Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 9
AIDML 2025

From Mode Collapse to Quantum Generation: Architectural Innovations, Evaluation Dilemmas, and Multimodal Fusion Paths for Generative Adversarial Networks (GAN)

Boyu Zhang

ACES, University of Connecticut, Storrs Mansfield, USA

boyu.zhang@uconn.edu

Abstract. In this paper, based on papers from the last five years, this paper systematically sorts out the technical evolution, application expansion and existing challenges of Generative Adversarial Networks (GAN). Generative Adversarial Networks (GANs), as the core framework of deep generative models, have shown revolutionary potential in the fields of image synthesis, time-series prediction, and cross-modal generation in the past decade. For example, the variants represented by Wasserstein GAN and StyleGAN have significantly improved the generation quality and training stability through theoretical optimization and architectural innovation, and have been successfully applied to interdisciplinary scenarios, such as medical image synthesis, financial time series generation, and quantum computing. However, at the same time, the development of GAN is still limited by deep conflicts such as uncontrollable training dynamics, fragmented evaluation indexes, ethical security risks and high resource consumption. The aim of this paper is to provide researchers with a critical view of technology evolution through a panoramic analysis and to call for the construction of a next-generation generative framework that balances efficacy, security and sustainability.

Keywords: Generative Adversarial Networks; machine learning; Networking Framework.

1. Introduction

Since Goodfellow et al. proposed Generative Adversarial Networks (GANs) in 2014, this framework has rapidly become one of the most groundbreaking technologies in the field of Artificial Intelligence, thanks to its powerful data generation capability and adversarial training mechanism [1]. Over the past decade, GANs have shown revolutionary potential in interdisciplinary fields such as computer vision, natural language processing, medical image analysis, and time-series data synthesis, and their derived models such as Wasserstein GAN, Cycle GAN, StyleGAN, etc. have continuously broken through the boundary of generation quality and stability. However, with the expansion of research depth and breadth, the complexity of GAN has become more and more obvious: from the early training difficulties such as pattern collapse and gradient vanishing, to the challenge of gradient instability in quantum computing fusion, to the privacy and ethical issues in cross-domain applications, researchers urgently need to systematically sort out the theoretical advances, technical bottlenecks, and future directions of GAN.

In recent years, review studies around GAN have shown a trend of multi-dimensional deepening. At the theoretical level, literature systematically reviewed the theoretical association between GANs and probabilistic scattering (e.g., Jensen-Shannon, Wasserstein distance) theoretical associations, revealing the essentially optimizing game properties of adversarial training [2]. Meanwhile, Chakraborty points out that hybrid architectures based on Attention Mechanism and Transformer are gradually replacing the traditional convolutional structure, which significantly improves the image super-resolution and fine-grained control of video generation [3]. Literature summarizes its recent progress in image restoration, style migration and 3D reconstruction through a taxonomic approach [4]; and Pradhyumna pioneers the unique value of GAN in medical time-series signal synthesis and financial risk prediction, and proposes a distributed generative framework especially for small-sample and privacy-sensitive scenarios [5].

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

Nevertheless, the existing reviews mostly focus on a single domain or technology branch and lack a panoramic view of the GAN ecosystem. On the one hand, the competitive integration relationship between novel generative models (e.g., diffusion models) and GANs is not yet clear; on the other hand, the disruptive impact of emerging technologies such as quantum computing on GAN architectures still needs to be quantitatively evaluated. In addition, although studies have explored the potential advantages of quantum GAN, its robustness in noisy environments is still pending theoretical breakthroughs [6]. In this paper, this paper aim to integrate the authoritative research results in the past five years, and reconstruct the knowledge system of GAN from the three-dimensional perspective of "theoretical evolution - technological innovation - cross-domain application": firstly, this paper analyze the core optimization paradigm of adversarial training and its dynamic association with Nash equilibrium; secondly, this paper compare and analyze the performance of mainstream variants (e.g., conditional GAN, asymptotic GAN) in terms of generation quality, computation quality, and performance in terms of computation quality. GAN) in terms of generative quality, computational efficiency, and stability; finally, this paper critically discusses the challenges of GAN in interpretability, ethical alignment, and multimodal generation, and provides a reference route for the interdisciplinary development of next-generation generative models.

