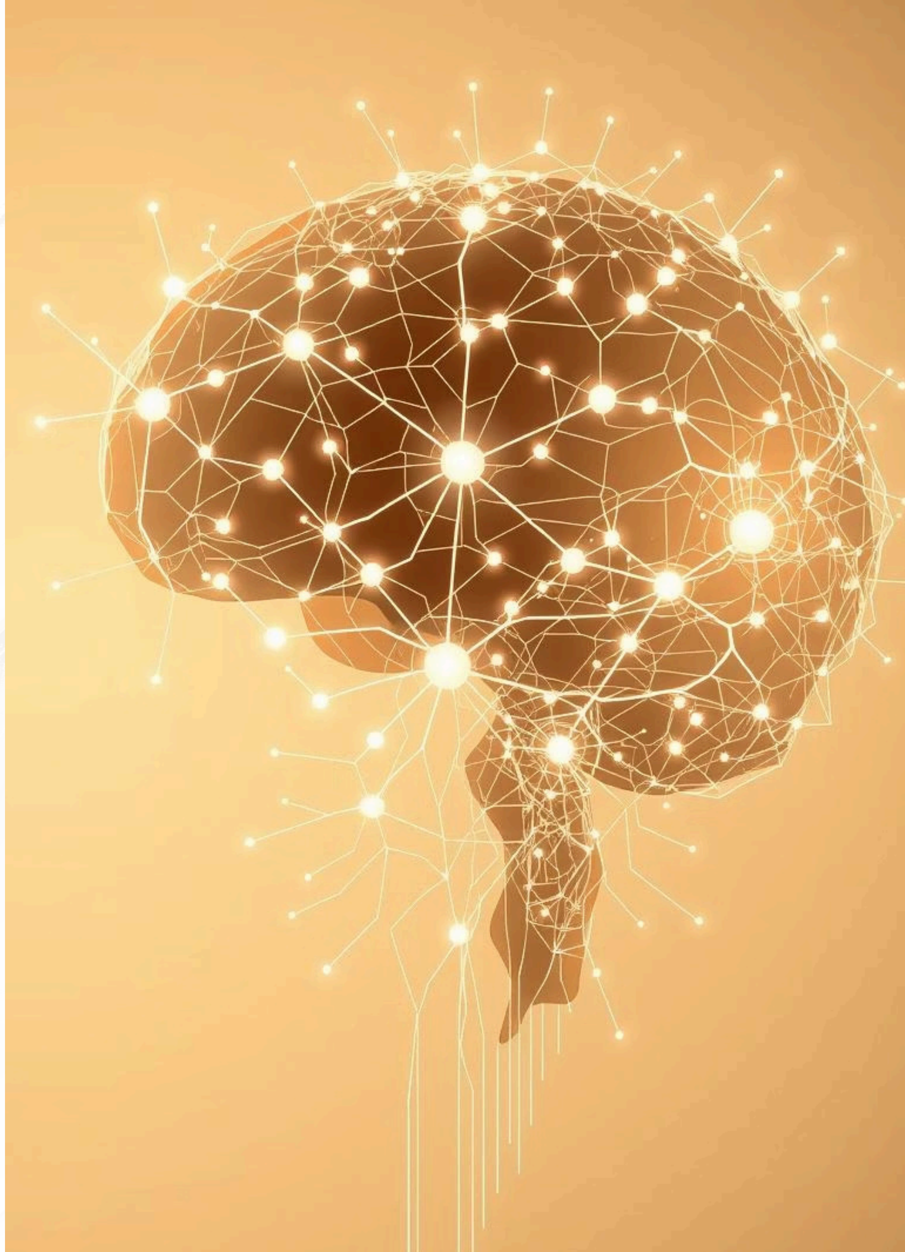
Generative Models for Data Augmentation – Internship Report

Presented by: Oumaima Bourmech

Company: CRNS

Supervisor: Madam Mouna Zouari

Date: July 1 to July 31



Made with GAMMA

Project Plan Overview

1. **Introduction & Problem Statement**

   Define the scope and objectives of using generative models for data augmentation, outlining the core problem the internship aims to address.

2. **Exploring Generative Models**

   Deep dive into foundational generative models: Generative Adversarial Networks (GANs) and Denoising Diffusion Probabilistic Models (DDPMs).

3. **Theoretical Models & Architectures**

   Analyze specific GAN architectures, including WGAN, WGAN-GP, and StyleGAN, and their theoretical underpinnings.

4. **Conditional DDPM for Control**

   Investigate the application of Conditional DDPM for targeted and controlled data generation tasks.

5. **Dataset Utilization**

   Identify and prepare relevant datasets (CIFAR-10, Fashion-MNIST, MNIST, Plastic) for model benchmarking and practical application.

6. **Evaluation & Performance**

   Assess the performance of implemented models on selected datasets, analyzing output quality and training stability.

7. **Summarize & Conclude**

   Synthesize key findings, discuss challenges encountered, and propose future directions for research and development in this field.

1.Introduction & Problem Statement

Challenge: Class Imbalance & Small Data

Real-world datasets often suffer from **class imbalance**, where some categories have very few samples. This makes models biased and less accurate. Combined with **small data**, it leads to poor generalization, unstable training, and low-quality results.

The Solution: Data Augmentation with Generative Models

To address this, we explore synthesizing new, representative data. Our goal is to use advanced generative models to rebalance datasets, thereby improving overall model performance.

Our Focus: GANs & DDPMs

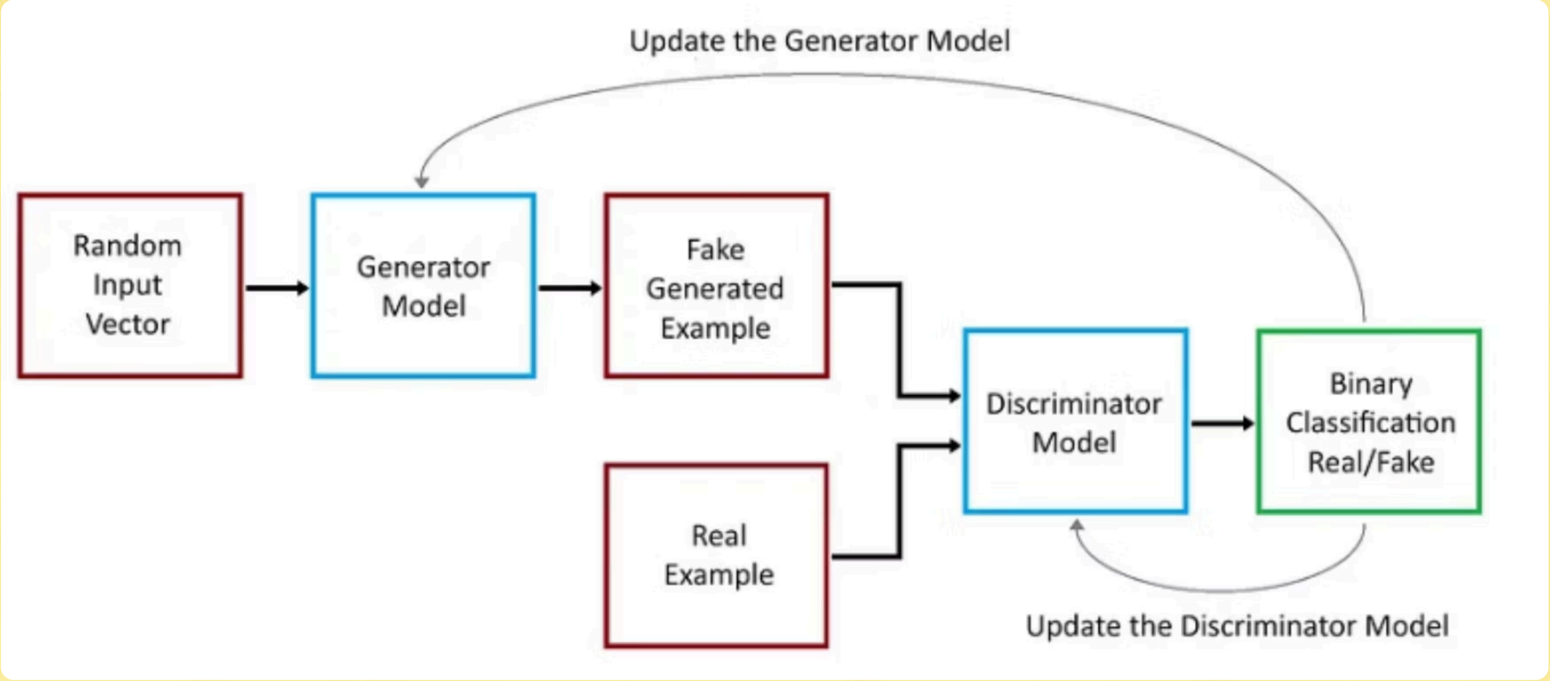
This report delves into the implementation and application of **Generative Adversarial Networks (GANs)** and **Denoising Diffusion Probabilistic Models (DDPMs)** to create synthetic data effectively.

