# Final Report: Enhancing DCGANs for Representation Learning and Image Generation

Yahav Cohen - 207261983
Avital Yoffe Vasiliev – 316444702

## Abstract

This project investigates targeted modifications to Deep Convolutional Generative Adversarial Networks (DCGANs), aiming to improve their stability, representation learning capabilities, and output quality under constrained computational resources. By integrating architectural changes, activation function replacements, alternative loss functions, and hyperparameter tuning, we aimed to evaluate their influence on two datasets: MNIST and CelebA. We demonstrate that hinge loss and self-attention mechanisms improve output sharpness and semantic coherence.

## Introduction

Generative Adversarial Networks (GANs), introduced by Goodfellow et al. (2014), have become a foundational framework in generative modelling. Despite their theoretical elegance, GANs are difficult to train due to issues such as vanishing gradients, mode collapse, and sensitivity to hyperparameters. Radford et al. (2016) proposed Deep Convolutional GANs, introducing a convolutional architecture combined with batch normalization and tailored activation functions (ReLU in the generator, LeakyReLU in the discriminator). DCGAN significantly improved training stability and demonstrated compelling results for unsupervised representation learning.
However, DCGAN still faces several practical limitations:
- Loss Function Instability: Binary cross-entropy (BCE) loss can saturate when the discriminator becomes overconfident, resulting in weak gradients for the generator.
- Lack of Global Context Modelling: Standard convolutional architectures struggle to capture long-range dependencies, which are vital for generating coherent and realistic images.
- Hyperparameter Sensitivity: GANs are highly sensitive to optimizer settings and learning rates, often requiring careful tuning for stable training.
In this project, we revisit the DCGAN architecture and propose several lightweight, theoretically motivated modifications that try to improve the issues described earlier.

## Methods

We utilized the publicly available DCGAN implementation and introduced several targeted modifications:

- Hyperparameter Tuning- Adjusted the Adam optimizer parameters by reducing momentum $\beta_1$ from 0.5 to 0.3 aiming to smooth out oscillations and

increasing the learning rate from 0.0001 to 0.0002 to encourage faster training and test the stability limits of the model.

- Loss Function Refinement- Replaced the BCE loss with hinge loss to address vanishing gradients and provide more consistent gradient feedback during training.
- Architectural Improvements-
    - A Self-Attention Layer was added to the generator to capture global spatial dependencies, improving consistency in image structure and symmetry.
    - ReLU was replaced with the Sigmoid Linear Unit (SiLU), offering smooth gradients and avoiding dead neurons.
    - We evaluated the effect of replacing BatchNorm with LayerNorm, inspired by its success in transformer-based architectures.

Experiments were conducted on MNIST and CelebA datasets, limiting training to five epochs per experiment (due to computational constraints). For MNIST, we first assessed the visual quality of generated '9' digits, then applied PCA and KMeans to cluster intermediate features. Performance was measured using silhouette score (separation) and compactness (cohesion). For CelebA, we evaluated results primarily through visual inspection, focusing on facial symmetry, realism, and coherence. In addition, we computed the Fréchet Inception Distance (FID), which compares the distribution of generated and real images using features from a pre-trained classifier.
All code modifications and experiments are documented and available here - **https://github.com/TaliVas/DGN-Final-Project**

# Results

### MNIST

| Model Variant | Silhouette Score | Compactness |
|---|---|---|
| Original DCGAN | 0.0615 | 0.1269 |
| Reduced Momentum ($\beta_1$=0.3) | 0.0674 | 0.1693 |
| Increased Learning Rate (0.0002) | 0.0943 | 0.0030 |
| Hinge Loss | 0.1074 | 0.0049 |

*Table 1:Clustering performance on MNIST for different DCGAN configurations.*

The original DCGAN served as a baseline, producing digits that were recognizable but blurry, and showed weak clustering performance.
Reducing the momentum parameter $\beta_1$ to 0.3 slightly stabilized training, resulting in clearer digit shapes and better-separated clusters. However, the increase in compactness suggests more dispersed points within each cluster.

Increasing the learning rate to 0.0002 yielded excellent numerical clustering metrics, especially in compactness, but caused a mode collapse, a phenomenon where the generator maps many latent vectors to a small set of outputs, the generated images were unrecongizable.
Replacing BCE with hinge loss led to the best overall clustering results. Generated digits appeared sharp and clear, but some reduction in variability suggested latent space over-compression.

