Dreaming of Bedrooms: A Study of GANs and DDPMs in Bedroom Image Generation

Maojun SUN *

March 25, 2025

Abstract

Generative models have witnessed significant advancements in recent years, with Generative Adversarial Networks (GANs) and Denoising Diffusion Probabilistic Models (DDPMs) emerging as two prominent approaches. This study provides a comparative analysis of GANs and DDPMs, focusing on their practical performance in generating high-quality image data. Through a case study utilizing the LSUN bedroom dataset, we evaluate the efficiency and effectiveness of both models, using the Fréchet Inception Distance (FID) score as a key metric for image quality assessment. Our results show that DDPM achieves a significantly lower FID score of 8.5 compared to 16.2 for GAN, indicating superior image fidelity and realism. These findings suggest that DDPMs generate more structurally coherent and visually realistic images, whereas GANs exhibit greater difficulty in preserving fine details. We discuss the implications of these results, highlighting the strengths and limitations of each approach and providing insights into their potential applications in generative tasks.

1 Introduction to GAN and DDPM

Generative models have undergone significant advancements since their inception. Historically, early generative models such as Gaussian Mixture Models (GMMs) and Hidden Markov Models (HMMs) laid the foundation for subsequent developments. The introduction of more powerful models like Variational Autoencoders (VAEs) and normalizing flows expanded the capabilities of generative learning. Among these models, Generative Adversarial Networks (GANs), introduced by Goodfellow et al. (2020) and Denoising Diffusion Probabilistic Models (DDPMs), proposed by Ho et al. (2020), represent two highly influential approaches in the field.

The primary difference between GANs and DDPMs lies in their underlying mechanisms and theoretical foundations. GANs leverage a game-theoretic approach involving two competing neural networks, while DDPMs operate through

*Hong Kong Polytechnic University. Email: mj.sun@connect.polyu.hk

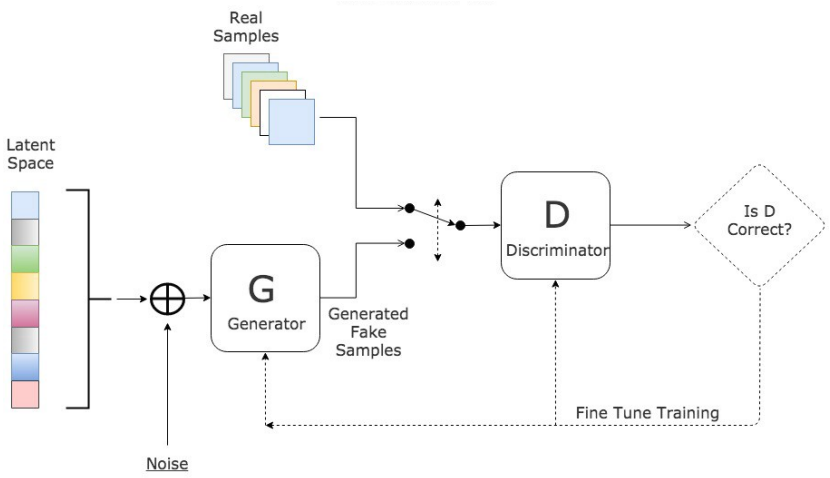

Figure 1: The structure of GAN.

an iterative denoising process grounded in diffusion processes. Each model offers unique advantages and challenges, which will be explored in this study.

1.1 Generative Adversarial Networks

Generative Adversarial Networks (GANs) consist of two core components: the generator G and the discriminator D. The generator $G(z)$ maps a random noise vector z to the data space, generating synthetic data samples. The discriminator $D(x)$, on the other hand, evaluates these samples to distinguish between real data from the training set and data generated by G.

Mathematically, the objective of GANs is to solve a minimax game between G and D, formalized as:

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

Here, $\mathbb{E}$ denotes the expectation, $p_{data}(x)$ represents the real data distribution, and $p_z(z)$ is the prior distribution from which the noise vector z is sampled. The generator aims to produce data that maximizes the probability of the discriminator making a mistake, while the discriminator strives to accurately classify real and fake samples (Goodfellow et al., 2020). The structure of GAN is illustrated in Figure 1.

