Report on GAN and VAE Implementations

Introduction

In this report, two of the generative models have been described as follows: Generative Adversarial Networks and Variational Autoencoder. These models were used in generating artificial images from the MNIST and Fashion-MNIST data sets. This includes the architectures of both models, the training procedures, some samples of generated images, t-SNE visualizations of the latent space and a quantitative and qualitative comparison of both models. Further, there is an analysis of using VAE in fraud analysis as an example of its application.

1. Handmade Architecture of Models

GAN Architecture

Generator:

The generator in a GAN learns to map random noise to realistic-looking images. It consists of fully connected layers that progressively upscale a noise vector into an image.

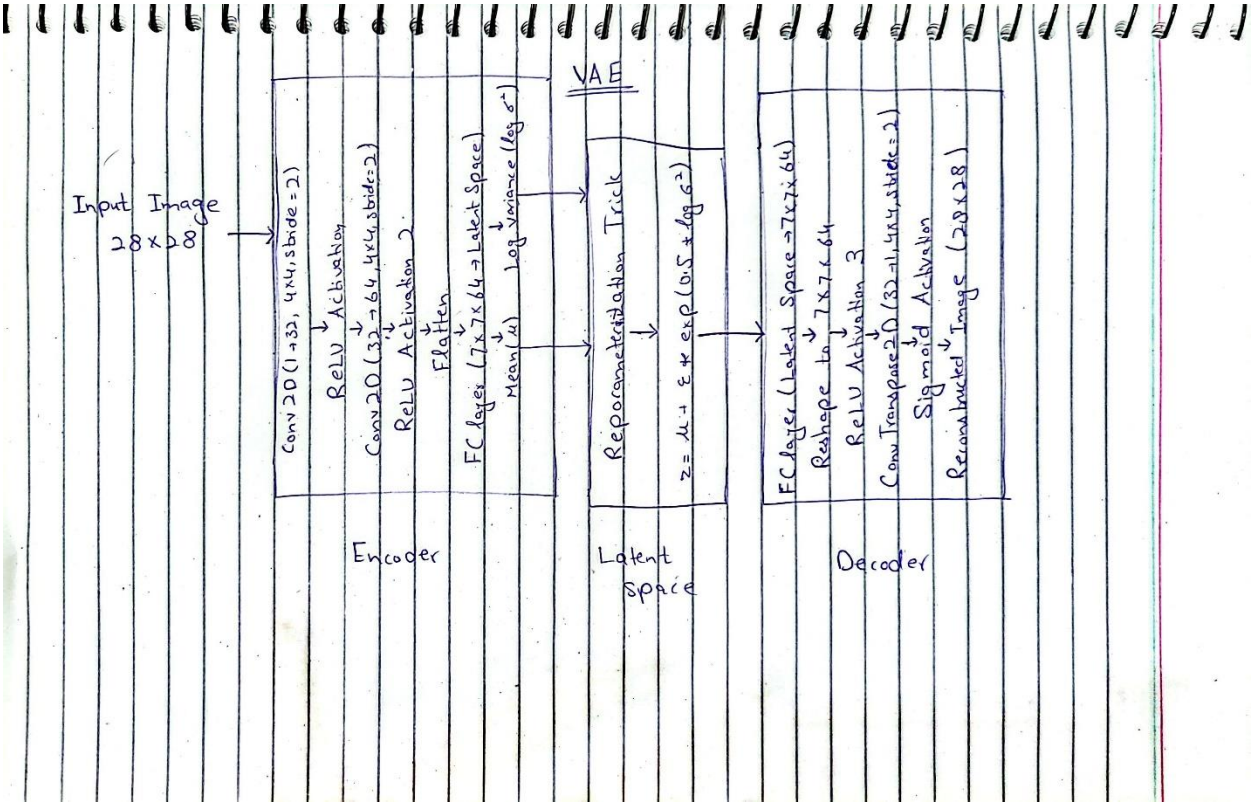
- Input Layer: 100-dimensional random noise vector.
- Fully connected layer 1 with 256 units and ReLU activation function on it.
- Hidden Layer 2: Fully Connected and has 512 neurons to the next hidden layer with ReLU activation function.
- Output Layer: 784 nodes or Fully Connected with the ReLU activation function and reshaped the output to have a dimension of 28*28 images.

Discriminator:

The discriminator is a model that helps to distinguish between a real image and a fake image generically created by the generator.

- The first layer comes out as an input layer, a 784-dimensional vector since the image is 28 x 28 and flattened.
- First hidden layer: This layer has 512 neurons, and uses the ReLU activation function as its non-linearity as it is fully connected.
- Layer 3 – Fully Connected – 256 units with ReLU activation function.
- Output Layer: Fully Connected (1 unit) + Sigmoid activation (probability output).

VAE Architecture



Variational Autoencoder aims at encoding input image and then reconstruct it back, or in other words, it tries to decode it.