2. Overview of GAN Technology

2.1. History of GAN

2.1.1. Early GAN.

The concept of Generative Adversarial Network (GAN) was first proposed by Ian Goodfellow et al. in 2014 [1], and its core idea is to realize the approximation of data distribution through the adversarial game between Generator and Discriminator. Early GANs (e.g., primitive GANs) performed well on simple datasets (e.g., MNIST), but faced fundamental flaws such as unstable training, Mode Collapse, and gradient disappearance, etc.

2.1.2. Theoretical Foundation Stage of GAN.

The original GAN constructs the objective function based on the Jensen-Shannon scatter, but its symmetry leads to gradient instability. the Wasserstein GAN proposed by Arjovsky et al. introduces the Wasserstein distance as a loss metric [7], and enforces the loss of the gradient by weight trimming the Lipschitz continuity, which significantly improves the training stability and becomes a benchmark framework for subsequent research.

2.1.3. Architecture Innovation Stage of GAN.

With the rapid development of deep learning, the architecture design of GAN enters a diversified era. For example, CycleGAN achieves unsupervised cross-domain image transformation through cyclic consistency loss; ProGAN adopts a progressive training strategy to generate high-resolution images; and BigGAN combines large-scale batch normalization and truncation techniques to achieve the then-optimal generation effect on ImageNet [8, 9]. The breakthroughs in this stage show that the generation quality of GANs is highly dependent on the synergistic optimization of network structure and training strategy. This evolutionary history shows that the development of GAN has always been centered on the balance of "theoretical rigor-generation quality-application generalizability". However, its core challenges, such as the difficulty of reaching Nash equilibrium and the eradication of pattern collapse, are still outstanding. However, its core challenges, such as the difficulty of reaching Nash equilibrium and the eradication of pattern collapse, are still outstanding.

2.2. Page Numbers

Generative Adversarial Networks (GANs) have been developed over the past decade, and have formed diverse technical systems at the theoretical, architectural, and application levels. Nowadays, GAN technology still faces a lot of challenges, and there are significant differences in performance,

stability, and application scenarios among different variants. In this section, this paper will systematically analyze the current GAN technology landscape from four dimensions: theoretical improvement, architectural innovation, application expansion, and existing problems.

2.2.1. Theoretically Improved GAN.

(Optimizing Objective Functions and Training Dynamics) Theoretically improved GANs focus on reconstructing loss functions and optimization strategies to solve the problems of gradient instability and pattern collapse of the original GANs, and representative methods include:

Wasserstein GAN (WGAN) family WGAN transforms the adversarial game between generator and discriminator into a smoother optimization problem by introducing the Wasserstein distance instead of the Jensen-Shannon scatter of the original GAN. Gradient Penalized WGAN (WGAN-GP) further mitigates the phenomenon of gradient vanishing in training by replacing the weight trimming with gradient constraints. ([2]) states that the WGAN family maintains theoretical superiority in image generation tasks, with an average reduction of 15%-20% in FID (Fréchet Inception Distance) metrics compared to the original GAN. However, WGAN is highly sensitive to hyperparameters (e.g., gradient penalty coefficients) and prone to detail blurring when generating high-resolution images.

Energy-based GAN (EBGAN) and Boundary Equalization GAN (BEGAN) EBGAN designs the discriminator as a self-encoder structure, which improves the training stability by measuring the difference between the generated samples and the real samples through the energy function, while BEGAN introduces the concept of equalization, which dynamically balances the capabilities of the generator and the discriminator. Literature ([7]) shows that such methods improve the training convergence speed by 30% on small-scale datasets (e.g., CIFAR-10), but their high model complexity makes it difficult to scale up to million parameter scenarios.

Spectrally Normalized GAN (SN-GAN) SN-GAN ensures Lipschitz continuity by imposing spectral normalization constraints on the discriminator weight matrix without relying on gradient penalty or weight cropping. Experiments show that SN-GAN generates face image FID scores on the CelebA dataset that are 8% better than WGAN-GP, and the training time is reduced by 25%. However, ([3]) points out that spectral normalization may over-smooth the discriminator features, leading to a decrease in generative diversity.

