GANs in Image Generation and Denoising: From Infrastructure to Real-World Applications

Haoyang Sun

School of Information Engineering, Minzu University of China, Beijing, China
albertmusk6@gmail.com

Abstract. Generative Adversarial Networks(GANs) are currently the mainstream generative models, widely used in image, audio, and other data processing. Over the course of ten years, many GAN models have emerged that are adapted to different use cases. With the constant improvement of GAN model structure, they have achieved very good performance in image reconstruction and resolution enhancement. Image generation and image denoising are currently two core application scenarios in the field of artificial intelligence and are widely used in the fields of medicine and art. Image generation depends on denoising technology to improve quality, and image denoising depends on generation technology to broaden its space. Two kinds of tasks often cooperate. This article will systematically review the development and application of GAN in image generation and image denoising, introduce typical GAN variant and technological evolution in two application scenarios, analyze their technological innovation and applicable scenario. Finally, looking forward to the possible development trend in the future. The importance of this part is that the article focuses on the two main core application scenarios of GAN in the image field to trace the technological development, so as to help readers select an appropriate technical solution.

Keywords: GAN, Image Generation, Image Denoising.

1 Introduction

Generative models are an important class of machine learning methods that generate new samples and understand new samples by understanding the intrinsic distribution of data. In the early days, some traditional machine learning models were applied to the field of content generation, such as Hidden Markov Models and Bayesian Networks. Due to their poor ability to model complex nonlinear data distribution and limited ability to model continuous variables, they have gradually been excluded from the mainstream generative model category.

As core application scenarios in the field of generation, image generation and image denoising have been constantly innovating in recent years. Early image generation mainly depends on texture synthesis and example-based completion, which can only handle specific types of images.

K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_65

After 2014, the development of image generation is booming. There are many kinds of models with different structures and good performance appearing, such as the variational autoencoders(VAE) and denoising diffusion probability models(DDPM), which make the application scenarios of image generation more extensive. The early image denoising just depended on some simple algorithms like filters. After the deep learning emerges, deep convolutional neural networks(DnCNN) and so on have been used in image denoising. With the further exploration of image generation and image denoising, more and more new models and application scenarios will appear, which provides more room for the application of artificial intelligence in the field of image processing.

In 2014, Ian Goodfellow and others proposed Generative Adversarial Network (GAN), and opened up the idea of "adversarial learning". Unlike other traditional generative models which directly model the probability distribution of data, GAN trains the model by letting the "generator" and the "discriminator" play a zero-sum game [1]. This way of designing model structure avoids the problem of complex probability calculation in traditional generative models. And through the deep learning framework, the model learns the deep features of data by using large amounts of data, and then gets the sample very close to the real data. And in just a few years, it has become a research hotspot in the field of computer vision, natural language processing, speech synthesis and other fields.

GANs, which have significantly contributed to image generation and denoising, represent one of the most important advancements in artificial intelligence. By learning image distribution features through adversarial game theory, GANs are better at capturing image details than Visual Algorithms (VAEs). Furthermore, after more than a decade of development, many GAN variants have achieved precise control over the generated content through specific network structures. The following sections will discuss GAN models applied in image generation and denoising, introducing their structural optimizations and application scenarios.

2 Image Generation

2.1 Style-Based Generative Adversarial Network

Traditional GAN models can generate clear, high-resolution images. However, if go a step further and modify certain features of the image, such as changing someone's hair color from black to blonde, this is impossible with traditional GAN models. This is because the neural network is a black box; the generator network G is only responsible for generating the image, so it's impossible to know which value affects the hair color. If someone try to change some of the values to change the hair color, this approach won't be very effective because there is coupling between features. Changing some feature values will affect other features as well. For example, if the model is adjusted to change the hair color, the skin color may also change along with it.

In order to address the problem of processing local features, three researchers from NVIDIA proposed Style Generative Adversarial Networks [2] in 2019, which modified the structure of the traditional generative network G. Starting with a constant input, a nonlinear network maps the vector z in the input normal distribution space Z to w∈W. Through an affine transformation, w is specialized into the pattern y= (ys,yb), which is used to control the adaptive instantiation (AdaIN) operation after each convolutional layer in the synthetic network g, thereby achieving direct control over the intensity of image features at different scales. Meanwhile, a single-channel image composed of uncorrelated Gaussian noise is introduced as noise input and provided to each layer of the synthesis network. The corresponding convolution output is added to generate an image with more details, as shown in Figure 1.

