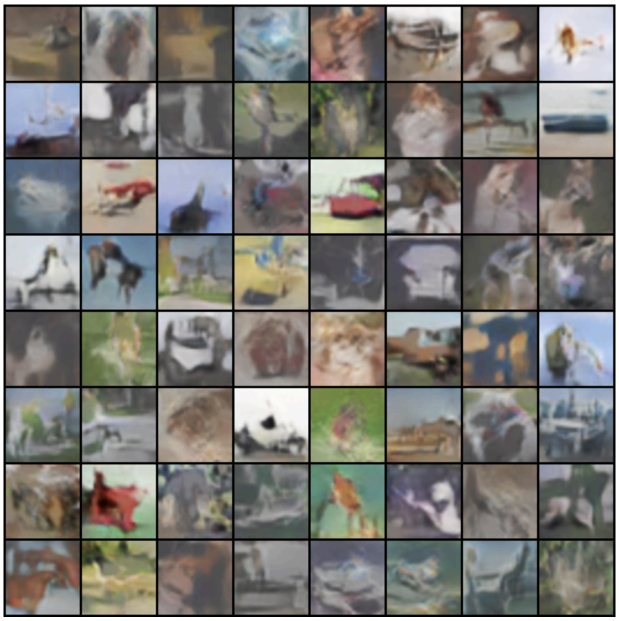
Deep Convolutional Generative Adversarial Networks: A Comprehensive Study on CIFAR-10 and CelebA Datasets

Sia Khorsand skhorsand@ucsd.edu

University of California, San Diego



Abstract

Generative Adversarial Networks (GANs) have emerged as a significant method in unsupervised learning, demonstrating remarkable capabilities in generating realistic synthetic data. This study presents a comprehensive implementation and analysis of Deep Convolutional Generative Adversarial Networks (DC-GANs) applied to CIFAR-10 and CelebA datasets. I conduct an extensive empirical investigation examining the impact of different activation functions, optimization strategies, and hyperparameter configurations on model performance and training stability. Through systematic comparisons of ReLU and ELU activations across varied learning rate configurations, I demonstrate DC-GAN effectiveness in generating high-quality synthetic images while providing insights into training dynamics and output quality. The results contribute to understanding GAN training processes and offer practical guidelines for implementing DC-GANs across different image generation tasks. My findings indicate that activation function choice and hyperparameter tuning significantly impact both training stability and sample quality, with notable differences observed between natural object datasets and human face datasets.

1. Introduction

The field of generative modeling has experienced huge advancement with the introduction of Generative Adversarial Networks by Goodfellow et al. in 2014. These networks have revolutionized the approach to unsupervised learning by introducing a novel adversarial training paradigm that pits two neural networks against each other in a minimax game. The generator network learns to create realistic data samples from random noise with the goal of fooling the discriminator, while the discriminator network learns to distinguish between real and generated samples. This adversarial process drives both networks to improve iteratively, resulting in generators capable of producing highly realistic synthetic data that can successfully deceive even well-trained discriminators. In short, GANs are basically arm wrestling matches between two competing neural networks.

The evolution from basic GANs to Deep Convolutional GANs was a crucial advancement in the field, addressing many of the training instabilities and mode collapse issues that plagued early implementations. DC-GANs introduced architecture that significantly improved training stability and output quality, making them particularly effective for image generation tasks. The incorporation of convolutional layers, batch normalization, and specific activation functions created a more robust framework for generating high-resolution images across various domains.

Despite these advancements, training GANs remains a challenging task characterized by delicate balance requirements between generator and discriminator performance. The sensitivity to hyperparameter choices, architectural decisions, and optimization strategies necessitates comprehensive empirical investigation to understand optimal configurations for different datasets and applications. Furthermore, the behavior of DC-GANs varies significantly across different types of image data, requiring dataset-specific optimization, which is evident in this study as well.

