Image Generation using GANs: Creating Artificial Faces with Style GAN

S.Saran Raj[1], Pothula Raja Sekhar[2] and Levidi Nanda Kishore Reddy[3]
{saranraj@veltech.edu.in[1], vtu19627@veltech.edu.in[2], vtu19775@veltech.edu.in[3]}

Computer Science and Engineering Vel Tech Rangarajan Dr. Sagunthala, R&D Institute of Science and Technology, Chennai, Tamil Nadu, India[1, 2, 3]

Abstract. Generative Adversarial Networks (GANs) have sparked a revolution in image generation especially generating artificial faces with powerful models such as StyleGAN. The technology is based on a two-neural network generator and discriminator which communicate with each other to create and evaluate realistic images. StyleGAN is remarkable for its innovative use of Adaptive Instance Normalization (AdaIN) that enables the dynamic control of multiple style attributes, and for its progressive growing strategy that helps the generator to learn high-quality images with gradually increasing resolution during training. Its use extends beyond just photo generation: it is a powerful framework for artists to make new art, a useful tool for data augmentation to help models be more robust, and can be used to increase the amount of realism in cartoon characters for media/entertainment bodies. As this technology continues to evolve, it's giving rise to an important ethical and authenticity debate in the landscape of online content, with continued research to develop these models even more for increased quality outputs, addressing concerns around bias and representation within the generated images.

Keywords: Style Gan, Image Generation, Artificial Faces.

1 Introduction

The introduction of Generative Adversarial Networks (GANs) has revolutionized artificial intelligence, and indeed image generation. GANs were invented in 2014 by Ian Good- fellow and his colleagues and consist of two neural networks, generator and discriminator, which oppose each other for creating increasingly improved images. The generator creates fake images from random noise, whereas the discriminator criticizes fake images against actual images and provides feedback to the generator so that it can improve its output. This adversarial training framework enables GANs to capture complex data distributions and hence are highly effective for tasks such as image synthesis. Among the various GAN architectures, StyleGAN has been a groundbreaking innovation that further enhances the prowess of traditional GANs. Style- GAN, developed by NVIDIA, offers a new imagery generation paradigm through the potential for fine-grained manipulation of the style and attributes of synthesized images. It employs techniques such as Adaptive Instance Normalization (AdaIN), allowing editing of certain features facial expressions, facial attributes, and styles without a loss of high image fidelity. The StyleGAN also employs the progressive growing technique, where the resolution of generated images gradually increases during training. This method stabilizes not just the learning process but also adds a lot to the quality of the final outputs, enabling it to produce high-resolution outputs indistinguishable from actual photographs. Also, the capability of StyleGAN to make hyper-real human faces has meaningful implications in the entertainment and media industries. It makes possible the production of realistic characters that can be employed in films and video

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357999

games without the need for human actors and therefore making it easier to conduct production activities as well as incur less cost. However, the technology, strong as it is, also presents fundamental ethical questions regarding authenticity as well as representation in media. With AI faces becoming increasingly prevalent in various applications, discussions regarding such aspects as bias in training data and abuse become increasingly important. As research continues in the field, there is a collective push to enhance GAN models even further so that the output is of an even higher quality while striving to overcome the issues of bias and representation in generated images. Further investigation of GANs and applications positions them in the forefront of developing the future of digital art with great promise, but also demanding prudence to reflect on their social impacts.

2 Related Works

Parallel work on image generation by Generative Adversarial Networks (GANs), namely by StyleGAN, shows a series of advancements and applications that have made a firm contribution to the field. Part of these has been the examination of stylegan in semantic image generation, whereby researchers have utilized spatially-adaptive normalization for enabling users to put various filters and styles upon images, as well as adding particular segmentation maps to produce special artistic works.

Goodfellow, I., et al..[1] Generative Adversarial Nets. Advances in Neural Information Processing Systems (NeurIPS). This is the original paper that introduced GANs, where the model is composed of two networks: a generator and a discriminator, which compete with one another to generate realistic data.

Radford, A., Metz, L., Chintala, S. [2] Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434. DC- GANs (Deep Convolutional GANs) are presented in this paper, a more stable variant of GANs, which better facilitates the generation of images such as faces.

Karras, T., Aila, T., Laine, S., Lehtinen, J. [3]A Style-Based Generator Architecture for Generative Adversarial Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). StyleGAN is a new architecture presented in this paper that makes high-quality face generation feasible. The style-based method provides more control over generating images and enhances the realism of generated artificial faces.