2.Generative Adversarial Networks (GANs)

A **Generative Adversarial Network** (GAN) operates on a principle of game theory, pitting two neural networks against each other in a dynamic competition:

- **Generator (G):** Tasked with creating synthetic data that mimics the distribution of real data.
- **Discriminator (D):** Aims to accurately distinguish between real data samples and the synthetic data generated by G.

The **training objective** involves a continuous adversarial process: G strives to produce data realistic enough to fool D, while D simultaneously refines its ability to detect fakes. This iterative competition ultimately drives the Generator to produce **high-quality, realistic synthetic images**.

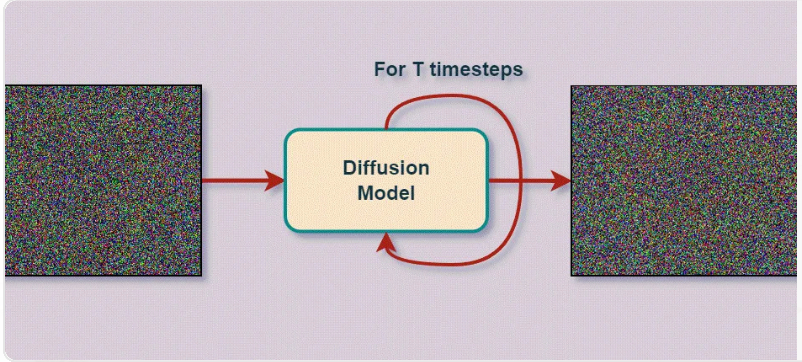


3.Denoising Diffusion Probabilistic Models (DDPMs)

DDPMs represent a distinct class of generative models that operate through a two-step process:

- **Forward Process:** Gradually adds Gaussian noise to an image over multiple steps, transforming it into pure noise.
- **Reverse Process:** Learns to systematically reverse this noise addition, effectively denoising the image step by step to reconstruct the original.

This methodology enables DDPMs to generate **high-fidelity, diverse images**. However, their primary limitations include **very slow training times** and a **high sensitivity to the quantity and quality of training data**.

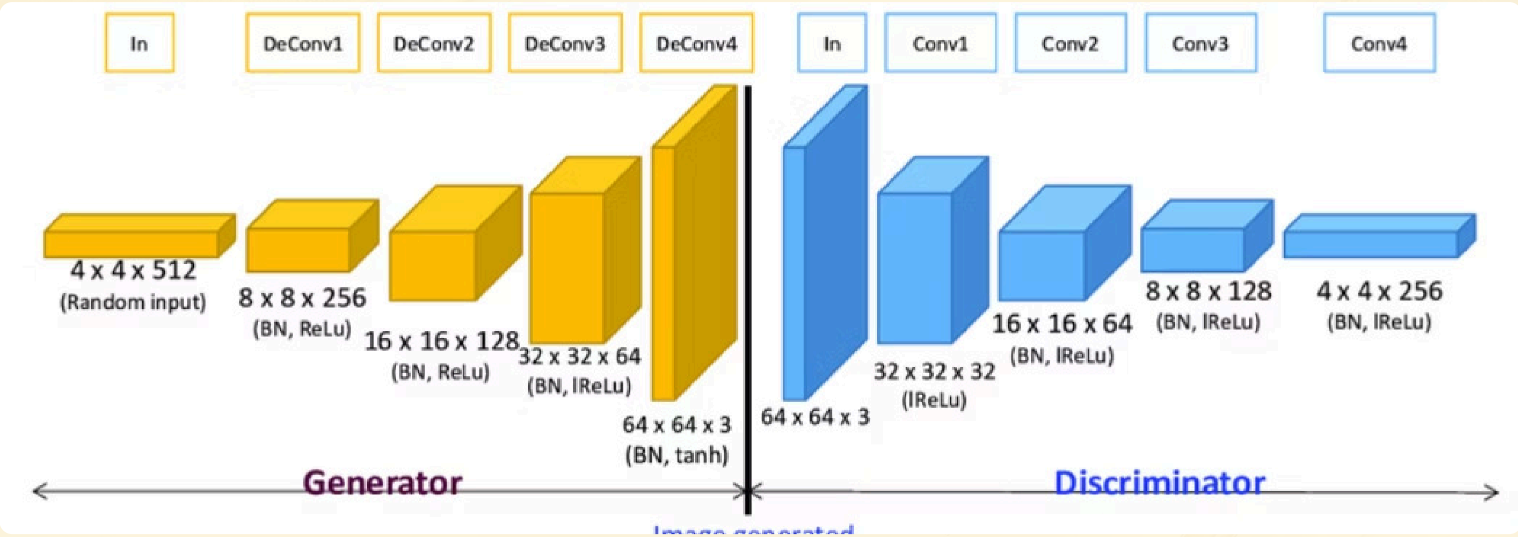


learnopencv.com

4.Theoretical Models: GAN Variants & Architectures

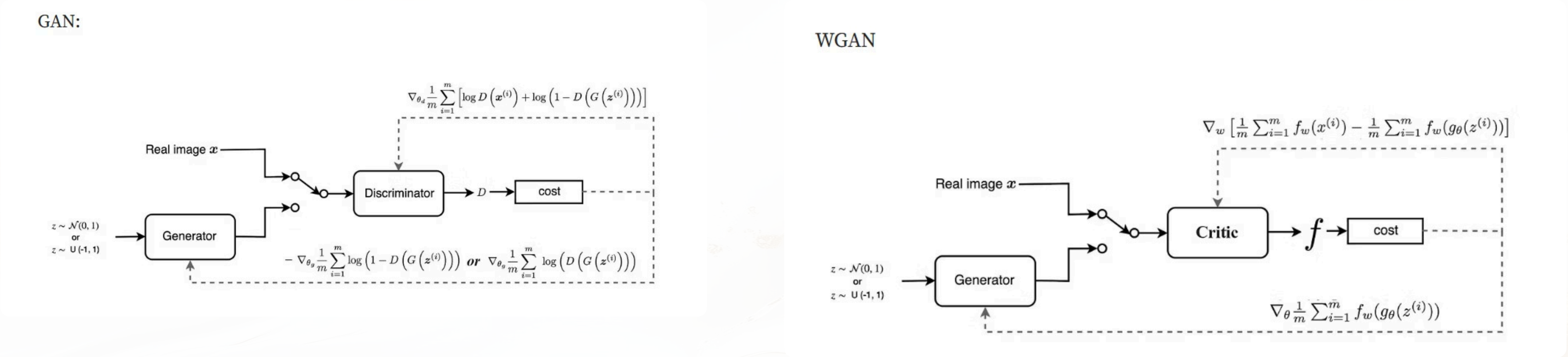
1.DCGAN (Deep Convolutional GAN)

DCGAN is a GAN architecture that uses convolutional layers (especially ConvTranspose2D in the generator) and Batch Normalization to generate images. It was one of the first GANs to produce stable and visually coherent results, especially on simple datasets.