1.2 Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models (DDPMs) follow a fundamentally different approach. These models involve a forward process that gradually adds Gaussian noise to the data, producing noisy versions of the original data over

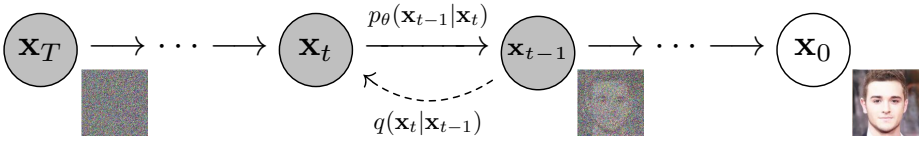

Figure 2: The structure of GAN.

T steps. This forward process can be described as a Markov chain where each step slightly perturbs the data (Khazrak et al., 2024).

The reverse process, which is the core of DDPMs, aims to iteratively denoise the data, step by step, to reconstruct the original data from the noisy input. The objective function of DDPMs is grounded in variational inference and aims to minimize the Kullback-Leibler (KL) divergence between the forward and reverse processes (Ho et al., 2020). The structure if DDPMs is illustrated in Figure 2.

The forward process is defined as:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \tag{2}$$

where β_t denotes the variance schedule, $\mathbf{I}$ is the identity matrix, and $\mathcal{N}$ indicates a normal distribution.

The reverse process then seeks to approximate:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \sigma_\theta(\mathbf{x}_t, t)\mathbf{I}), \tag{3}$$

with parameters θ optimized to minimize the variational lower bound:

$$L_{vlb} = \mathbf{E}_q[D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]. \tag{4}$$

The purpose of this study is to provide a comparative examination of GANs and DDPMs, focusing on their practical performance in generating high-quality bedroom image data. Specifically, this study will first implement GANs and DDPMs models and train them on the LSUN bedroom dataset (Yu et al., 2015) with a resolution of 256 $\times$ 256 pixels. Then, we analyze the efficiency and effectiveness in the training process and evaluate it in generation manner based on FID score. The study try to discern the strengths and limitations of each approach and offer insights into their application in various generative tasks.

2 Methods

2.1 Dataset

The Large-scale Scene Understanding (LSUN) Bedroom dataset (Yu et al., 2015) is a prominent collection used extensively in imagery and generative model research. The dataset comprises over 700 k high-resolution images of bedrooms, providing a diverse set of indoor scenes. Each image in the dataset comes with a resolution of 256 $\times$ 256 pixels, which allows for detailed and high-fidelity

training of image generation models. In our experiments, we randomly scale down the 700k dataset to 100k, as we believe that the diversity of the 100k dataset is sufficient for our analysis.

2.2 Modeling

2.2.1 Training Details for GAN

GAN are trained using a generator and a discriminator, both of which exhibit distinct architectural details. Our GAN employs a Deep Convolutional GAN (DCGAN) structure, consisting of several convolutional layers, batch normalization, and Leaky ReLU activation functions for the discriminator. The generator, conversely, makes use of transposed convolutions (also known as fractionally-strided convolutions) to upsample from a lower-dimensional latent space to the desired image resolution.

For training, the hyperparameters utilized are shown in Table 1. The learning rate is set to 2×10^{-4}, with a batch size of 128 and 500 training epochs. The generator and discriminator feature maps are set to 64 and 64, respectively, with an image size of 256×256 pixels.

Table 1: Key Hyperparameters for Traning GAN

Parameter	**Value**
Image size	256×256
Batch size	128
Latent vector size	100
Learning rate	2×10^{-4}
Training steps	20k
GPUs	$4 \times$ NVIDIA A100 (80GB)

2.2.2 Training Details for DDPM

DDPM follow a markedly different setup, involving a forward diffusion process that progressively adds noise to the training data and a reverse denoising process that attempts to reconstruct the original data. The architecture of our DDPM consists of a U-Net-based model, which employs attention mechanisms at multiple scales to refine the details of the generated images. Specifically, the model architecture employs a U-Net with 256×256 image resolution and 128 base channels, utilizing 2 residual blocks per resolution level. Self-attention layers are incorporated at 16×16 resolution with 1 attention head. The diffusion process follows a 1000-step linear noise schedule, with learnable variance prediction. Training uses a batch size of 128 and learning rate of 2×10^{-5}, optimized without scale-shift normalization in both model and diffusion components. A summary of the key hyperparameters for training the DDPM is provided in Table 2.

