Advances in fidelity-preserving GANs for small datasets: focus on StyleGAN-ADA and its variants

Zhongxuan Mei [1*]

[1]Artificial Intelligence, Faculty of Science and Technology, Beijing Normal-Hong Kong Baptist University, Zhuhai, Guangdong, China

Abstract. It seems challenging to train a generative adversarial network with limited data as issues like mode collapse and discriminator overfitting occurs. In this way, the appearance of StyleGAN2-ADA had solved the problems with adaptive discriminator augmentation technique. It works by dynamically adjusting data augmentation to prevent overfitting. This review analyses StyleGAN2-ADA as well as its variants including StyleGAN3 and ViT-StyleGAN2-ADA. Furthermore, it compares models based on CIDAR-10 dataset. The result demonstrates that the performance of StyleGAN2-ADA is better than StyleGAN2. Vit-StyleGAN2-ADA further enhances the performance by combining vision transformer into discriminator. Even though, StyleGAN3 performs well on high-resolution images, it may be not suitable for training on limited data. This review also considers some challenges such as high computational costs and limitations of evaluation metrics. Eventually, this review recommends to use StyleGAN2-ADA for cases with small dataset. StyleGAN3 is suitable for artistic image generation while ViT-StyleGAN2-ADA is suitable for natural image synthesis. Future work ought to concentrate on finding more efficient architectures and developing evaluation metrics for limited data.

1 Introduction

Generative Adversarial Networks (GANs) have revolutionized high-fidelity image synthesis, enabling applications from medical imaging to industrial inspection. Basically, its core lies in learning from complex features of training data and accurately regenerating new images with high quality and realism [1]. This technology has demonstrated significant potential in computer vision tasks, including realistic face synthesis, style transfer, image super-resolution, and missing content restoration. [2-4]. However, small datasets (e.g., <1k samples) cause critical issues:

(1) Mode collapse reduces diversity

(2) Discriminator overfitting degrades stability.

In particular, GANs often leads to problems such as insufficient diversity and mode collapse under the constraints of limited data[5]. Moreover, overfitting discriminator, which

*Corresponding author: s230034035@mail.uic.edu.cn

can lead to unstable training and degraded image quality is another major problem [6]. In practical applications such as medical imaging and industrial defect detection, these problems are particularly significant [7-8]. In these cases, obtaining large numnber of high-quality data is usually costly and time-consuming. Therefore, it is crucial to discover potential GAN architectures that can adapt to limited data environments.

Based on this context, the appearance of StyleGAN2-ADA has become a milestone in the field of high-fidelity image synthesis [9]. This model introduces the adaptive discriminator augmentation (ADA) mechanism, which can dynamically adjust the strength of data augmentation, further effectively alleviating the overfitting problem of the discriminator caused by small datasets. Through this innovation, StyleGAN2-ADA achieves almost the same effect of StyleGAN2 on large datasets under limited data conditions, significantly improving image fidelity and creating new possibilities for the application of GANs in data-constrained environments.

This review systematically examines the following aspects:

1. StyleGAN-ADA's Adaptive Augmentation Mechanism: An in-depth analysis of its core principle.

2. Variants for Improving Realism and Diversity: A systematic review and evaluation of various models proposed to improve upon StyleGAN-ADA.

3. Remaining Challenges in Small-Data Synthesis: A discussion of the key issues that current techniques still face.

2 Methodology

2.1 StyleGAN2

StyleGAN2 is a model constructed based on original StyleGAN and some innovative techniques, including demodulation, path length regularization, and non-saturating loss with R1 regularization. It aims to eliminate artifacts in generated images and further improve image quality and training stability [10].