2.WGAN (Wasserstein GAN)

Instead of the usual loss function, WGAN uses something called Wasserstein distance. This helps it train better and prevents it from getting stuck on just a few types of images. It needs a special step called "weight clipping" to work correctly.

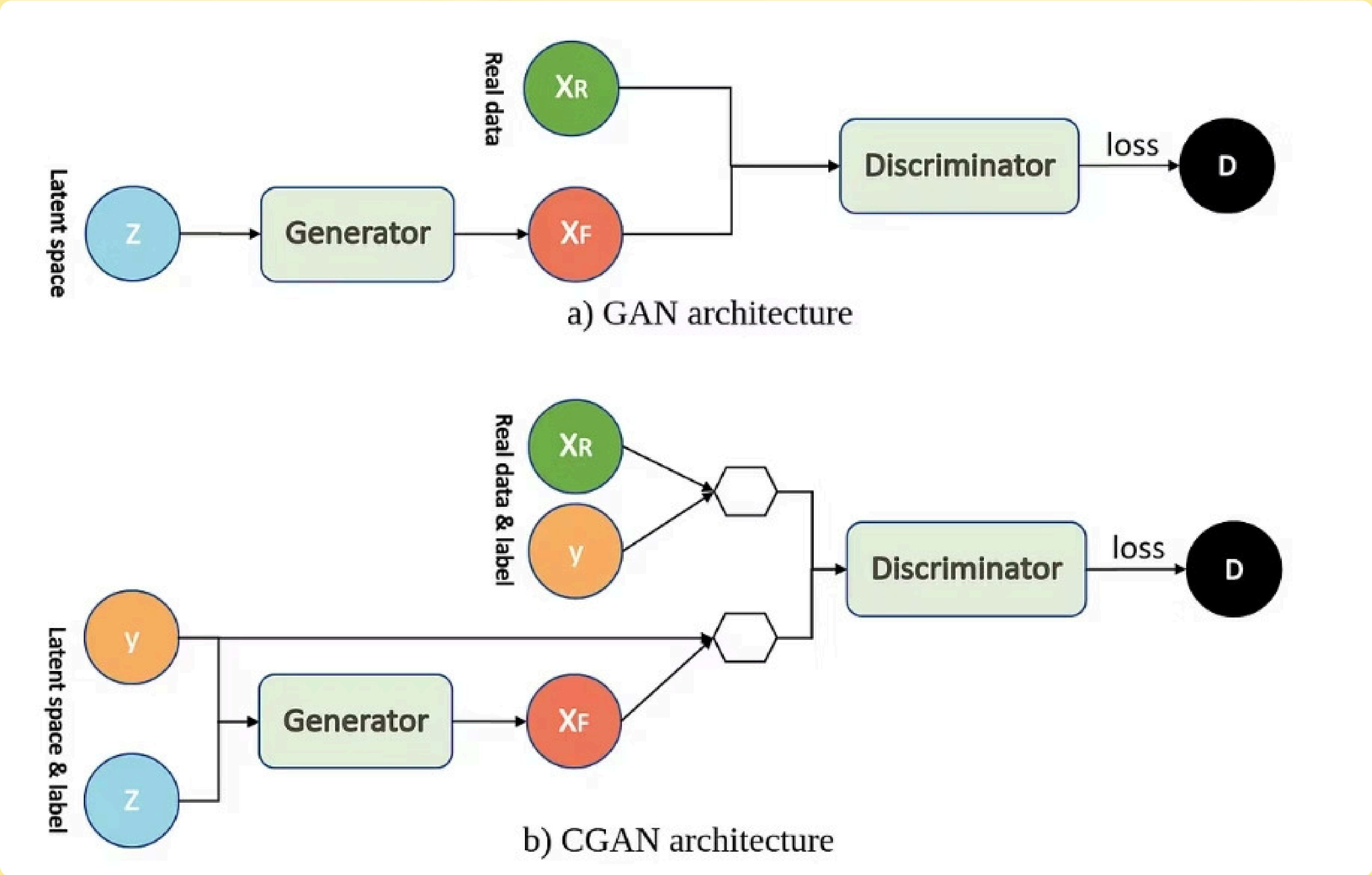


3.WGAN-GP (Wasserstein GAN with Gradient Penalty)

This version improves WGAN by using a "gradient penalty" instead of weight clipping. This makes it more stable and less sensitive to how you set its training parameters.

4.cGAN (Conditional GAN)

cGANs let you control what kind of image they create. They do this by giving information (like class labels) to both the Generator and Discriminator, so you can tell them to generate specific things (e.g., only cats, or only dogs).



5.StyleGAN (Style-based Generator Architecture)

StyleGAN introduces a mapping network and adaptive instance normalization (AdaIN) to control different aspects of the image (like shape, texture, color). It also uses progressive training (starting from low to high resolution), which helps generate high-quality and detailed images.

5.Conditional DDPM for Controlled Generation

- **Conditional DDPM**

Conditional DDPM extends the denoising diffusion probabilistic model by incorporating **class labels** directly into the image generation process. This is achieved by integrating label embeddings within the **UNet architecture** that governs the denoising steps.

The primary strength of this approach lies in its ability to offer **class-aware, high-quality generation**, allowing for precise control over the attributes of the synthesized images.

⚠ **Limitations:** Despite its capabilities, Conditional DDPM suffers from **slow convergence** and demands **massive, meticulously clean datasets** for effective training. In our specific application to the Resized Plastic dataset, its relatively small size and inherent noise prevented the DDPM from converging adequately.

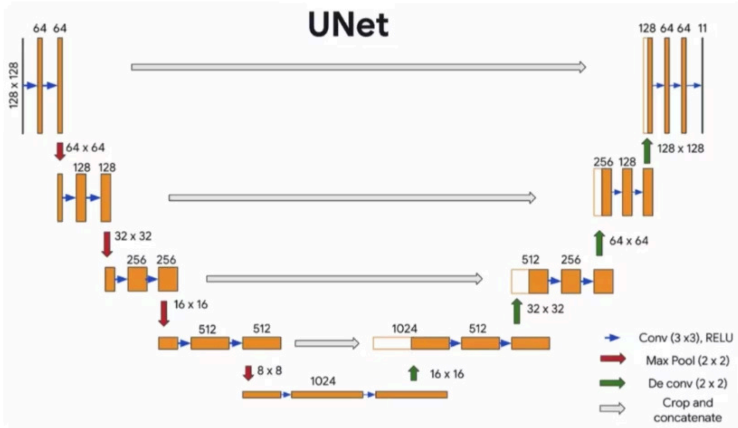


Figure: U-Net Architecture Overview

6.Datasets Utilized for Benchmarking and Application

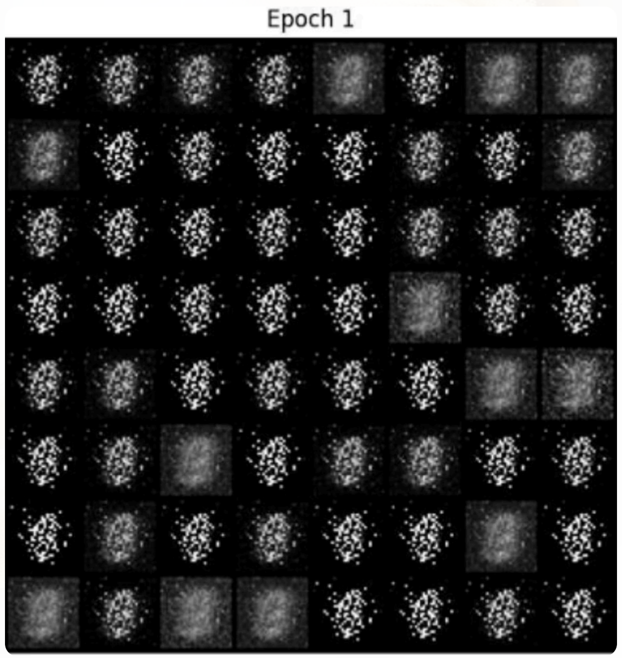
To thoroughly evaluate the efficacy of our generative models, we employed a suite of diverse public datasets for initial benchmarking. This phased approach allowed us to validate model performance under controlled conditions before transitioning to our specific real-world application.

