Structure Analysis and Performance Comparison of Image Generation Methods Based on Generative Adversarial Networks

ZiRui Wang[a]
College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin, 300457, China

Keywords: Generate Adversarial Network (GAN), Image Generation, StarGAN, CycleGAN.

Abstract: In recent years, Generative Adversarial networks (GAN) have made remarkable progress in image generation. This paper reviews various GAN-based image generation methods, including CycleGAN, Pix2pix, and StarGAN models, focusing on their performance on different tasks and data sets. The advantages and limitations of each model are discussed by comparing structural similarity (SSIM) and peak signal-to-noise ratio (PSNR). Comprehensive experimental data analysis results show that different GAN models behave differently in specific application scenarios, CycleGAN performed well on image diversity tasks, Pix2pix has an advantage in high-fidelity scenes, while StarGAN shows excellent performance in face image generation. In this paper, the characteristics and application scope of each model are summarized, and the development direction of image generation technology in the future prospects, including model fusion, high-resolution image generation, multimodal fusion, and so on. This study is designed to act as a guide for researchers and practitioners in the field of image generation and to encourage the expansion and implementation of generative adversarial network technology.

1 INTRODUCTION

Image generation technology has always played an important role in the field of computer vision. From the initial rule-based approach to the deep learning-based approach, its application and development have undergone significant changes. Traditional image generation methods mainly rely on hand-designed features and rules, which have significant limitations in dealing with complex image structure and content. With the rise of deep learning, especially the extensive application of Convolutional Neural Networks (CNNS), image generation technology has ushered in a breakthrough. By automatically learning image features, CNN greatly improves the quality and efficiency of image generation, which includes convolution layer, pooling layer and fully connected layer. Its training process requires a large amount of data and computing resources, so the processing speed of CNNs for style transfer is slow, and the diversity and detail processing of generated images need to be improved (Jiao & Zhao, 2019).

The emergence of Generative Adversarial Networks (GANs) has revolutionized image generation. GAN is composed of Generator and Discriminator. Through the adversarial training of the two, the generator can generate realistic images, while the discriminator can improve the ability to detect false images, to continuously improve the quality of the generated images (Chakraborty, KS, & Naik et al., 2024). Since GAN was proposed, many improved and variant models have emerged, such as CycleGAN, Pix2pix, and StarGAN, which show unique advantages in different tasks of image generation.

Xu proposed an effective immunohistochemical pathology microscopic image generation method, which adopted CycleGAN as the basic framework of the model to realize style conversion between different image domains without the need for paired training data (Xu, Li, & Zhu et al., 2020). With this model, Converting H&E staining images into synthetic IHC images was completed by Xu with the same level of detail and structural information as the original images. This method has important application value in pathological image analysis

[a] https://orcid.org/0009-0004-8996-6326

Wang and Z.
Structure Analysis and Performance Comparison of Image Generation Methods Based on Generative Adversarial Networks.
DOI: 10.5220/0013487100004619
In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)*, pages 45-50
ISBN: 978-989-758-754-2

because it can generate high-quality synthetic images, reduce the need for manual labeling, and improve the efficiency of pathological diagnosis.

A study by Li demonstrated the ability of Pix2Pix model to generate high-fidelity data sets(Li, Guan, & Wei, et al., 2024). The results show that the model can accurately render complex urban features, indicating its wide potential for practical applications. This work provides a scalable solution to the shortage of high-quality real images, which is of great practical significance.

This paper aims to review a variety of GAN-based image generation methods and analyze their structure and performance in detail. By comparing the experimental results of different models on multiple data sets, the advantages and disadvantages of each model are discussed, and the development direction and research focus of image generation technology in the future are proposed. This paper aims to serve as a valuable reference for researchers and practitioners in the area of image generation and to foster the ongoing development and application of this sector.

2 METHOD

2.1 Generative Adversarial Network (GAN)

GANs were proposed by Goodfellow et al in 2014. Gans performs adversarial training by training two neural networks, namely generators and discriminators. The generator is in charge of producing realistic images, and the discriminator is in charge of separating the generated images from the authentic ones (Goodfellow, Pouget-Abadie, & Mirza et al., 2014). It works by the generator receiving random noise as input and attempting to produce realistic images. The discriminator receives images (both generator-generated images and real images) as input and attempts to distinguish whether these images are generated or real. The generator and discriminator work against each other during training, with the generator improving to produce a more realistic image and the discriminator improving to more accurately distinguish the authenticity of the image. The flowchart of its working principle is shown in Figure 1.

