterfeiters fabricate products with the intention of deceiving appraisers, while appraisers are tasked with verifying the genuineness of an object. When the appraisers determines that the item is a forgery by the counterfeiter, then the counterfeiter has to improve his counterfeiting skills. Conversely, if the counterfeiter succeeds in fooling the appraiser, then the firm will enhance his or her appraisals. They both are trained simultaneously and compete with each other [4]. The two networks continuously enhance themselves until they achieve a state of Nash equilibrium. That's what "generative" and "adversarial" mean.

We suppose that the generator is denoted as $G(x,\theta_2)$, which is multilayer perceptrons. The θ_2 here represents the parameter of the generative model. Firstly, we define an input noise variable $p_z(z)$ and it is mapped to the data space by the generative model to generate a synthetic sample and and passing it to the discriminator. The discriminator here is also denoted as $D(x,\theta_1)$ which is used to determine whether the input data originates from the generative model or the training data. The output of D represents the probability that x is a sample from the training data. So If the sample received by the discriminator is a real sample x , its output can be represented as $D(x)$. Conversely, if the sample is a synthetic sample generated by the generator, its output is represented as $D(G(x))$ .Finally, the discriminator is trained to accurately distinguish the data source [1], while the generator is trained to produce data that closely matches the distribution p_g of the training data. This is the overall structure of a generative adversarial network.

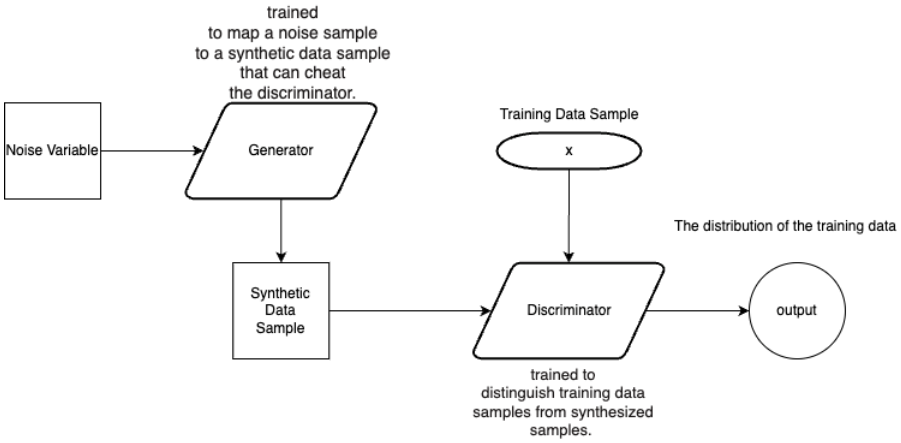

Figure 1. Generative Adversarial Network model

2.1.2. Loss functions. The adversarial objectives of the generator and discriminator make up the loss function in the original GAN model. Maximum ability to distinguish between generated and genuine data is the aim of the discriminator, and minimum ability to distinguish between generated and actual

DOI: 10.54254/2755-2721/92/20241743

data is the aim of the generator. In concrete terms, the discriminator *D* aims to maximize the following equation:

$$E_{x\sim p_{data}(x)}[log\ D(x)] + E_{z\sim p_z(z)}[log\ (1 - D(G(z)))]$$

where *x* is a sample from a real data distribution, and *z* is the random noise vector. *G*(*z*) is the sample generated by the generator. The generator *G* tries to minimize the second term of the above equation, thereby deceiving the discriminator into not being able to distinguish between generated and true samples [1]. This loss function can be thought of as a game of minima maxima where the discriminator and the generator are always trying to maximize their own goals.

Nevertheless, training directly with the original GAN loss function can result in the generator's gradient disappearing, which would make the training process unstable. Researchers have suggested various enhanced loss functions and optimization techniques to help with this issue.

2.1.3. Wasserstein loss function for WGAN. To address the common difficulties of gradient vanishing and pattern collapse in the original GAN, Wasserstein GAN (WGAN) proposes a new loss function. WGAN uses the Wasserstein distance to measure the difference between generated and real data distributions. Specifically, the loss function of WGAN is:

$$E_{x\sim p_{data}(x)}[D(x)] - E_{z\sim p_z(z)}[D(G(z))]$$

In this situation, the discriminator tries to maximize this distance, whereas the generator tries to minimize it [5]. Shearing the discriminator's weights to ensure they fit the continuity condition is an important aspect of WGAN. This strategy does very well at reducing instability during training.