2.1.1 Demodulation

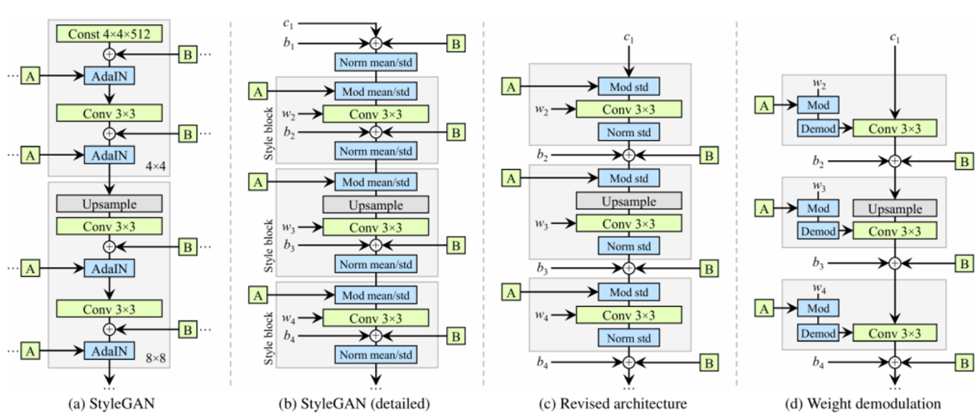

Fig. 1. Architecture of the StyleGAN synthesis network and StyleGAN2 synthesis network.

The graphs c and d from Figure 1 illustrate that StyleGAN2 redesigns the AdaIN operation, transforming the normalization and modulation process into a demodulation step directly handled within the convolution layer. This change prevents the generator from creating overly strong signal correlations at each operation, further avoiding artifacts.

2.1.2 Path Length Regularization (PLR)

Path Length Regularization is a technique that achieves smooth and consistent mapping of the generator. It encourages the generator to produce perceptually uniform changes in the image space for equidistant changes in the latent space, making the latent space more regularized. In this way, it helps the generator learn smoother interpolations and a better latent space structure, enhancing image quality and diversity.

2.1.3 Non-saturating Loss with R1 Regularization

StyleGAN2 also utilizes the classic non-saturating adversarial loss combined with R1 gradient penalty, an effective method from Wasserstein GAN with Gradient Penalty for stabilizing discriminator training.

2.2 StyleGAN2-ADA

StyleGAN2-ADA is a crucial advancement in the StyleGAN series, which is designed to address the challenges of training stability in limited datasets. It is constructed upon the foundational StyleGAN2 model with a mechanism: Adaptive Discriminator Augmentation (ADA).

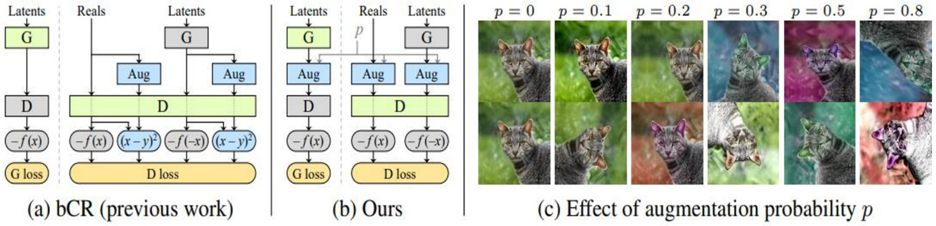

Fig. 2. Comparison of Discriminator Enhancement Methods in GAN Training and Their Effects. (a) and (b) stands for the flowcharts of balanced consistency regularization and stochastic discriminator augmentations. [11-12]. (c) The effect of augmentation probability.

ADA is a dynamic data augmentation strategy that prevents the discriminator from overfitting, which is a common problem when trai Unlike traditional GANs, augmentations are applied to all images fed to the discriminator, including both real and generated images. These augmentations are carefully designed to be non-leaking, ensuring they leave no recognizable traces in the image. ning with small datasets. As shown in chart (b) in Figure 2, ADA seamlessly and dynamically integrates data augmentation operations into the GAN training process. This forces the discriminator to learn higher-level semantic features, rather than simply inverting the augmentations to recognize the original image.