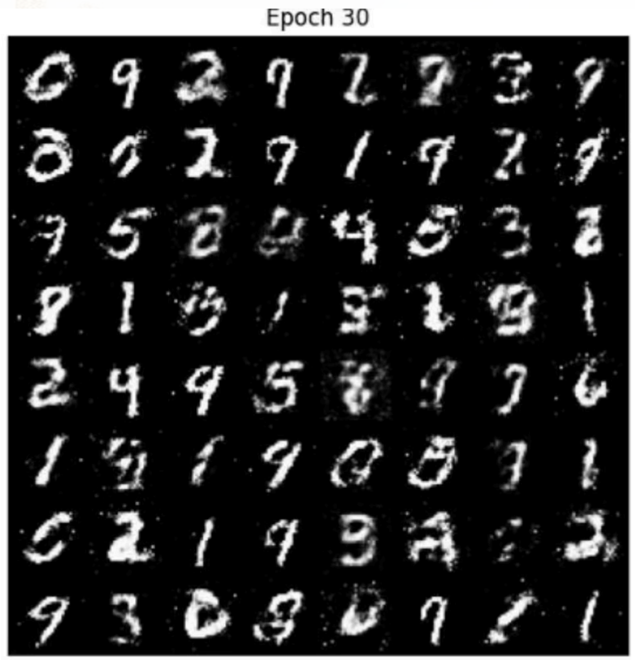
Dataset	Description
MNIST	28×28 grayscale images of handwritten digits.
Fashion-MNIST	28×28 grayscale images of various clothing items.
CIFAR-10	32×32 RGB images across 10 distinct categories (e.g., cars, animals).
Ours (Plastic)	Custom dataset of plastic-waste images, uniformly resized.

Our methodology involved a progressive application: starting with established public datasets for robust benchmarking, and subsequently applying the models to our proprietary, real-world **Resized Plastic** dataset to address specific challenges in waste classification.

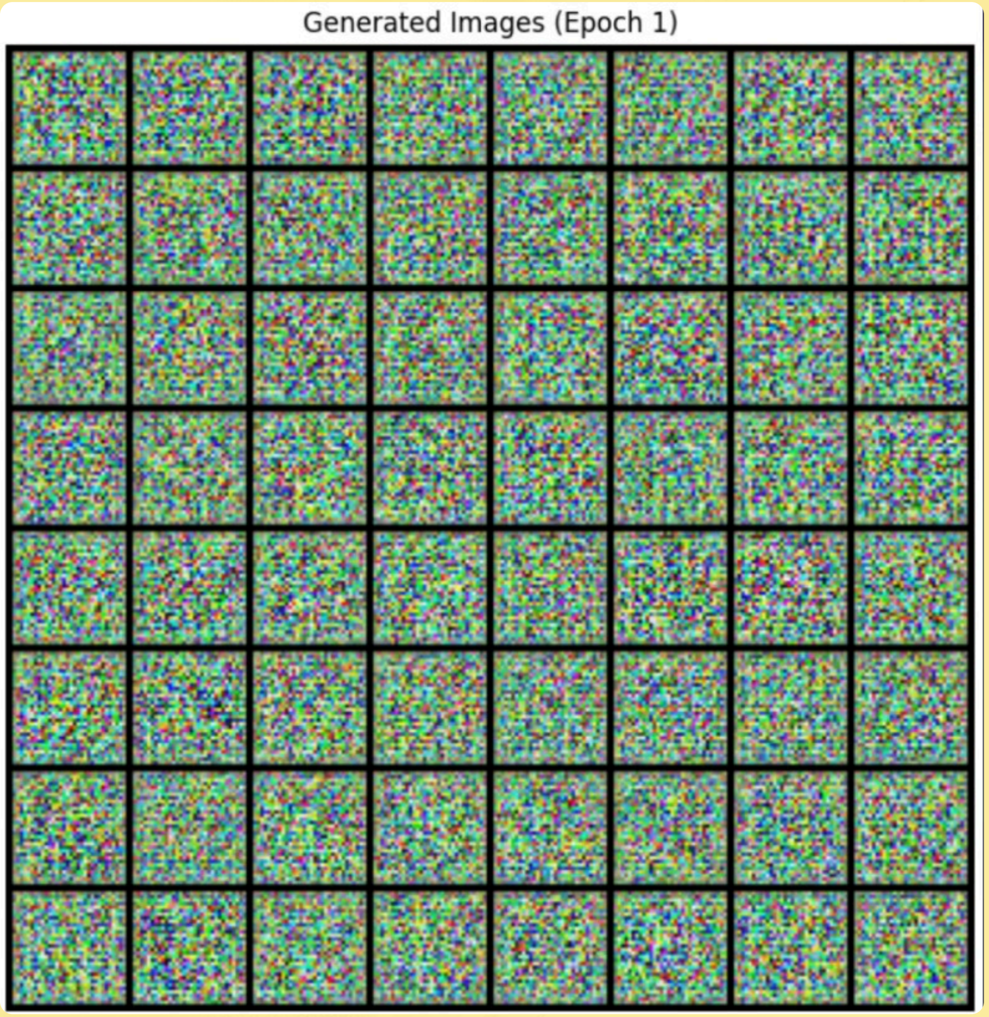
7.Evaluation: Results & Model Performance

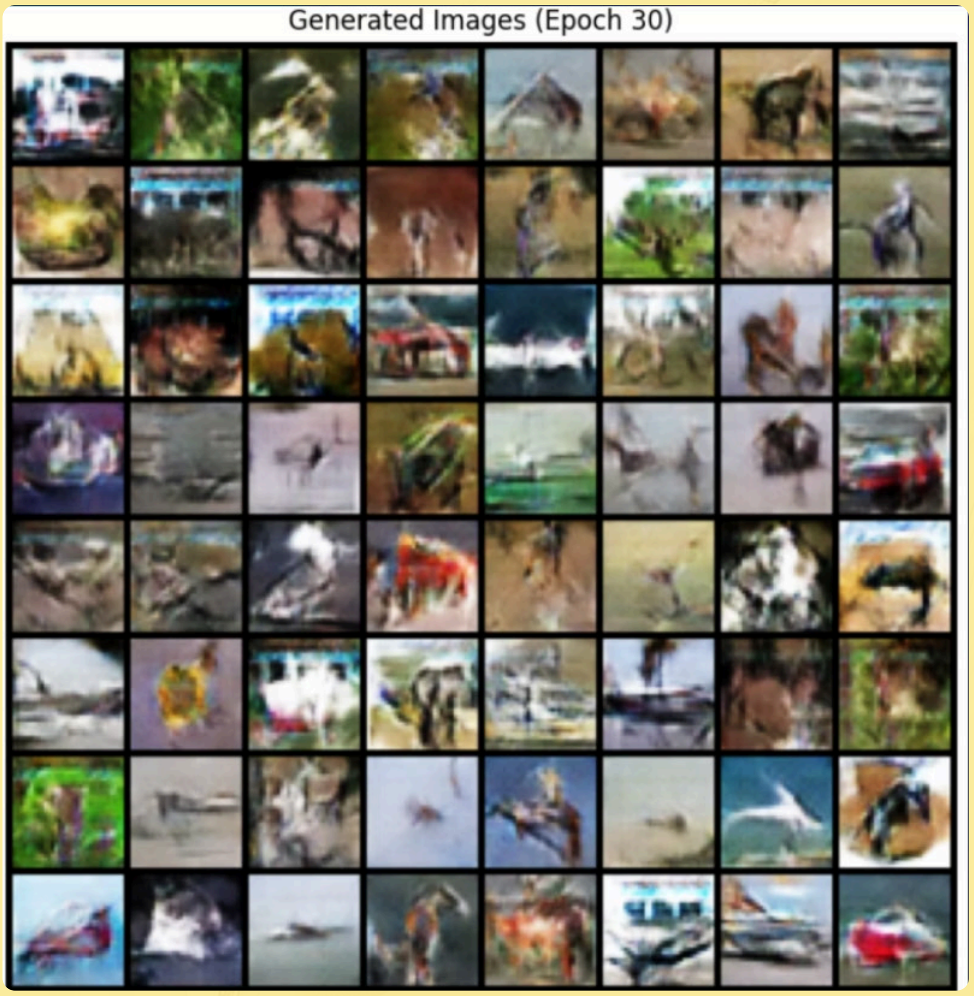
- **DCGAN on MNIST (50 epochs)**: Generated digits with decent quality. Training was stable, but some images appeared blurry due to limited model complexity.