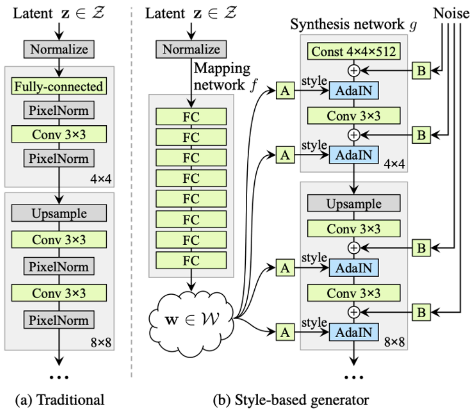

Fig. 1. Core architecture of StyleGAN [2].

StyleGAN generates images with two typical problems. One of them is that even if the final image is not obvious, there are always abnormal regions similar to water droplets in the intermediate features of the generator. To address this problem, Tero Karras et al. found that the adaptive instance normalization (AdaIN) used in StyleGAN normalizes the mean and variance of each feature map separately, destroying the intensity information between features. In order to bypass this normalization, the generator will actively generate local strong signal peaks, forming water droplet artifacts. Another typical problem is that the details of the generated image, such as the eyes and mouth, are fixed in position and cannot move with the pose of the person. The article points out that as StyleGAN is trained progressively, different resolution stages will briefly serve as the output resolution to generate the stage's maximum frequency details, causing the network to have inertia for the position of details and be unable to coordinate the features of each part well under global high resolution [3].

In order to solve these two typical problems, based on StyleGAN (Figure 2), the authors replaced AdaIN with "weight demodulation" to eliminate the teardrop artifacts, while preserving style controllability and improving training speed; by adding path length regularization, the mapping from the latent space W to the image is made smoother, improving the consistency of image content; and a new architecture is proposed to replace the progressive architecture, using the generator with output jumper and the discriminator with residual connection to solve the problem of fixed details. The computation frequency of the regularization term is reduced to reduce memory and computation overhead [3].

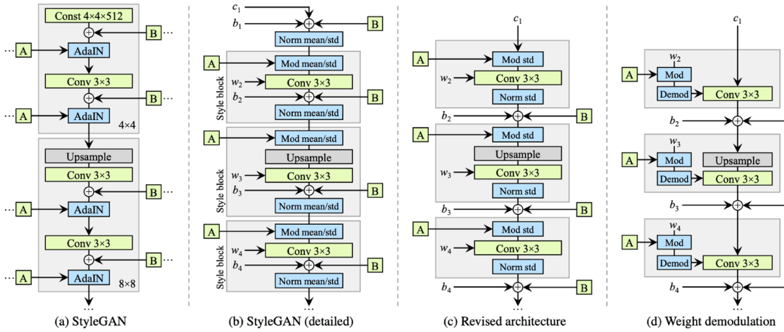

Fig. 2. StyleGAN variant architecture [3].

After one iteration, Tero Karras et al. still found that the generative model in StyleGAN and StyleGAN2 depends on fixed pixel coordinates, that is, features such as teeth and hair are attached to coordinate points rather than the surface of the object. This will cause horizontal stripes to appear instead of naturally following when animal hair is translated. When a person's head moves, the teeth do not follow the head. This situation is caused by the fact that the sampling operation will cause the network to obtain absolute coordinate information by "distinguishing odd pixels", and the point-by-point nonlinearity such as ReLU and Swish will introduce high-frequency noise. After multi-scale superposition, the network can construct a texture template fixed to the screen coordinates [4].

In response to this phenomenon, researchers proposed to treat all signals in the grid as continuous band-limited functions. Discrete feature maps are only used as sampling results of continuous signals. Continuous signals can be reconstructed through ideal interpolation filters, ultimately enabling all layers of the generator to have continuous equivariance to "sub-pixel translation/rotation", ensuring that the position of object details is determined only by the coarse features at the bottom layer, without relying on absolute pixel coordinates. Researchers first replaced the fixed constant input of StyleGAN2 with Fourier features to avoid leakage of initial coordinate information; and improved the upsampling/downsampling module, using a sinc filter modulated by a Kaiser window to suppress high-frequency noise by more than 100dB. By reducing the filter cutoff frequency of low-resolution layers through non-critical sampling, aliasing was completely eliminated; at the same time, nonlinear operations such as ReLU were wrapped in the upsampling and downsampling operations, using temporary high resolution to suppress ultra-high frequency signals, and customized CUDA kernel fusion operations to improve efficiency; finally, by adjusting the filtering parameters and rotation equivariance architecture according to geometric progression, translation and rotation equivariance were achieved, forcing the network to generate details at natural levels, thus completely solving the texture stickiness problem [4].

