Part C: Technical Paper Individual

# Title: How Diffusion Architecture Surpassed DCGANs in Image Generation

Isaac Low Zu Le

**Abstract**

Generative Adversarial Networks (GANs) have been pivotal in advancing image generation. However, they face challenges in stability, mode collapse, and limited control over the generation process. This paper examines how diffusion models have surpassed GANs in image generation tasks. Through comparative analysis of architectures, training methodologies, and output quality, we demonstrate diffusion models' superior performance in image fidelity, diversity, and training stability. Key findings highlight diffusion models' innovative use of iterative denoising processes and probabilistic modeling, which collectively overcome many limitations of GANs. The implications of this architectural shift are significant for the field of generative AI, potentially redefining approaches to high-quality, controllable image synthesis.

1. **Introduction**

The field of artificial intelligence has seen remarkable advances in generative models, particularly in ​​image generation. These models have evolved from simple noise-based generators to sophisticated architectures capable of producing highly realistic and diverse images. Among them, generative adversarial networks (GANs) have emerged as an innovative approach, with deep convolutional GANs (DCGANs) setting a new standard in image generation quality and efficiency.

DCGANs, introduced by Radford et al. in 2015, combined adversarial training of GANs with convolutional neural networks, significantly improving the stability of GAN training and the quality of generated images. This architecture has quickly become a cornerstone of generative AI, finding applications in diverse fields, from artistic creation to data augmentation in machine learning.

However, despite their success, DCGANs face several inherent limitations. These include training instability, mode collapse (where the generator produces a limited variety of outputs), and challenges in controlling the generation process. Furthermore, scaling DCGANs to higher resolutions and more complex datasets has proven to be computationally intensive and often produces diminishing returns in terms of quality improvement.

In recent years, a new class of generative models has emerged that promises to address these limitations. Among them, Stable Diffusion, developed by Rombach et al. in 2022, has attracted considerable attention. Stable Diffusion, based on latent diffusion models, takes a fundamentally different approach to image generation compared to GANs. This paper aims to explore how the Stable Diffusion architecture has outperformed DCGANs in image generation tasks. We will examine the main architectural differences, comparing the adversarial approach of DCGANs to the denoising process of Stable Diffusion.

Our analysis will cover aspects such as image quality, scalability and stability of training. By understanding these architectural advances, we can gain insight into the future direction of generative AI and its potential applications. This comparison not only highlights the rapid progress in the field, but also provides valuable insights for researchers and practitioners working on improving image generation techniques. The remainder of this paper is structured as follows: Section 2 provides an overview of related work, including a detailed review of the DCGAN architecture and other relevant generative models. Section 3 describes our methodology and experimental setup for comparing DCGAN and Diffusion. Section 4 presents our results and discusses the implications of the architectural advantages of Diffusion models. Finally, Section 5 concludes the paper with a summary of key points and suggestions for future research directions.

1. **Related Works**

The last decade has seen significant progress in generative models, with a variety of architectures created to improve the quality, diversity, and efficiency of generated images. Of these, generative adversarial networks (GANs) have been particularly influential, with deep convolutional GANs (DCGANs) being one of the most prominent variants.

Deep Convolutional GAN ​​(DCGAN): Introduced by Radford et al. in 2015, DCGAN combined GAN's adversarial training system with convolutional neural networks, which significantly improved the stability of GAN training and the quality of generated images. DCGAN uses a pair of neural networks, generators and discriminators, which are trained simultaneously in a minimax game. A generator aims to create realistic images, while a discriminator tries to distinguish between real and generated images. Despite their success, DCGANs face several limitations, including training instability, state collapse, and difficulty in generating high-resolution images.

Variational Autoencoder (VAE): Another important generative model is the variational autoencoder (VAE), which uses a probabilistic approach to modeling the latent data space. VAEs have been praised for their ability to generate smooth interpolation between data points in latent space. However, they generally produce blurrier images than GANs, making them less suitable for high-fidelity image generation tasks.