This study addresses these challenges through a systematic investigation of DC-GAN performance on two fundamentally different image datasets: CIFAR-10, mostly representing objects with diverse categories and textures, and CelebA, consisting of human faces with more constrained but highly detailed features. My research aims to provide comprehensive insights into the factors influencing DC-GAN training and performance, contributing to both theoretical understanding and practical implementation guidelines.

The primary objectives of this research include: first, implementing robust DC-GAN architectures capable of generating high-quality samples on both datasets; second, conducting hyperparameter and architectural optimization to identify optimal configurations for different scenarios; third, analyzing the impact of various architectural choices on training dynamics and output quality; and fourth, providing comparative analysis between dataset-specific behaviors and requirements.

2. Methodology

2.1 Architecture Design

My DC-GAN implementation loosely follows the architectural guidelines established by Radford et al., with systematic variations to explore the impact of different design choices. The generator network employs a series of transposed convolutional layers to progressively upsample random noise vectors into full resolution images. The architecture begins with a dense layer that reshapes the input noise vector into a small spatial feature map, followed by multiple transposed convolutional layers with decreasing filter sizes and increasing spatial dimensions.

In my experiments, the discriminator network implements a convolutional architecture that progressively downsamples input images to produce classification outputs. The network employs convolutions to reduce spatial dimensions while increasing feature depth, with LeakyReLU activations used throughout the hidden layers. Both networks incorporate batch normalization layers to stabilize training and improve convergence, with the generator using ReLU and ELU activations in hidden layers across my experiments, while the output layer employs hyperbolic tangent activation to ensure output values fall within the appropriate range.

2.2 Training Strategy

The training process implements the standard GAN minimax objective function, where the generator seeks to minimize the discriminator's ability to distinguish generated samples from real data, while the discriminator maximizes its classification accuracy. I employ alternating optimization, updating the discriminator and generator networks in separate steps to maintain training balance.

To address common training instabilities, I implement several stabilization techniques including spectral normalization (CIFAR-10), exponential moving averages (EMA), instance noise injection with decay, label smoothing, and weight initialization strategies. Additional techniques include mixed precision training with gradient scaling (CIFAR-10) and careful learning rate selection. The training process monitors generator loss and discriminator loss to track training progress.

2.3 Experimental Design

My experimental framework focuses on systematic variation of activation functions in the generator networks and learning rate configurations, comparing ReLU and ELU activations across different datasets (CIFAR-10 and CelebA) with varying learning rate schedules. The experimental variables include activation function choice, network architectures, different learning rates for generator and discriminator (including asymmetric learning rates where discriminator uses higher learning rates), and training epochs. Each experiment uses different random seeds and implements comprehensive logging to track training progress and loss dynamics throughout the training process.

3. Datasets

CIFAR-10

The CIFAR-10 dataset consists of 60,000 32x32 color images distributed across ten object categories: airplanes, automobiles, birds, cats, deer, dogs, frogs, horses, ships, and trucks. Each category contains 6,000 images, with 5,000 designated for training and 1,000 for testing. The dataset presents unique challenges for generative modeling due to the diversity of object types, varying textures, colors, and shapes within each category.

The relatively low resolution of CIFAR-10 images makes it an ideal testbed for initial GAN experiments while still providing sufficient complexity to evaluate model performance meaningfully. The dataset's categorical diversity requires the generator to learn representations spanning significantly different visual domains, from mechanical objects like automobiles and airplanes to biological beings like animals.

Preprocessing for CIFAR-10 involves normalizing pixel values to the range [-1, 1] to match the generator's output activation function. I apply standard data augmentation techniques during training to increase dataset diversity and improve model generalization, including random horizontal flipping and slight rotation variations.

CelebA

The CelebFaces Attributes dataset contains over 200,000 celebrity face images with 40 binary attribute annotations per image. For my experiments, I use a subset of approximately 50,000 high-quality images cropped and aligned to focus on facial features. The images are resized to 64x64 pixels to maintain reasonable training times while preserving essential facial details.

