STM JOURNALS

Journal of
Multimedia Technology &
Recent Advancements

ISSN: 2349-9028
Volume 13, Issue 1, 2026
January–April
DOI (Journal): 10.37591/JoMTRA

https://journals.stmjournals.com/jomtra

Research JoMTRA

A Study of DCGAN-Based Generative Models for Anime Character Face Generation

Aaditya Mittal[1], Harsh Bathija[2], Yamini Sharma[3], Shreya Agarwal[4,*]

Abstract

Artificial intelligence, or AI, has in recent years moved from simple rule-based systems to models that are now fully capable of creative content generation and are referred to as generative AI. One such approach for content generation, introduced in the year 2014, is called generative adversarial networks (GANs), which consists of training a generator to create fake content that tries to mimic real content as closely as possible and a discriminator that tries to differentiate whether the content is real or fake, both contending against each other. Although GANs did show great potential for content generation, they lacked any stable way to generate images since their shallow architecture led to high volatility in training, with poor convergence, and poor image quality. To overcome these challenges, deep convolutional generative adversarial networks (DCGANs) were introduced in the year 2015, which were built upon traditional GANs by including convolutional and batch normalization layers in their architecture that allowed for stable and spatially coherent training for image generation. This project will research deeply into the performance of DCGAN and its advanced versions, with further optimizations in its architecture to overcome the limitations of GANs in generating images. In this study, a DCGAN model was trained to generate anime character images, exploring the benefits of convolutional architectures that allow stable training of adversarial models to improve the generated output qualitatively.

Keywords: Deep convolutional generative adversarial network, discriminator, generator, mode collapse

INTRODUCTION

AI has recently grown from simple prediction and rule-based models to generating creative content in the form of text, images, audio, videos, etc. Due to such advancements in AI, a branch called generative AI has emerged.

Since the development of generative AI, many approaches and architectures for generating content have been introduced, some of which are unique and creative. One such creative approach is the use of generative adversarial networks (GANs), which were introduced by Goodfellow et al. (2014) and consist of two different parts: a generator and a discriminator. The former attempts to generate realistic content, while the discriminator attempts to distinguish between the fake content produced by the generator and the real content [1]. They play against each other to increase the effectiveness of the generator and the discriminator during training

***Author for Correspondence**
Shreya Agarwal
E-mail: shreya.agarwal@jecrcu.edu.in

[1]Student, Department of Computer Science and Engineering, JECRC University Jaipur, Rajasthan, India
[2]Student, Department of Computer Science and Engineering, JECRC University Jaipur, Rajasthan, India
[3]Student, Department of Computer Science and Engineering, JECRC University Jaipur, Rajasthan, India
[4]Assistant Professor, Department of Computer Science and Engineering, JECRC University Jaipur, Rajasthan, India

Received Date: January 12, 2026
Accepted Date: February 09, 2026
Published Date: March 20, 2026

Citation: Aaditya Mittal, Harsh Bathija, Yamini Sharma, Shreya Agarwal. A Study of DCGAN-Based Generative Models for Anime Character Face Generation. Journal of Multimedia Technology & Recent Advancements. 2026; 13(1): 31–41p.

in a stable manner, and the generator produces realistic but creative output, such as that on which it was trained.

However, earlier GAN architecture faced several difficulties in image generation. Their training tends to be unstable, and they can be extremely sensitive to hyperparameters, as even a slight change in them can lead to drastically different output qualities. Generators can also suffer from mode collapse, a condition in which the generator produces only a limited variety of output because it exploits a temporary weakness within the discriminator.

To address these issues, Radford et al. proposed the deep convolutional generative adversarial network (DCGAN) architecture in 2015, which replaces the fully connected layers with deep convolutional layers and uses practices such as "strided/fractional-strided convolutions and batch normalization" in both the generator and discriminator [2]. The model can now learn hierarchical and spatial image features stably during adversarial training, thus producing better results for image generation.