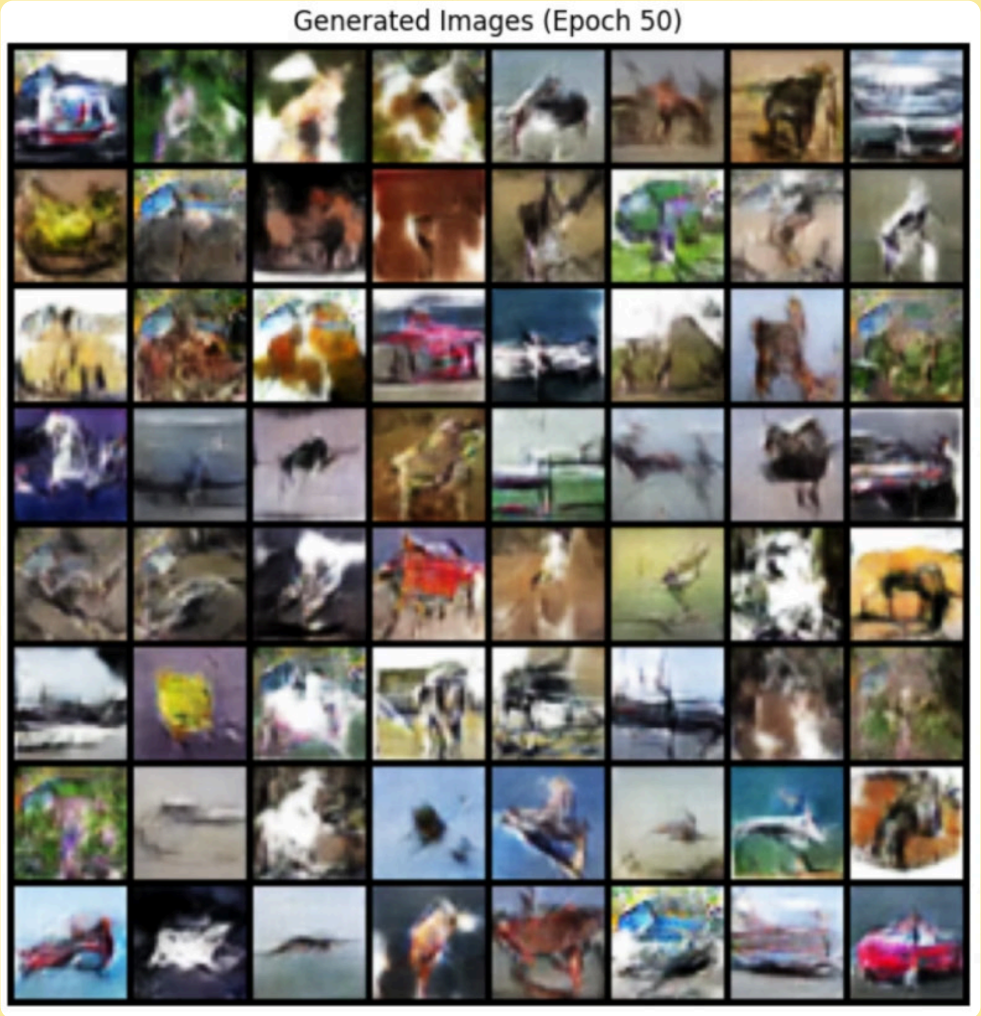




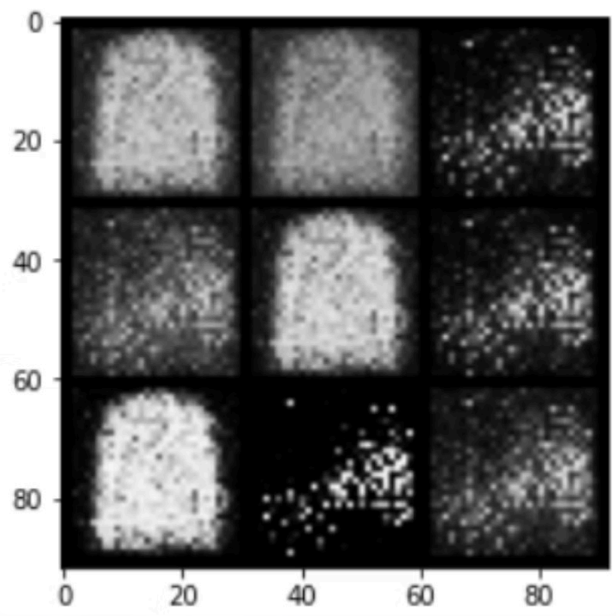
- **DCGAN on CIFAR-10 (50 epochs)**: Struggled with complex and colorful images. The training was unstable and outputs were low quality and incoherent.

