Applications of Generative Adversarial Network Technologies for Image Generation

Zhixuan He

School of Mathematics and Statistics, Shandong University, Shandong, 264209, China
202200820051@mail.sdu.edu.cn

Abstract. Image generation is crucial in the digital age, revolutionizing the way that people perceive and interact with visual content. It is a cornerstone for various industries, from advertising and marketing to film production and gaming. By enabling the creation of images, it fuels creativity, enriches storytelling, and enhances user engagement. The content of this article is to summarize and analyses the advancement generative adversarial network (GAN)-based image generation. GAN has produced impressive achievements and serves a variety of purposes in the realm of images. It is a deep learning model, which consists of a generator and a discriminator, and the adversarial methods are used in training this model. In this paper, the basic theory of GAN is described, and then the various applications of GAN in the image field are introduced in detail, including the applications of image generation, image modification, image style transfer and image restoration. Finally, some existing problems of the model are discussed, and its development direction and trend are forecasted.

Keywords: Generative Adversarial Network, Image Generation, Deep Learning.

1 Introduction

Recently, alongside the advancement of artificial intelligence, the area of generative models has also continued to advance. Generative Adversarial Network (GAN), a famous deep learning model, can be traced back to 2014 and proposed by researcher Ian J. Goodfellow et al. and has been continuously developed till now [1]. Following previous arts, lots of models are proposed. In the field of deep learning, generation-based research has grown in importance, attracting the attention of a large number of researchers, and showing considerable application potential in many fields.

The generator and discriminator neural networks make up the majority of a GAN. To "fool" the discriminator, the generator aims to provide an authentic sample. In order to force the generator to provide high-quality images with sufficiently similar appearance to the real ones, discriminators must be able to discriminate between the real sample and the sample from the generator. This process is done through an adversarial competition between the two.

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_39

Researching the applications of GAN in the area of generation is very valuable. With the advancement of technology, the field of generative model has also been developed, which provides a material basis for its application in image generation. On the demand of the industry: in the game industry, movie special effects, monitoring equipment, medical images and many other fields, the demand for high-quality graphics generation is increasing; On the practical value: generative adversarial network has many functions, including but not limited to image generation, image restoration, image enhancement and so on, and its practical value in image generation cannot be ignored. All in all, it is significant and valuable to study the application of GAN in the task of creating images.

The paper discusses and summarizes the many functionalities of the GAN model in the area of picture generation, as well as outlining its prospect.

2 Basic Theory of GAN

In this section, two important components of GAN, the generator and discriminator, will be sequentially elaborated [1]. The following content displays GAN's construction.

The generator is designed for delivering samples with similar appearance to the real ones. During training, the generator network (denoted G) outputs a generated sample (denoted $G(z)$) from its hidden space random variable z. The training purpose is to make the fake sample $G(z)$ generated by it as similar as possible to the real sample X. Mathematically, it reduces the distributional differences between p_G and p_{data}, respectively denoting the distribution of fake and real samples. Improving discriminator's capacity to discern between authentic and fraudulent samples is the discriminator's aim. In the training process, the discriminator network (denoted D) simultaneously receives real sample X and generates sample $G(z)$, and outputs the discriminant result $D(x)$. Through training, the differences between synthesized and real samples are minimized and finally the whole model can produce better quality data samples.

The goal of training is to improve the respective capabilities of generator and discriminator. As the learning process could be regarded as a minimax game, that is, enhancing the capability of generator G will make it more difficult for discriminator D to differentiate true and false samples. The objective is denoted as:

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{x \sim p_{data}(x)}[logD(x)] + \mathbb{E}_{z \sim p_z(z)}\left[\log\left(1 - D(G(x))\right)\right] \tag{1}$$

$V(D,G)$ is the objective function. Taking as input a real image x, $D(x)$ outputs the judgement. $D(G(x))$ is the output value after the discriminator receives the false sample $G(z)$, that is, the probability of judging the false sample as true. $\mathbb{E}_{x \sim p_{data}(x)}[logD(x)]$ denotes the expectation. p_{data} and $\mathbb{E}_{z \sim p_z(z)}\left[\log\left(1 - D(G(x))\right)\right]$ represents the expected value calculation for the vector z. It is sampled from the generator input distribution $p_z(z)$ as the random noise. After continuous alternating optimization between discriminator and generator, the quality of the model is getting better and better.