In this study, we trained a DCGAN generator on 10,000 images of a curated dataset of anime character faces running for 1,000 epochs. The results show the capability of DCGANs in generating detailed and varied outputs, showing the edge that convolutional adversarial architectures have in the domain of image generation tasks.

LITERATURE REVIEW

GANs, first introduced by Goodfellow et al. (2014), consist of a generator and a discriminator that contend with each other; the generator generates some content, for instance, an image, text, video, or text, by trying to imitate the original data on a noise vector, while the discriminator predicts whether the data it receives are original, real, or generated by the generator or fake, both of which improve through this contention. While GANs produce great results, their training is often highly unstable, and with issues such as mode collapse and vanishing gradients, they lose their appeal, especially for image generation [1, 3, 5].

DCGAN architecture was first proposed by Radford et al. in 2015 to establish a set of design principles based on the GAN architecture. These architectural changes were targeted toward achieving a design that excluded both pooling and fully connected layers, using "strided and fractional-strided convolutions," batch normalization, and rectified linear unit (ReLU) and leaky activations in the generator and discriminator, respectively [2, 4]. Later, several studies proved that such architectural choices drastically reduced the chances of mode collapse, thereby increasing the training stability and enhancing the spatial coherence and sharpness of the generated images [2, 5].

The presence of stylized features and exaggerated proportions commonly seen in anime character images raises serious challenges for GANs in generating the faces of anime characters. Jin et al. (2017) noted that training a simple GAN on anime datasets often fails to yield quality images. Hence, they introduced DRAGAN and some techniques for refining the dataset for stable quality anime character image generation [6]. Li et al. (2021) further explored this aspect and proposed AniGAN, a style-guided architecture using adaptive normalization and attention mechanisms, resulting in improved color coherence and consistency in the generated anime character images [7].

In addition to anime, numerous domain-specific GANs have been proposed to achieve better performance. The accomplished work can perform more advanced tasks, including image-to-image translation and domain adaptation. For example, CycleGAN and Coupled Generative Adversarial Networks (CoGAN) can convert real photos into anime-style images [8, 9]. StyleGAN further developed an even finer manipulation of the latent space and achieved finer details with more style control for images [10, 11]. Domain-specific image generation GANs target convolution-based models based on the basic principles of DCGAN, as shown in Table 1.

Table 1. Summary of some existing GAN architectures for image generation.

Author(s) and year	Architecture	Key contribution	Relevance
Goodfellow et al. (2014) [1]	GAN	Introduced the generative adversarial network (GAN) framework.	Laid the foundation for generative modelling.
Radford et al. (2015) [2]	DCGAN	Introduced convolutional layers, batch normalization, and activation design to stabilize GAN training and improve image quality.	Core architecture used in this project for anime-face generation.
Arjovsky et al. (2017) [3]	Wasserstein Generative Adversarial Network (WGAN)	Replaced JS divergence with Earth Mover's distance to stabilize training and improve convergence.	Points out the stability issues that DCGAN partially resolves architecturally.
Salimans et al. (2016) [4]	Improved GANs	Proposed feature matching and minibatch discrimination.	Supports the analysis of stability and image diversity in DCGAN.
Gulrajani et al. (2017) [5]	WGAN-GP	Introduced a gradient penalty for better stability and to avoid the mode collapse problem.	Places architectural and loss-based stabilization into context.
Jin et al. (2017) [6]	DRAGAN	Addressed anime-face generation instability by incorporating regularization and curated datasets.	Directly relates to the anime- face synthesis problem studied here.
Zhu et al. (2017) [8]	CycleGAN	It introduced unpaired image-to-image translation for domain transfer, such as photo ↔ anime.	Illustrates the domain flexibility of GAN architectures.
Liu and Tuzel (2016) [9]	CoGAN	Proposed coupled GANs for multi-domain learning without paired data.	Demonstrates early efforts related to domain adaptation applicable to stylized data.
Karras et al. (2019) [10]	StyleGAN	Introduced the style-based generator that allows for control over the visual features and learning hierarchical representations.	Represents a major step beyond DCGAN in fine-grained generative control.
Karras et al. (2020) [11]	StyleGAN2	Improved normalization and path length regularization for higher fidelity in Enhanced StyleGAN.	Demonstrates an evolution of stable, high-quality synthesis from DCGAN foundations.
Zhang et al. (2019) [14]	Self-Attention Generative Adversarial Network (SAGAN)	Integrated self-attention into GAN to model long-range dependencies, yielding improvement in global coherence.	Forms a basis for later attention-based anime generation models.
Li et al. (2023) [15]	BaMSGAN	Implemented self-attention with blur and memory mechanisms to generate high-quality anime faces.	This showcases better anime image synthesis based on DCGAN.
Dubey and Singh (2024) [18]	Transformer-based GANs	Transformer-based GAN that models self-attention along with global context.	Highlights present frontier trends which go beyond DCGAN toward transformer- driven synthesis.