Furthermore, the core of ADA is its adaptive control mechanism, which dynamically adjusts the augmentation strength through an augmentation probability p shown in Chart 3 in Figure 2. It adjusts based on the degree of discriminator overfitting. When the discriminator overfits, p is increased to apply stronger augmentation. In contrast, it is reduced to ensure training stability and generalization.

2.3 Key Variant Models

2.3.1 StyleGAN3

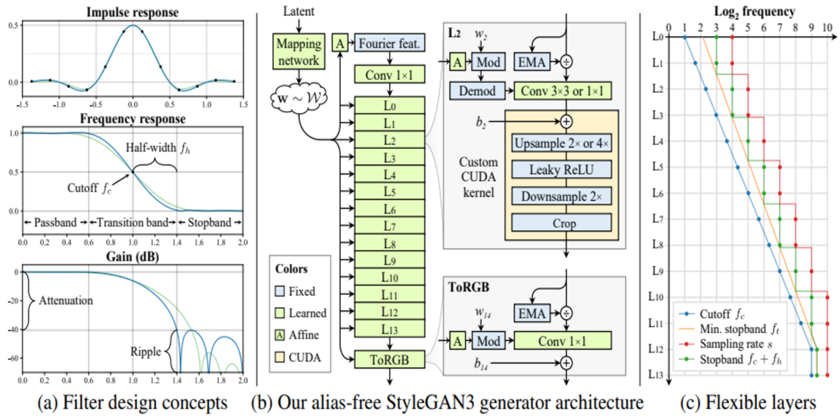

Fig. 3. Filter design and layer configuration for the StyleGAN3 alias-free generator architecture [13].

StyleGAN3 is an improved version of StyleGAN2-ADA, which aims to achieve more consistent and controllable generation by thoroughly addressing the issues of aliasing and equivariance. According to Figure 3, its generator has been redesigned with an aliasing-free architecture, integrating low-pass filtering from classical signal processing into the network. This ensures that high-frequency information that can cause artifacts is removed after each upsampling and convolution, resulting in smoother images. Moreover, StyleGAN3 attains equivariance through the development of a continuous internal signal representation within the network architecture. This approach enables transformations applied in the latent space to be directly and smoothly manifested in the image space, thereby effectively mitigating the unnatural jitter and distortion observed in earlier iterations.

2.3.2 VIT-StyleGAN2-ADA

VIT-StyleGAN2-ADA represents a novel approach that incorporates Vision Transformers (ViT) within the StyleGAN2-ADA architecture. The objective is to improve both the performance and regularization ability in the synthesis of high-resolution images. This model integrates a ViT into the discriminator, enabling the utilization of a self-attention mechanism to effectively capture the global contextual information of images.

Unlike traditional convolutional layers that mainly focus on extracting local features, the ViT-based discriminator assesses overall spatial consistency and realism of the image. This approach represents a fusion of deep learning techniques as it combines convolutional neural networks for image generation with Transformer architectures for more comprehensive global discrimination [14].

2.4 Potential Challenges

Although StyleGANs have made significant achievement, there are still some potential challenges.

2.4.1 Computational Demands and Scalability Barriers

Training a StyleGAN model requires a huge amount of computational power. Particularly, the computation cost is even larger for generating high-resolution images such as 1024x1024 pixels or larger. This involves powerful GPUs, large amounts of memory and long training time. It could last from several days to weeks on multiple advanced GPUs. Basically, the internal design of both generator and discriminator as well as the repetitive training process of GANs leads to these substantial resource demands [15].

2.4.2 Applicability and Limitations of Evaluation Metrics

Evaluating images produced by GANs continues to be a major challenge. Commonly used metrics include the Fréchet Inception Distance (FID) and Inception Score (IS). These metrics depend on pre-trained image classification models like InceptionV3 to extract features and assess the distribution differences between real and generated images. While these metrics are helpful, they have inherent drawbacks. They may not perfectly correlate with human perceptual judgment. Apart from that, they could be sensitive to dataset size and quality as well. What is more, they might not fully capture the mode coverage or structural coherence. For instance, a high FID score does not guarantee the complete absence of artifacts or the presence of all modes within the generated distribution.