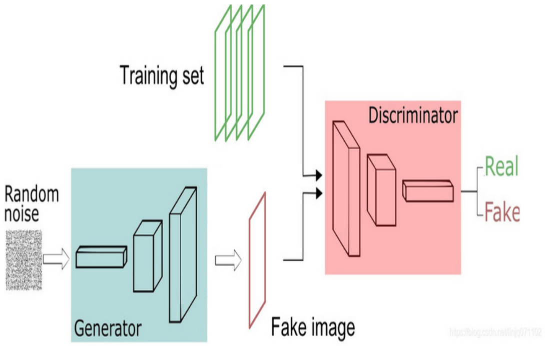

Figure 1: Basic structure diagram of GAN (Duke666, 2020)

2.2 Pix2Pix

Pix2Pix is a conditional GAN-based image conversion model capable of image-to-image conversion, such as from sketches to photos, from satellite images to maps, etc. With the U-Net architecture, the input image is passed through the encoder and decoder to generate the target image, and then the discriminator is responsible for determining whether the input image pair (the original image and the generated image) is real. Because the U-Net architecture is used, the lower-level features are preserved through skip connections, and the quality of the generated images is improved. The model structure is shown in Figure 2.

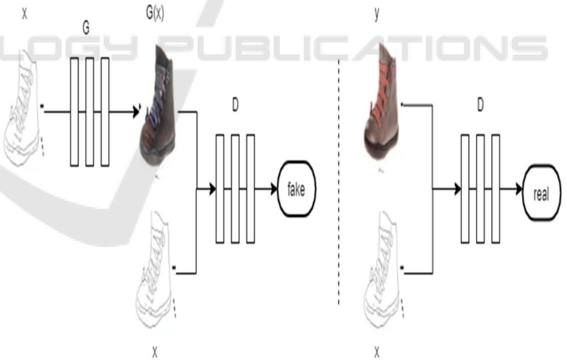

Figure 2: pix2pix structure diagram(Isola, Zhu, & Zhou et al., 2017)

2.3 CycleGAN

CycleGAN is a model capable of image-to-image conversion without paired data, suitable for situations where paired training data is not available. Two generators convert the image from domain X to domain Y, and from domain Y back to domain X, respectively. Two discriminators are responsible for determining whether the image of domain X and domain Y is real (Xie, Chen, & Li, et al., 2020).

Ensure that the image can be restored after two conversions, improve the stability and consistency of the conversion. This eliminates the need for pairs of training data and enables image generation between different domains. The model structure is shown in Figure 3.

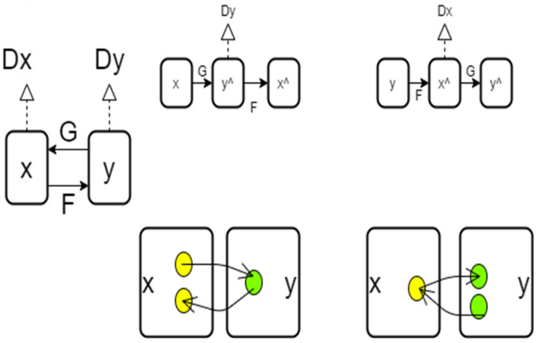

Figure 3: CycleGAN structure diagram (Picture credit: Original).

2.4 StarGAN

StarGAN is a multi-domain image transformation model that can transform images between multiple domains by sharing a generator and a discriminator. The model architecture of multi-domain image conversion is simplified, the number of parameters and the consumption of computing resources are reduced. The domain classifier ensures that the generated image matches the characteristics of the target domain. The model structure is shown in Figure 4.

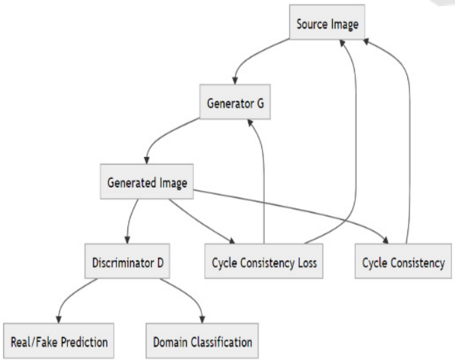

Figure 4: StarGAN structure diagram (Picture credit: Original).

2.5 StyleGAN

StyleGAN is a generative adversarial network model capable of generating high-quality, realistic images, mainly used to generate images with complex structure and detail. Generator: Use the style module to achieve fine-grained control of the generated image by controlling different levels of input noise. StyleGAN introduces a style control mechanism (AdaIN) into the generator, which allows each layer to adjust the characteristics of the generated image according to the style vector w, allowing for more granular control. This method can produce images with variety and high quality. The model structure is shown in Figure 5.