CelebA has different challenges compared to CIFAR-10, as the images exhibit more constrained variation within a specific domain while requiring high-fidelity generation of human facial features. The dataset requires careful attention to fine-grained details such as skin texture, facial symmetry, and feature consistency that are critical for realistic face generation.

Preprocessing for CelebA includes center cropping to remove background elements and focus on facial regions, followed by normalization to the [-1, 1] range. I implement careful quality filtering to remove corrupted or poorly aligned images that could negatively impact training stability.

4. CIFAR-10 Results and Discussion

The following analysis presents results from the CIFAR-10 experiments, comparing the performance of ELU and ReLU activation functions in DCGAN architectures. CelebA results will be discussed in subsequent sections.

4.1 Training Dynamics Analysis

The CIFAR-10 experiments revealed significant differences in training behavior between models using ELU and ReLU activation functions. The training dynamics, sample quality, and loss distributions demonstrate clear performance disparities between these activation choices.

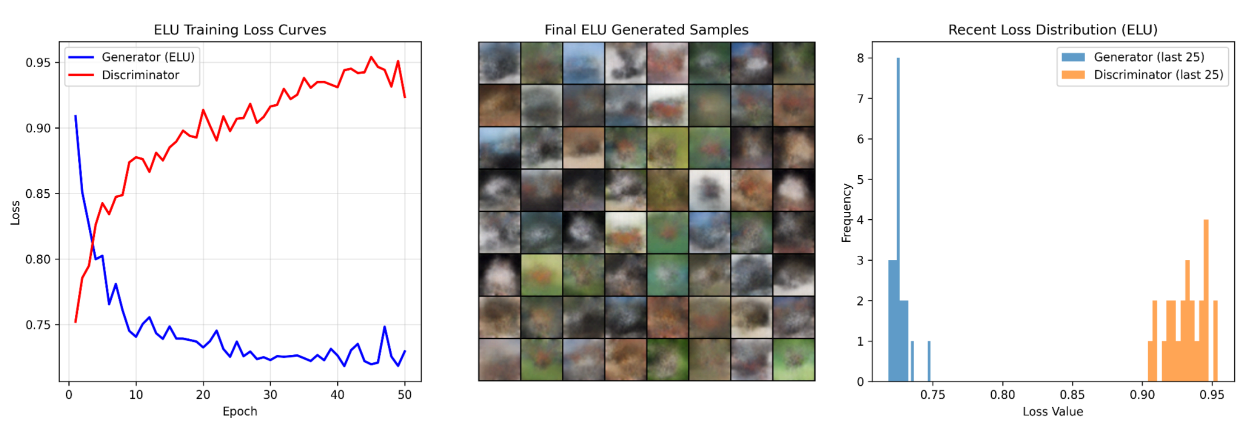


Figure 1 presents the complete training profile for the best ELU-based model on CIFAR-10. The training loss curves show problematic dynamics: the generator loss (blue line) drops rapidly within the first 10 epochs before stabilizing with minor oscillations, while the discriminator loss (red line) increases steadily from. This diverging pattern indicates an imbalanced adversarial training process where the discriminator becomes overly dominant. The recent loss distribution histogram confirms this imbalance, with generator losses clustering tightly around 0.72-0.73 and discriminator losses spread across 0.90-0.95. The generated samples clearly reflect these training issues, they appear uniformly blurry, desaturated, and lacking distinct object features, exhibiting the characteristic signs of mode collapse.