Table 2: Key Hyperparameters for Training DDPM

Parameter	**Value**
Image size	256×256
Diffusion steps	1000
Learning rate	2×10^{-5}
Batch size	128
Residual blocks per level	2
Attention resolutions	16×16
Attention heads	1
Training steps	100k
GPUs	4 × NVIDIA A100 (80GB)

2.3 Evaluation Metric

Evaluation of generative model performance relies heavily on the Fréchet Inception Distance (FID) score, which quantifies the similarity between real and generated images in terms of their feature representations. Lower FID scores indicate higher fidelity and realism in the generated images.

The FID score is calculated using the formula:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}),$$

where (μ_r, Σ_r) represent the mean and covariance of the feature embeddings of the real images, and (μ_g, Σ_g) represent those of the generated images.

In this study, we sample 1000 generated samples after training both GAN and DDPM models and evaluate based on the FID scores.

3 Experimental Results

With the training of both GAN and DDPM models completed, we proceed to analyze the loss during training, and evaluate their performance based on the FID score. The loss during the training process is presented at Figure 3 for GAN and Figure 4 for DDPM. The results of the FID on 1k images are presented Table 3, highlighting the efficiency and effectiveness of each model in generating high-quality images. Besieds, the generated images by DDPM and GAN are shown in Figure 5 and Figure 6 respectively.

The loss curves for GAN training, as shown in Figure 4, depict the behavior of both the discriminator loss (D loss) and the generator loss (G loss) over the training steps. Initially, the generator loss is high, while the discriminator loss fluctuates significantly, indicating the adversarial nature of training where the two networks compete against each other. Over time, the generator loss stabilizes but remains relatively high, suggesting that the generator continuously struggles to produce realistic images that can deceive the discriminator.

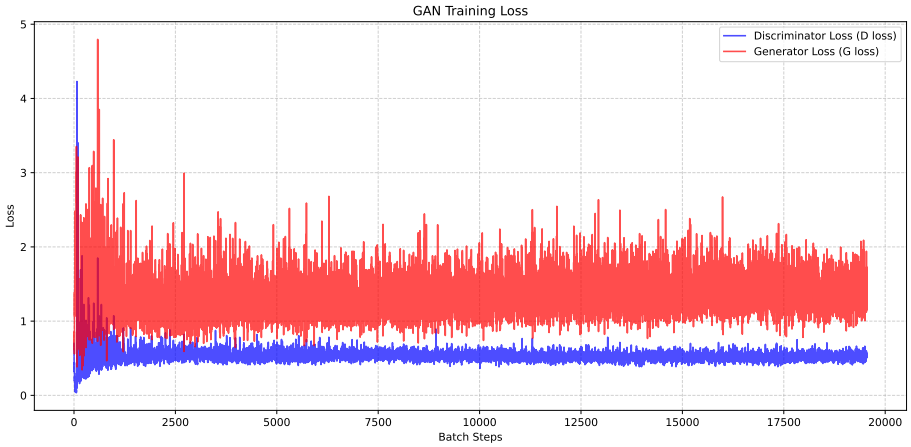

Figure 3: The loss in GAN's training process.

Meanwhile, the discriminator loss converges to a lower value, indicating that it maintains a relatively strong ability to distinguish real from fake samples. However, the presence of oscillations, especially in the generator loss, suggests instability in training, which is a common challenge in GANs due to the adversarial training dynamics and potential mode collapse.

The loss curves for DDPM training, as illustrated in Figure 4, exhibit a markedly different pattern. The total loss, shown on a logarithmic scale, decreases smoothly and consistently over the training steps, indicating stable and gradual convergence. Unlike GANs, where the loss oscillates due to adversarial interactions, DDPM follows a denoising process where the model learns to iteratively refine noisy images. The lower-right plot shows the Mean Squared Error (MSE) loss, which also follows a downward trend, reinforcing the stability of the training process. Additionally, the gradient norm stabilizes over time, suggesting that the model's updates remain controlled, avoiding issues such as vanishing or exploding gradients.