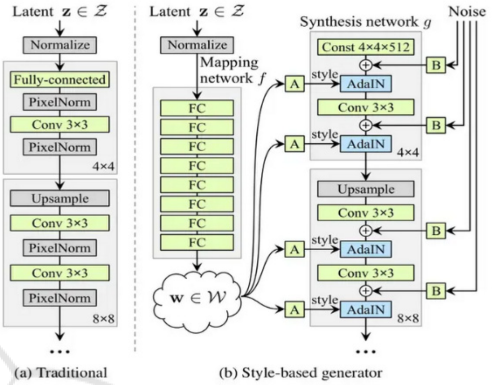

Figure 5: StyleGAN structure diagram (Karras, Laine, & Aila, 2019).

3 RESULT

3.1 Correlation Dataset

To compare the efficiency of several mentioned methods, the following data sets were integrated:

(1) Animal Faces data set (AFHQ): It is a high-quality image dataset containing 15,000 animal face images, divided into three categories of cats, dogs, and wildlife, with 5,000 images each. These images have a resolution of 512x512 pixels.

(2) North US Birds data set: Contains about 5,000 to 6,000 bird images. The images cover a wide variety of bird species across North America, with different postures and backgrounds.

(3) CUHK1 data set: It consists of 188 pairs of students' sketches and corresponding face images (Wang & Tang, 2008). The cropped version was used in the test, and 100 pairs were used for training and 88 pairs were used for testing.

(4) Facade data set: Contains 400 images. The data set contains front photos of buildings and corresponding structural label images. Each photo includes an actual photograph of the building and a simplified drawing of its structure.

(5) CelebA Dataset: It is a large dataset of facial

attributes, containing more than 200,000 celebrity facial images, each with 40 attribute labels, such as gender, age, expression, hairstyle, etc.

3.2 Interpretation of the Result

As can be seen in Table 1, the table shows the performance indicators (SSIM, PSNR, IS, FID) of different models (CycleGAN, Pix2pix, StarGAN) on different data sets. The following is a brief explanation of each indicator: SSIM (Structural similarity): Measures the similarity of the image structure, the greater the value, the better. PSNR (peak signal-to-noise ratio): A measure of image quality, the larger the better.IS (Score of Generated Adversarial Network): Measures the diversity and quality of the generated images, the greater the value, the better. FID (Score of generated image): The smaller the measurement of similarity between the generated image and the real image, the better.

As can be seen from Table 1, the SSIM and PSNR of CycleGAN on the CUHK-Student dataset are relatively high, 0.6537 and 28.6351 respectively, indicating that Cyclegan performs well in the quality of reconstructed images. The traditional GAN model has a relatively low SSIM and PSNR on the Facades dataset, 0.1378 and 27.9706 respectively, which indicates that the image it generates on this dataset is of poor quality. On the North-US-Birds dataset, CycleGAN has an IS of 25.28 compared to StarGAN's of 18.94, indicating that CycleGAN performs better in terms of the diversity and authenticity of the images it generates. The FID for StarGAN is 260.04, higher than the 215.30 for CycleGAN, indicating that the quality of the image produced by StarGAN is even more different from the real image. On the Animal-Faces dataset, StarGAN's FID is 198.07, slightly higher than CycleGAN's 197.13, indicating a small difference in performance between the two on this dataset. On the CelebA dataset, the SSIM of StarGAN is 0.788, which is significantly higher than other models, indicating that StarGAN performs better in processing the CelebA dataset.

Table 1: Performance comparison table

Model	Dataset	SSIM	PSNR	IS	FID
CycleGAN	AFHQ	-	-	7.43	197.13
StarGAN	AFHQ	-	-	6.21	198.07
CycleGAN	NUB	-	-	25.28	215.30
StarGAN	NUB	-	-	18.94	260.04 (Wang & Tang, 2008)
GAN	CUHK1	0.5398	28.3628	-	-
Pix2Pix	CUHK1	0.6056	28.5989	-	-
CycleGAN	CUHK1	0.6537	28.6351 (Kancharagunta & Dubey, 2019)	-	-
GAN	Facades	0.1378	27.9706	-	-
Pix2Pix	Facades	0.2106	28.0569	-	-
CycleGAN	Facades	0.0678	27.9489 (Kancharagunta & Dubey, 2019)	-	-
Pix2pix	CelebA	0.767	21.463	-	-
CycleGAN	CelebA	0.749	20.686	-	-
StarGAN	CelebA	0.788	22.752(Xu, Chang, & Ding, 2022)	-	-