METHODOLOGY

Dataset and Preprocessing

The dataset on which this project was carried out is an open-source collection of anime character facial images found on Kaggle. The dataset was preprocessed to maintain uniformity and effectively train the model. All images were resized to 64×64 pixels, converted into Red, Green, and Blue (RGB) format, and normalized in the range [−1, 1] for stability in training and good convergence [2]. Preprocessing operations, including resizing, tensor conversion, and normalization, were performed using the PyTorch Transforms library.

Generator Architecture

The generator network was implemented based on the DCGAN architecture introduced by Radford et al. [2]. Given a 100-dimensional latent vector drawn from a standard normal distribution, the generator produces an image of size 64×64 pixels by successive applications of transposed convolutional layers. Each layer increases the spatial resolution and decreases the feature map depth. Batch Normalization was introduced after each convolutional layer to stabilize learning and speed up convergence, as stated by Ioffe and Szegedy [12]. For the hidden layers, ReLU activations promote gradient flow and nonlinearity. The consequent pixel values for the output layer were mapped to the normalized range $[-1, 1]$ using the Tanh activation function, as shown in Figure 1.

Discriminator Architecture

By differentiating between the produced and real images, the discriminator functions as a binary classifier. Batch Normalization and LeakyReLU were the activation functions, and convolutional layers with a stride of two were used to downsample the image data [13]. After the last layer, it produces a single probability with Sigmoid activation, which determines the probability of the input image being real (Figure 2).

DCGAN Framework

The adversarial learning setup for DCGAN incorporates a “minimax approach” where the generator tries to convincingly deceive the discriminator, and the discriminator tries to achieve a correct differentiation between real and generated images [1, 2]. It does not use fully connected or pooling layers but is rather supported solely by convolutional and transposed convolutional operations to maintain spatial information and allow for high visual realism.

Training Configuration

PyTorch Lightning was used because of its modularity, reproducibility, and support for logging. The training was performed for 1,000 epochs with a batch size of 128 on an NVIDIA T4 GPU. The Adam optimizer was implemented in both the generator and the discriminator; the learning rates were 5×10^{-4} and 1×10^{-5}, respectively, and the momentum parameters were $\beta_1 = 0.5$ and $\beta_2 = 0.999$. During training, model checkpoints were periodically saved, and TensorBoard was utilized to visualize loss patterns to track the stability and convergence of the model, as shown in Figure 3.

Loss Function

This work was implemented as an adversarial objective using BCE loss, based on the standard formulation from Goodfellow et al., adapted for convolutional networks by Radford et al. [1, 2]. The model is trained such that the discriminator, D, correctly classifies the real images as “valid” (1) and generated images as “fake” (0), whereas the generator, G, learns to produce synthetic images that maximize the likelihood that the discriminator will classify them as real Figure 4.