No single metric comprehensively captures all desirable aspects of GAN performance—fidelity, diversity, and semantic consistency. This makes objective comparison between models difficult and can lead to models optimizing for a specific metric rather than truly achieving holistic generative quality. The lack of robust metrics also complicates the identification of specific failure modes beyond simple visual inspection.

3 Comparative Analysis

This section provides a systematic comparative analysis of four prominent StyleGAN variants: StyleGAN2, StyleGAN2-ADA, StyleGAN3, and VIT-StyleGAN2-ADA in 3 dimensions: Quantitative performance on benchmark datasets, Training stability under data limitations, and Effectiveness in specific applications.

3.1 Quantitative Performance on Benchmark Datasets

In order to demonstrate the effectiveness of models in small datasets, this review attempts to compare the performance of models with the dataset CIFAR-10. And the metrics utilized in evaluation are FID and IS. FID indicates the statistical distance between the feature distributions of target images and generated images. This distance is calculated using features extracted from a pre-trained Inception-v3 network. Therefore, lower FID values indicate higher quality and diversity in the generated images. IS also uses a pre-trained classification network to evaluate the generated images. It calculates the class discriminability of the generated images and diversity.

As Table 1 demonstrates, model performance has gradually improved with the evolution of its architecture. The basic StyleGAN2 model achieved an FID of 8.23. By introducing ADA technology, the StyleGAN2-ADA model significantly reduced this to 4.76, representing a 42% decrease. It also improved the Inception Score (IS), effectively validating the effectiveness of ADA on small datasets and demonstrating that it can drastically enhance both generation quality and diversity. The ViT-StyleGAN2-ADA further improved performance by reducing the FID to 3.57, representing a 25% decrease. Additionally, it attained an increased Inception Score (IS) of 10.68. This demonstrates that integrating technologies such as the ViT into StyleGAN2-ADA can make the outputs of the model more closely align with real-world data distributions. However, it is important to note that this performance gain comes at the cost of a 40% increase in computational resources. Since StyleGAN3 was originally designed to solve the aliasing problem in high-resolution image generation, its official performance evaluation mainly focuses on large-scale, high-resolution datasets such as FFHQ. No results have been reported for StyleGAN3 on small datasets, which indicates that its design priorities are centered on high-resolution scenarios rather than small data training.

Table 1. Performance on CIFAR-10 (50k tiny images), data from [9][14].

Methods	CIFAR-10	
	FID ↓	IS ↑
StyleGAN2	8.23	9.12
StyleGAN2-ADA	4.76	9.98
StyleGAN3	/	/
ViT-StyleGAN2-ADA	3.57	10.68

3.2 Training Stability under Data Limitations

As mentioned above, StyleGAN2 has a serious overfitting problem when the amount of data is limited. This results in the case that the generator can suffer from mode collapse or generate memorized images that are highly similar to real images but lack diversity. This instability makes StyleGAN2 poorly trained on small datasets. In order to solve the instability problem of StyleGAN2 when data is limited, StyleGAN2-ADA was proposed. In the early stages of training, the discriminator is weak and does not perform much augmentation on real images. As training progresses, if the discriminator becomes too strong, the strength of data augmentation is gradually increased. This adaptive augmentation strategy effectively prevents overfitting of the discriminator, thereby providing a stable gradient for the generator and greatly improving the training stability on small datasets. ViT-StyleGAN2-ADA builds on StyleGAN2-ADA by replacing the traditional convolutional discriminator with the ViT architecture. This improvement further enhances the model's performance in data-constrained environments. The ViT architecture better captures long-range dependencies in images, facilitating more effective image authenticity assessment. Combined with ADA technology, ViT-StyleGAN2-ADA is more stable on small datasets and achieves superior performance to StyleGAN2-ADA on benchmarks such as CIFAR-10, demonstrating further improvements in data efficiency and training stability. StyleGAN3 primarily processes high-resolution and high-complexity images, but its training stability in data-constrained environments has not been fundamentally improved. Official research has not focused on its performance on small datasets such as CIFAR-10.

