GENERATIVE SHOWDOWN: FROM GAN FAILURE TO DIFFUSION SUCCESS

Group 4 - LB09

Arieldhipta Tarliman - 2702234636
Cornelia Nathania Pramudita - 2702241603
Geraldus Jeremy Chandra - 2702368236
Vincenzo Jason Carter - 2702242700
Sherly Vaneza - 2702222163

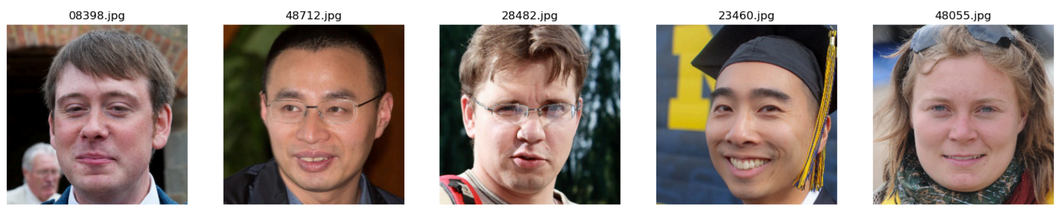
1 Introduction

The demand for **unique visual content** across diverse sectors often outpaces the traditional, resource-intensive methods of producing it, creating a barrier to innovation and tailored communication. Our **generative models**, especially the sophisticated **Stable Diffusion**, offer a **transformative solution** by enabling the rapid, on-demand creation of specific, high-quality imagery, effectively **democratizing visual content generation** and accelerating creative and technical workflows.

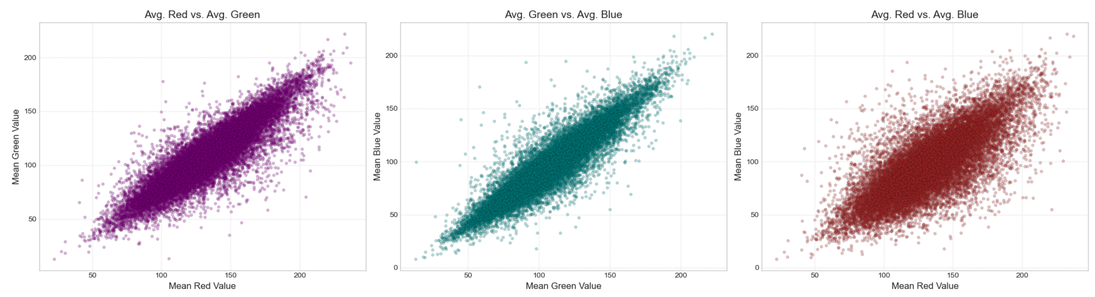
2 Data Preview

- **Dataset Source: 140,000** real and artificially generated facial images from Kaggle.
- **Images Used: 20,000 images**, subset of 50,000 real images from the training set.
- The images used are **256×256 pixels** with an **RGB** color type.

3 Exploratory Data Analysis



The **5 sample images** above show **real faces** with diverse characteristics (e.g., varying ages, genders, hairstyles, and accessories). These real images could serve as the training data or the **target distribution** for the model.



- From the color distribution (the purple, green, and red point clouds), it is apparent that there's a **strong positive linear correlation** between the average intensities of the paired **RGB color** channels.
- This indicates that as the intensity of one color channel increases, the intensity of the other channel in the pair generally **increases proportionally**.

4 Data Preprocessing

1. **Delete all potential blurry and grainy images**
2. **Normalize Pixel Values**
3. **Define Data Loader**
4. **Apply Batch, Shuffle, Prefetch**
5. **Example using Data Loader**

5 Models

DCGAN

Epoch: 0210 | **Learning Rate G:** 0.0002 ; **D:** 0.0001

Generator

- **Input:** Random noise vector
- **Core Layers:**
  - Multiple blocks of Conv2DTranspose
- **Output:** (256, 256, 3) image
- **Total Parameters:** ~19.17 million

Discriminator

- **Input:** (256, 256, 3) image
- **Core Layers:**
  - Multiple blocks of Conv2D
- **Output:** Probability value
- **Total Parameters:** ~17.49 million

Stable Diffusion Model

- **Input:**
  - Latent (N, 4, H/8, W/8)
  - promt (N, 77, 768)
  - time
- **Core Layers:**
  - Conv Layer
  - Resblock & Attention Layer
  - Downsampling & Upsampling
- **Output:** Latent (N, 4, H/8, W/8)
- **Total Parameters:** ~859 million

- **CLIP TEXT ENCODER**
  - Pretrained Model that build by Open AI that allow to connect text to image
- **VARIATIONAL AUTO ENCODER**
  - compress data & distributed data according to multivaraite distribution

