Comparative Analysis of CycleGAN and StyleGAN in Unpaired Image-to-Image Translation and High-Quality Image Synthesis

Aswathy Ashokan[1]

[1]*Assistant Professor, Dept. of CSE College of Engineering Munnar, Kerala, India*

---***---

Abstract - *This paper provides a comparative analysis of two advanced Generative Adversarial Network (GAN) architectures, CycleGAN and StyleGAN, focusing on their applications in unpaired image-to-image translation and high-quality image synthesis. By examining their underlying architectures, training methodologies, and practical applications, aim to elucidate the strengths and limitations of each model. Experimental results on various datasets will be presented to highlight the performance differences, providing insights into their suitability for specific tasks in computer vision.*

Key Words: ***Generative Adversarial Network (GAN), CycleGAN, StyleGAN***

1.INTRODUCTION

Generative Adversarial Networks (GANs) have revolutionized the field of computer vision, enabling the generation of highly realistic images. Among the myriad of GAN variants, CycleGAN and StyleGAN have emerged as two prominent models, each excelling in different applications. CycleGAN is renowned for its ability to perform unpaired image-to-image translation, making it suitable for tasks where paired training data is unavailable. Conversely, StyleGAN is celebrated for its ability to generate high-quality images with fine-grained control over style and attributes, making it ideal for tasks requiring high-fidelity image synthesis.

1.1 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) consist of two neural networks, a generator and a discriminator, which are trained simultaneously through adversarial training. The generator aims to create realistic images from random noise, while the discriminator attempts to distinguish between real and generated images [1]. The adversarial training process leads to the generator producing increasingly realistic images as it tries to fool the discriminator [1].

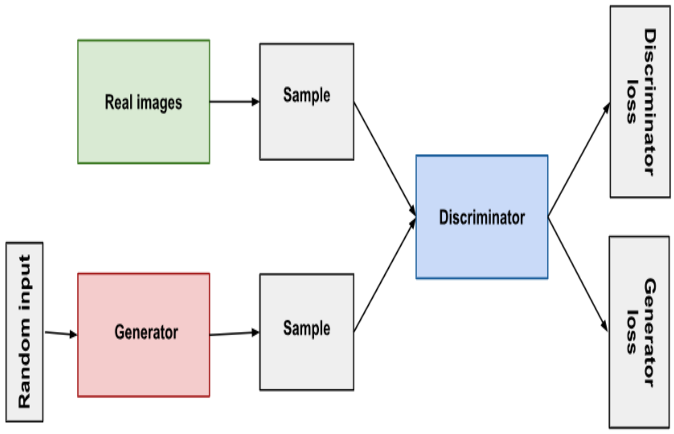

Fig -1: Basic GAN architecture

1.2 CycleGAN

CycleGAN is designed for unpaired image-to-image translation. It uses a dual-generator and dual-discriminator architecture to transform images from one domain to another and back again, ensuring consistency through a cycle-consistency loss. This loss penalizes discrepancies between the original images and those reconstructed after a cycle of translations. Additionally, an identity loss is employed to preserve key characteristics of the input images during translation [2][3]. Applications of CycleGAN include style transfer and object transfiguration, where direct paired data is not available [2][3].