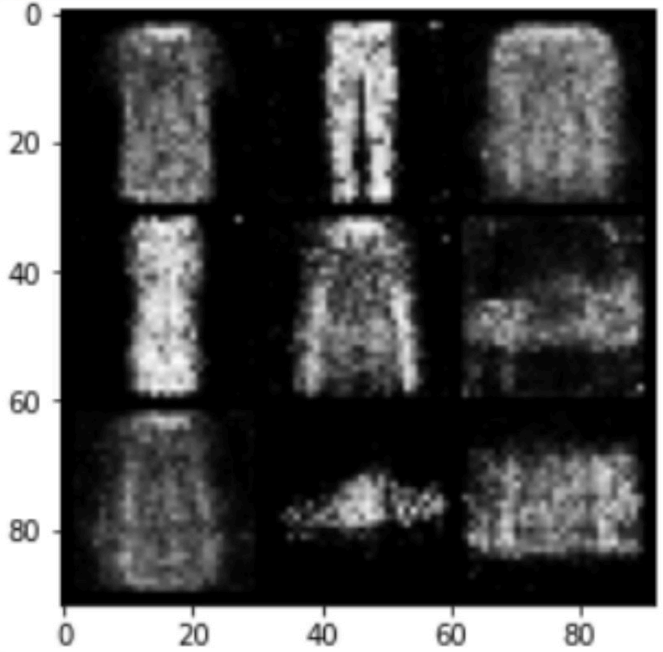


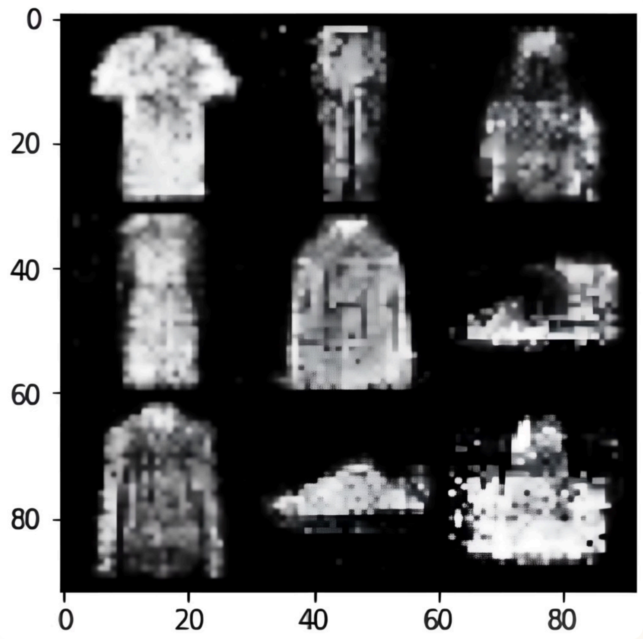




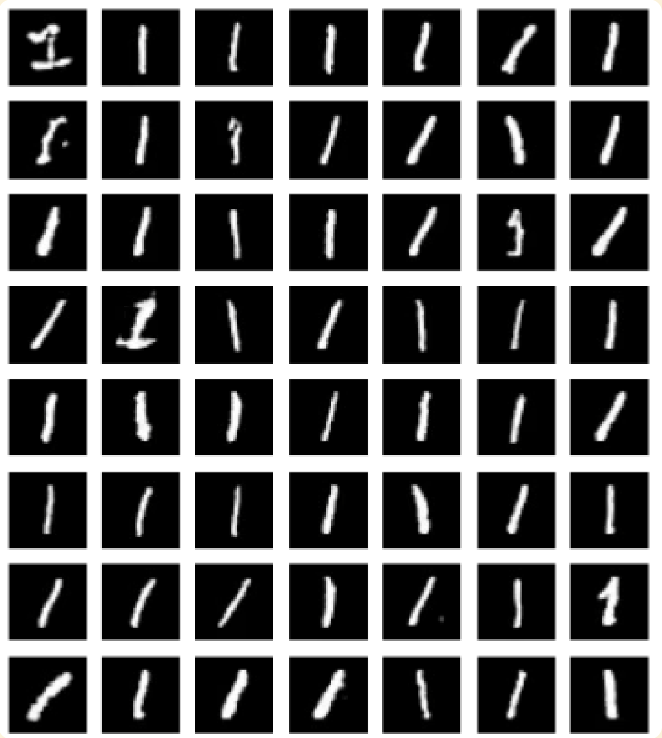
- **cGAN on Fashion-MNIST (50 epochs)**: Generated labeled clothing items with reasonable accuracy. The images were recognizable but lacked sharpness.

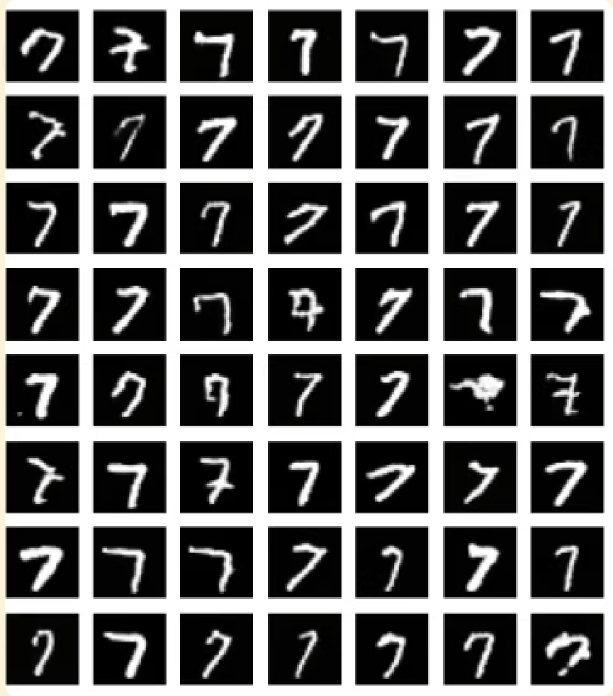


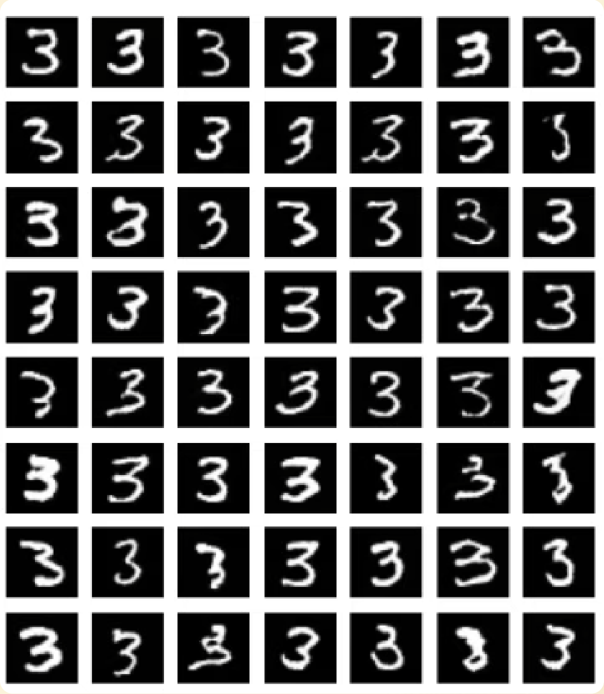




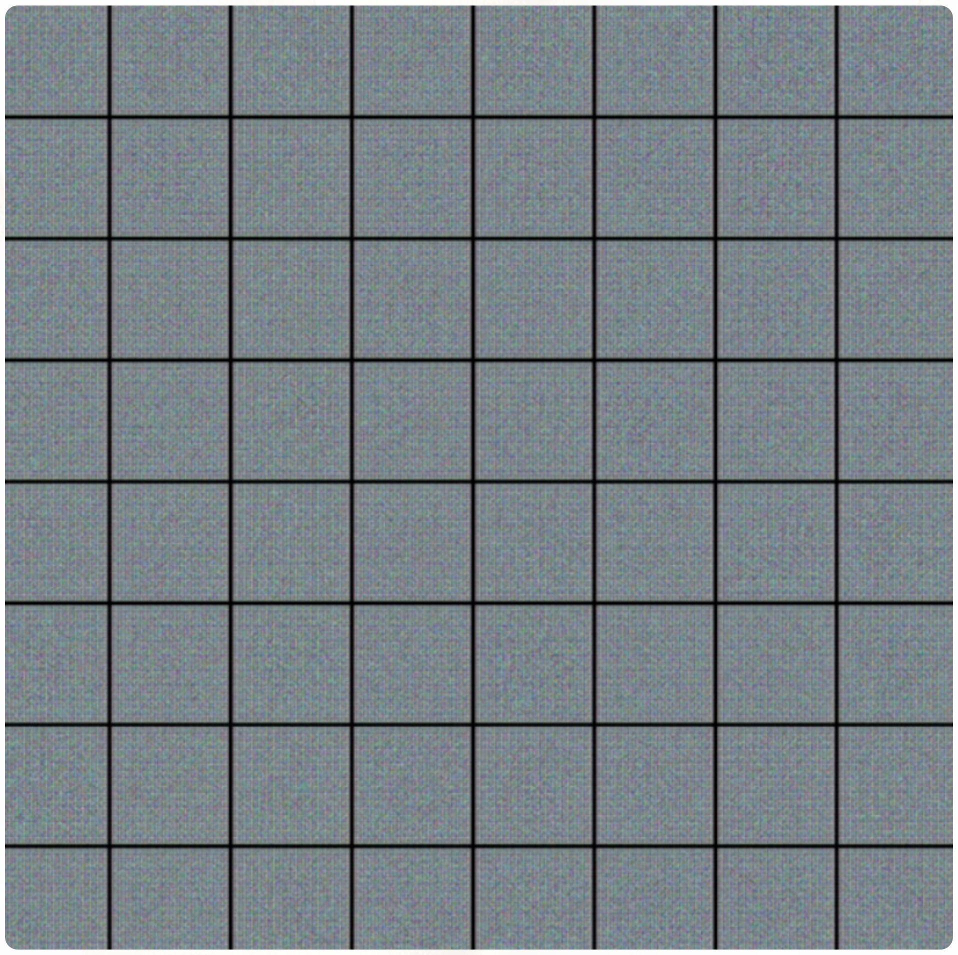
- **StyleGAN on MNIST (70 epochs)**: Delivered high-fidelity and realistic digit images. The architecture is complex, but the results were visually excellent.(exemple of final output for the classes "1" , "7" and "3")