Table 2: Model comparison table

Model	Structural advantage	Applicable scene
CycleGAN	With circular consistency, there is no need to pair data.	Style transfer and domain conversion tasks without paired images.
StarGAN	Multi-domain transformation learning is realized by sharing generators and discriminators.	Multi-domain image editing, such as face attribute editing, and cross-domain image conversion.
Pix2Pix	The use of GAN conditions, is suitable for high-precision paired data image generation.	Paired image tasks such as image repair, denoising, and synthetic image generation.
GAN	The structure is simple and flexible, and high-quality images are generated through adversarial training.	High-quality image generation and data enhancement tasks in a single domain.

Therefore, in combination with Tables 1 and 2, it can see that different models have their advantages in different data sets, and the selection of appropriate models needs to be determined according to specific task requirements. StarGAN shows an advantage in processing face datasets, Pix2pix is suitable for scenes requiring high fidelity, and CycleGAN has an advantage in image diversity. Taking these factors into consideration, it can better select and apply a generative adversarial network model to image generation.

3.3 Vision of the Future

Looking to the future, there are still many directions worth exploring and improving for GAN-based image generation technology: try to combine the advantages of multiple GAN models and develop new hybrid models. For example, combine the diversity of CycleGAN and the high quality of Pix2pix to create an image generation model with high diversity and high quality. In addition, although existing models have made remarkable progress in generating low-resolution images, high-resolution image generation still faces challenges. Future research could further optimize the model structure and training strategies to improve the quality and efficiency of high-resolution image generation. Future research on image generation can try to combine multiple data modes (such as text, speech, video, etc.) to achieve cross-modal image generation. For example, the corresponding image is generated by input text description, or the corresponding visual content is generated by input audio.

4 CONCLUSION

This paper briefly introduces the development of generative adversarial networks in image generation and analyzes several main GAN-based image generation models in detail, including CycleGAN, Pix2pix, and StarGAN. By comparing the performance of these models in different image generation tasks, it is found that each model has its unique advantages and limitations on specific tasks and data sets. Different GAN models have their advantages in specific application scenarios, and researchers and practitioners should choose the right model according to their specific needs.

The study also mentioned the future development direction, through the exploration and research of these directions, image generation technology is expected to play an important role in more practical applications, such as film and television production, game development, virtual reality, and other fields, to promote the continuous progress and innovation of visual content creation and processing technology.

REFERENCES

Chakraborty, T., KS, U. R., Naik, S. M., Panja, M., & Manvitha, B. 2024. Ten years of generative adversarial nets (GANs): A survey of the state-of-the-art. *Machine Learning: Science and Technology*, 5(1), 011001.

Duke666. 2020, August 20. Generative adversarial network. CSDN. https://blog.csdn.net/linjq071102/article/details/107979158.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems* (Vol. 27).

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1125-1134).

Jiao, L., & Zhao, J. 2019. A survey on the new generation of deep learning in image processing. *IEEE Access*, 7, 172231-172263.

Kancharagunta, K. B., & Dubey, S. R. 2019. CSGAN: Cyclic-synthesized generative adversarial networks for image-to-image transformation. *arXiv preprint arXiv:*1901.03554.

Karras, T., Laine, S., & Aila, T. 2019. 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).

Li, Z., Guan, B., Wei, Y., Zhou, Y., Zhang, J., & Xu, J. 2024. Mapping new realities: Ground truth image creation with pix2pix image-to-image translation. *arXiv preprint arXiv:*2404.19265.

Liu, M. Y., Huang, X., Mallya, A., Karras, T., Aila, T., Lehtinen, J., & Kautz, J. 2019. Few-shot unsupervised image-to-image translation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10551-10560).

Wang, X., & Tang, X. 2008. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11), 1955-1967.

Xu, X., Chang, J., & Ding, S. 2022. Image style transferring based on StarGAN and class encoder. *International Journal of Software & Informatics*, 12(2).

Xie, X., Chen, J., Li, Y., Shen, L., Ma, K., & Zheng, Y. 2020. Self-supervised CycleGAN for object-preserving image-to-image domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference,* Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16 (pp. 498-513). Springer International Publishing.

Xu, Z., Li, X., Zhu, X., Chen, L., He, Y., & Chen, Y. 2020. Effective immunohistochemistry pathology microscopy image generation using CycleGAN. *Frontiers in Molecular Biosciences*, 7, 571180.