Applications of GAN-Based Extension Methods in Image Processing

Yang Ding

School of Science, Jiangnan University, Wuxi, Jiangsu, China
`1131230205@stu.jiangnan.edu.cn`

Abstract. In the era of rapid AI advancement, Generative Adversarial Networks (GANs) stand as one of the most disruptive innovations in deep learning, deeply integrated into every facet of human society. This paper focuses on the optimization pathways within the GAN technology system, systematically reviewing research progress on GAN-based improvement methods. It categorizes and analyzes these advancements across four dimensions: innovative loss function design, generator-discriminator architecture optimization, dynamic training strategy control, and functional module expansion and enhancement. By examining typical application scenarios, particularly in image processing, the paper explores practical implementation strategies. By establishing a three-tier classification framework-"core components, training mechanisms, and extended architectures”-this paper fills a gap in existing reviews regarding the systematic categorization of GAN improvement methods. It reveals the technical essence and applicability boundaries of different optimization strategies. The findings provide methodological guidance for researchers selecting suitable approaches or designing novel improvements tailored to specific application scenarios, offering theoretical support for the engineering implementation and performance breakthroughs of GAN technology.

Keywords: Adversarial Generative Networks, Image Style Transfer, Deep Learning.

1 Introduction

Generative Adversarial Networks (GANs) stand as a highly influential and powerful technology within the field of deep learning, demonstrating exceptionally broad application prospects in areas such as image generation, data augmentation, image translation, and text-to-image conversion. Its core training mechanism involves optimizing the generative model by constructing two adversarial neural networks. These networks play the roles of Generator and Discriminator, respectively, competing and co-evolving throughout the training process.

The Generator's primary responsibility is to deeply learn and capture the intrinsic distributional features of input sample data. Based on these features, it generates new data that closely resembles real data while exhibiting subtle differences. The Discriminator, essentially a binary classification model, is tasked with classifying input

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_25

data, attempting to accurately distinguish between real data and the “forged” data generated by the Generator.

During GAN training, the optimization processes of the generator and discriminator alternate, forming a dynamic game mechanism. The generator aims to refine its parameters so that its generated data can successfully “fool” the discriminator, making it difficult to distinguish authenticity. The discriminator, conversely, strives to enhance its classification capabilities to more accurately identify differences between real and generated samples. This process is achieved by introducing adversarial loss, abstracting the entire optimization problem into a minimax game. During training, an alternating fixed strategy is employed: parameters of one network are fixed while the other's parameters are adjusted and updated iteratively, progressively improving model performance. Ultimately, a well-trained generator produces high-quality data.

As demands diversify—such as expanding style variations or demanding higher image fidelity—generative adversarial network technology has spawned numerous extensions and improvement methods. This paper aims to systematically and comprehensively summarize existing GAN-related approaches from four widely recognized and applied improvement directions: loss function design, generator and discriminator architecture design, training strategy optimization, and module extension optimization. Additionally, this paper introduces practical applications of GAN technology in specific scenarios, filling a research gap in the classification and categorization of GAN extension methods across different improvement directions. It provides valuable references and insights for researchers selecting GAN improvement paths, thereby advancing the field's development.

2 Improvements in Loss Function Design Approach

This enhancement method focuses on addressing the requirements for the final generated style transfer images. By introducing different loss functions and modifying the edges and artistic features of generated images, it resolves limitations in image style and expands the application scope of generative adversarial networks (GANs). Existing approaches for improving loss functions include: MSGAN [1], AnimeGAN [2], CycleGAN [3], LSGAN [4], CartoonGAN [5], etc. Among these, MSGAN introduces a novel loss function to guide network optimization and utilizes saliency information for seasonal style transfer tasks. AnimeGAN proposes three new loss functions—grayscale style loss, grayscale adversarial loss, and color reconstruction loss—while adding animated texture patterns to enhance the visual stylization of images. CycleGAN incorporates a cycle consistency loss function to prevent the generator network from producing images with different styles that are unrelated to the original samples. LSGAN utilizes an L2 least-squares loss function to assist in optimizing both the generator and discriminator, thereby enhancing image quality and training stability. CartoonGAN employs two new loss functions—semantic content loss and edge enhancement adversarial loss—to address the significant differences between cartoon images and real images while preserving sharp edges. These improved methods are primarily applied to generating images of different styles.