2.1.4. Least Squares Loss Function for LSGAN. By substituting a least squares loss function for the original GAN's cross-entropy loss, Least Squares GAN further enhances the quality of the image produced by the generator. The loss function of LSGAN is defined as follows:

For the discriminator:

$$\frac{1}{2}\mathrm{E}_{x\sim p_{data}(x)}[(D(x) - b)^2]\frac{1}{2}\mathrm{E}_{z\sim p_z(z)}[(D(G(z)) - a)^2]$$

For the generator:

$$\frac{1}{2}\mathrm{E}_{z\sim p_z(z)}[(D(G(z)) - c)^2]$$

where *a* and *b* are the discriminator's target output values for the real and generated samples, respectively. *c* is the target output value of the generator. LSGAN decreases this least squares error to improve the quality of samples generated and decrease instability in GAN training [6] .

The selection of an optimization method is as important to the GAN training process as the design of the loss function. Most GAN models are trained using the Adam optimizer because it can alter the learning rate to ensure convergence and stability in complex tasks [7].The Adam optimizer improves the generator and discriminator, and the GAN training process is faster.

2.2. Training Difficulties and Improvements in GAN

2.2.1. Mode Collapse. Mode collapse is a typical issue in GAN training, in which the generator produces only a limited number of sample patterns during the training process, failing to cover the whole distribution of the training data. This results in a lack of diversity in the generated samples, and the model struggles to create high-quality photos from various classes [1]. The mode collapse happens because the generator's optimization is unduly reliant on discriminator feedback, which does not supply sufficient information to encourage the generator to explore a more diverse sample space.

DOI: 10.54254/2755-2721/92/20241743

Researchers have offered a number of improvement strategies to address the mode collapse problem. Wasserstein GAN, for example, decreases pattern collapse by measuring the difference between the generated and real distributions with the Wasserstein distance [5].The fundamental improvement to WGAN is to produce smoother gradients during training by clipping the discriminator weights to ensure continuity.

2.2.2. Training instability. GAN training is typically unstable. Generator and discriminator parameter updates becoming unbalanced, which makes convergence challenging. This instability is induced in part by GANs' min-max game structure, in which the discriminator and generator alternately optimize their respective objective functions. It causes issues such as vanishing or gradient explosion.

To solve this issue, Least Squares GAN suggests replacing the original cross-entropy loss function with the least squares error to produce smoother gradients and lessen training instability [6]. This technique increases training stability and sample quality. It also results in a more balanced optimization process for the generator and discriminator.

3. Image Generation

3.1. Overview of Image Generation Models

The image generating models of the Generative Adversarial Network (GAN) family have evolved through numerous stages. Starting with the original Deep Convolutional Generative Adversarial Network (DCGAN) and moving to Progressive Generative Adversarial Networks (ProGAN) and, more recently, StyleGAN. Each of these models has significantly improved the clarity, resolution, and diversity of the images produced.

3.1.1. Deep Convolutional Generative Adversarial Networks. Deep Convolutional Generative Adversarial Network is one of the first models to add Convolutional Neural Networks (CNNs) into GANs, whose goal is to increase generated image quality by leveraging the benefits of deep learning [3]. The fundamental innovation of DCGAN is the removal of entirely linked layers in traditional GANs and the use of convolutional and inverse convolutional layers to generate images. This structure not only improves the stability of the generated samples, but also enables the model to produce higher-quality images.

DCGAN is excellent at producing low-resolution photographs with lots of detail and texture, but it has limitations when it comes to producing high-resolution photos. However, the creation of DCGAN sets the stage for the design of upcoming models.

3.1.2. Progressive Generative Adversarial Networks. Progressive Generative Adversarial Networks (ProGAN) greatly improve the quality and resolution of generated images by employing a training strategy that steadily raises image resolution [2]. Different with conventional GAN models, ProGAN first trains the discriminator and generator at a low resolution before progressively raising it to reach the desired resolution.

This progressive training method significantly increases the detail and authenticity of the generated images while also resolving the instability that arises when training GAN models at high resolution. ProGAN produces images with much higher quality and much more variance than DCGAN. It makes it possible to create richer and more varied samples.

3.1.3. Style Generating Adversarial Network. Style Generative Adversarial Networks (StyleGAN) are another key improvement in GAN modeling in the field of image generation. It provides for fine-grained control over the features of generated images with the introduction of a style control module [8]. A distinguishing feature of StyleGAN's generator architecture is that latent vectors are transformed to an intermediate space before being introduced via the style control module into different generative network levels.

DOI: 10.54254/2755-2721/92/20241743

By using this method, StyleGAN may independently alter several aspects of the picture, such color, texture, form, and so on, producing images with greater quality and a wider range of styles. Moreover, StyleGAN improves the realism and consistency of the generated images by addressing the problem of unstable picture structure, which can occasionally arise in ProGAN.