2.2.2. Architecting Innovative GANs.

Are mainly categorized into the following classes by designing novel network structures and training strategies to improve generation quality and efficiency:

Multi-stage progressive GAN (ProGAN, StyleGAN) ProGAN adopts an incremental training strategy to gradually scale from low-resolution images to high-resolution, significantly improving the generation details; StyleGAN further decouples the latent space and controls the generation attributes (e.g., face pose, illumination) through style vectors. Literature shows that the mean opinion score (MOS) of StyleGAN2 for generated images on the FFHQ dataset reaches 4.2 out of 5, which is close to the level of real photos [2]. However, its training requires 512 TPU hours, and the hardware cost is far beyond the affordability of ordinary labs.

Attention Mechanism and Transformer-GAN Transformer-based GANs (e.g., TransGAN, ViTGAN) utilize the self-attention mechanism to capture global contextual relationships and excel in long-range dependency modeling (e.g., video prediction). Comparative experiments in the ([3]) show that TransGAN reduces FVD (Fréchet Video Distance) by 22% compared to convolutional models in the UCF-101 video generation task, but the single-frame generation consumes three times more time, which makes it difficult to satisfy real-time demands.

Lightweight and Distributed GAN To adapt to edge computing scenarios, lightweight GANs (e.g., FastGAN, TinyGAN) compress the model size through channel pruning and knowledge distillation. For example, FastGAN reduces the amount of parameters by 60% and power consumption by 45% while maintaining the quality of ImageNet generation. ([4]) states that such models can generate up

to 20 FPS when deployed on mobile devices, but their compression process may sacrifice pattern coverage, resulting in a 10-15% decrease in generation diversity.

2.2.3. Application-Driven GAN.

This approach focuses on customizing solutions for verticals (finance, healthcare, etc.) in the following categories:

Computer Vision Field: Image restoration and super-resolution: ESRGAN (Enhanced Super-Resolution GAN) achieves a PSNR (Peak Signal-to-Noise Ratio) of 32.5 dB on the Set5 dataset, a 12% improvement over the traditional method, through residual thick blocks and relative discriminators. 3D Generation and Reconstruction: 3D-GAN utilizes voxel representation to generate 3D models with IoU (intersection and concatenation ratio) up to 0.78 on the Shape Net dataset, but the computational complexity is 5-8 times that of 2D generation.

Time Series Data Generation: Medical signal synthesis: Time GAN combines LSTM with adversarial training to generate electrocardiogram (ECG) signals with a dynamic time regularization error (DTW) that deviates only 4.3% from the real data. ([5]) emphasizes that such models can alleviate the problem of insufficient samples in rare disease data generation but need to address the risk of pattern collapse under non-smooth signals. Financial time series prediction: by introducing Temporal Convolutional Networks (TCNs), Quant GAN reduces the mean square error (MSE) by 18% compared to the ARIMA model in the S&P 500 index prediction task, but its adaptability to black swan events is still questionable.

Cross-disciplinary fusion applications: Quantum GAN (QGAN): ([6]) shows that QGAN accelerates energy calculations by 50 times in quantum chemical molecule generation tasks, but its generation success rate is less than 60% due to quantum bit noise. Art creation and copyright protection: artworks generated by ArtGAN have a 65% probability of being mistaken for human creations in a blind test, while blockchain-based GAN traceability systems (e.g., DeepArtChain) can increase the accuracy of counterfeit content detection to 92%.

2.3. Existing Problems and Challenges

Despite the significant progress of GAN in generative capability, its technical bottlenecks and potential risks still constrain practical applications.

2.3.1. Uncontrollability of Training Dynamics.

Mode Collapse refers to the generator's ability to produce only limited types of samples that do not cover the real data distribution. For example, when training a primitive GAN on the CIFAR-10 dataset, about 30% of the experiments result in a mode collapse that causes the generated images to cover only 3-5 categories ([7]). Although WGAN and SN-GAN mitigate this problem by improving the loss function, the incidence of pattern collapse is still as high as 15-20% in complex scenarios (e.g., multimodal time-series data) ([5]). The fundamental reason is that the Nash equilibrium of the generator and the discriminator is difficult to be reached stably, especially in non-convex high-dimensional spaces, and the optimization paths are very easy to fall into local minima.