Mathematically, the losses for the discriminator and generator are computed as:

$$L_D = \frac{1}{2}\left[BCE(D(x_{real}),1) + BCE\big(D(G(z)\big),0\right]$$

$$L_G = BCE\big(D\big(G(z)\big),1\big)$$

Where, D(xreal) and D(G(z)) denote the discriminator outputs for real and generated images, respectively. This formulation encourages the generator to minimize the distance between the distributions of the real and generated samples.

In practice, a batch of fake images is first generated in every training step. The discriminator is first trained to minimize the average loss between the real and fake predictions. After updating the discriminator, the generator is trained to fool the discriminator by maximizing the probability of the generated images being classified as real.

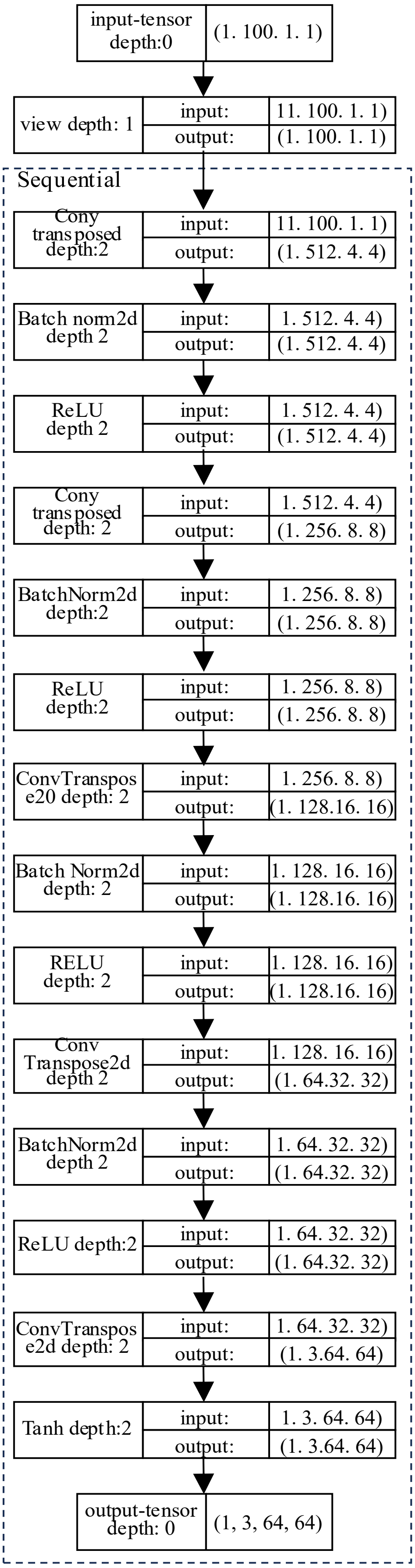

Figure 1. DCGAN generator network architecture.