After two iterations, StyleGAN can not only generate high-quality images of faces, animals, etc., but also control specific scales of images by manipulating style vectors, such as changing lip color or hair shape. There are many applications of StyleGAN now. In terms of hairstyle generation and custom hairstyle, a local generator is introduced to independently model the local semantics of hairstyle such as hair strands and contours. Combined with a local segmentation mask feature fusion unit, high and low resolution features are naturally fused, and low-error hairstyle images are successfully generated. A multi-encoder group and a hierarchical feature modulation module are designed to map the face, hairstyle and hair color reference images to the W latent space of StyleGAN respectively. The detail layer is precisely adjusted by combining the CBAM attention mechanism to ensure that the overall recognition of the person does not change much while adjusting the local features. Finally, the generated image has no obvious artifacts and the hairstyle and hair color are accurately modified. The identity consistency (IDS) can reach up to 0.92 [5]. There are also dataset expansions using the StyleGAN model. For example, focusing on the automated detection of grape black rot and black scab, addressing the pain points of low efficiency of traditional manual identification, easy confusion of diseases with similar phenotypes, difficulty in detecting early small lesions, and scarce samples, the problem of insufficient samples is solved by expanding the StyleGAN2-ADA data, and the lesion characteristics are highlighted by MSRCP image enhancement. The improved YOLO v7 is used to detect and achieve rapid and accurate identification of early grape diseases, with a detection accuracy of 94.1% [6].

2.2 Cycle-Consistent Adversarial Networks

Cycle-Consistent Adversarial Network (CycleGAN) as one of the most well-known GAN models, has the core innovation of achieving bidirectional mapping between two domains through cycle consistency constraints in the absence of paired data. That is, there are two generators and two discriminators. Adversarial loss and cycle consistency loss are calculated and the two types of losses are weighted and combined for optimization [7]. In this way, images can be generated without manual labeling of the input-output correspondence, which greatly expands the application scenarios of image generation.

CycleGAN is most widely used for image transformation. It can transform winter photos into summer photos and horse images into zebra images. The model can accurately modify the texture and color of objects while keeping the background and pose unchanged by training with only unpaired images of two types of objects. It can also perform image style transfer, converting Monet paintings into real landscape photos and ensuring color consistency by adding identity loss [7].

In summary, CycleGAN is a landmark work in the field of unsupervised cross-domain image transformation. Its "cyclic consistency constraint" design solves the mapping ambiguity problem under unpaired data. It has been widely used in image generation applications such as medical care and autonomous driving.

3 Image Denoising

3.1 CycleGAN

There are many GAN models that can achieve denoising effects, and there are corresponding GAN models for different image denoising requirements. The CycleGAN mentioned above can also complete the denoising task without a clean reference image. Usually, two generators and two discriminators are constructed, and denoising mapping is achieved through bidirectional cyclic constraints. For example, for denoising ocean echoes from high-frequency ground wave radar, a dual generator and dual discriminator structure is used. The generator includes an encoder, a converter, and a decoder. The total loss function is optimized by generating adversarial loss and cyclic consistency loss [8]. CycleGAN's image denoising capability is also used in the medical field. By using dual generators, dual discriminators, and cyclic consistency loss, the anatomical structure of the generated image is ensured to be reasonable, effectively solving the contradiction between the radiation risk and diagnostic accuracy of low-dose PET [9].

3.2 Conditional Generative Adversarial Network

A few months after the release of GAN, Conditional Generative Adversarial Network (CGAN) was proposed to make up for the lack of directionality in the generation process of basic GAN. It takes "conditional information y" as an additional input and feeds it to both the generator (G) and the discriminator (D) to achieve the effect of directional generation [10]. In the field of image denoising, information such as noise level and noise category can be added to the generator so that the generator can accurately control the denoising process according to the conditions. The generator takes the noisy image and the conditional signal as input and generates a denoised image; the discriminator receives the conditional signal and the image, distinguishes between the real clean image and the generated denoised image, and finally ensures that the effect meets the conditional requirements.