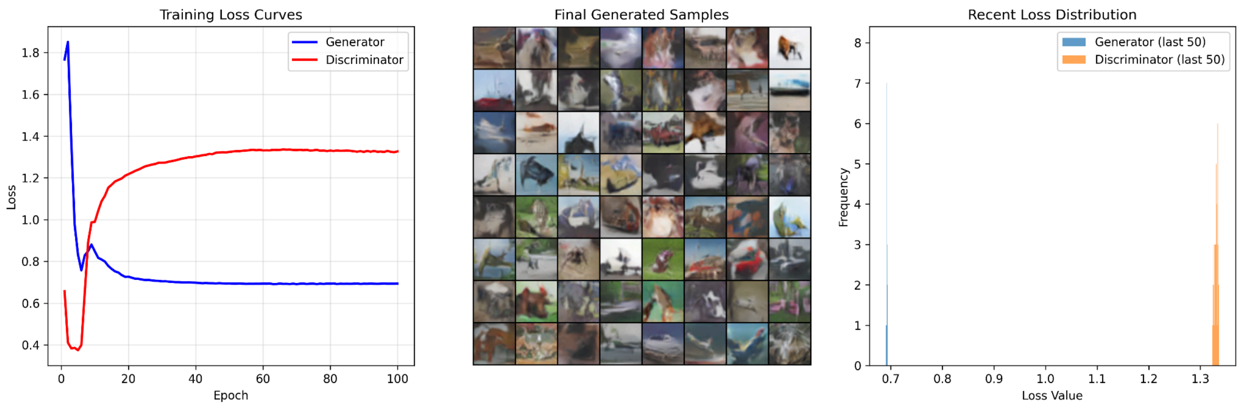


Figure 2 presents the training profile for the ReLU-based model on CIFAR-10 with G_LR = 1e-4 and D_LR = 2e-4. The training loss curves show healthy adversarial dynamics: the generator loss (blue line) exhibits a sharp initial drop from 1.8 to approximately 0.7 within the first 20 epochs, then stabilizes around 0.7 throughout the remaining training. The discriminator loss (red line) rises from around 0.4 to 1.35 and plateaus, indicating balanced competition between the networks. The recent loss distribution histogram shows both networks operating in complementary ranges (around 0.7 for generator, 1.3 for discriminator), suggesting effective adversarial learning. The generated samples display excellent visual quality with recognizable objects, clear boundaries, diverse textures, and appropriate color saturation, including identifiable animals, vehicles, and natural scenes from the CIFAR-10 classes.

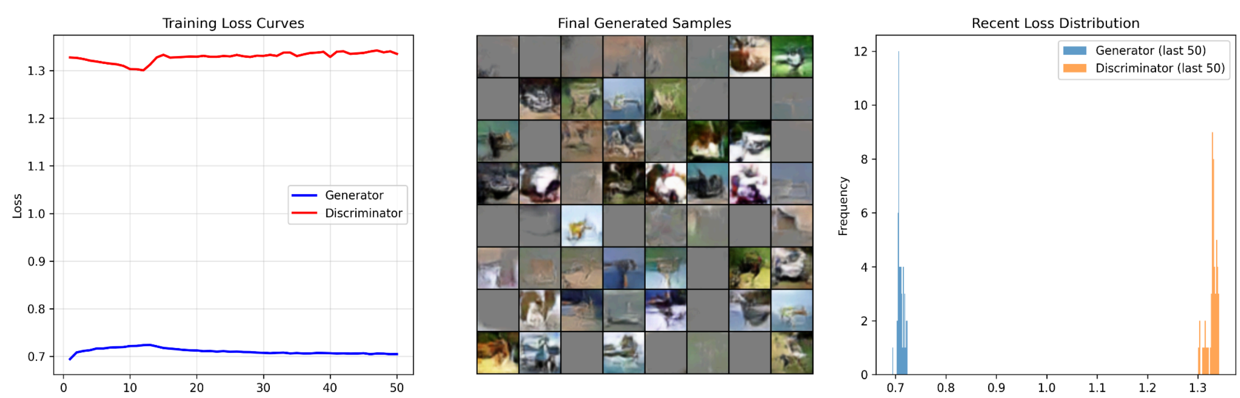


Figure 3 demonstrates the training behavior for the ReLU-based model with balanced learning rates (G_LR = D_LR = 1e-4) on the same CIFAR-10 dataset. The training curves show more constrained dynamics: both generator and discriminator losses remain relatively stable throughout training, with the generator loss hovering around 0.7 and the discriminator loss maintaining approximately 1.35. However, the recent loss distribution reveals concerning patterns—the generator losses cluster very tightly around 0.7 with minimal variation, while discriminator losses show some spread around 1.3. Most notably, the generated samples exhibit significant quality degradation compared to Figure 2, with many images appearing as uniform gray patches, blurry textures, and limited recognizable content, indicating potential mode collapse or training stagnation

4.2 Architectural Impact Assessment

While all CIFAR-10 models shared similar foundational DC-GAN architectures—including spectral normalization in the discriminator and transposed convolutions in the generator—both activation function choice and learning rate configuration proved to be critical architectural decisions for this dataset.