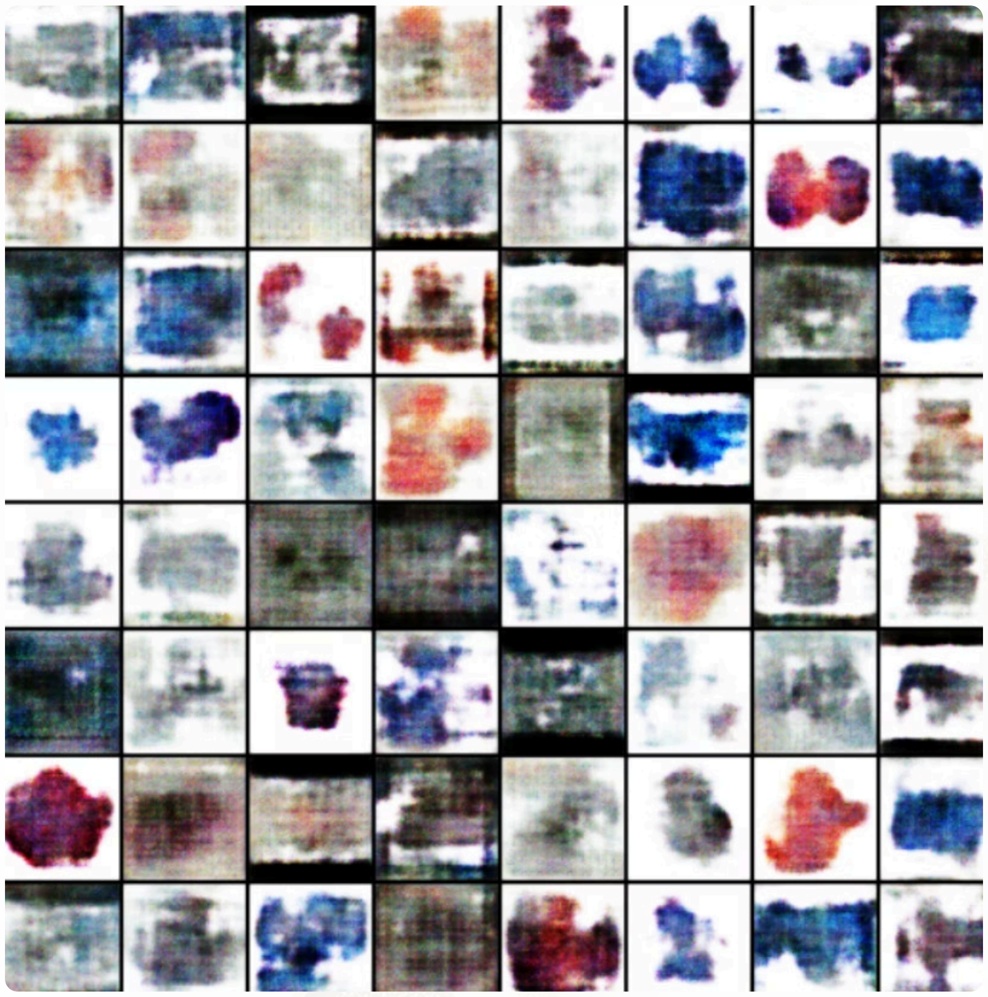




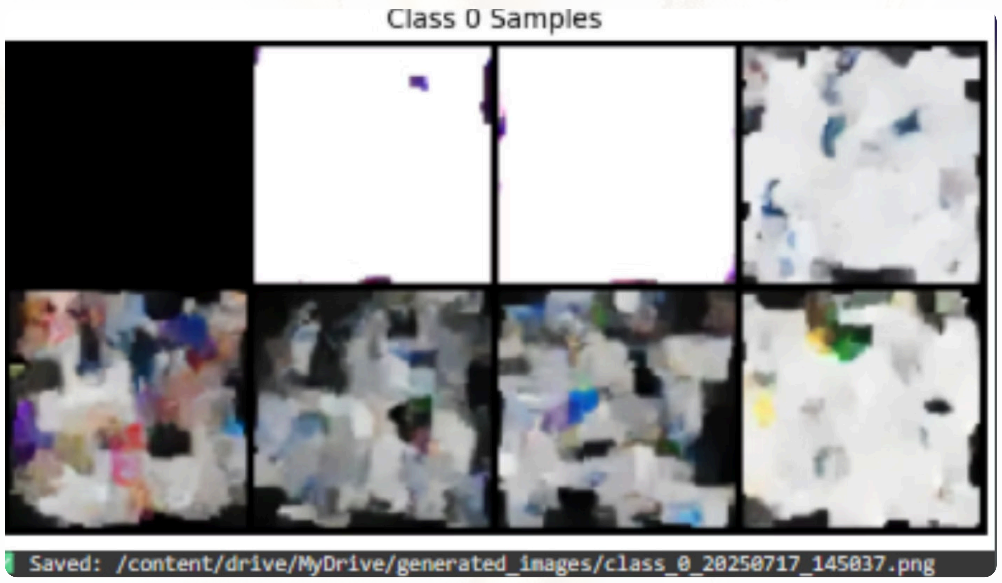


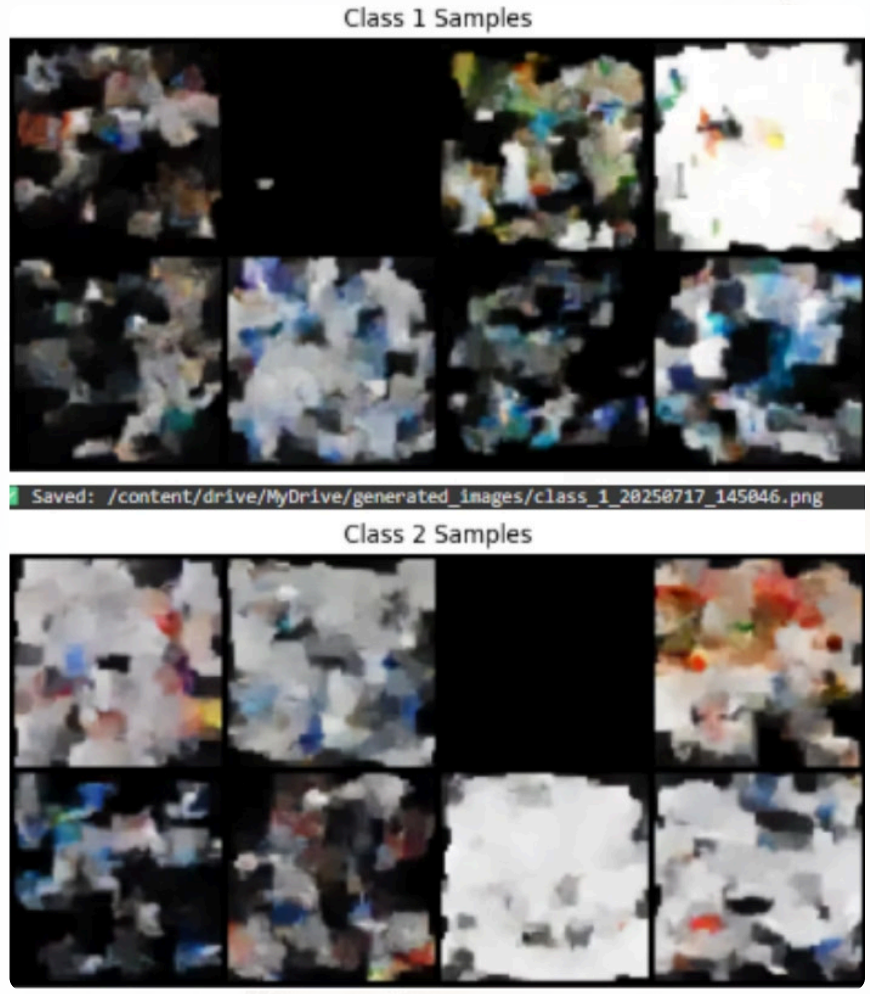
- **WGAN-GP on Plastic (PVC class) (5000 epochs)**: Stable training and good-quality outputs, though some images were still blurry.