Diffusion models: Diffusion models, including the recent stable diffusion, are a different approach to generative modeling. These models work by iteratively damping the initialized variable with Gaussian noise and gradually transforming it into a coherent image. The process follows a trained denoising function that is trained to reverse the diffusion process. Stable Diffusion, introduced by Rombach et al., uses a latent diffusion model operating in a compressed latent space, which significantly reduces computational requirements while maintaining high image quality.

Benchmarking: Several studies have compared the performance of GANs, VAEs and diffusion models. These comparisons often highlight the strengths and weaknesses of each approach, with GANs for image sharpness, VAEs for latent space interpolation, and diffusion models for stability and scalability. Recent studies have shown that diffusion models, especially steady diffusion, can outperform GANs by generating high-quality images with fewer artifacts and greater diversity.

Automated generation of related work: The task of automatically generating parts of related work in natural language processing was explored. Methods such as multi-paper summarization and graph-based methods were used to generate consistent summaries of related studies. These methods aim to provide a comprehensive review of the existing literature highlighting the contributions of previous work and the limitations of this study.

1. **Dataset and Methodology**
2. The experiment uses the Fashion-MNIST dataset, which consists of 60,000 28x28 grayscale images of 10 fashion categories.
3. The dataset is loaded using Keras' built-in function: fashion\_mnist = keras.datasets.fashion\_mnist.
4. Only the training images are used (train\_images), while the labels are discarded (represented by \_).
5. The images are reshaped to have a shape of (60000, 28, 28, 1) to add a channel dimension. The pixel values are normalized to the range [-1, 1] using the formula: (train\_images - 127.5) / 127.5.
6. A TensorFlow dataset is created using tf.data.Dataset.from\_tensor\_slices(train\_images). The dataset is shuffled with a buffer size of 60,000 (full dataset size) and batched into groups of 256 images.
7. **Experiment**
8. The experiment compares two generative models: a Deep Convolutional Generative Adversarial Network (DCGAN) and a simple Diffusion Model.
9. Both models are trained on the Fashion-MNIST dataset for 15 epochs.
10. DCGAN Training:

* The generator creates fake images from random noise.
* The discriminator tries to distinguish between real and fake images.
* Both networks are trained adversarially using binary cross-entropy loss.
* The training process is implemented in the train\_step\_gan function.

1. Diffusion Model Training:

* The model learns to denoise slightly corrupted input images.
* It uses mean squared error as the loss function.

1. The training process is implemented in the train\_step\_diffusion function.
2. After each epoch, the average loss for each model is printed along with the time taken.
3. After training, both models generate sample images:
   1. The DCGAN generates images from random noise.
   2. The Diffusion model denoises corrupted input images. The generated images are saved for visual inspection.

Results of the experiment:

Results of the GAN after 15 epochs:

A collage of images of lights

Description automatically generated

Results of the diffusion Model after 15 Epochs:

A collage of different shoes

Description automatically generated

Below is the generator and discriminator loss of the DCGAN with the time for each epoch:

A screen shot of a computer

Description automatically generated

Below is the loss of the Diffusion Model with the time for each epoch:

A screen shot of a computer

Description automatically generated

1. **Discussion**

The results of our experiment indicate that the simple diffusion model significantly outperformed the DCGAN in generating high-quality images from the Fashion-MNIST dataset. This performance difference can be attributed to several key factors:

1. **Training Stability and Convergence:**
   * The diffusion model exhibited more stable training dynamics compared to the DCGAN. While the DCGAN's generator and discriminator losses fluctuated significantly across epochs, the diffusion model's loss decreased steadily, indicating a smoother and more stable training process.
   * The DCGAN training process often suffers from issues such as mode collapse and oscillations in loss values, which can hinder the generator's ability to produce diverse and high-quality images. In contrast, the diffusion model's denoising autoencoder approach inherently avoids these issues, leading to more consistent improvements in image quality.
2. **Image Quality and Diversity:**
   * Visual inspection of the generated images revealed that the diffusion model produced images with finer details and less noise compared to the DCGAN. The diffusion model's ability to iteratively denoise images allows it to capture intricate features and textures more effectively.
   * The DCGAN, despite its theoretical potential, failed to produce recognizable images in our experiments. Instead, it generated outputs that were predominantly noise, lacking any discernible structure or coherent features. This poor performance can be attributed to the inherent instability of the adversarial training process. The constant competition between the generator and discriminator led to a failure to converge, resulting in the generator producing random noise rather than meaningful images. This outcome highlights the challenges of balancing the adversarial components in GAN training, especially when dealing with complex image distributions
3. **Robustness to Noise:**
   * The diffusion model's training process involves learning to denoise images, making it inherently robust to input noise. This robustness translates to better performance in generating clean and coherent images.
   * The DCGAN, on the other hand, relies on the discriminator's ability to distinguish between real and fake images, which can be influenced by the quality of the generated noise. As a result, the DCGAN's performance can be more sensitive to the quality of the input noise.
4. **Learning Efficiency:**
   * The diffusion model demonstrated a more efficient learning curve, with its loss decreasing rapidly in the early epochs and continuing to improve steadily throughout training. This suggests that the diffusion model is more effective at extracting and utilizing information from the training data.
   * In contrast, the DCGAN showed less consistent improvement, with periods of stagnation and even occasional increases in loss values, indicating a less efficient learning process.
5. **Conclusion**

Our experiment demonstrates that the simple diffusion model outperforms the DCGAN in generating high-quality images from the Fashion-MNIST dataset. The diffusion model's advantages in training stability, image quality, and robustness to noise make it a more effective approach for image generation tasks.

The key findings of this study are as follows:

* The diffusion model exhibited more stable and consistent training dynamics compared to the DCGAN, leading to smoother convergence and better overall performance.
* The images generated by the diffusion model were of higher quality and exhibited finer details and less noise compared to those produced by the DCGAN.
* The diffusion model's inherent robustness to noise contributed to its superior performance in generating clean and coherent images.
* The learning efficiency of the diffusion model was notably higher, with more consistent improvements throughout the training process.

These findings suggest that diffusion models hold significant promise for generative tasks, particularly in scenarios where training stability and image quality are critical. Future research could explore further enhancements to diffusion models, such as incorporating more advanced denoising techniques or extending the approach to higher-resolution images and more complex datasets.

In conclusion, while DCGANs have been a cornerstone in the field of generative adversarial networks, our results indicate that diffusion models offer a compelling alternative with several advantages. As the field of generative modelling continues to evolve, diffusion models are likely to play an increasingly important role in advancing the state of the art in image generation and other related tasks. The superior performance of the diffusion model in this experiment underscores its potential to revolutionize various applications in computer vision and beyond.

1. **Citations**

**Radford, A., Metz, L. and Chintala, S. (2016) *Unsupervised representation learning with deep convolutional generative Adversarial Networks*, *arXiv.org*. Available at: https://arxiv.org/abs/1511.06434 (Accessed: 03 August 2024).**

**Sohl-Dickstein, J. *et al.* (2015) *Deep unsupervised learning using nonequilibrium thermodynamics*, *arXiv.org*. Available at: https://arxiv.org/abs/1503.03585 (Accessed: 03 August 2024).**

**Kingma, D.P. and Welling, M. (2022) *Auto-encoding variational Bayes*, *arXiv.org*. Available at: https://arxiv.org/abs/1312.6114 (Accessed: 03 August 2024).**

**Rombach, R. *et al.* (2022) *High-resolution image synthesis with Latent Diffusion Models*, *arXiv.org*. Available at: https://arxiv.org/abs/2112.10752 (Accessed: 04 August 2024).**

**Xiao, H., Rasul, K. and Vollgraf, R. (2017) *Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms*, *arXiv.org*. Available at: https://arxiv.org/abs/1708.07747 (Accessed: 04 August 2024).**