The loss of generator could be calculated as an expression:

$$-\frac{1}{m}\sum_{i=1}^{m} logD(G(z^{(i)})) \tag{2}$$

, where $G(z)$ represents the i-th false sample corresponding to the i-th noisy vector, $D(G(z))$ is a discriminant result for this false sample $G(z)$, representing the probability that a false sample is judged as true (this value is within the range of 0 to 1). The likelihood that the discriminator judge fake as real increases with decreasing loss function, indicating that the generator produces high-quality samples. Therefore, the quality of samples generated by the generator can be improved by continuously optimizing the generator loss function.

In order to evaluate GAN models, it is necessary to use evaluation indicators, that is, to evaluate whether GAN models are effective in image generation, whether they are high quality, and whether they are realistic enough.

Two primary evaluating techniques, including Frechet Inception Distance (FID) and Inception score (IS) are introduced, although there are a few more novel approaches as well.

IS validates the variety and quality of samples produced by GANs [2]. IS uses the Inception v3 deep convolutional architecture pre-trained on ImageNet. IS validates image qualities according to the classification probability predicted by inception v3 and an evaluator of generated samples. The higher the quality of the sample, the more likely it is to be classified into the corresponding class. The general form of the IS function is the formula shown in the formula:

$$IS(G) = \exp\left(\mathbb{E}_{x\sim p_g} D_{KL}(p(y|x) \parallel p(y))\right) \tag{3}$$

On the one hand, the high IS value demonstrates that the generated samples are more uniform in category distribution, covering a variety of categories; On the other hand, it also means that fake samples are believable.

FID is another index employed for validating image quality [3]. It compares the difference between the real and fake feature distributions. The general form of the FID indicator is the formula shown in the formula:

$$FID = |\mu - \mu_W|^2 + tr(\Sigma + \Sigma_w - 2(\Sigma\Sigma_w)^{1/2}) \tag{4}$$

The quality of the created picture increases with decreasing value of the FID indicator, since it shows a smaller discrepancy between the feature distribution between fake and real samples.

3 Applications of GAN in Image Generation

The technique of creating images by computer algorithms is referred to as image generation. This process can be rule-based, or it can be based on machine or deep learning. GAN is widely used in image field. The purpose of this section is to give an overview of the typical applications of Gans in the field of images, which mainly fall into the following categories: image generation, image editing, image restoration, image style conversion.

3.1 Image generation

The image generation function of GAN, refers to the process of learning the sample input for GAN and output the generated picture by the generator. The image generation function is a typical function of GAN.

A method called Age-CGAN is able to generate faces of different ages [4]. This method uses conditional generation adversarial network CGAN, which can introduce external conditional information in the generation process, so that the generation process is more flexible and controllable with stable learning process. In Age-CGAN, first input a face image x and set the condition of the image as Age y0, thus generating a reconstructed face image x', which is close to the original face image; Then change the given condition to the target age y^T, and the result will be an aged face image x^T. The characteristic of this model lies in the application of a potential vector optimization method of "identity preservation", which can preserve the identity of the original character in the reconstruction process.

Multi-view image generation is a function of GAN. That is, by an image A, some photo of this image A is generated from different perspectives. A method called Two-pathway GAN (TP-GAN) can realize the function of face correction, that is, the front face image of a person is generated from the side face image, and even the image from different angles [5]. This approach is inspired by the process of human view synthesis, which focuses on both local details and global structures, so the model structure is mainly composed of two convolutional neural networks: the global path handles the global information, and the patch paths handle the local features of the face. To combine global and local information, the model combines four output features into one tensor without losing spatial resolution. The final composite output is then produced by simply fusing the feature tensors of each path together to create a fused feature tensor, which is then supplied to the convolution layer.

Furthermore, the Vari-GAN model can realize multi-view image generation, that is, a single 2D image can generate multiple 3D view images of the object [6]. Three components make up the model's architecture: a condition discriminator, a fine image generator, and a coarse image generator. The coarse image generator creates a low-resolution (LR) image first in the training process, using the target image, condition image, and target view as the conditions, as well as the target view's representation of the object's basic shape. This LR image is then fed into the fine image generator, which creates a high-resolution (HR) image via appending details to the coarse image and fixing some defects. Ultimately, the HR image and Conditional image serve as negative pairs and are sent into the Conditional Discriminator for evaluation alongside positive pairs (the target picture and the conditional image).