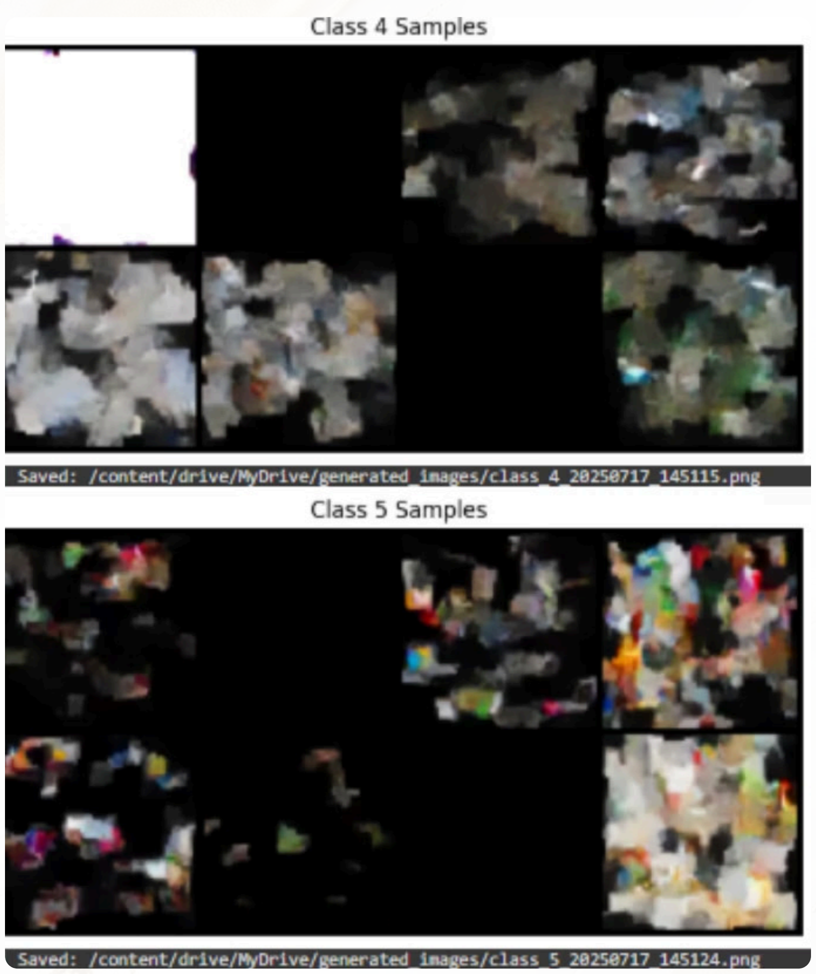


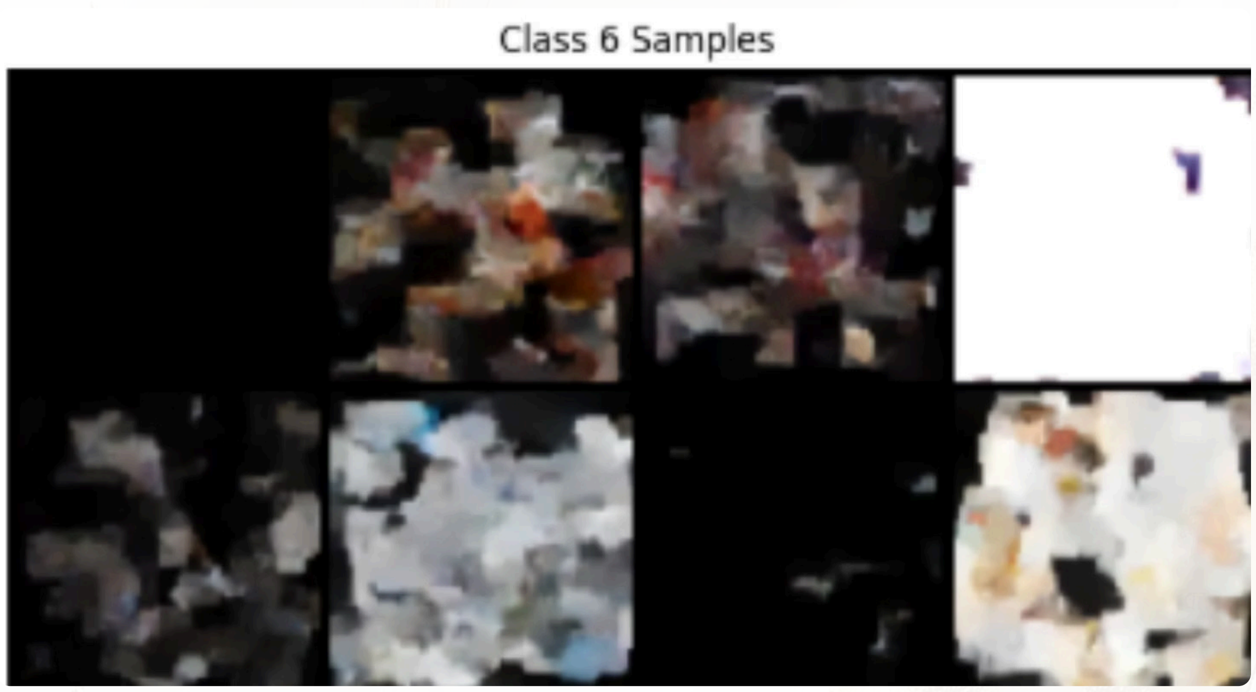


- **DDPM on Resized Plastic dataset (1200 epochs)**: Performed poorly. The model failed to converge properly, likely due to limited and unbalanced training data.









Summarize

Model	Dataset	Epoch	Visual Quality	Notes
DCGAN	MNIST	50	Good	Stable but images could be blurry.
DCGAN	CIFAR-10	50	Poor	Performed poorly; unstable training and low-quality, incoherent images.
cGAN	Fashion-MINIST	50	Good	Goals realized, but images were less sharp.
StyleGAN	MNIST	70	Excellent	Complex architecture, but high fidelity.
WGAN-GP	A class from our plastic dataset (PVC)	5000	Good	also Stable but images could be a little bit blurry.
DDPM	Resized Plastic	1200	Poor	Did not converge effectively with limited data.

8.Conclusion & Future Directions

1

WGAN: Optimal Balance

WGAN demonstrated the most effective balance between generated image quality and computational training speed, making it highly practical for many applications.

2

Diffusion Models: High Potential

While **Diffusion Models** show immense potential for high-fidelity synthesis, their current implementation requires substantially more training time and larger, cleaner datasets.

3

Benchmark vs. Custom Datasets

Our generative pipeline produced strong results on **benchmark, well-structured datasets** like MNIST and Fashion-MNIST. However, performance dropped on our **custom plastic dataset**, which was relatively small, unstructured, and visually diverse (containing 8,365 varied images). This highlights the challenge of training generative models when the data lacks visual consistency and domain uniformity.

Final Note: WGAN and StyleGAN delivered the most promising visuals and are expected to improve further with additional epochs and hyperparameter tuning . DDPM, despite its current limitations, holds significant promise for future investment and could potentially outperform other models with sufficient training and data resources.

Thank You

For your attention and insights.

Questions?