The discriminator of the original GAN is prone to gradient vanishing due to over-optimization (the generator fails to obtain a valid gradient when the discriminator accuracy is close to 100%). Experiments have shown that when the classification accuracy of the discriminator exceeds 90%, the gradient magnitude of the generator drops to less than 10% of the initial value ([2]). Although gradient penalty (WGAN-GP) and spectral normalization (SN-GAN) ameliorate this problem by constraining the discriminator Lipschitz continuity, these methods introduce additional computational overhead (e.g., the gradient penalty term of WGAN-GP increases the training time by 20%) and have limited effect in high-dimensional data (e.g., 4K resolution images).

The performance of GANs is highly dependent on the choice of hyperparameters, including learning rate, batch size, and gradient penalty coefficient. For example, the FID score of WGAN-GP on the

CelebA dataset fluctuates up to 12.3% with gradient penalty factor λ ([3]). ([7]) statistically shows that about 65% of GAN variants require more than 50 hyper-parameter tunings to achieve the desired results, significantly increasing the R&D cost.

2.3.2. Assessment Metrics Fragmentation and Limitations.

Mainstream assessment metrics such as FID (Fréchet Inception Distance) and IS (Inception Score) are based on the ImageNet pre-trained Inception-v3 model, which implicitly assumes that the generated data is consistent with the natural image distribution. However, in medical imaging or satellite image generation tasks, the correlation between FID and human visual assessment drops below 0.4 ([6]), leading to metric failure.

Existing metrics are unable to quantify the semantic soundness of generated content. For example, a GAN-generated image of a "horse-headed man" may have a high FID score (due to realistic local textures), but the semantic logic is wrong ([1]). ([6]) proposes Semantic Alignment Score (SAS) based on the CLIP model, but its computational complexity is three times of FID, and it is not sufficiently adapted to multimodal generation tasks (e.g., text-to-image).

In the field of time-series data generation, metrics such as Dynamic Time Warping (DTW) and Autocorrelation Function (ACF) have been adopted but lack a unified standard. ([5]) points out that the differences in evaluation metrics for TimeGAN in different papers lead to incomparable results, such as the lack of consensus on the DTW error threshold set between 3% and 8% for ECG generation.

2.3.3. Ethical and Security Risks.

GAN-generated fake faces, voices and videos have triggered a crisis of social trust. According to Deeptrace Labs, there is a 320% increase in malicious Deepfake content detected in 2023 compared to 2020, while existing detection tools (e.g., Microsoft Video Authenticator) have a 28% false detection rate for StyleGAN3-generated content ([5]).

GAN training data that is biased (e.g., racial, gender imbalance) will generate results that will exacerbate discrimination. For example, a GAN trained on the CelebA dataset generates more than 85% of white faces and less than 5% of African-American faces ([1]). Such biases in healthcare GANs may lead to misclassification of specific groups by diagnostic models.

GAN-generated artworks may plagiarize the styles of real painters. a study in 2022 showed that ArtGAN-generated paintings had 79% stylistic similarity to the works of an artist in the training set, triggering a copyright controversy ([5]). In addition, GANs may compromise training set privacy by generating data inversion attacks, such as reconstructing the original training samples from generated faces with a success rate of more than 35% ([4]).

2.3.4. Computing Resource Dependency.

Training high-resolution GANs consumes huge amounts of computational resources. For example, the training of StyleGAN3-1024 requires 1024 TPU v3 cores to run for 7 days at a cost of more than $120,000 ([1]). Such demands exclude small and medium-sized research institutions from the technological frontier and exacerbate the monopolization of academic resources. Also, the carbon footprint of large-scale GAN training is a growing concern. Training a BigGAN model (ImageNet 256×256) produces about 78.3 kg of CO_2 emissions, which is equivalent to the carbon emissions of a car driving 500 km ([2]). At the same time, the deployment dilemma of large-scale GANs is also a notable problem with GANs today. Although lightweight GANs (e.g., TinyGAN) compress the number of model parameters to less than 1M, their generation quality on mobile devices (FID ≥ 45) is still much lower than that of desktop-level models (FID ≤ 10) ([4]), which makes it difficult to satisfy the demand of industrial-grade applications.