In addition, GAN can realize the generation function from text to image. Text to image GAN produces images with text input [7]. Its principle is roughly as follows: Firstly, it samples noise z and encodes text query by text encoder φ. The text encoding and noise z are connected in series and then input into the convolutional network, from low dimension to high dimension, and generate image X. In the discriminator, the generated image X is transformed from high dimension to low dimension by convolutional network, and connected with the text encoding φ(t), and then discriminated. At the same time, the model also proposes other functions, such as Matching aware discriminator, manifold interpolation learning, inverting style conversion generator, and some others.

3.2 Image Modification

The image modification function of GAN refers to the process of applying GAN to change, adjust or process the image to achieve the change of the content, appearance and quality of the image. Image modification is a typical function of GAN.

GAN is capable of image coloring, and from black and white sketches, a model known as auto-painter may produce colorful cartoon graphics [8]. The model is based on CGAN and uses “Wasserstein” distance loss for better results. In the generator, in order to create the final color cartoon image, "U-net" is used to connect the corresponding layers in both encoder and decoder. The discriminator has only an encoder unit compared to the generator, and determining if the input picture is "true" or "false" is the aim.

When it comes to image coloring, there is another Deep Convolutional GAN (DCGAN)-based architecture that is capable of coloring near-infrared (NIR) images [9]. The generator network of the model accepts a near-infrared image as the input image and is characterized by the application of a triple-based coloring model, which uses triples to independently learn channel-wise image and recombine channels as RGB image. The discriminator validates the probability that a sample is not synthesized.

GANs are capable of image editing, a vector arithmetic property. DCGAN, developed by Alec Radford et al., can perform vector algorithms on face samples [10]. For example, a vector "King" -vector "Man" + vector "Woman" can generate a vector close to "Queen".

3.3 Image restoration

The image restoration function of GAN refers to the process of repairing and restoring damaged, missing or damaged digital images. Among them, there are different repair methods according to different damage conditions, such as filling the missing area of the image to make it complete; Enhance image clarity and quality, etc.

GAN can enhance the resolution and sharpness of the image. One common use of GAN is in Super resolution technologies, that is, while preserving the original details of the image, it generates a higher resolution image and improves the resolution of the image. Super-Resolution Generative Adversarial Network (SR-GAN) introduced by Ledig et al., is a generative adversarial network model [11]. It trains a function G for producing an estimated HR image, depending on a LR sample. It also proposes a perceptual loss, including both adversarial and content parts. One of the innovations in the content loss function is that, rather than pixel-level loss, it uses a loss closer to perceptual similarity. The Mean Opinion score test is used to confirm its excellent perception performance.

Wang et al. proposed a Generative Facial Prior-GAN (GFP-GAN) model and achieved results in restoring the sharpness of face images [12]. Structurally, in order to remove degradation such as blur and noise, a degradation removal module composed of U-Net structure is designed to reduce the burden of other modules [13]. The pre-trained face GAN can generate facial priors and provide a variety of facial details. The two are connected through channel segmentation and latent code mapping and. During training, the model uses intermediate restoration losses for removing degradation. Moreover, to enhance facial details a facial component loss is exploited together with an identity retention loss to preserve facial identity. The model can balance fidelity and authenticity, that is, it can maintain the original features and enhance the image quality.

GAN also has several variants. For example, a Probabilistic Diverse -GAN (PD-GAN) probabilistic variety network proposed by Liu et al. has made progress in filling holes in images, using prior information to debug random noise to produce different repair effects [14]. The model uses spatially probabilistic diversity normalization (SPDNorm) to diversify the generated samples. Moreover, to diversify the generated images, the model uses Perceptual Diversity Loss. Furthermore, a Convolutional Neural Network (CNN) and GAN proposed by Dhamo et al. can remove certain objects from an image and keep only the background [15].