3.3 Effectiveness in Specific Applications

StyleGAN2 performs well on large and diverse datasets, particularly in high-resolution, high-fidelity tasks such as face generation (FFHQ) and Animal generation (AFHQ). Its architecture is stable and produces high-quality images. With the function of ADA technology, StyleGAN2-ADA can extend the powerful capabilities of StyleGAN to small-scale datasets that were previously incapable of processing. For instance, in medical imaging, the model is still feasible as generating a small number of X-rays and CT images is sufficient for the model training. In artistic style transfer and generation, the model can learn the style of a small number of works of art and generate new works of art. ViT-StyleGAN2-ADA is more suitable for natural image generation than StyleGAN2-ADA. ViT-StyleGAN2-ADA achieves excellent performance on natural image classification datasets such as CIFAR-10. Its ViT discriminator can better understand the overall structure of the image, thereby generating more globally consistent images. Furthermore, in cases where it is necessary to capture the overall context or complex composition of an image, like a landscape, its performance can be better than convolutional models that rely purely on local features. The primary application cases of StyleGAN3 are areas that require high equivariance and detail accuracy, particularly in animation and video generation. Its anti-aliasing design eliminates artifacts that occur when images are scaled, rotated, or translated, making generated images smoother and more consistent under these transformations.

Table 2. Multi-Model Visual Turing Test Results (Medical Imaging), data from [15].

Methods	FID ↓
StyleGAN2	10.43
StyleGAN2-ADA	5.06
ViT-StyleGAN2-ADA	/

As demonstrated in Table 2, the StyleGAN2-ADA model significantly reduced the FID score to 5.06, representing an approximate 50% decrease. This validates the effectiveness of ADA in the field of medical imaging. Since ViT-StyleGAN2-ADA is the latest model, no results related to medical imaging are currently available.

3.4 Technical Limitations and Scenario Adaptation

3.4.1 Key Limitations

One of the primary challenges with StyleGAN models is the significant computational resources they require for training, especially for generating high-resolution images (e.g., 1024x1024 pixels or higher). This process needs considerable GPU power, large memory capacity, and can take days or even weeks to complete on multiple high-end GPUs. The complexity of the generator and discriminator, along with the iterative nature of GAN training, contributes to these high demands. Another major limitation is the evaluation of generated images. Common metrics like FID and IS use pre-trained classification networks to compare feature distributions. However, these metrics have inherent limitations and may not align perfectly with how humans perceive image quality. They can be sensitive to dataset size and quality, and may not fully capture aspects like mode coverage or structural coherence.

3.4.2 Scenario Adaptation Suggestions

Table 3. Scenario-Based Recommendations for StyleGAN Variants

Scenario	Recommended Model	Reason
Medical/Industrial Small Samples	StyleGAN2-ADA	Balancing performance and computational cost
Artistic Creation	StyleGAN3	Best texture detail optimization
Global consistency of natural images	ViT-StyleGAN2-ADA	Attention mechanism captures long-range dependencies.

Based on the features of various models, this review offers recommendations for model selection in different scenarios. Details are presented in Table 3.

