Generative Anomaly Detection Using GANs and VAEs

Eyad Salman

February 20, 2025

1 Introduction

This project explores generative approaches to anomaly detection in image datasets using Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). These models learn the underlying distribution of normal data and flag outliers by evaluating reconstruction error or latent deviation. We evaluate and compare both approaches on two canonical datasets: MNIST and Fashion MNIST.

2 Objective

- Implement and compare GANs and VAEs for generative anomaly detection.
- Train on MNIST Digits and Fashion MNIST datasets.
- Generate new samples, visualize latent space, and analyze anomaly detection performance.
- Apply VAE-based anomaly detection to a real-world domain with financial or operational impact.

3 Datasets

1. **MNIST Digits:** Handwritten digit images (0-9), grayscale, 28×28 pixels.
2. **Fashion MNIST:** Grayscale clothing images (e.g., sneakers, shirts), 10 categories, 28×28 pixels.

4 Exploratory Data Analysis (EDA)

- Loaded both datasets using torchvision and displayed sample images.
- Verified class balance across both datasets.
- Preprocessed datasets by normalizing pixel values to [0,1].

5 Generative Adversarial Networks (GANs)

Model Architecture

- **Generator:** Converts random noise to fake images using linear/conv layers with ReLU/Tanh activations.
- **Discriminator:** Distinguishes real vs. fake images via conv layers and sigmoid output.

Training and Results

- Trained on MNIST for 50 epochs using Binary Cross-Entropy loss.
- Generated 10 synthetic digits, and 5 class-specific samples (digit '3').
- Trained on Fashion MNIST (e.g., sneaker class) and generated samples successfully.

6 Variational Autoencoders (VAEs)

Model Architecture

- **Encoder:** Encodes image into mean and log-variance of latent vector.
- **Decoder:** Reconstructs image from latent sample using ConvTranspose layers.
- **Loss:** Binary Cross-Entropy + KL Divergence.

Training and Results

- Trained on MNIST and Fashion MNIST for 50 epochs.
- Visualized latent space using t-SNE and PCA.
- Generated 10 generic images and 5 samples of digit '2'.
- Successfully reconstructed and generated shoe class from Fashion MNIST.

7 Comparison: GAN vs VAE

- **Image Quality:** GANs produced sharper images, VAEs smoother but more diverse.
- **Training Stability:** VAEs are easier to train; GANs required careful tuning to avoid mode collapse.
- **Latent Space:** VAEs learn structured latent representations, enabling interpolation and controlled generation.

8 Saving the World with VAE: Real-World Anomaly Detection

Domain: Financial Fraud Detection
Undetected fraud costs institutions $42B annually. We propose using VAE-based anomaly detection to identify unusual transaction patterns based on:

- Latent deviation from learned normal transaction profiles.
- Reconstruction error as a fraud indicator.

Why VAE?

- Learns compact and meaningful representation of normal behavior.
- Flags deviations without requiring labeled anomalies.

Other Applications

- **Cybersecurity:** Early intrusion detection based on access patterns.
- **Healthcare:** Detecting misdiagnosis through medical image deviation.
- **Manufacturing:** Predictive maintenance through abnormal sensor data.

9 Conclusion

This project demonstrates how generative models, specifically GANs and VAEs, can be used not only to synthesize realistic images but also to detect anomalies in both synthetic and real-world domains. Future work includes hyperparameter tuning, multi-class anomaly detection, and real-time deployment.

References

- MNIST Digits: `http://yann.lecun.com/exdb/mnist/`
- Fashion MNIST: `https://github.com/zalandoresearch/fashion-mnist`
- PyTorch: `https://pytorch.org/`