**CelebA**

| Model Variant | FID Score |
|---|---|
| Original DCGAN | 128.01 |
| Self-Attention Layer | 139.96 |
| SiLU Activation Function | 237.24 |
| LayerNorm | 265.96 |

*Table 2: CelebA – FID Scores for Different DCGAN Model Variants*

The original DCGAN produced faces with basic structure, but many images were noisy or distorted.
Introducing a self-attention layer, inspired by SAGAN (Zhang et al., 2019), enhanced global spatial consistency and facial symmetry. Although the FID increased slightly, this likely reflects reduced sample diversity, a known limitation of FID when outputs become more consistent but less varied.
Replacing ReLU with SiLU, a smooth non-monotonic activation, yielded more coherent facial features but backgrounds were inconsistent, leading to a substantially higher FID. This suggests that SiLU helps with fine details but may hurt learning the overall image structure.
Swapping BatchNorm with LayerNorm significantly degraded output quality. Images lacked structure and appeared noisy. LayerNorm is less suited to convolutional architectures due to its disruption of spatial inductive bias, a fact reflected in its high FID score of 265.96.

# Conclusion

Our project set out to revisit the DCGAN framework with a critical lens, examining whether small, theoretically motivated modifications could improve training stability and output quality. By systematically testing adjustments, we found that several of these changes led to meaningful differences, both positive and negative. In the MNIST experiments, reducing $\beta_1$ and using hinge loss improved class separation and diversity, as reflected in silhouette and compactness scores. Notably, hinge loss produced more distinct clusters in KMeans space, suggesting better latent structure, though with some class overlap. Increasing the learning rate improved these metrics as well, but led to a collapse in image quality, likely due to unstable generator updates. Self-attention improved CelebA generation considerably, producing sharper and more coherent faces. SiLU activations resulted in more realistic facial structure

but slightly degraded lighting and texture, while LayerNorm consistently underperformed. The contradiction between FID and visual coherence in some experiments points to the nuanced nature of generative model evaluation. FID, while popular, may not fully reflect human-perceived quality, especially when diversity is sacrificed for fidelity.

Despite promising qualitative results, we were constrained by limited computational resources, training each model for only five epochs. This likely prevented the models from reaching their full potential. In GAN literature, it is widely acknowledged that training typically requires at least 100 epochs for convergence (Radford et al., 2016; Kurach et al., 2019). Given these limitations, our experiments focused more on trends and relative improvements than on achieving state-of-the-art visuals.

Ultimately, our results show that even with lightweight interventions and short training times, it is possible to influence GAN behavior in meaningful ways. However, the full impact of these changes can only be assessed with longer training and more robust evaluation metrics.

## Bibliography


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. Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv:1511.06434*.
3. Kurach, K., Lucic, M., Zhai, X., Michalski, M., & Gelly, S. (2019). A Large-Scale Study on Regularization and Normalization in GANs. *arXiv:1807.04720*.
4. Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2019). Self-Attention Generative Adversarial Networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 97, 7354–7363.

# Appendix

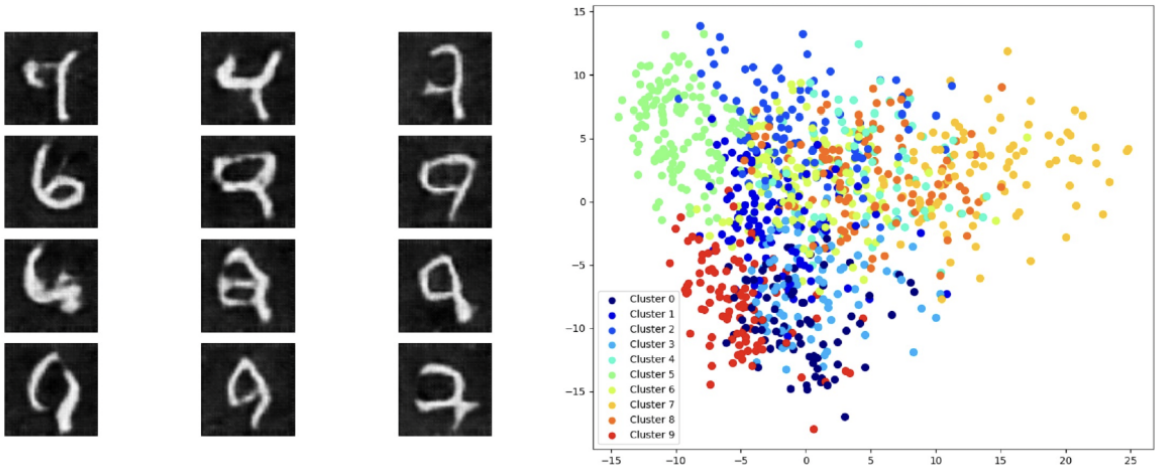


*Figure 1: Results for DCGAN with $\beta_1 = 0.3$. Left: Generated digit '9' samples. Right: Clustering Results.*