Table 3: FID Scores with Confidence Intervals on 1000 Sampling Images

Model	**FID Score**
GAN	16.2 ± 0.5
DDPM	8.5 ± 0.5

The FID scores, presented in Table 3, provide a quantitative measure of the quality of images generated by the GAN and DDPM models. The DDPM model achieves an FID score of 8.5, which is significantly lower than the 16.2 obtained by the GAN. This suggests that the DDPM model produces images that more closely resemble the real data distribution compared to those generated by the

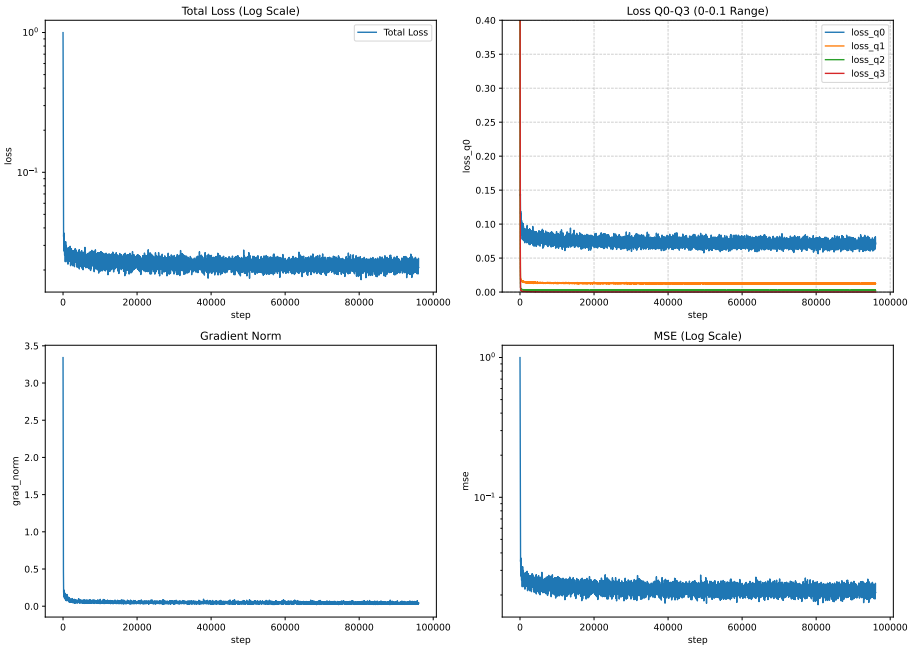

Figure 4: The loss in DDPM's training process.

GAN.

The superior performance of DDPM can be attributed to its iterative denoising process, which enables finer control over image generation and reduces artifacts commonly observed in GAN-generated samples. In contrast, GAN training, despite its efficiency in generating images in a single pass, often suffers from instability and mode collapse, leading to lower sample diversity and potentially degraded image quality. The relatively higher FID score of GAN indicates that the generated images may exhibit noticeable differences from the real data distribution. However, during inference, we observed that when sampling 1,000 samples, GAN required significantly less time compared to DDPM.

Overall, these results highlight a key trade-off between the two generative models: while GANs are typically faster in inference due to their direct generation process, DDPMs demonstrate a more stable training process and superior image quality, albeit at the cost of increased computational complexity.

From the generated samples (Figure 5 and Figure 6), it is evident that the images produced by DDPM are of significantly higher quality compared to those generated by GAN, which aligns with the conclusions drawn from the FID scores. In the case of GAN, Figure 6 (a) represents real samples, while Figure 6 (b) displays the corresponding generated images. Although the generated images capture the overall style of the real data distribution, they exhibit considerable inconsistencies in fine details. Objects such as beds, nightstands,

Figure 5: Samples from DDPM on LSUN bedroom datasets.

tables, and pillows appear disorganized and lack clear structural definition, indicating the model's struggle to accurately reproduce intricate features.

In contrast, the samples generated by DDPM demonstrate substantial improvements over those of GAN. Firstly, the overall room aesthetics and style closely resemble the real data distribution, ensuring higher visual fidelity. Secondly, most objects within the generated images are well-preserved with distinct structural integrity, enhancing the realism of the synthesized scenes. However, minor imperfections remain, particularly in the transitions between different objects, such as the connection between beds and pillows or between beds and the floor, where certain details are not fully reconstructed.