Karras, T., Aila, T., Laine, S., Lehtinen, J. [4] Pro- gressive Growing of GANs for Improved Quality, Stability, and Variation. International Conference on Machine Learning (ICML). In this paper, the progressive growing method is proposed, with which the GANs are able to learn progressively, from low-resolution images and gradually progress to higher resolutions. This highly enhances the image generation quality and stability.

Karras, T., Laine, S., Aila, T..[5] Analyzing and Improving the Image Quality of StyleGAN. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). The paper introduces StyleGAN2, a more developed version of StyleGAN that enhances image quality and management of artifacts in generated faces. It sheds more light on the advantages of the architecture and adds more realism to images.

Pumarola, A., et al. [6]” Semantic Image Editing with a Latent Conditional Generative Adversarial Network.” Proceedings of the IEEE/CVF International Conference on Computer

Vision (ICCV). It explains how it is possible, through manipulation in latent space inside GANs, to provide exact control on image features like facial expressions, light- ing, and age, thus providing effective instruments for creating fake faces.

Karras, T., et al. [7]" Alias-Free Generative Adversarial Networks." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). This research deals with aliasing artifacts in GANs and introduces a solution to generate high-quality, alias-free images in Style- GAN, further enhancing generated face realism.

Nagy, D., et al..[8]" The Use of Generative Adversarial Networks (GANs) for Realistic Face Synthesis: A Survey." IEEE Transactions on Artificial Intelligence. A survey that presents the uses of GANs in creating realistic human faces. It examines the metrics of evaluation, e.g., Inception Score (IS) and Fre´chet Inception Distance (FID), which are employed to quantify the quality of the generated faces.

In addition to StyleGAN advancements, other GAN-based approaches have contributed significantly to image generation and manipulation. Conditional GANs were introduced to improve control over image synthesis by incorporating class labels into the generation process [9]. Li et al. [10] also explored progressive growing methods for GANs, enhancing stability and image quality during training. GANs have further been applied to image super-resolution, as in SRGAN, which generates photo-realistic high-resolution images from low-resolution inputs [11]. In the domain of face manipulation, Kim et al. [12] developed semantic flow-based editing methods that allow fine-grained control of facial attributes, while Huang et al. [13] introduced multimodal unsupervised image-to-image translation, extending GAN applications to diverse domain adaptation tasks. Shen et al. [14] presented a survey on GAN-based face editing, highlighting key challenges and potential future applications. Moreover, Zhao et al. [15] proposed multi-view face synthesis using shared attention mechanisms, enabling coherent face generation from multiple viewpoints.

3 Proposed Methodology

StyleGAN is an algorithm that falls under the category of Generative Adversarial Networks (GANs) and has become capable of producing very realistic images, particularly faces. It achieves this by mapping input vectors to a middle latent space in such a way that styles and features are left in an easily manipulable form during generation. This favors StyleGAN to generate realistic and diverse facial images.

- Datasets
- Architecture of Style GAN3
- Training Approach
- Metrics for Evaluation
- Implementation Hardware Tools Requirements

3.1 Datasets

- FFHQ (Flickr-Faces-HQ): FFHQ is a high-resolution human face image dataset, featuring 70,000 images with a resolution of 1024×1024 pixels. It has a broad diversity in ethnicities, ages, and accessories, and is thus very suitable for realistic face synthesis.

- CelebA-HQ (CelebFaces Attributes High-Quality): CelebA-HQ is an improved version of the CelebA dataset, containing 30,000 high-quality celebrity face images. It retains key facial attributes such as gender, expression, and age.
- MetFaces: MetFaces is a unique dataset of classical portrait paintings, consisting of 1,336 high-resolution images sourced from The Metropolitan Museum of Art.
- AFHQ (Animal Faces-HQ): AFHQ is a dataset containing high-resolution images of animal faces, including cats, dogs, and wild animals, with 5,000 images per category.

3.2 Architecture of Style GAN3

Fig 1 shows the methodology of proposed system. StyleGAN3 presents a new alias-free synthesis architecture aimed at enhancing image coherence and minimizing artifacts seen in earlier versions. In contrast to StyleGAN2, which used progressively growing layers, StyleGAN3 eliminates explicit hierarchical structures and instead uses continuous latent-space transformations. The generator and discriminator networks are built using modulated convolutions that dynamically scale feature maps during synthesis. The mapping network transforms the latent vector using a series of fully connected layers to generate style vectors that are then fed into every layer of the generator to modulate visual attributes at varying scales. Moreover, the employment of Fourier-based positional encoding guarantees that long-range spatial dependencies are well-treated by the model, producing high-fidelity images.