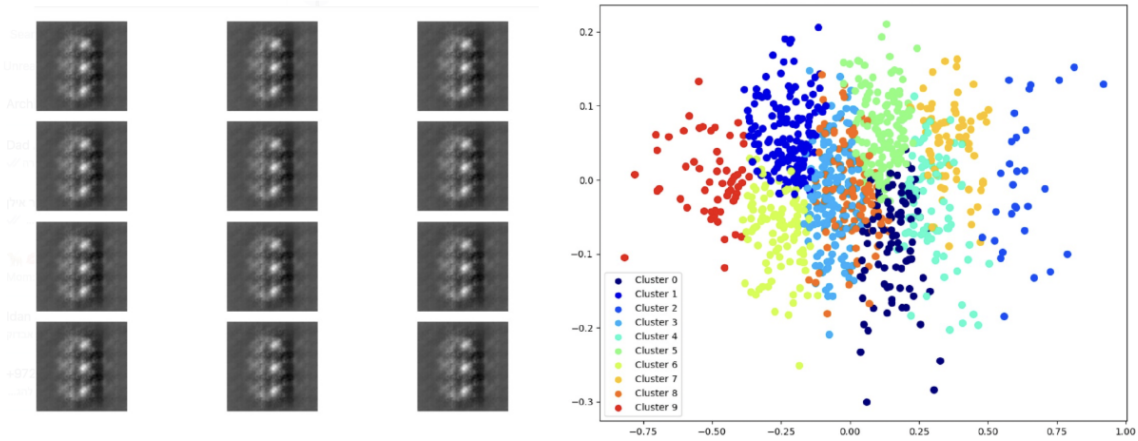


*Figure 2: Results for DCGAN with an increased learning rate (0.0002). Left: Generated digit '9' samples. Right: Clustering Results.*

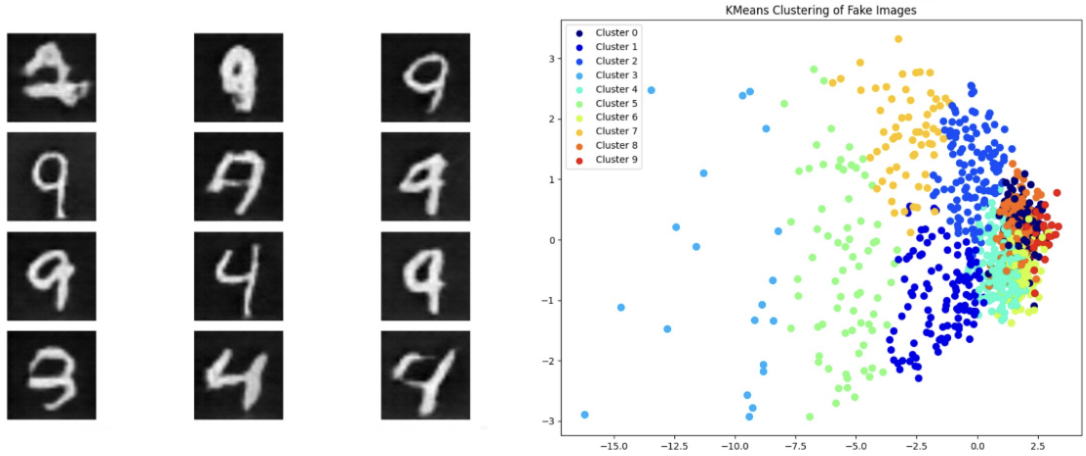


*Figure 3: Results for DCGAN with hinge loss. Left: Generated digit '9 samples'. Right: Clustering Results.*

*Figure 4: Generated CelebA face samples from four different model configurations. Each row corresponds to a different variant of DCGAN: (top to bottom) the original DCGAN, DCGAN with self-attention, DCGAN with SiLU activation, and DCGAN with LayerNorm.*