This paper will focus on the more widely applied CycleGAN. Beyond standard-format images, this method can also be applied to infrared images. By combining it with YOLOv3 and using CycleGAN to uniformly transform the illumination intensity of infrared images, the expanded infrared sample library becomes richer. Ultimately boosting transformer recognition rates for infrared images by 3% [6]. This approach can also enhance AUV underwater object detection. To acquire large-scale high-quality sonar data, CycleGAN generates sonar images from remote sensing imagery, overcoming limitations of traditional generative models while reducing reliance on real sonar data [7]. Furthermore, this approach was applied to speckle noise reduction in medical ultrasound images. CycleGAN performed style transfer between noisy speckle data domains and noise-free domains, partially restoring ultrasound images to aid disease diagnosis and treatment [8]. In remote sensing mapping, CycleGAN can integrate with transfer learning models to establish semantic segmentation methods, enabling higher-precision, larger-scale mapping based on existing data [9].

3 Improvements in Generator-Discriminator Design Approaches

This improvement direction focuses on the architectural design of generators and discriminators. Based on different image requirements, the generator and discriminator networks are reconfigured to enhance training stability and further improve the quality of generated images. Existing approaches to enhancing GAN architecture include StarGAN [10], PGGAN [11], SNGAN [12], DCGAN [13], and StyleGAN [14]. StarGAN modifies the generator's process—which previously generated images by inputting style and outputting style-modified images—to instead output images conditioned on target domain labels. It introduces an auxiliary classifier to enable the discriminator to control multiple domains. These core component improvements facilitate the transition from learning two domains to learning multiple domains. PGGAN progressively increases the resolution of both the generator and discriminator through sequential training stages, thereby enhancing image resolution and quality. SNGAN employs spectral normalization to constrain discriminator weights, reducing gradient discrepancies between generator and discriminator to stabilize training. DCGAN pioneered the systematic use of transposed convolutions as an upscaling method for the generator while simultaneously utilizing a convolutional network as the discriminator. In deeper architectures, removing fully connected hidden layers enhances training stability and image quality. StyleGAN redesigns the generator's internal structure while introducing a mapping network and adaptive instance normalization (AIN) to improve spatial decoupling, enabling precise control over generated results and attribute separation.

This paper primarily explores StyleGAN's applications. First, StyleGAN can generate and edit human faces. Its architecture enables creation of non-existent faces while allowing modification of numerous facial features—age, hair, smiles, etc[15]. Beyond this, StyleGAN finds utility in medicine. To address the significant challenge posed by dermatoscope image complexity to classification, a decision fusion method based on StyleGAN was proposed. Incorporating multi-block decision-making, enhances model generalization, robustness, and stability. Generating high-quality images alleviates the scarcity and uneven distribution of dermatoscopic datasets, thereby improving classification accuracy[16]. StyleGAN also demonstrates utility in engineering applications. For instance, in 3D reconstruction of digital rocks, integrating StyleGAN with transformers and incorporating attention mechanisms enhances feature extraction from multi-scale training images. This approach generates data highly similar to original training samples, significantly boosting reconstruction efficiency and reliability[17].

4 Improvements Based on Training Strategy Optimization

This improvement direction focuses on the model's learning process, refining training methods to address inherent issues in the original training process such as instability, pattern collapse, and vanishing gradients. This makes the training process more stable and reliable, resulting in higher-quality generated images. Existing methods for improving training strategy optimization include GAL-GAN [18], WGAN [19], EBGAN [20], etc. Among these, GAL-GAN differs from traditional adversarial training strategies by employing mutually independent training approaches. It performs weighted mapping of image features during the global mapping process to enhance the network's extraction of global cartoon style characteristics. Subsequently, it utilizes specialized high-pass filters to extract local detail features for fine-grained learning. This ultimately achieves accurate capture and transformation of information within cartoon images, dynamically learning both global and local style information to generate high-quality cartoon images. WGAN replaces the original Jensen-Shannon divergence with Wasserstein distance and truncates discriminator parameters by taking their absolute values after each update to prevent excessive values. This fundamentally addresses mode collapse and gradient explosion issues, ensuring diversity in generated samples and improving GAN training stability. EBGAN provides an energy interpretation for GANs, introducing a pull-away term to prevent mode collapse, enhance training stability, and generate high-resolution images.