For future work, further refinements can be explored by modifying the loss function to enhance the accuracy of these object boundaries and improve the seamless integration of different elements within the generated images (Go et al., 2024).

4 Discussion

The detailed comparative analysis reveals several factors contributing to the disparity in performance. Firstly, the sequential denoising process in DDPMs permits finer control over the generated image details, whereas GANs may sometimes suffer from mode collapse, leading to less diverse and lower fidelity outputs. Secondly, the incorporation of attention mechanisms in DDPMs allows for better handling of global and local image features, which may explain their enhanced performance in terms of FID.

(a) Real Images (b) Generated Images

Figure 6: The real images and generated images by GAN.

4.1 Implications of Results

The superior performance of DDPMs over GANs can be attributed to several critical factors. Primarily, the inherent robustness of DDPMs, which stems from their iterative refinement process, allows them to progressively generate high-fidelity images. This iterative process, contrasted with the adversarial training methodology of GANs, appears to mitigate issues such as mode collapse and training instability commonly associated with GANs (Huang et al., 2024). The implications of these findings are profound, suggesting that future research and development in the field of generative modeling might increasingly favor diffusion models. The demonstrated effectiveness of DDPMs could inspire a shift toward methodologies that leverage structured iterative refinement, potentially leading to more advanced and reliable generative models.

4.2 Additional Thoughts

Reflecting on the comparative strengths and limitations of GANs and DDPMs, several key points emerge. GANs, with their competitive and adversarial training framework, have historically set the benchmark for image generation. However, their susceptibility to issues like mode collapse and sensitivity to hyperparameters limits their robustness (Khazrak et al., 2024). On the other hand, DDPMs, by virtue of their probabilistic and incremental approach, exhibit stronger robustness and consistently high performance in image fidelity metrics. Despite this, DDPMs often demand more computational resources and longer training times, underscoring the need for optimization in their application (Kuznedelev et al., 2024). Exploring avenues for improving the efficiency and scalability of DDPMs presents a promising area for future research. Additionally, this comparison raises important questions regarding the potential

hybridization of GANs and DDPMs, harnessing the strengths of both models to overcome their individual limitations.

5 Conclusion

In this study, we conducted a comparison of GANs and DDPMs, focusing on their performance in generating high-quality images from the LSUN bedroom dataset. Our findings highlight the superior generative capabilities of DDPMs, as evidenced by their lower FID scores and higher visual fidelity in generated samples. While GANs offer faster inference and simpler architectures, their training instability and susceptibility to mode collapse limit their effectiveness. In contrast, DDPMs demonstrate robust training dynamics and produce images with greater realism, albeit at the cost of increased computational complexity. These results underscore the potential of diffusion models as a powerful alternative to GANs for high-fidelity image generation tasks. Future research could explore optimizing DDPMs for efficiency and investigating hybrid approaches that combine the strengths of both models.

References

Go, Y., Torbunov, D., Rinn, T., Huang, Y., Yu, H., Viren, B., Lin, M., Ren, Y., and Huang, J. (2024). Effectiveness of denoising diffusion probabilistic models for fast and high-fidelity whole-event simulation in high-energy heavy-ion experiments. *arXiv preprint arXiv:2406.01602*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851.

Huang, N., Gokaslan, A., Kuleshov, V., and Tompkin, J. (2024). The gan is dead; long live the gan! a modern gan baseline. *Advances in Neural Information Processing Systems*, 37:44177–44215.

Khazrak, I., Takhirova, S., Rezaee, M. M., Yadollahi, M., Green II, R. C., and Niu, S. (2024). Addressing small and imbalanced medical image datasets using generative models: A comparative study of ddpm and pggans with random and greedy k sampling. *arXiv preprint arXiv:2412.12532*.

Kuznedelev, D., Startsev, V., Shlenskii, D., and Kastryulin, S. (2024). Does diffusion beat gan in image super resolution? *arXiv preprint arXiv:2405.17261*.

Yu, F., Seff, A., Zhang, Y., Song, S., Funkhouser, T., and Xiao, J. (2015). Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*.