International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

A Comparative Study of GANs and VAEs for Image Generation

Anushka Narula
Georgia Institute of Technology
Email: anarula31@gatech.edu

Abstract—Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) are two of the most prominent generative models for image synthesis. This paper provides a comprehensive comparison of GANs and VAEs, focusing on their architectures, training methodologies, and performance in image generation tasks. We also evaluate their differences in terms of stability, quality, and diversity of outputs, supported by quantitative metrics such as Frechet Inception Distance (FID) and´ Inception Score (IS). A detailed analysis of their applications in various domains is presented, along with a discussion on their limitations and future directions.

I. Introduction

Generative models are fundamental in computer vision and machine learning, enabling tasks such as image generation, style transfer, and data augmentation. These models aim to learn the underlying distribution of data and generate new samples that are indistinguishable from the original dataset. Among these, GANs and VAEs have emerged as leading frameworks due to their impressive capabilities.

GANs, introduced by Goodfellow et al. [1], leverage adversarial training between a generator and a discriminator to produce high-quality samples. VAEs, proposed by Kingma and Welling [2], use probabilistic methods to model latent spaces, ensuring a structured representation of data. Despite their shared goal of generating data, their methodologies, strengths, and weaknesses differ significantly. This paper explores these differences and evaluates their suitability for various applications.

II. Generative Adversarial Networks (GANs)

GANs consist of two neural networks, a *generator* $G(z)$ and a *discriminator* $D(x)$, trained in a zero-sum game. The generator aims to generate realistic data, while the discriminator attempts to distinguish real data from generated samples. The objective function is given by:

$$\min_G \max_D \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log(1-D(G(z)))]. \quad (1)$$

A. Challenges in GANs

GANs suffer from instability during training and mode collapse, where the generator produces limited diversity in outputs. Techniques like Wasserstein GAN (WGAN) [3], Gradient Penalty [4], and Spectral Normalization [5] have been introduced to mitigate these issues. Moreover, the evaluation of GANs often requires human judgment to assess the quality of generated samples, adding subjectivity to the process.

B. Applications of GANs

GANs are widely used in tasks such as image-to-image translation [6], super-resolution [14], and data augmentation for training deep models. For example, GAN-based methods have achieved state-of-the-art results in generating photorealistic human faces [13]. Furthermore, GANs have been employed in creative domains, including artwork generation and music composition.

III. Variational Autoencoders (VAEs)

VAEs are probabilistic generative models that learn a latent space representation of data. The encoder maps input x to a latent distribution $q(z|x)$, while the decoder reconstructs x from z. The loss function is given by:

$$L = -\mathbb{E}_{q(z|x)}[\log p(x|z)] + KL(q(z|x)\|p(z)), \quad (2)$$

where KL represents the Kullback-Leibler divergence.

A. Advantages of VAEs

VAEs ensure a continuous and interpretable latent space, enabling smooth interpolation between data points. This property is particularly useful in applications requiring structured latent spaces, such as molecular design [11]. Additionally, VAEs are less prone to mode collapse compared to GANs, making them suitable for applications demanding high output diversity.

B. Applications of VAEs

VAEs have been used in anomaly detection, generative design, and molecular generation due to their structured latent spaces. For example, VAEs have been applied to detect anomalies in medical imaging [12]. They are also utilized in natural language processing tasks, including text generation and language modeling, showcasing their versatility.

IV. Comparison of GANs and VAEs

A detailed comparison is provided in Table I. In addition, Figure 1 visualizes the differences in latent space structures.

V. Evaluation Metrics

Evaluating GANs and VAEs requires a combination of quantitative and qualitative metrics.