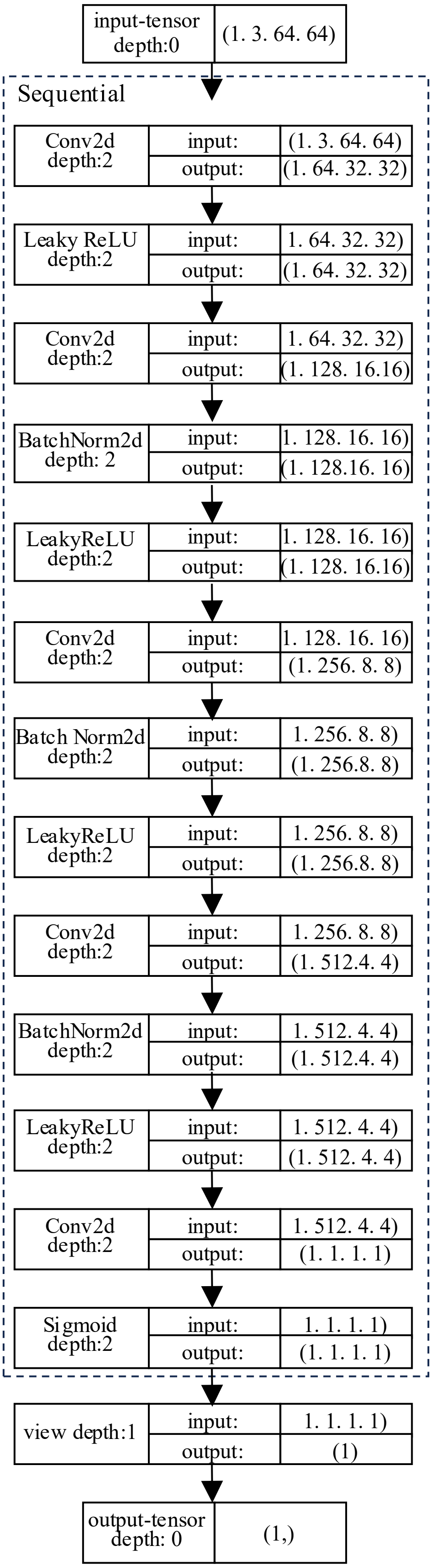

Figure 2. DCGAN discriminator network architecture.

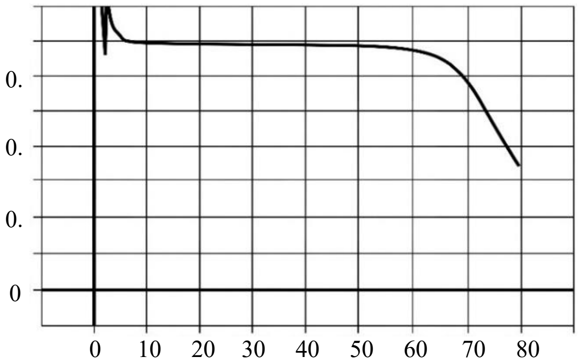

Figure 3. Discriminator loss per step.

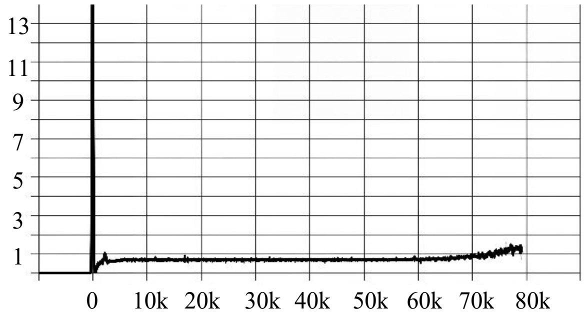

Figure 4. Generator loss per step.

Both optimizers, optD and optG, are updated alternately using manual gradient control to maintain stability during adversarial training [3, 5]. This two-step optimization closely follows the "adversarial minimax objective" described by Goodfellow et al., while the use of convolutional structures and BCE-based stabilization follows the DCGAN training guidelines established by Radford et al. [1, 2].

Model Saving and Evaluation

The generator model was saved separately after training. 32 random anime-style faces were then created in an 8×4 grid. The generated samples were denormalized to the [0, 1] range for visualization. The output was then assessed based on quality and variation.

RESULTS

The model generated showed a clear increase in quality and variation of output with an increase in the number of training epochs. Major improvements in output occur at critical training steps: approximately the 100th, 600th, and 1,000th epochs (Figures 5–7).

At the 100th epoch, the generated output faces showed clear outlines and features, such as eyes, mouth, and hair. This shows that the generator learned the important spatial and textural distribution of the dataset. However, the output lacked variation in the images produced.

By the 600th epoch, there was a significant improvement in the structure and diversity of the images generated. The face began to vary from one to another, which may be in terms of the orientation of the face, hairstyle, or even facial expression. It had definitely improved upon generating different samples that captured a better understanding of data distribution. However, some images retained a degree of distortion: sometimes the facial proportions were unbalanced, while other samples exhibited blurred or distorted textures. This indicates that the generator was still refining its inner representations and had not yet achieved full adversarial stability with the discriminator.