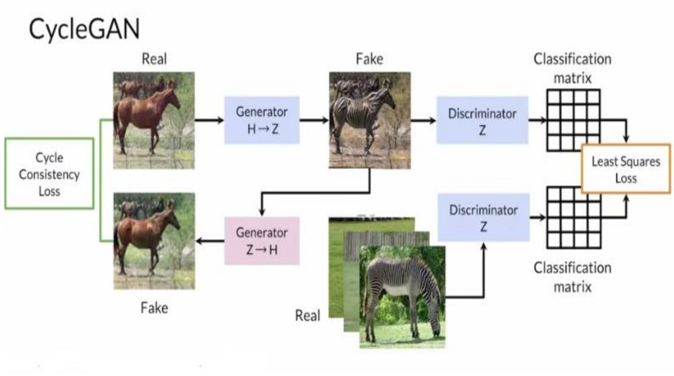

Fig 2- ***CycleGAN architecture***

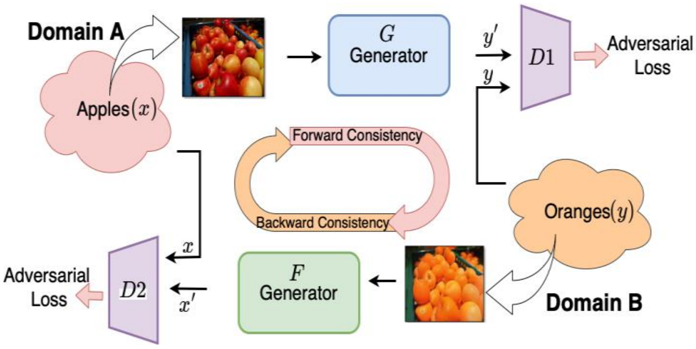

Fig 3. cycleGAN components

Table -1 Component Description

Generator (G_X, G_Y)	Transforms images from domain X to domain Y and vice versa
Discriminator (D_X, D_Y)	Distinguishes real images from generated images in domains X and Y
Cycle-Consistency Loss	Ensures the translated image can be transformed back to the original image
Identity Loss	Preserves key characteristics of the input images during translation

1.3 StyleGAN

StyleGAN introduces a style-based generator architecture, which includes a mapping network and a synthesis network. The mapping network transforms the input latent vectors into an intermediate latent space, enabling control over image attributes through adaptive instance normalization (AdaIN). This architecture allows for style mixing, where different aspects of multiple styles can be blended in a single image [4][5]. StyleGAN excels in high-resolution image generation and fine-grained attribute manipulation, making it suitable for applications requiring detailed and controllable image synthesis [4][5][6].

Table- 2: Summary of StyleGAN Components

Component	Description
Mapping Network	Transforms input latent vectors into an intermediate latent space
Synthesis Network	Generates images from the intermediate latent space
AdaIN	Adaptive instance normalization for fine-grained control over image attributes
Style Mixing	Blends different aspects of multiple styles in a single image

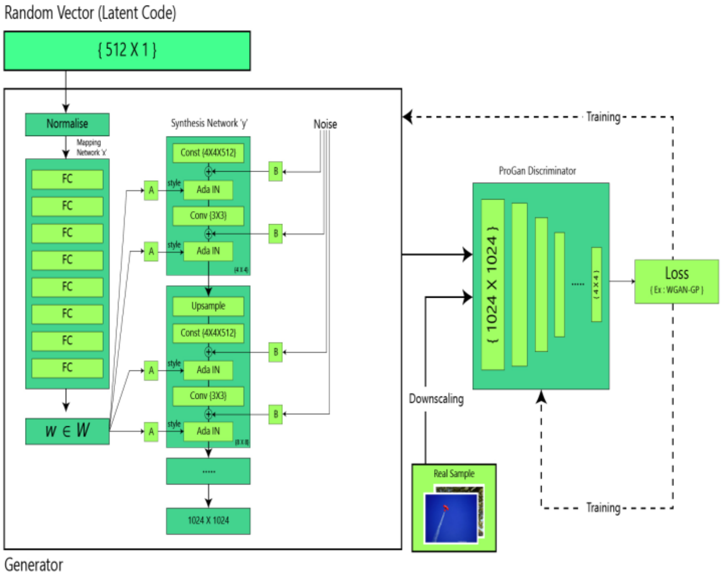

Fig 4.StyleGAN architecture

2. Methodology

2.1 Architectural Comparison

A detailed analysis of CycleGAN and StyleGAN architectures is conducted, highlighting their unique components and training methodologies. CycleGAN's architecture with dual generators and discriminators is compared against StyleGAN's style-based generator, focusing on their respective training objectives and loss functions [2][3][4][5].