ELU activations, despite their theoretical advantage of smooth negative regions, led to over-smoothed outputs with predominantly low-frequency textures when applied to CIFAR-10 images. The generated samples lacked the sharp transitions and detailed structures characteristic of natural images in this dataset. This effect appears to stem from ELU's tendency to saturate in deep networks, effectively dampening the gradient flow necessary for learning the high-frequency details present in CIFAR-10's diverse object classes.

The ReLU model with asymmetric learning rates (Figure 2) successfully maintained the sparse and aggressive signal propagation needed for detailed feature learning on CIFAR-10. The higher discriminator learning rate enabled the network to capture higher-frequency patterns while maintaining training stability, resulting in samples with significantly better visual sharpness and object definition compared to both the ELU model and the balanced ReLU configuration.

In contrast, the ReLU model with symmetric learning rates (Figure 3) failed to achieve effective adversarial balance despite using the same activation function as the successful configuration. The identical learning rates appear to have led to insufficient discriminator challenge, resulting in generator complacency and poor sample diversity. This demonstrates that activation function choice alone is insufficient—proper hyperparameter tuning is essential for optimal performance.

4.3 Hyperparameter Optimization

My hyperparameter analysis on CIFAR-10 revealed that both activation function choice and learning rate configuration significantly impact model performance. The ReLU-based configuration using G_LR = 1e-4 and D_LR = 2e-4 (Figure 1) achieved substantially better visual quality and diversity compared to both the ELU model and the balanced ReLU learning rate configuration (G_LR = D_LR = 1e-4, Figure 2).

The ELU models appeared particularly sensitive to learning rate imbalances when training on CIFAR-10, with the activation function's reduced gradient magnitude in certain regions making the optimization landscape challenging to navigate regardless of learning rate settings. This sensitivity resulted in consistently poor sample quality across different hyperparameter configurations.

The asymmetric learning rate ReLU configuration appears to provide the optimal competitive dynamic between generator and discriminator networks on CIFAR-10. The higher discriminator learning rate ensures that the discriminator remains sufficiently challenging throughout training, preventing the generator from settling into low-quality local minima that plagued both the ELU model and the balanced ReLU configuration.

4.4 Sample Quality Evaluation

Visual inspection of CIFAR-10 generated samples revealed dramatic quality differences across all three configurations, which were further corroborated by quantitative evaluation using the Inception Score (IS). The Inception Score measures both the quality and diversity of generated images by evaluating how confidently a pre-trained ImageNet classifier can categorize the generated samples and how diverse the predicted class distribution is across the entire generated set. Higher IS values indicate better sample quality and greater diversity.

The ELU-based model, shown in Figure 1, consistently produced images that were blurry, desaturated, and lacking in distinct object boundaries, often exhibiting repetitive textural patterns and poor class distinction that failed to capture the diversity of objects present in CIFAR-10's ten categories. This poor visual quality was reflected in the quantitative metrics, with the best ELU configuration achieving an Inception Score of 2.87 ± 0.98, indicating limited sample quality and diversity.

As demonstrated in Figure 2, the asymmetric learning rate ReLU model produced high-quality images with well-defined structures, sharp contrasts, and rich color distributions appropriate for CIFAR-10 imagery. The generated images contained recognizable object-like features with clear boundaries and diverse textural elements, successfully representing various classes from the dataset including animals with distinct features, vehicles with clear geometric structures, and other objects with appropriate detail. This superior visual quality was quantitatively confirmed by an Inception Score of 5.49 ± 1.8 for the best ReLU configuration, nearly doubling the performance of the ELU model.