TABLE I
Comparison of GANs and VAEs

Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Aspect	GANs	VAEs
Training Stability	Unstable	Stable
Output Quality	Sharp, Realistic	Blurry, Smooth
Diversity of Outputs	Mode Collapse Risk	High Diversity
Latent Space Structure	Unstructured	Structured
Applications	Style Transfer, SuperRes	Anomaly Detection, Design

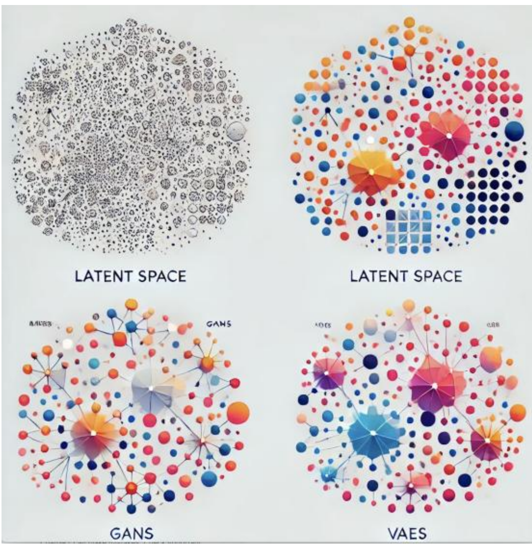

Fig. 1. Visualization of latent space structures in GANs and VAEs.

A. Inception Score (IS)

The Inception Score evaluates the quality and diversity of generated samples using a pre-trained classifier:

$$IS = \exp(\mathrm{E}_x[KL(p(y|x)||p(y))]), \qquad (3)$$

where $p(y|x)$ is the conditional label distribution, and $p(y)$ is the marginal distribution [7]. Higher IS values indicate better image quality and diversity.

B. Frechet Inception Distance (FID)´

FID measures the similarity between real and generated data distributions:

$$FID = ||\mu_r - \mu_g||^2 + \mathrm{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}), \qquad (4)$$

where μ_r, Σ_r and μ_g, Σ_g are the means and covariances of real and generated data in feature space [8]. FID has become a standard metric for generative models due to its robustness and interpretability.

C. Reconstruction Error

VAEs are often evaluated using reconstruction error, typically measured as Mean Squared Error (MSE) between input and output. This metric directly quantifies the model's ability to capture data features.

D. Perceptual Similarity

Metrics like LPIPS [9] assess perceptual similarity by comparing deep features rather than pixel-wise differences. This approach aligns more closely with human visual perception, making it a valuable tool for evaluating generated images.

VI. Conclusion and Future Work

GANs and VAEs each have unique strengths and weaknesses. While GANs excel in generating sharp, high-quality images, VAEs provide structured latent spaces for meaningful interpolation. Future work could explore hybrid models combining the strengths of both approaches, as seen in VAEGAN [10]. Additional research is needed to improve training stability, enhance output diversity, and develop better evaluation metrics. Moreover, applying these models to emerging domains such as video synthesis, reinforcement learning, and trustworthy AI remains a promising direction.

References

[1] I. Goodfellow et al., "Generative adversarial nets," in *NeurIPS*, 2014.
[2] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *ICLR*, 2014.
[3] M. Arjovsky et al., "Wasserstein GAN," in *ICML*, 2017.
[4] I. Gulrajani et al., "Improved Training of Wasserstein GANs," in *NeurIPS*, 2017.
[5] T. Miyato et al., "Spectral Normalization for GANs," in *ICLR*, 2018.
[6] P. Isola et al., "Image-to-Image Translation with Conditional GANs," in *CVPR*, 2017.
[7] T. Salimans et al., "Improved Techniques for Training GANs," in *NeurIPS*, 2016.
[8] M. Heusel et al., "GANs Trained by a Two Time-Scale Update Rule," in *NeurIPS*, 2017.
[9] R. Zhang et al., "Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in *CVPR*, 2018.
[10] A. B. Larsen et al., "Autoencoding beyond pixels using a learned similarity metric," in *ICML*, 2016.
[11] B. Gomez-Bombarelli et al., "Automatic Chemical Design using Varia-´ tional Autoencoders," in *Science*, 2018.
[12] C. Baur et al., "Autoencoders for Unsupervised Anomaly Segmentation in Brain MR Images," in *Medical Image Analysis*, 2021.
[13] T. Karras et al., "A Style-Based Generator Architecture for Generative Adversarial Networks," in *CVPR*, 2019.
[14] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *CVPR*, 2017.