In addition, a model Mobile-Inpaint-GAN (MI-GAN) developed by Sargsyan et al. can be applied to Mobile devices [16]. An important feature of this model is that it has a lighter volume and superior performance. For lowering the complexity of the model and avoid sacrificing the output quality, on the one hand, the generator of the model is composed of two branches in structure: A main branch contains bilinear resizing operations, depth-separable convolutions and linear projection layers; A painting branch that mimics the painting process of experts. Besides, knowledge distillation and adversarial learning are used in training, and Co-Mod-GAN is selected as the corresponding teacher network. This model has better performance to achieve the goal of deployment in mobile devices.

3.4 Image Style Transfer

The image style transfer function of GAN refers to applying the style of an image X (such as art style, painting style) to another image Y. The process means creating a new image that appears in a certain style while retaining the structural elements of the old image.

GAN can realize the style transfer of images. Among them, an architecture called Pix2pix is a GAN-based architecture [17]. A major feature of this model is that the architecture is based on CGAN framework, adopts supervised learning method, and uses paired image data for training. Another feature is that its generator uses a U-Net, which preserve the details of the input image; And a Markovian discriminator called PatchGAN is designed, which can be applied on arbitrarily large images while improving the running speed. This model can be used to reconstruct objects from edge maps, change the background of the image from day to night, color the image, and so on.

The Cycle-GAN model,a model created by Zhu et al. attempts to train a function to map from image X to image Y [18]. Different from pix2pix, this model adopts unsupervised learning, that is, it learns the transformations from domain to domain without paired input-output examples. The model realizes the bidirectional transformation from X domain to Y domain by introducing cyclic consistency loss, which ensures the semantic consistency to the original image when returning from one domain to another, so as to make the transformation more reliable. The model has achieved good results, such as realizing the seasonal change of the background of the image, transforming the flower field photos into different artists' styles and so on.

4 Discussion

There are still many problems in generating adversarial network models, such as Mode collapse, Gradient Descent Problem, Non-Convergence problem and so on. Among them, the mode collapse problem is usually used to describe the fact that the generator tends to generate similar or repeated samples in the training process, instead of covering the entire distribution, ignoring the generation diversity, which will lead to the generated samples being too simple and the diversity of the original data is lost, thus affecting the effect and quality of the generated model. The problem of Gradient Descent Problem refers to that, because of the intricate design of GANs, it is necessary for both generators and discriminators to maintain equilibrium and work together to learn. In the training process, gradients propagate in each layer of the neural network, which may drastically reduce the situation, resulting in learning stagnation. The Training instability problem is that GAN training is usually more difficult, prone to training instability. For example, the adversarial process could be regarded as a "zero-sum game" state, which means that oscillations or non-convergence (that is, the performance indicators of the model fluctuate and fail to converge to the near optimal value) problems may occur during the training of the mode, resulting in difficult model optimization and large cost. In addition, there are many other problems, etc., and researchers have also made improvements in the corresponding areas, such as developing new models, optimizing algorithms, optimizing functions, etc., to contribute to solving these problems.

The application of GAN in the field of image has made good achievements, although it is still faced with some problems such as the above mentioned, but GAN itself still shows a greater application value, with huge development potential. The following is a discussion of the possible future development direction of GAN:

Firstly, improve the quality: This includes not only making the images of higher quality and closer to the real image, but also making the images of variety, that is, to solve the problem of pattern collapse as much as possible.

Secondly, optimize model efficiency: Researchers can improve the loss function and optimization algorithm. It is helpful for enhancing the learning ability, stabilize the training process, and possibly reduce the cost of training.

Thirdly, introduction of other techniques: Researchers can enhance the quality and efficiency of the model by introducing other techniques in combination with Gans, such as by introducing CNN (Convolutional neural networks), VAE variational autoencoders, reinforcement learning and other techniques.

Fourthly, expand the application field: GAN can expand the application field of image generation, such as assisting creators to realize creativity in art creation, enhancing medical images to assist the treatment of diseases in the medical field, enhancing visual effects in movies, games and other industries, etc., which are closely related to image generation.

Fifthly, ethics: Improper use of GAN may lead to adverse effects, such as generating malicious images, forging news, generating false information, revealing personal privacy and other social problems. Researchers should consider ethical implications in their development, and develop Deepfake Detection Algorithms to identify false information; Policy makers should also introduce regulations to regulate the use of the technology.

All in all, through the efforts of researchers, Generative adversarial networks are increasingly being used in the field of image generating. Not only are the types of GAN models becoming more and more diverse, but also the different applications of GAN in the image field are becoming more and more extensive. With the continuous deepening of GAN technology, it is expected to see GAN's role and status in the future artificial intelligence field becoming more and more important.