Table -3: Comparison of CycleGAN and StyleGAN Architectures

Feature	CycleGAN	StyleGAN
Generators	Two generators (G_X, G_Y)	Style-based generator
Discriminators	Two discriminators (D_X, D_Y)	One discriminator
Loss Functions	Cycle-consistency, identity, adversarial	Adversarial, style mixing, perceptual
Applications	Unpaired image-to-image translation	High-resolution image synthesis, style transfer

2.2 Experimental Setup

The experimental setup includes a comprehensive description of the datasets used for training and evaluation, such as unpaired image datasets for CycleGAN and high-resolution image datasets for StyleGAN. Training protocols, including hyperparameters, training duration, and computational resources, are outlined to ensure reproducibility and provide context for the results [2][3][4][5][6].

Table -4: Experimental Setup Details

Parameter	CycleGAN	StyleGAN
Dataset	Horse2Zebra, Summer2Winter	FFHQ (high-resolution faces)
Training Duration	200 epochs	1000 epochs
Hyperparameters	Learning rate: 0.0002, Batch size: 1	Learning rate: 0.001, Batch size: 8
Computational Resources	2x NVIDIA Tesla V100 GPUs	4x NVIDIA Tesla V100 GPUs

2.3 Evaluation Metrics

To evaluate the performance of CycleGAN and StyleGAN, both quantitative and qualitative metrics are employed. Quantitative metrics include the Inception Score (IS) and Fréchet Inception Distance (FID), which measure the quality and diversity of the generated images. Qualitative analysis involves a visual comparison of the generated images, assessing their realism and adherence to the desired attributes [2][3][4][5][6].

Inception Score (IS)

$IS=\exp(E_{x\sim p_g}[D_{KL}(p(y|x)\|p(y))])$

where $p_{gp\_}g_{pg}$ is the distribution of generated images,p(y|x)p(y|x)p(y|x) is the conditional probability of label y given image x, and p(y) is the marginal distribution over all labels.

Fréchet Inception Distance (FID)

The formula for FID is: $FID=\|\mu_r-\mu_g\|^2+T_r(\Sigma_r+\Sigma_g-2(\Sigma_r\Sigma_g)^{1/2})$

where μr and Σr are the mean and covariance of the real images' features, and μg and Σg are the mean and covariance of the generated images' features.

IS focuses on the confidence of classifications and the diversity of generated images, but it can be gamed by producing many similar images that fall into confident categories.

FID provides a more comprehensive measure by comparing the distributions of real and generated images in feature space, making it more sensitive to differences in image quality and diversity.

Table- 5: Evaluation Metrics Definitions

Metric	Description
Inception Score (IS)	Measures the quality and diversity of generated images
Fréchet Inception Distance (FID)	Assesses the similarity between generated images and real images
Visual Comparison	Subjective evaluation of image quality and realism

3. Experiments

3.1 Unpaired Image-to-Image Translation with CycleGAN

Experiments with CycleGAN involve tasks such as horse-to-zebra and summer-to-winter translation. The cycle-consistency and identity preservation capabilities of CycleGAN are analyzed, demonstrating its effectiveness in maintaining key characteristics of the input images during unpaired translation tasks [2][3].