3.1.4. Comparison and summary. The three crucial phases in the creation of GAN models for picture production are DCGAN, ProGAN, and StyleGAN. DCGAN's use of convolutional neural networks significantly improved image quality. ProGAN addressed the instabilities that occur during high-resolution generation. It dramatically improves picture resolution and detail reproduction. StyleGAN's innovative style control module design gives users complete control over the characteristics of produced pictures so that it can leads in considerable improvements in image quality, diversity, and high-resolution creation. he expansion of these models illustrates the ongoing successes and developments of GANs in the area of image production. They have sparked the explosive expansion of associated applications and established the foundation for contemporary picture creation technology.

3.2. Evaluation metrics for GAN-generated images

3.2.1. Fréchet Inception Distance. The Fréchet Inception Distance (FID) is a measure of the similarity between the distribution of the produced image and the true image, proposed by Heusel et al. in 2017 [9]. FID is derived by running the generated and real images through the Inception network separately to extract features, and then calculating the Fréchet distance between the mean and covariance matrices of the features. Specifically, FID is calculated as:

$$FID = \| \mu_r - \mu_g \|^2 + Tr\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{\frac{1}{2}}\right)$$

where μ_r and μ_g represent the mean vectors of real and generated image features respectivel. Σ_r and Σ_g denote their covariance matrices, respectively.

FID can better reflect the similarity between the generated image and the real image at the distribution level. A low FID value usually indicates that the generated image is of high quality and the sample distribution is close to the real data.

FID is sensitive to sample diversity and quality, although it uses the Inception network for feature extraction. As a result, while dealing with various forms of data, the network structure may limit it. In addition, FID requires a larger number of image samples for computation to ensure the reliability of the assessment results.

3.2.2. Inception Score. Inception Score (IS) is another widely used GAN evaluation metric proposed by Salimans et al. in 2016 [10]. The basic principle of IS is to categorize the generated images by Inception network, and use the category distribution of the output to measure the diversity and quality of the images.

Specifically, the formula for IS is:

$$IS = exp\left(\mathbb{E}_x KL(p(y|x) \| p(y))\right)$$

In this equations, $p(y \mid x)$ denotes the conditional probability that image x belongs to category y . $p(y)$ denotes the average class distribution of all generated images. A high IS value indicates that there are many categories and a high quality image produced.

IS can reflect the diversity and fidelity of the generated images to a certain extent, but it relies only on the classification results of the Inception network in its evaluation, which may be affected by the training bias of the network itself.

IS cannot directly assess the distributional similarity between generated images and real images, so in practical applications, it is usually combined with other metrics for comprehensive evaluation. In

DOI: 10.54254/2755-2721/92/20241743

addition, IS is prone to produce higher scores when dealing with low-diversity samples, leading to a distorted evaluation of image diversity.

IQA metrics can directly reflect the visual quality of images and are suitable for assessing the detail fidelity and visual perception of generated images. However, IQA methods usually lack a comprehensive assessment of the overall quality of the image, especially in terms of diversity anddistributional similarity, and their limitations are more obvious.

3.3. Application Case Studies

The wide application of Generative Adversarial Networks in the field of image generation has demonstrated their powerful potential and diverse capabilities. In practice, GAN has been used for many tasks, including face generation, art image generation, style migration, etc. The following demonstrates the application of GAN in these fields through specific cases and compares the generation effects of different models under the same task.

3.3.1. Face Generation. Face generation is one of the most widely used areas of GANs. Early generative adversarial networks have been able to generate realistic face images, but still lack in detail and realism [3]. With the advancement of GAN technology, Progressive Generative Adversarial Networks (ProGAN) have significantly improved the quality and resolution of face generation, and through the training method of gradually increasing the image resolution, ProGAN generates high-resolution face images that are more detailed and lifelike, and are close to the quality of real photos [2].

However, it is the Style Generative Adversarial Network (StyleGAN) that has really taken face generation to the next level. styleGAN is able to independently adjust individual features of a face image, such as hairstyle, skin color, and facial expression, through its unique style control mechanism to generate highly diverse and naturally lifelike face images [8].

In practical applications, StyleGAN has been widely used in various face generation tasks such as avatar creation and video game character design. The face images it generates are extremely difficult to distinguish from real faces in visual perception. It demonstrates its great potential in high-quality image generation.