After 1,000 epochs were completed, the balance between the generator and discriminator was relatively good; therefore, the generated anime faces became visually coherent and realistic. These produced highly varied yet structurally correct images with clear facial outlines and appropriate color distributions. In contrast to the previous steps, the output during this stage was not distorted in general, with large variances among the samples. In other words, the model learned to generate distinct high-quality images from random noise. Improvement in the stability and variation of the final results might be due to the convolutional and deconvolutional structure of DCGAN, batch normalization layers, and the use of such ReLU and LeakyReLU activation functions that allow for stable gradient flows without mode collapses [2, 4].

Figure 5. Generated images at 100 epochs.

Figure 6. Generated images at 600 epochs.

Figure 7. Generated images at 1,000 epochs.

The learning pattern of the model presented in this study is similar to that of previous studies on anime and artistic face synthesis. The first epochs grasp coarse spatial features, whereas later epochs enhance the stylistic and semantic coherence of the output [15, 19].

The fact that the images generated evolve steadily across training epochs serves as an indicator of the efficiency of the DCGAN framework for image generation tasks [20]. The results confirmed that DCGANs learn complex image representations to output visually appealing samples with structural accuracy and artistic variation. This further supports the hypothesis that DCGANs yield a more stable and efficient generative process than traditional GANs, especially for domains with rich visual details, such as anime-face generation.

Other works, such as SAGAN and BlendGAN, further extended the work of DCGAN with attention mechanisms and style blending, respectively, to achieve higher semantic consistency or even more artistic control in the output [14, 16, 17].

CONCLUSION

This study explores the advantages of DCGANs in generating high-quality anime character facial images. The early GAN architectures faced several issues, such as unstable training, mode collapse, and distortions in the output produced during generation. These issues were resolved by DCGAN using convolutional and deconvolutional layers with batch normalization and appropriate activation functions. This makes DCGAN a more stable and effective training method for generating visual content.

As the DCGAN model started training, gradual improvement was observed in terms of visual quality and the variation of the output generated over 1,000 epochs. Initially, the output images were simple and repetitive but still recognizable as faces. Gradually, the output images began to show variations with an increase in the number of training epochs. Finally, during the last stages of training, the output images were not only of higher quality but also varied sufficiently to show their capability to learn stylistic features representative of anime imagery. This steady progress indicates the model's ability to learn hierarchical representations using a convolutional architecture.

These improvements were achieved through several important architectural and optimization choices. The convolution framework functions as a means to abstract features through multiple levels of spatial hierarchy. Transposed convolutions in the generators allow for a smooth upsampling process. Batch normalization greatly stabilizes the flow of gradients and reduces the problems related to internal covariate shift, while using LeakyReLU and ReLU resolves the vanishing gradient problem. Finally, using the Adam optimizer allows for efficient convergence, improving the balance between the generator and discriminator, which is an important factor for adversarial balance during training.

Some other more advanced models, such as WGAN and StyleGAN, have pushed realism and control even further; however, DCGAN remains one of the basic architectures that balances simplicity, computational efficiency, and image quality.

In this context, the present study asserts that DCGAN is a robust and powerful generative model for anime-face synthesis. The experimental results prove that the DCGAN models hierarchical spatial features with stable adversarial training and coherence with diversity in the output. Such improvements are due to architectural novelties, such as convolution operations, batch normalization, and optimized activation dynamics, which enhance the stability of the gradient and the quality of visualization.

Compared to earlier GANs, DCGAN have better stability, less mode collapse, and higher representational efficiency; hence, they are suitable for structured visual domains, such as anime-face generation. More recent models, including WGAN, StyleGAN, and transformer-based GANs, have further advanced the frontiers of realism and control. Despite all these advancements, the DCGAN is still a milestone architecture that connects early adversarial designs and modern generative paradigms.

The balance between simplicity and stability, along with expressive quality, is highly relevant in the evolving field of generative AI.