Vanilla GAN

Epoch: 0573 | **Learning Rate:** 0.0001

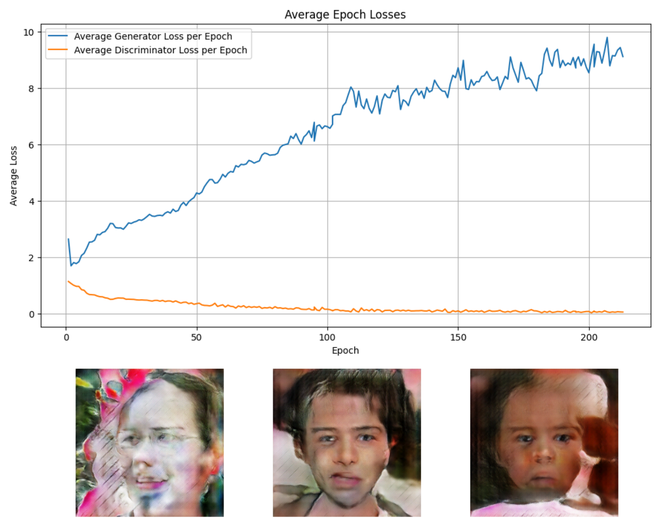
Generator

- **Input:** Random noise vector
- **Core Layers:**
  - Multiple blocks of Dense
- **Output:** (64, 64, 3) image
- **Total Parameters:** ~13.28 million

Discriminator

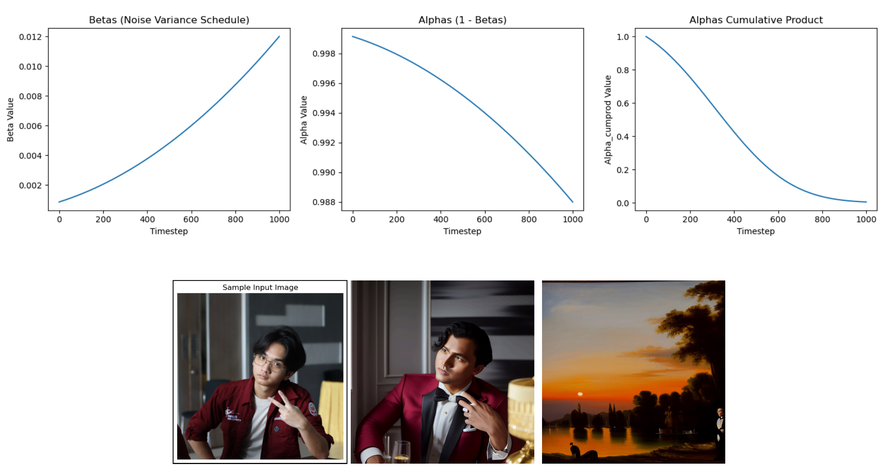
- **Input:** (64, 64, 3) image
- **Core Layers:**
  - Multiple blocks of Dense
- **Output:** Probability value
- **Total Parameters:** ~13.24 million

6 Evaluation

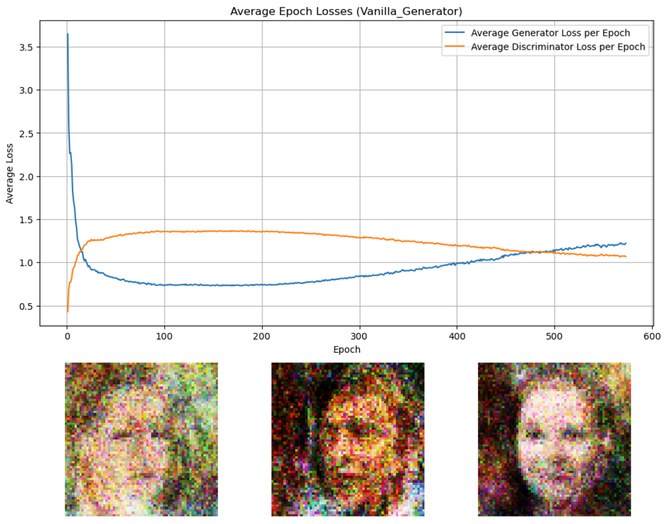


FID Score: 158.92

This DCGAN model exhibits **severe training imbalance**, where the discriminator's rapid dominance (low loss) prevents the **generator from learning effectively** (wildly increasing loss). Consequently, the **high FID score of 158.92** confirms the generator produces poor-quality or non-diverse images due to this disequilibrium.



Visually, our **Diffusion Model clearly surpasses older architectures like Vanilla GANs and DCGANs.** It generates images with significantly **higher photorealism and structural coherence**, avoiding classic GAN failures such as mode collapse. This qualitative evidence strongly suggests that the iterative denoising process is a more advanced and effective approach for creating high-fidelity images.



FID Score: 282.9644

The Vanilla GAN showed **severe training instability** with **diverging losses** (generator increasing, discriminator decreasing) post-Epoch 300, indicating discriminator dominance. This resulted in **lack of convergence** and **mode collapse**, producing limited image variety and poor quality, as confirmed by high and **fluctuating FID scores 282.9644** at Epoch 0573.

7 Conclusion

While our evaluation showed traditional GANs fail due to severe instability and mode collapse, our Stable Diffusion model consistently produced images with superior photorealism and structural coherence. This project confirms **Stable Diffusion is a transformative solution**, capable of meeting the modern demand for rapid, high-quality custom content where older methods are inadequate