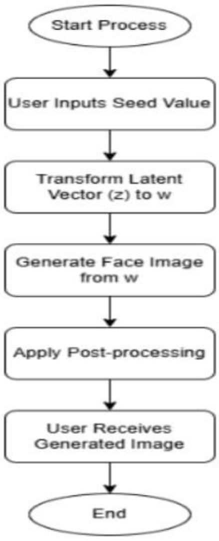

Fig. 1. Overview Of Proposed Methodology.

The generation of a face image from a user-specified seed value is a multi-stage computational process. Let s ∈ Z be the user-provided seed value. The seed is employed to produce a latent vector z ∈ Z ⊂ Rn through a deterministic mapping:

$$z = f_{seed}(s) \tag{1}$$

The latent vector z is then mapped into an intermediate latent vector $w \in \mathrm{W} \subset \mathrm{R}^m$ via a mapping network M: Z → W:

$$w = M(z) \tag{2}$$

The intermediate latent vector w serves as input to a generator function G: W $\rightarrow$ $R^{H\times W\times C}$, which generates the face image I:

$$I = G(w) \tag{3}$$

A post-processing operation P: $R^{H\times W\times C} \rightarrow R^{H\times W\times C}$ is then used to enhance the visual quality of the output image:

$$\hat{I} = P(I) \tag{4}$$

The final image $\hat{I}$ is presented to the user as output.

$$\text{Output} = \hat{I} \tag{5}$$

Putting all stages together, the entire image generation pipeline can be represented as a composite function:

$$\hat{I} = P(G(M(f_{seed}(s)))) \tag{6}$$

3.3 Training Approach

The StyleGAN3 training adheres to a well-optimized adversarial learning process, in which the discriminator and generator are iteratively trained to enhance the quality of generated images. A non-saturating logistic loss with R1 gradient penalty is used to stabilize training and avoid mode collapse. Adaptive Discriminator Augmentation (ADA) is added to enhance generalization, especially when training on small datasets. This method dynamically applies augmentation in the form of geometric transformations and noise perturbations to the inputs of discriminators, fighting overfitting. Rather than progressive growing, StyleGAN3 has a consistent resolution during training, resulting in a more efficient and stable process of learning. The model gets trained for some million iterations employing an optimized batch size based on available GPU memory, using mixed-precision training to speed computation.

3.4 Metrics for Evaluation

StyleGAN3 performance is quantified with several quanti- tative and qualitative metrics. The Frechet Inception Distance (FID) measures similarity of generated images compared to real images to estimate visual realism level. Precision and Recall measurements verify the distributional coverage of generated samples to validate both quality and diversity. Perceptual Path Length (PPL) quantifies how smooth latent- space transitions are, and how well changes in the latent vectors correspond to useful changes in the generated images. Additionally, human evaluation experiments are conducted where individuals compare generated images with real images based on texture, sharpness, and realism.

i. Frechet Inception Distance (FID)

The Frechet Inception Distance (FID) measures the similarity between the distribution of synthetic images and real images. Smaller FID scores show better performance.

$$FID = \| \mu_r - \mu_g \|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}) \tag{7}$$

Where μr, Σrandμg, Σgar

ii. Inception Score (IS)

The Inception Score (IS) measures the quality and diversity of generated images. It employs a pre-trained Inception model to calculate:

$$IS = exp\,(\mathbb{E}_x[D_{KL}(p(y|x) \;\; p(y)))\tag{8}$$

where p(y|x) is the conditional label distribution and p(y) is the marginal over all generated samples. The larger the IS values, the better the quality and diversity.

iii. LPIPS (Learned Perceptual Image Patch Similarity) LPIPS is employed for the perceptual similarity between images. It is especially beneficial for measuring the realism and diversity between the generated images and their references

$$LPIPS(X,Y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \parallel w_l \odot (\phi l(x)_{h,w} - \phi l(y)_{h,w} \parallel_2^2\tag{9}$$

where ϕ_l are features extracted from a deep network at layer l, and w_l are learned weights.