This paper primarily explores WGAN applications. First, WGAN can be applied in image processing to remove motion blur and enhance image quality. A WGAN framework was developed to simultaneously address these tasks by pixel-level matching of target images and integrating perceptual style loss into the generator objective. This approach ultimately eliminates motion blur while enhancing color effects [21]. Beyond this, WGAN finds applications in medical fields. For instance, by implementing WGAN within an advanced pixel segmentation U-Net architecture, it addresses traditional pattern collapse and gradient explosion issues. This approach enhances model performance for classifying lesions in laryngoscopic images, improving cancer detection accuracy and helping physicians reduce diagnostic errors [22]. WGAN also contributes to astrophysical research. For instance, it computes and simulates particle resolutions in dark matter simulations, generating high-resolution statistics, power spectra, and halo mass functions. This significantly reduces computational costs in cosmological and astrophysical modeling [23].

5 Improvements Based on Module Expansion Optimization

This enhancement method focuses on refining GANs by introducing new theoretical frameworks, feature utilization approaches, and attention mechanisms. Ultimately, it aims to enhance the model's efficiency in processing more information and improve the quality of generated images. Existing approaches in this optimization direction include CR Module [24], U-GAT-IT [25], and EnlightenGAN [26], etc. CR Module adds a cascaded suppression module to the discriminator's final layer, extracting multiple non-overlapping features through iterative vector rejection operations to enable better utilization of feature space information. U-GAT-IT introduces auxiliary classifiers to generate attention maps (CAM) and adaptive normalization controls normalization methods, ultimately achieving flexible control over shape and texture variations during image transformation, thereby enhancing unsupervised image-to-image conversion performance. EnlightenGAN pioneers the application of unpaired training in low-light image enhancement, employing a self-regularized attention mechanism. Through comprehensive experimental methods for artistic processing, it improves visual image quality while offering simple and flexible adaptation to process real-world low-light images across diverse domains.

This paper primarily explores EnlightenGAN's applications. First, it enables multimodal human recognition. To address the lack of facial and body-based human recognition studies in ultra-low-light conditions, an enhanced EnlightenGAN converts ultra-low-light images to normal-illumination images. Recognition is then performed by fusing facial and body feature matching scores from the converted images. This approach ultimately enhances human recognition accuracy under extremely low-light conditions [27]. Additionally, EnlightenGAN finds utility in construction engineering. For instance, to address the challenge of shadow-obscured concrete crack images in engineering, a framework improvement based on EnlightenGAN is implemented. Combined with an enhanced super-resolution generative adversarial network, this approach first increases image resolution. Through the integration of multiple modules, the shadow removal capability of EnlightenGAN is enhanced, ultimately improving the quality of shadowed crack images [28]. EnlightenGAN also holds promise in medical applications. For instance, in diabetic retinopathy screening, color retinal images serve as the most prevalent imaging data in healthcare. To address poor retinal scan quality caused by improper exposure, abnormal device parameter settings, and operator error, EnlightenGAN can be employed to enhance fundus image quality. Subsequent classification through convolutional transfer learning modules then achieves robust and efficient detection [29].

6 Current Limitations and Future Prospects

Generative Adversarial Networks (GANs) have undergone years of development, achieving significant progress in numerous aspects. Researchers have continuously explored and innovated, proposing a series of improvement methods that have alleviated traditional limitations such as training instability, convergence difficulties, mode collapse, and vanishing gradients to a certain extent. However, GAN technology still faces several pressing challenges:

Regarding evaluation and measurement systems, a unified, reliable, and objective assessment metric has yet to be established. Existing evaluation metrics often fail to comprehensively and accurately reflect a model's true performance, leading to a heavy reliance on researchers' subjective judgments during model evaluation. This subjectivity not only increases uncertainty in evaluation results but also hinders fair comparisons and effective communication between different models, limiting the broader application and promotion of GAN technology.