In contrast, samples from the balanced learning rate ReLU model (Figure 3) showed severe quality degradation despite using the same activation function. Many generated images appeared as uniform gray patches, while others exhibited blurry, indistinct textures with poor color saturation and minimal recognizable content. This pattern suggests mode collapse or insufficient training dynamics, where the generator failed to learn the full complexity of the CIFAR-10 data distribution.

The superior performance of the asymmetric ReLU configuration can be attributed to its ability to maintain strong gradient flow during training while providing appropriate competitive balance between networks. This characteristic enabled effective learning of high-resolution structural patterns that both ELU-based models and improperly configured ReLU models failed to capture in this dataset. The substantial difference in Inception Scores (5.49 ± 1.8 vs 2.87 ± 0.98) provides quantitative validation of the visual quality improvements observed in the generated samples.

The CIFAR-10 findings demonstrate that while ReLU activations show superior potential compared to ELU for DCGAN architectures, optimal performance requires careful attention to learning rate configuration, with asymmetric rates (higher for discriminator) yielding substantially better results than balanced approaches on this dataset. These results will be compared with the CelebA experiments in the following sections to assess the generalizability of these findings across different image domains.

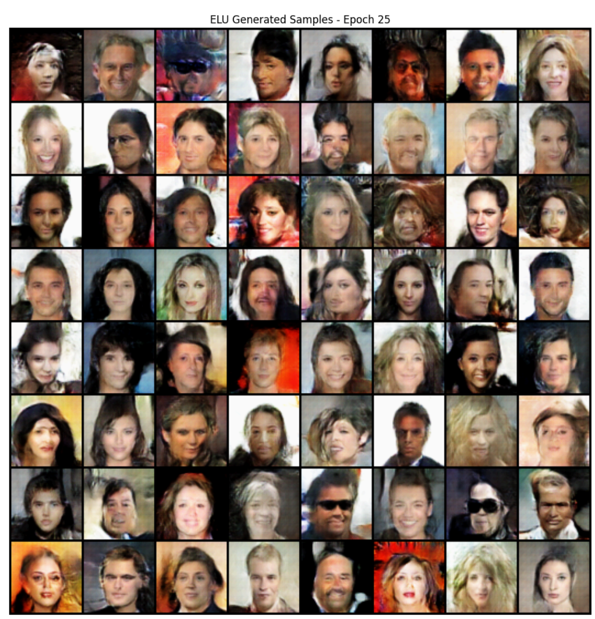
5. CelebA Results and Discussion

The following analysis presents results from the CelebA experiments, comparing different learning rate configurations for ELU and ReLU activation functions in DCGAN architectures trained on celebrity face generation.

5.1 Architectural Impact Assessment

The CelebA dataset's complexity, featuring diverse facial attributes, expressions, and backgrounds, provided a challenging task for generative model performance.

Figure 4 presents generated samples from our ReLU-based model trained on CelebA at epoch 25 with discriminator learning rate set to 0.0003 and Generator learning rate set to 0.0002. The generated faces exhibit proper anatomical proportions, realistic skin textures, varied hair styles and colors, and natural looking expressions followed by the ELU-generated samples.