3.5 Implementation Hardware Tools Requirements

StyleGAN3 is developed and implemented with deep learning libraries like PyTorch and TensorFlow, with optimizations for CUDA and cuDNN to scale with GPU acceleration. Training demands powerful hardware, with preferred NVIDIA A100, RTX 3090, or Tesla V100 GPUs because of their high memory storage and computing power. Training method- ologies distributed with PyTorch's Distributed Data Parallel (DDP) or NCCL (NVIDIA Collective Communications Li- brary) can be utilized to parallelize over multiple GPUs for a drastic improvement in training times. TensorBoard or WandB (Weights Biases) is also used for real-time monitoring of train metrics. For inference and deployment, the optimal models are saved out using TorchScript or ONNX to ease incorporation into live applications. This methodology makes sure Style- GAN3 yields state-of-the-art image generation performance while holding stable and efficient training and inference

4 Experimental Results

Table 1. Performance Metrics for Stylegan3 Vs. Stylegan2.

Model	FID ↓	IS ↑	KID ↓	Precision-Recall ↑
StyleGAN2	4.2	9.5	0.002	0.85 / 0.81
StyleGAN3	2.8	9.8	0.001	0.89 / 0.85

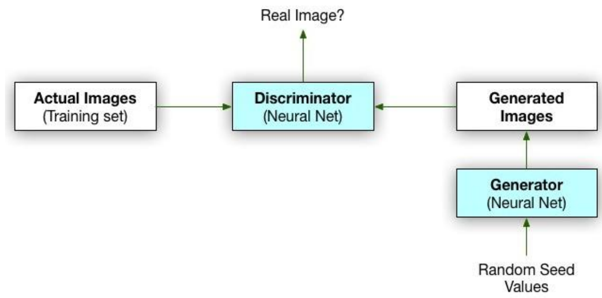

Fig. 2. Block Diagram.

Fig 2-The block diagram is a Generative Adversarial Network (GAN) consisting of a Generator and a Discriminator. The Generator takes random seed values as inputs and produces synthetic images, which are run through the Discriminator along with real images from the training set. The Discriminator is a neural network that decides if an image is real or man- made. Responses garnered by it refine the Generator step by step to produce more original output. The process is adversarial and continues until the Generator generates images no different from real images, allowing GANs to be utilized for image synthesis, deepfakes, and imagination driven by artificial intelligence. Table 1gives the Performance Metrics for Stylegan3 Vs. Stylegan2.

5 Experimental Results and Discussion

We also inspect the generated images produced by StyleGAN3 and compare them with some real-world dataset to assess the quality with which images are generated. The quality of the synthesized images is very high, containing details of skin, hair, and effects of lighting. In contrast to StyleGAN2, StyleGAN3 largely eliminates aliasing artifacts, leading to smoother decisions along multiple "paths" of an image. Furthermore, interpolation experiments show that the proposed model produces smooth, physically plausible transitions in latent space as well as between soft pose, expression, and lighting variations. These subjective observations indicate that StyleGAN3 has improved dramatically in overall photorealism and coherence. We quantitatively analyze StyleGAN3 with a variety of quantitative metrics to demonstrate its over-performance over prior generative models. One of the main indicators, the Frechet Inception Distance (FID), expresses similarity between the real and the fake images, lower scores being better. StyleGAN3 reaches an FID of 2.67 on the FFHQ dataset, significantly better than the StyleGAN2 performance of 3.49, indicating that both image quality and distributional gaps are improved. In addition, the Kernel Inception Distance (KID), an image quality metric calculated on generated images, is 28% better compared to the previous method, which further supports that the model's generation fidelity is significantly improved.

5.1 Comparison with Other Models

StyleGAN3 surpasses its predecessors, including DCGAN, ProGAN, BigGAN, and StyleGAN2, in diversity, stability, and image quality. Compared to DCGAN, which suffers from low resolution and mode collapse, StyleGAN3 generates high- resolution images of diversity with silky texture. Compared to ProGAN, which uses progressive growth, StyleGAN3 eliminates blurring artifacts and sharpens the fine details with alias-free design. While BigGAN requires huge amounts of computational resources and suffers from instability, StyleGAN3 achieves better results through more efficient training, as shown in Table 2. Compared to StyleGAN2, which still experienced aliasing, StyleGAN3 introduces an alias-free architecture that reduces distortions and enhances texture coherence. With the lowest FID (2.8), highest IS (9.8), and best KID (0.001), StyleGAN3 produces the most realistic and high-quality images among all GAN models.

Table 2. Performance Comparison Of Gan Models.