3.3.2. Artistic Image Generation. Art image generation is an important application of GAN in the creative field. By generating images of artistic styles through GAN, researchers are able to explore the migration between different styles and generate artworks with unique visual effects.CycleGAN is a typical example. It makes it possible to transform between several artistic styles, for example, turning a photograph into a painting a la Van Gogh [11].

StyleGAN does very well when it comes to creating artistic images. Thanks to its style control module, StyleGAN is not only able to generate images that conform to a specific art style, but also to generate innovative and diverse artworks by mixing different styles. For example, StyleGAN can generate images that are both impressionistic and characterized by modern abstract art [8]. This generative capability has important applications in fields such as digital art creation and cultural heritage preservation. *3.3.3 Image Style Migration.* The technique of creating a new image by incorporating the style of one image into the content of another is known as image style migration. In this assignment, GANs demonstrate special strengths. Pix2Pix is a GAN-based conditional model. To enable precise picture style migration, it learns the mapping relationships between an input image and a target image [12].

Compared to Pix2Pix, StyleGAN is more flexible in style migration tasks.StyleGAN can control different layers of features in an image to enable multi-level style migration. For example, users can use StyleGAN to generate more natural and stylistically unique images by changing only the surface texture or color style of an image while maintaining its overall structure [8]. This ability makes StyleGAN promising for a wide range of applications in fields such as advertising design and movie special effects.

DOI: 10.54254/2755-2721/92/20241743

4. Future developments

Generative Adversarial Networks (GANs) have made significant achievements in the field of image generation, but there are still many challenges and unsolved problems. Future directions will center around model diversity, complexity of generated images, and integration with other deep learning techniques. The following are a few possible research directions, including multimodal style migration and 3D image generation.

4.1. Multimodal style migration

Multimodal style migration is an important direction for future GAN research. Existing style migration models usually only support the conversion of a single style, while multimodal style migration aims to apply multiple styles simultaneously or generate new styles from multiple styles. This approach can extend the generative capabilities of existing models to enable them to handle more complex visual tasks [11]. For example, combining features from different art genres to generate innovative new artworks, or applying data from different modalities to improve diagnostic accuracy in the field of medical image processing.

4.2. 3D image generation

3D image generation is another research direction with great potential. With the development of Virtual Reality (VR), Augmented Reality (AR) and 3D printing technologies, generating realistic 3D images and models becomes more and more important. 2D image generation is the primary application of traditional GAN models. There are additional difficulties with 3D generation, like generation speed and spatial consistency. GAN-based 3D generative models have shown their ability to generate 3D scenes and shapes in recent years [13].

Future research may combine graphics methods with deep learning techniques. This combination across fields could advance 3D image generation techniques for wider applications in fields such as entertainment, healthcare, and industry design.

5. Conclusion

This thesis delves deeply into the use of Generative Adversarial Networks (GAN) in picture production and style migration, investigating GAN's development history, key technologies, evaluation metrics, and real-world applications. We demonstrate how GAN has greatly enhanced image quality, resolution, and diversity by evaluating the performance of multiple image generation models, focusing on DCGAN, ProGAN, and StyleGAN data [3][2][8]. These results advance computer vision research and open up new possibilities for application sectors such as virtual reality, medical image processing, and art production.

The key contribution of this paper is to completely classify the progress of GAN technology in picture production and demonstrate how it can be applied in a variety of applications using case studies. Furthermore, we look at the metrics used to judge the quality of GAN-generated images, such as FID (Fréchet Inception Distance) and IS (Inception Score), as well as their limitations and applications in diverse situations [9][10]. By examining these evaluation measures, this article serves as a reference for future research into how to more scientifically evaluate the quality of GAN-generated images.

However, this study has certain disadvantages. We were unable to extensively investigate various technical concerns and alternative GAN models due to space constraints. In the application case study, we demonstrate the use of GAN in several tasks but do not go into detail about how each task is implemented.

Overall, Generative Adversarial Networks is a powerful generative model with a wide range of applications. Despite these challenges, GANs will remain important for research and applications in the future as technology progresses and new approaches are discovered.

DOI: 10.54254/2755-2721/92/20241743

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[2] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[3] A Radford. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[4] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1):53–65, 2018.

[5] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[6] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017.

[7] D Kinga, Jimmy Ba Adam, et al. A method for stochastic optimization. In *International conference on learning representations (ICLR)*, volume 5, page 6. San Diego, California;, 2015.

[8] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019.

[9] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[10] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

[11] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[12] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[13] Thu Nguyen-Phuoc, Chuan Li, Lucas Theis, Christian Richardt, and Yong-Liang Yang. Hologan: Unsupervised learning of 3d representations from natural images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7588–7597, 2019.