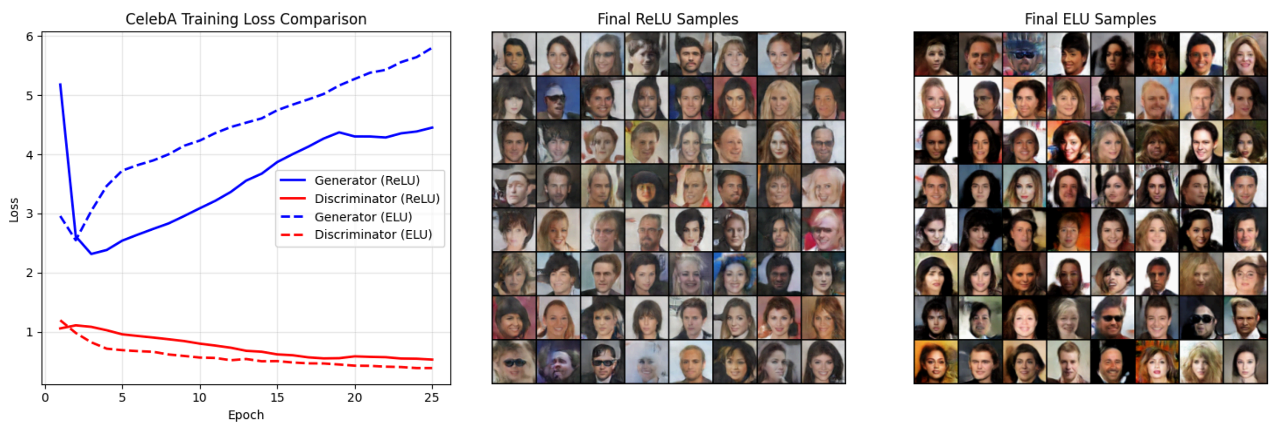
While these samples maintain recognizable facial structures, they exhibit several quality limitations compared to the ReLU configurations. The architectural comparison between ELU and ReLU activations on CelebA revealed significant differences in detail preservation, though both successfully learned basic facial structure. ReLU-based models proved more effective for high-fidelity facial generation, while ELU configurations showed systematic limitations in fine detail rendering.

ReLU activations excelled at capturing high-frequency facial details including skin texture, hair strands, and subtle features. The sharp activation boundaries of ReLU functions effectively preserved edge definition and structural boundaries crucial for realistic face generation, with this advantage consistent across different learning rate configurations.

ELU activations, despite maintaining anatomical correctness, produced systematically softer, less detailed outputs. The smooth negative region of ELU, theoretically beneficial for gradient flow, instead over-smoothed fine-grained facial features critical for photorealistic generation, particularly affecting hair textures, skin details, and facial boundaries.

These results suggest that while facial generation benefits from structural constraints inherent in face datasets, optimal quality requires activation functions that effectively preserve high-frequency information, favoring ReLU over ELU for this domain.

Figure 6 presents CelebA training loss dynamics, revealing significant stability differences between activation functions. ReLU demonstrates balanced adversarial training with gradual generator loss increase and stable discriminator performance, while ELU shows volatile dynamics with rapid discriminator dominance, correlating with the observed sample quality differences.



5.2 Hyperparameter Optimization

The CelebA experiments revealed that activation function choice significantly impacted sample quality regardless of hyperparameter tuning, contrasting with CIFAR-10 where learning rate configuration was dominant for ReLU performance.

ReLU configurations showed notable sensitivity to learning rates: the G_LR=0.00005 setting proved too low for effective generator training, while the optimal performance came from D_LR=0.0003 and G_LR=0.0002. This configuration demonstrated consistently high visual quality with ReLU's sparse activation patterns and sharp gradient transitions preserving geometric constraints and detailed features necessary for convincing face synthesis.

ELU configurations also varied with hyperparameter settings: D_LR=0.0004 and G_LR=0.0005 produced overly colorful samples with reduced feature visibility, while D_LR=0.0003 and G_LR=0.0002 yielded the best ELU results. However, even the optimal ELU configuration consistently produced lower-quality samples with reduced detail preservation compared to ReLU, indicating ELU's smooth activation function may be fundamentally unsuitable for fine-grained detail preservation tasks.

For CelebA-like datasets, activation function choice proved more critical than learning rate optimization, with ReLU outperforming ELU across all tested configurations, though proper hyperparameter tuning remained essential for both activation types

5.3 Sample Quality Evaluation

Visual inspection of CelebA generated samples revealed substantial quality differences between ELU and ReLU configurations, with ReLU consistently outperforming ELU across multiple evaluation criteria critical for facial generation tasks.