5 Conclusion

This study presents GAN and its associated applications in the image generation. On the introduction of generating adversarial network GAN itself, this paper introduces the architecture of GAN, including generator and discriminator and their respective functions. At the same time, the evaluation indicators of GAN are introduced, including the mainstream IS and FID indicators and other indicators.

In an introduction to the functionality of the image domain, this article introduces some popular examples as well as newer ones. In this paper, four categories—image production, image alteration, image restoration, and image style transfer—are used to group the functions of the image field in this study.: In terms of image generation, this paper introduces Age-CGAN model that can generate faces, TP-GAN and Vari-GAN models that can generate multi-view images, and Text to image GAN model that can generate images according to text. In the image modification, the auto-painter model of image coloring, a DCGAN-based NIR image coloring method, and a vector operation method of DCGAN are introduced. In image restoration, this work introduces SR-GAN with super resolution technology, GFP-GAN which can restore image sharpness, and PD-GAN which can fill the missing area of image. In terms of image style transfer, pix2pix architecture and Cycle-GAN model are introduced.

In these models, it could be observed that these models have different characteristics: for example, using existing models such as CGAN and DCGAN as the basis to develop their own models, such as optimizing the structure and algorithm to enhance the efficiency and quality of the model, etc., all of which have achieved good results in different aspects.

References

1. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, etc.: Generative adversarial nets. In: Advances in neural information processing systems, vol. 27, pp. 1-13. MIT Press, Montreal (2014).
2. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X.: Improved techniques for training GANs. In: Advances in neural information processing systems, vol 29, pp. 1-9. MIT Press, Barcelona (2016).
3. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in neural information processing systems, vol. 30, pp. 1-9. MIT Press, Long Beach (2017).
4. Antipov, G., Baccouche, M., & Dugelay, J. L.: Face aging with conditional generative adversarial networks. In 2017 IEEE international conference on image processing, pp. 2089-2093. IEEE, Beijing (2017).
5. Huang, R., Zhang, S., Li, T., & He, R.: Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In Proceedings of the IEEE international conference on computer vision, pp. 2439-2448. IEEE, Venice (2017).
6. Zhao, B., Wu, X., Cheng, Z. Q., Liu, H., Jie, Z., & Feng, J.: Multi-view image generation from a single-view. In Proceedings of the 26th ACM international conference on Multimedia, pp. 383-391. ACM Digital Library, Seoul (2018).
7. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H.: Generative adversarial text to image synthesis. In International conference on machine learning, pp. 1060-1069. PMLR, New York City (2016).
8. Liu, Y., Qin, Z., Wan, T., & Luo, Z. Auto-painter: Cartoon image generation from sketch by using conditional Wasserstein generative adversarial networks. Neurocomputing, 311, 78-87 (2018).
9. Suárez, P. L., Sappa, A. D., & Vintimilla, B. X.: Infrared image colorization based on a triplet dcgan architecture. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 18-23. IEEE, Hawaii (2017).
10. Radford, A., Metz, L., & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015).
11. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., etc.: Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681-4690. IEEE, Hawaii (2017).
12. Wang, X., Li, Y., Zhang, H., & Shan, Y.: Towards real-world blind face restoration with generative facial prior. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9168-9178. IEEE, Virtual (2021).
13. Ronneberger, O., Fischer, P., & Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention, pp. 234-241. Springer International Publishing, Munich (2015).

14. Liu, H., Wan, Z., Huang, W., Song, Y., Han, X., & Liao, J.: Pd-gan: Probabilistic diverse gan for image inpainting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9371-9381, IEEE, Virtual (2021).
15. Dhamo, H., Tateno, K., Laina, I., Navab, N., & Tombari, F.: Peeking behind objects: Layered depth prediction from a single image. Pattern Recognition Letters, 125, 333-340 (2019).
16. Sargsyan, A., Navasardyan, S., Xu, X., & Shi, H.: Mi-gan: A simple baseline for image inpainting on mobile devices. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7335-7345. IEEE, Paris (2023).
17. Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A.: Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134. IEEE, Hawaii (2017).
18. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pp. 2223-2232. IEEE, Venice (2017).