3.3 Pixel-to-Pixel Generative Adversarial Network

Pixel-to-Pixel Generative Adversarial Network (Pix2Pix GAN) is based on the understanding of conditional generative networks and is mainly used for image-to-image conversion tasks with paired data. The generator adopts a U-Net architecture with skip connections, the discriminator adopts PatchGAN, and the loss function combines adversarial loss and L1 reconstruction loss to avoid generating realistic but unrelated images [11]. This supervised image-to-image mapping can be used to convert noisy images into noise-free images. Such a denoising mode is often used in the medical field. For example, it can solve the contradiction between radiation risk and image quality in myocardial perfusion (MP) SPECT imaging and enhance the presentation effect. Some studies have designed Pix2Pix GAN models and compared them with traditional filtering and 3D CAE. The results show that Pix2Pix GAN is superior to the comparison methods in terms of NMSE, SSIM, CV, FWHM and RSD, especially in low-dose scenarios, and can effectively reduce noise in low-dose MP SPECT images and reduce patient radiation exposure [12].

4 Conclusion

Since its inception in 2014, GANs have spawned hundreds of related models. Their widespread application in fields such as medicine and art has inspired many students and researchers outside of artificial intelligence to enter the field. The adversarial learning framework they embody has become mainstream for generative models.

Image generation and image denoising are the fastest-growing areas in artificial intelligence and the mainstream applications of GAN models, representing the generation of images from scratch and the reconstruction of images from existing ones to a higher quality. They also share a high degree of commonality in model structure and training mechanisms. Many GAN models can handle both image generation and image denoising tasks simultaneously.

In some typical application scenarios, the technical significance of GANs is even more intuitive in autonomous driving. At present, the training data of autonomous driving decision-making models need a large amount of highly reliable labeled data. While some features in actual road conditions are very sensitive to environmental factors such as weather, collecting data that affects the entire scene is very expensive - a problem that the image generation ability of GANs can perfectly solve. However, due to the instability of GAN training and the low detail recovery rate of generated images, basic GAN models cannot meet the requirements of autonomous driving in terms of scene realism and feature differentiation. Therefore, specially modified GAN variants have become the key to technological innovation.

In summary, this paper is based on GAN models in image generation and image denoising. It reviews the main research progress from the perspective of technical significance, reviews the main application scenarios and models and their corresponding technical performance in typical image generation and image denoising scenes. Finally, it provides some practical technical application suggestions on "how to choose suitable GAN variants and optimize training strategies" in different scenarios, hoping to provide researchers and engineers engaged in vision fields with a complete reference framework from "technical understanding" to "application implementation".

References

1. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y.: Generative adversarial nets. Advances in Neural Information Processing Systems 27 (2014)
2. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4401–4410 (2019)

3. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8110–8119 (2020)
4. Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., Aila, T.: Alias-free generative adversarial networks. Advances in Neural Information Processing Systems 34, 852–863 (2021)
5. Huang, X.: Research and Application of Improved StyleGAN Hairstyle Generation Algorithm. Master's thesis, Northwest A&F University (2024)
6. Zhang, L., Bayintana, Z., Zeng, Q.: Early Detection Method of Grape Leaf Diseases Based on StyleGAN2-ADA and Improved YOLOv7. Transactions of the Chinese Society for Agricultural Machinery 55(01), 241–252 (2024)
7. 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 (2017)
8. Zhou, H., Li, B.: Research on Ocean Echo Denoising Based on High-Frequency Ground Wave Radar Using CycleGAN Network. Modern Electronics Technology 48(03), 56–60 (2025)
9. Zhou, L., et al.: Supervised learning with CycleGAN for low-dose FDG PET image denoising. Medical Image Analysis 65, 101770 (2020)
10. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
11. 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 (2017)
12. Sun, J., et al.: Pix2Pix Generative Adversarial Network for Low-Dose Myocardial Perfusion SPECT Denoising. Quantitative Imaging in Medicine and Surgery 12(7), 3539 (2022)