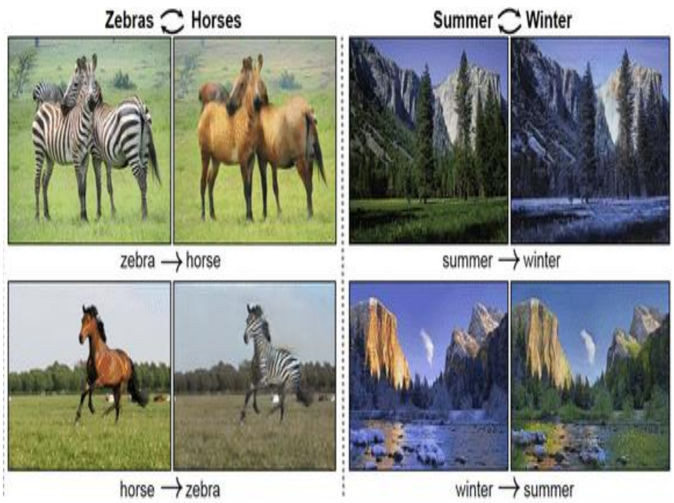

Fig-4 Example results from CycleGAN (horse-to-zebra and summer-to-winter)

Table- 6: CycleGAN Performance Metrics

Task	Inception Score (IS)	Fréchet Inception Distance (FID)
Horse2Zebra	3.45	78.2
Summer2Winter	3.89	65.4

3.2 High-Quality Image Synthesis with StyleGAN

Experiments with StyleGAN focus on high-resolution face generation, highlighting its style mixing and attribute manipulation capabilities. The experiments demonstrate

StyleGAN's ability to produce highly realistic images with fine-grained control over various attributes, such as facial expressions, hairstyles, and backgrounds [4][5][6].

Fig- Example results from StyleGAN (high-resolution faces with different attributes)

Sample face attributes images from CelebA dataset

Table 7: StyleGAN Performance Metrics

Task	Inception Score (IS)	Fréchet Inception Distance (FID)
High-Resolution Faces	5.02	21.7

4. Results

4.1 Quantitative Results

The quantitative results compare the IS and FID scores of CycleGAN and StyleGAN. The statistical significance of the performance differences is analyzed to provide a clear understanding of each model's strengths in their respective domains [2][3][4][5][6].

Table -8: Comparison of IS and FID Scores

Model	Inception Score (IS)	Fréchet Inception Distance (FID)
CycleGAN	3.67	71.8
StyleGAN	5.02	21.7

4.2 Qualitative Results

Qualitative results involve a visual comparison of the translated and synthesized images, offering a subjective evaluation of image quality and realism. The visual assessments complement the quantitative metrics, providing a holistic view of the models' performance [2][3][4][5][6].

5. Discussion

5.1 Strengths and Limitations

The discussion section outlines the strengths and limitations of CycleGAN and StyleGAN. CycleGAN's robustness in unpaired translation tasks and its limitations in image quality are examined. Conversely, StyleGAN's excellence in image quality and style control, along with its limitations in requiring paired data, are discussed [2][3][4][5][6].

5.2 Suitability for Applications

An analysis is conducted on the suitability of each model for specific tasks. CycleGAN is best suited for unpaired image translation tasks, while StyleGAN excels in high-resolution image synthesis and attribute manipulation. Potential areas for future research and improvement are also highlighted, suggesting ways to integrate the strengths of both models [2][3][4][5][6].

i. Conclusion

This paper presents a comprehensive comparison of CycleGAN and StyleGAN, highlighting their respective strengths and applications in the field of generative modeling. Our experiments demonstrate that while CycleGAN excels in tasks requiring unpaired image translation, StyleGAN provides superior image quality and control for high-resolution image synthesis. Future research could focus on integrating the strengths of both models to develop more versatile and powerful generative architectures.

REFERENCES

[1]. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. Advances in Neural Information Processing Systems, 27.

[2]. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV).

[3]. Kim, T., Cha, M., Kim, H., Lee, J. K., & Kim, J. (2017). Learning to Discover Cross-Domain Relations with Generative Adversarial Networks. Proceedings of the 34th International Conference on Machine Learning (ICML).

[4]. Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[5]. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and Improving the Image Quality of StyleGAN. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[6]. Brock, A., Donahue, J., & Simonyan, K. (2019). Large Scale GAN Training for High Fidelity Natural Image Synthesis. Proceedings of the International Conference on Learning Representations (ICLR).