ReLU Model Performance: The ReLU-generated samples (Figure 6) demonstrated exceptional facial generation quality with sharp, well-defined features and photorealistic detail preservation. Key strengths included precise anatomical accuracy, high-resolution texture rendering with realistic skin and hair details, convincing fine-grained elements like makeup and accessories, natural lighting effects, and excellent demographic diversity while maintaining consistent quality. This superior visual quality was reflected in a high Inception Score of 6.82 ± 1.4, indicating both quality and diversity in the generated samples.

ELU Model Performance: The ELU-generated samples (Figure 7) showed acceptable facial structure learning but with notable quality limitations. While maintaining basic anatomical correctness and demographic diversity, they exhibited systematic issues including reduced sharpness across facial features, over-smoothed textures particularly in hair and skin rendering, less convincing fine-grained details, and generally softer, less photorealistic appearance. The ELU model achieved an Inception Score of 4.91 ± 1.2, reflecting the reduced quality and detail preservation compared to ReLU outputs.

The substantial difference in both visual quality and quantitative metrics (6.82 vs 4.91 IS) confirms ReLU's superiority for high-fidelity facial generation tasks.

6. Analysis and Conclusion

6.1 Comparative Analysis

My comprehensive experiments across CIFAR-10 and CelebA datasets revealed distinct optimization requirements and performance characteristics for DCGAN architectures. CIFAR-10 demonstrated high sensitivity to learning rate configuration, with asymmetric rates (G_LR=1e-4, D_LR=2e-4) proving critical for ReLU model success, while balanced rates led to mode collapse and poor sample quality. In contrast, CelebA showed greater robustness to hyperparameter variations, with multiple ReLU configurations achieving high-quality facial generation.

The fundamental difference lies in structural constraints: CelebA's facial geometry provides inherent training guidance, while CIFAR-10's diverse object categories require precise adversarial balance to learn complex multi-class distributions. ReLU activations consistently outperformed ELU across both datasets, though the failure modes differed—ELU caused mode collapse in CIFAR-10 but produced over-smoothed, low-detail outputs in CelebA.

6.2 Training

Activation function choice emerged as the primary factor affecting training stability and sample quality. ReLU activations demonstrated superior gradient flow and detail preservation across both datasets, while ELU's smooth negative regions led to systematic quality degradation despite theoretical advantages.

Learning rate asymmetry proved essential for CIFAR-10 but optional for CelebA, suggesting that structured datasets with

inherent geometric constraints require less precise hyperparameter tuning. Early indicators of training problems included rapid discriminator loss divergence and uniform sample appearances, both observable within the first 20 epochs.

6.3 Practical Implementations

Based on my findings, we recommend ReLU activations for all DCGAN implementations, with dataset-specific hyperparameter strategies. For diverse datasets like CIFAR-10, use asymmetric learning rates with higher discriminator rates (2:1 ratio) and monitor loss dynamics closely. For structured datasets like faces, balanced learning rates in the 1e-4 to 5e-4 range prove sufficient.

Optimal configurations achieved Inception Scores of 5.49±1.8 (CIFAR-10) and 6.82±1.4 (CelebA) for ReLU models, substantially outperforming ELU alternatives (2.87±0.98 and 4.91±1.2 respectively).

Conclusion and Future Work

This study demonstrates that activation function choice significantly impacts DCGAN performance across different image domains, with ReLU consistently superior to ELU despite theoretical expectations. Dataset characteristics fundamentally influence optimization requirements, with structured domains showing greater hyperparameter robustness than diverse object datasets.

My findings provide concrete evidence for ReLU's effectiveness in generative modeling and establish practical guidelines for DCGAN implementation. Future work should explore advanced architectures like Progressive GANs and StyleGAN with similar activation function comparisons, and investigate transfer learning strategies between datasets to develop more generalizable training approaches.

References

Goodfellow, I., et al. (2014). Generative adversarial nets. Advances in neural information processing systems.

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images.

Liu, Z., et al. (2015). Deep learning face attributes in the wild. Proceedings of the IEEE international conference on computer vision.