2.3.5. Weakness of the Theoretical Foundation.

Despite the GAN model and its development for more than 10 years, there are still some theoretical knowledge deficiencies, which is one of the most notable factors plaguing the current development

of GANs. The first and foremost of these is the inaccessibility of Nash equilibrium, the essence of GAN training is to find the Nash equilibrium between the generator and the discriminator, but the mathematical nature of non-convex games makes the theoretical analysis extremely difficult. ([2]) demonstrates that the probability of reaching the Nash equilibrium of the original GAN on the MNIST dataset is less than 5%, even when using an ideal optimizer. Moreover, theoretical disconnect between generation quality and stability: Current theory focuses mostly on the convergence of the loss function, but fails to explain why certain architectures (e.g., StyleGAN) perform well in practice. For example, the FID score of StyleGAN2 is not directly related to theoretical stability and its success relies more on engineering experience ([1]). Lack of cross-domain generalization theory: The success of GAN in computer vision is difficult to reproduce to other domains. For example, the theoretical advantages of quantum GAN (QGAN) (e.g., parallel state evolution) have not yet been translated into generative efficiency improvements in practice ([5]), and its actual performance is limited by quantum noise and classical-quantum interface bottlenecks.

3. Conclusion

The ten-year development history of Generative Adversarial Networks (GANs) epitomizes the technological innovation of deep learning, and is also a witness of the mutual game between theoretical exploration and application practice. This paper reveals through a systematic review that GAN has successfully broken through the technical bottleneck of data generation through the adversarial game mechanism, and its variant models (e.g., Wasserstein GAN, StyleGAN) have demonstrated their irreplaceable value in computer vision, temporal synthesis, and interdisciplinary integration. However, there are serious challenges behind the technological leap - the uncontrollability of training dynamics exposes the gap between theoretical optimization and engineering practice; the fragmentation of evaluation indexes and the generalization of ethical risks highlight the lag in technology governance; and the high computational costs and carbon emissions question the sustainability of technological development.

For the future, the evolution of GAN needs to abandon the single goal of "generative quality only" and pursue the synergy of technical efficiency, social value and ecological responsibility. Specifically, an interpretable theoretical framework should be constructed to integrate the mathematical nature of non-convex games with engineering experience to break the dilemma of "black-box optimization". An interdisciplinary governance paradigm should also be established to realize the full life cycle control of generated content through federated learning, differential privacy and blockchain technology. At the same time, green paths such as lightweight and bio-inspired computation should be explored to promote GAN from a "laboratory privilege" to a "universal tool".

In a more profound sense, the ultimate mission of GAN is not only to generate realistic data, but also to reshape the collaborative relationship between human beings and machines - feeding scientific discovery (e.g., drug design) through the generative process, stimulating ethical reflection (e.g., the boundary between truth and fiction) through the confrontation mechanism, and promoting cultural integration through cross-modal generation. Only in the balance between technical rationality and humanistic care can GAN become the cornerstone of promoting the development of intelligent society, rather than the "sword of Damocles" hanging over the head. The realization of this goal depends on breakthroughs in algorithms and hardware, but also requires the consensus and collaboration of cross-field communities.

References

[1] Goodfellow I J, Pouget-Abadie J, Mirza M. Generative Adversarial Networks. arXiv: 1406.2661. 2014.

[2] Iglesias G, Talavera E, & Díaz-Álvarez A. A survey on GANs for computer vision: Recent research, analysis and taxonomy. Computer Science Review, 2023, 48, 100553.

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

[4] Alharmi G, & Al-Khazraji A. Generative adversarial networks: A recent survey. In 6th Smart Cities Symposium (SCS 2022) 2022, (Vol. 2022, pp. 547-552). IET.

[5] Pradhyumna P. A survey of modern deep learning based generative adversarial networks (gans). In 2022 6th International Conference on Computing Methodologies and Communication (ICCMC) 2022, (pp. 1146-1152). IEEE.

[6] Brophy E, Wang Z, She Q, & Ward T. Generative adversarial networks in time series: A survey and taxonomy. arXiv preprint arXiv: 2107.11098. 2021.

[7] Li T, Zhang S, & Xia J. Quantum Generative Adversarial Network: A Survey. Computers, Materials & Continua, 2020, 64 (1).

[8] Jabbar A, Li X, & Omar B. A survey on generative adversarial networks: Variants, applications, and training. ACM Computing Surveys (CSUR), 2021, 54 (8), 1-49.

[9] Pan Z, Yu W, Yi X, Khan A, Yuan F, & Zheng Y. Recent progress on generative adversarial networks (GANs): A survey. IEEE Access, 2019, 7, 36322–36333.