4 Discussion and Future Work

This review systematically analyzes the progress of StyleGAN-ADA and its variants, highlighting StyleGAN2-ADA as a milestone for its ADA mechanism that addresses discriminator overfitting and mode collapse on small datasets. This innovation has made GANs valuable in areas such as medical and industrial imaging. ViT-StyleGAN2-ADA enhances performance by incorporating a ViT to better capture the overall image context, achieving superior results on benchmarks like CIFAR-10. StyleGAN3 performs well at high resolutions by removing aliasing. Despite significant progress, two main challenges exist: the high computational cost of generating high-resolution images and the limitations of evaluation metrics like FID and IS. The choice of StyleGAN model depends on the specific application. StyleGAN2-ADA is recommended for small datasets as it balances performance with computational efficiency. StyleGAN3 is suitable for artistic applications due to its anti-aliasing design. For natural image generation that demands global consistency, ViT-StyleGAN2-ADA is preferred because its ViT-based discriminator effectively captures long-range dependencies. Future research should focus on developing more advanced techniques for limited data cases. A key focus should be on creating evaluation metrics that is suitable for small datasets. They are not only required to satisfy simple accuracy, but also better assess model performance on unseen data. At the same time, advancements in self-supervised learning (SSL) methods for limited data are crucial. This includes designing new pretext tasks that allow models to learn robust feature representations from scarce and unlabeled data before fine-tuning with minimal labeled samples. Ultimately, to make powerful architectures like Vision Transformers more practical, it is important to develop lightweight ViT variants or hybrid models that reduce computational requirements and dependence on large datasets, thereby enhancing their usability in data-limited environments.

References

1. J. Lin, R. Zhang, F. Ganz, S. Han and J. Y. Zhu, "Anycost GANs for Interactive Image Synthesis and Editing," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), virtual, 2021, pp. 14986-14996, doi: 10.1109/CVPR46437.2021.01476.
2. J. Zhao et al., "Dual-Agent GANs for Photorealistic and Identity Preserving Profile Face Synthesis," Advances in Neural Information Processing Systems, Long Beach, CA, USA, 2017, pp. 6059-6068.
3. J. An et al., "Is Bigger Always Better? An Empirical Study on Efficient Architectures for Style Transfer and Beyond," 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2023, pp. 4073-4083, doi: 10.1109/WACV56688.2023.00407.
4. Y. Cui et al., "Reconstruction of Large-Scale Missing Data in Remote Sensing Images Using Extend-GAN," in IEEE Geoscience and Remote Sensing Letters, vol. 21, pp. 1-5, 2024, Art no. 5001105, doi: 10.1109/LGRS.2023.3317898.
5. T. Hoang, T. Tran, and S. Venkatesh, "On catastrophic forgetting and mode collapse in Generative Adversarial Networks," arXiv:1807.04015 [cs.LG], 2018.
6. H. Y. Tseng et al., "Regularizing Generative Adversarial Networks under Limited Data," arXiv:2104.03310 [cs.CV], 2021.

7. L. Maack, L. Holstein and A. Schlaefer, "GANs for generation of synthetic ultrasound images from small datasets," Current directions in biomedical engineering, vol. 8, no. 1, pp. 17-20, 2022.
8. Y. Cheng et al., "A Comprehensive Survey for Real-World Industrial Defect Detection: Challenges, Approaches, and Prospects," arXiv:2507.13378 [cs.CV], 2025.
9. T. Karras et al., "Training Generative Adversarial Networks with Limited Data," arXiv:2006.06676 [cs.CV], 2020.
10. T. Karras et al., "Analyzing and Improving the Image Quality of StyleGAN," arXiv:1912.04958 [cs.CV], 2019.
11. Z. Zhao, S. Singh, H. Lee, Z. Zhang, A. Odena, and H. Zhang, “Improved consistency regularization for GANs,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, pp. 11033-11041, May 2021.
12. I. J. Goodfellow et al., “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, vol. 27, 2014.
13. T. Karras et al., "Alias-Free Generative Adversarial Networks," arXiv:2106.12423 [cs.CV], 2021.
14. M. M. Rahman, B. Chen, and H. Zeng, "ViT-StyleGAN2-ADA for Limited-Data Training," 2025.
15. M. Woodland et al., "Evaluating the performance of StyleGAN2-ADA on medical images," in International Workshop on Simulation and Synthesis in Medical Imaging, 2022, pp. 142-153.