Regarding training costs, GANs demand exceptionally high quality and quantity of training data. Large volumes of high-quality training data are crucial for ensuring effective model generation, yet the processes of data collection, annotation, and organization are often time-consuming, labor-intensive, and costly. Simultaneously, training GAN models consumes substantial computational resources, with persistently high computational complexity and extremely demanding hardware requirements. This not only limits in-depth exploration of GAN technology by small-to-medium research institutions and individual researchers but also hinders the technology's rapid development and widespread adoption to some extent.

From a theoretical interpretability perspective, the machine learning training process of GANs exhibits typical “black box” characteristics. How the model makes decisions during training, how parameters interact, and the underlying logic of generated outputs are difficult to explain clearly and intuitively. This lack of interpretability increases the model's opacity, severely limiting GAN applications in practical scenarios—especially in fields demanding high security and reliability, such as healthcare and finance.

Despite these challenges, GAN technology continues to demonstrate vigorous development momentum. Based on current research progress and trends, future GAN research will continue to focus on multiple core directions. On one hand, efforts will concentrate on further enhancing training stability, delving into more robust and reliable theoretical foundations, and developing more stable and efficient training methods to address complex and variable data distributions and training scenarios. On the other hand, the pursuit of simpler, lighter, and more efficient architectural designs will aim to reduce computational complexity and hardware resource requirements, thereby improving deployment efficiency and scalability. Additionally, active exploration will be conducted into integrating GANs with other generative models to leverage their respective strengths and achieve stable, high-quality image generation. Simultaneously, efforts will break through the limitations of image processing by vigorously exploring GAN applications across diverse modalities such as text, audio, and video. This will enable broader multimodal mutual generation and conversion, paving new pathways for AI technology development in cross-modal domains.

7 Conclusion

This paper presents a comprehensive and in-depth exploration of extended improvement methods for generative adversarial networks (GANs). Focusing on four key enhancement directions — loss function design, generator-discriminator architecture, training strategy optimization, and module expansion—it meticulously categorizes and summarizes a series of relevant approaches tailored for image processing. Additionally, it details the application directions of these improvements across various domains based on specific use cases. Practical results demonstrate that these enhancements address multiple shortcomings of traditional GANs. They effectively improve the quality of generated data, significantly enhance training stability, and enable GANs to play a vital role across broader domains and diverse problem types.

The conclusion provides an in-depth analysis of existing limitations in GANs and outlines future development trends. Despite GAN technology's remarkable achievements and sustained momentum, inherent challenges like its “black-box” nature remain difficult to fully resolve in the short term. However, with ongoing research and technological innovation, there is reason to believe that more effective solutions can be designed to address these limitations, further breaking through technical bottlenecks and enhancing GANs' versatility and adaptability. This paper systematically organizes an introduction to extended and improved GAN methods, filling a gap in previous research perspectives that focused on classification approaches. It provides subsequent researchers with clear and comprehensive guidance on improvement directions, helping to stimulate their innovative thinking and prompt deeper consideration of the synergistic effects generated by combining multiple improvement methods. This will drive the continuous advancement of generative adversarial network technology.

References

1. Zhang, F., Wang, C.: MSGAN: Generative adversarial networks for image seasonal style transfer. IEEE Access 8, 104830–104840 (2020)
2. Chen, J., Liu, G., Chen, X.: AnimeGAN: A novel lightweight GAN for photo animation. In: Li, K., Li, W., Wang, H., Liu, Y. (eds.) Artificial Intelligence Algorithms and Applications. ISICA 2019. Communications in Computer and Information Science, vol. 1205, pp. 1–12. Springer, Singapore (2020)
3. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 2223–2232 (2017)

4. Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z., Smolley, S.P.: Least squares generative adversarial networks. In: IEEE Int. Conf. Comput. Vis. (ICCV), pp. 2813–2821. Venice, Italy (2017)
5. Chen, Y., Lai, Y.K., Liu, Y.J.: CartoonGAN: Generative adversarial networks for photo cartoonization. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 9465–9474. Salt Lake City, UT, USA (2018)
6. Du, X., Qi, J., Kang, J., Xie, J., Wang, C., Wang, J.: Style transfer and recognition of infrared images of transformers under different lighting based on CycleGAN. J. Phys.: Conf. Ser. 2741(1), 012016 (2024)
7. Zhou, J., Li, Y., Qin, H., Dai, P., Zhao, Z., Hu, M.: Sonar image generation by MFA-CycleGAN for boosting underwater object detection of AUVs. IEEE J. Ocean. Eng. (2024)
8. Liu, J., et al.: Speckle noise reduction for medical ultrasound images based on cycle-consistent generative adversarial network. Biomed. Signal Process. Control 86, 105150 (2023)
9. Zhang, T., Hu, D., Wu, C., Liu, Y., Yang, J., Tang, K.: Large-scale apple orchard mapping from multi-source data using the semantic segmentation model with image-to-image translation and transfer learning. Comput. Electron. Agric. 212, 108204 (2023)
10. Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J.: StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In: IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 8789–8797. Salt Lake City, UT, USA (2018)
11. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. Int. Conf. Learn. Represent. (ICLR) (2018). https://arxiv.org/abs/1710.10196
12. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. Int. Conf. Learn. Represent. (ICLR) (2018). https://arxiv.org/abs/1802.05957
13. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. Int. Conf. Learn. Represent. (ICLR) (2016). https://arxiv.org/abs/1511.06434
14. Bermano, A.H., et al.: State-of-the-art in the architecture, methods and applications of StyleGAN. Comput. Graph. Forum (2022)
15. Melnik, A., et al.: Face generation and editing with StyleGAN: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 46(5), 3557–3576 (2024)
16. Gong, A., Yao, X., Lin, W.: Dermoscopy image classification based on StyleGANs and decision fusion. IEEE Access 8, 70640–70652 (2020)
17. Zhang, T., Zhang, W.: 3D reconstruction of digital rocks based on StyleGAN and transformer. Int. J. Coal Sci. Technol. (2025)
18. Li, L., Zheng, L., Yang, C., Dong, Y.: GAL-GAN: Global styles and local high-frequency learning based generative adversarial network for image cartoonization. Comput. Electr. Eng. 123, 110164 (2025)
19. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein GAN. In: Proc. 34th Int. Conf. Mach. Learn. (ICML), pp. 214–223. Sydney, NSW, Australia (2017)
20. Zhao, J., Mathieu, M., LeCun, Y.: Energy-based generative adversarial network. Int. Conf. Learn. Represent. (ICLR) (2017). https://arxiv.org/abs/1609.03126
21. Feng, J., Qi, S.: Motion deblurring in image color enhancement by WGAN. Int. J. Opt. 2020, 1295028 (2020)
22. Kim, S., et al.: Enhanced WGAN model for diagnosing laryngeal carcinoma. Cancers (2024)
23. Brennan, J., Sreedhar, B., Regan, J., Power, C.: On the use of WGANs for super resolution in dark-matter simulations. arXiv (2025). https://doi.org/10.48550/arXiv.2501.13056

24. Yeo, Y.J., Shin, Y.G., Park, S., Ko, S.J.: Simple yet effective way for improving the performance of GAN. IEEE Trans. Neural Netw. Learn. Syst. 33(4), 1811–1818 (2021)
25. Kim, J., Kim, M., Kang, H., Lee, K.: U-GAT-IT: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. IEEE Trans. Image Process. 30, 2340–2349 (2021)
26. Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z.: EnlightenGAN: Deep light enhancement without paired supervision. IEEE Trans. Image Process. 30, 2340–2349 (2021)
27. Koo, J.H., Cho, S.W., Baek, N.R., Park, K.R.: Multimodal human recognition in significantly low illumination environment using modified EnlightenGAN. Mathematics 9(16), 1934 (2021)
28. Sun, R., Li, X., Law, S.S., Zhang, L., Hu, L., Liu, G.: An improved EnlightenGAN shadow removal framework for images of cracked concrete. Mech. Syst. Signal Process. 216, 111943 (2024)
29. K., M., S., N., Baig, T., S., A., Mukhopadhyay, A.: Explainable diabetic retinopathy detection using EnlightenGAN images with deep learning technique. Procedia Comput. Sci. (2025)