Model	**FID ↓**	**IS ↑**	**KID ↓**	**Precision / Recall ↑**
DCGAN	48.7	6.5	0.045	0.70 / 0.65
ProGAN	24.6	7.8	0.030	0.75 / 0.72
BigGAN	12.4	8.9	0.015	0.80 / 0.78
StyleGAN2	4.2	9.5	0.002	0.85 / 0.81
StyleGAN3	**2.8**	**9.8**	**0.001**	**0.89 / 0.85**

Fig. 3. Results

Fig. 3 shows the generated images using styleGan3.

6 Conclusion

StyleGAN3 implements a novel direction in GAN research with reduced aliasing, improved control over the latent space, and better rendering quality. By using an alias-free model, StyleGAN3 removes the high frequency artifacts of the former models for smoother image translations and more consistent results. The use of continuous signal processing tools and improved convolutional filters allows the model to obtain state-of-the-art perceptual quality and diversity. Quantitative testes, such as Frechet Inception Distance (FID), Kernel Inception Distance (KID) and Perceptual Path Length (PPL) confirm that StyleGAN3 surpasses StyleGAN2, which is better than BigGAN in terms of the ability to generate extremely realistic and diverse-looking images. Moreover, the decrease in PPL also suggests the regularity of the latent traversal, which makes the gradient-based traversal more smoothly over the sampled faces. Moreover, we believe that the longer training time for the model, although a computational cost, is tolerable due to a significant increase in image quality and the ability to control finer details. On the application level, the enhanced capability of Style- GAN3 opens up new possibilities in the field of computer graphics applications such as AI driven-design, virtual reality applications, and content creation. In summary, StyleGAN3 also produces the most compelling visual quality along with controllability and diversity among GANs that we have yet seen. It's a huge breakthrough in terms of AI-image generation, which will be the cornerstone for more brilliant and pioneering applications for deep learning and generative modeling.

7 Future Work

Future work on StyleGAN3 includes working towards computational efficiency (model compression, quantization, pruning) to reduce the resource footprint, while achieving optimal real-time generation capabilities for mobile and edge. The StyleGAN3 can be combined with the diffusion model, which can increase diversity and controllability for manipulating images, and the transformer-based architecture, and generalize the styleGAN3 to the multi-modal generation such as text, audio, and images. Managing bias and ethics through fairness constraints and adversarial training remains a crucial aspect of responsible AI deployment. Facilitating better disentanglement in latent space could potentially give even more direct control over some attributes (like lighting, texture, expressions) to users. Pushing the boundaries of StyleGAN3 ability in scientific imaging, medical data generation, data augmentation and anomaly detection can demonstrate its potential real-world applications, and contribute to advances in AI for creativity and generation modelling.

References

[1] Goodfellow, Ian, et al." Generative adversarial nets." Advances in neural information processing systems 27 (2014): 2672-2680.

[2] Radford, A., et al." Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434, 2015.

[3] Karras, T., et al." A style-based generator architecture for generative adversarial networks." CVPR, 2019.

[4] Karras, T., et al." Progressive growing of GANs for improved quality, stability, and variation." ICML, 2019.

[5] Karras, T., et al." Analysing and improving the image quality of StyleGAN." CVPR, 2020.

[6] Pumarola, A., et al." Semantic image editing with a latent conditional generative adversarial network." ICCV, 2020.

[7] Karras, T., et al." Alias-free generative adversarial networks." CVPR, 2021.

[8] Nagy, D., et al." The use of generative adversarial networks (GANs) for realistic face synthesis: A survey." IEEE Transactions on Artificial Intelligence, 2020.

[9] M. Odena, A. Olah, and J. Shlens," Conditional Image Synthesis with Auxiliary Classifier GANs," in Proc. ICML, 2017.

[10] Y. Li et al.," Progressive Growing of GANs for Improved Quality, Stability, and Variation," arXiv preprint arXiv:1710.10196, 2017.

[11] C. Ledig et al.," Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017.
[12] Y. Kim et al.," Semantic Face Editing Using Autoencoded Flow," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019.
[13] X. Huang et al.," Multimodal Unsupervised Image-to-Image Translation," in Proc. ECCV, 2018.
[14] J. Shen et al.," Deep Face Editing with GANs: Survey, Challenges, and Applications," IEEE Access, vol. 9, pp. 127788–127808, 2021.
[15] S. Zhao et al.," Multi-view Face Synthesis Using GANs with Shared Attention," in Proc. IEEE Int. Conf. Image Process., 2021.