REFERENCES

1. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, editors. Advances in Neural Information Processing Systems. Vol. 27. Red Hook (NY): Curran Associates Inc.; 2014.
2. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks [preprint]. 2015. arXiv:1511.06434. doi:10.48550/arXiv.1511.06434.
3. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: Precup D, Teh YW, editors. Proceedings of the 34th International Conference on Machine Learning; 2017 Aug 6–11; Sydney, Australia. Proceedings of Machine Learning Research. Vol. 70. PMLR; 2017. p. 214–223.
4. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X. Improved techniques for training GANs. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, editors. Advances in Neural Information Processing Systems. Vol. 29. Red Hook (NY): Curran Associates Inc.; 2016.
5. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of Wasserstein GANs. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems. Vol. 30. Red Hook (NY): Curran Associates Inc.; 2017.
6. Jin Y, Zhang J, Li M, Tian Y, Zhu H, Fang Z. Towards the automatic anime characters creation with generative adversarial networks [preprint]. 2017. arXiv:1708.05509. doi:10.48550/arXiv.1708.05509.
7. Li B, Zhu Y, Wang Y, Lin CW, Ghanem B, Shen L. AniGAN: Style-guided generative adversarial networks for unsupervised anime face generation. IEEE Trans Multimed. 2022;24:4077–4091. doi:10.1109/TMM.2021.3113786.
8. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy. 2017. p. 2242–2251. doi:10.1109/ICCV.2017.244.
9. Liu MY, Tuzel O. Coupled generative adversarial networks. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, editors. Advances in Neural Information Processing Systems. Vol. 29. Red Hook (NY): Curran Associates Inc.; 2016.
10. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. IEEE Trans Pattern Anal Mach Intell. 2021;43(12):4217–4228. doi:10.1109/TPAMI.2020.2970919.
11. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T. Analyzing and improving the image quality of StyleGAN. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA. 2020. p. 8107–8116. doi:10.1109/CVPR42600.2020.00813.
12. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [preprint]. 2015. arXiv:1502.03167. doi:10.48550/arXiv.1502.03167.
13. Kingma DP, Ba J. Adam: A method for stochastic optimization [preprint]. 2014. arXiv:1412.6980. doi:10.48550/arXiv.1412.6980.
14. Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. In: Chaudhuri K, Salakhutdinov R, editors. Proceedings of the 36th International Conference on Machine Learning. PMLR; 2019. p. 7354–7363.
15. Li X, Li B, Fang M, Huang R, Huang X. BaMSGAN: Self-attention generative adversarial network with blur and memory for anime face generation. Mathematics. 2023;11(20):4401. doi:10.3390/math11204401.
16. Liu M, Li Q, Qin Z, Zhang G, Wan P, Zheng W. BlendGAN: Implicitly GAN blending for arbitrary stylized face generation. In: Advances in Neural Information Processing Systems. Vol. 34. Red Hook (NY): Curran Associates Inc.; 2021. p. 29710–29722.

17. Chakraborty T, Reddy UKS, Naik SM, Panja M, Manvitha B. Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art. Mach Learn Sci Technol. 2024;5(1):011001. doi:10.1088/2632-2153/ad1f77.
18. Dubey SR, Singh SK. Transformer-based generative adversarial networks in computer vision: A comprehensive survey. IEEE Trans Artif Intell. 2024;5:4851–4867. doi:10.1109/TAI.2024.3404910.
19. Joshi D, Sonawane S, Bhat T, Jumbad V, Wasade P, Zod S. Anime face generation using DC-GANs. AIP Conf Proc. 2023;2981(1):020028. doi:10.1063/5.0182666.
20. Cao YJ, Jia LL, Chen YX, Lin N, Yang C, Zhang B, et al. Recent advances of generative adversarial networks in computer vision. IEEE Access. 2019;7:14985–15006. doi:10.1109/ACCESS.2018.2886814.