Encoder:

- Conv2D Layer 1: 32 filters, kernel size (4x4), stride 2, ReLU activation.
- Conv2D Layer 2: 64 filters, kernel size (4x4), stride 2, ReLU activation.
- Flatten Layer: This layer transforms feature maps into a vector that is now flattened.
- Fully Connected Layer for Mean (mu): Outputs a latent vector means.
- Fully Connected Layer for Variance (logvar): Outputs the log variance of latent variables.

Reparameterization Trick:

Get a sample from a latent vector using mean and variance adding Gaussian noise for the purpose of backpropagation.

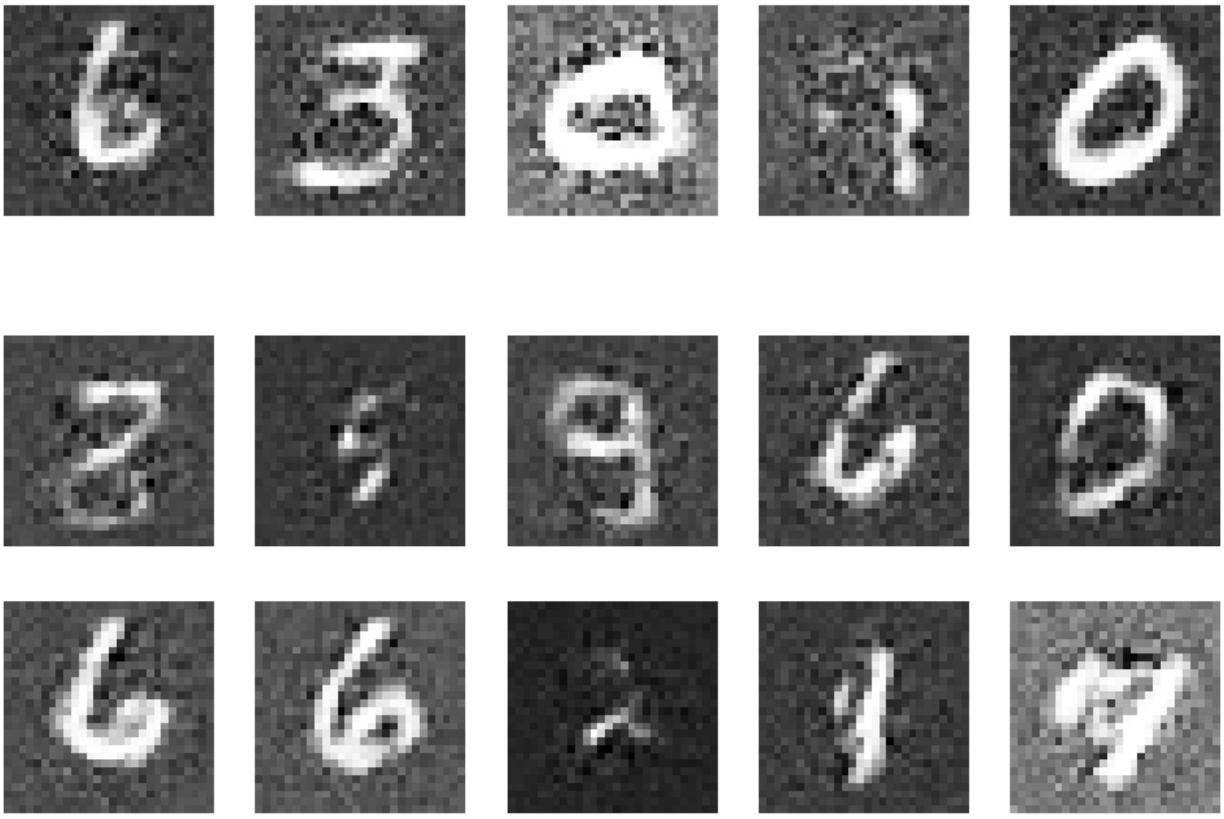
Decoder:

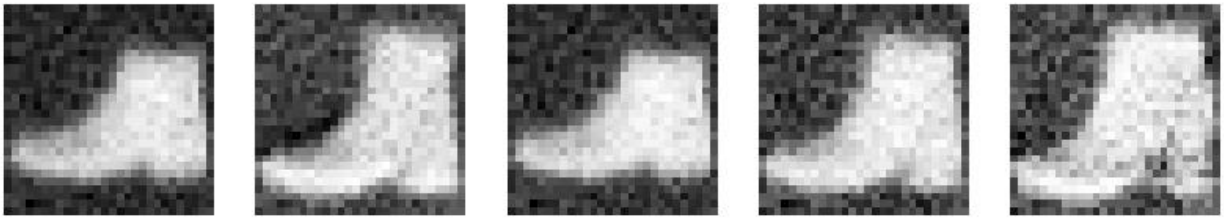
- Fully Connected Layer: Maps latent space to a 7x7x64 feature map.
- ConvTranspose2D Layer 1: 64 filters, kernel size (4x4), stride 2, ReLU activation.
- ConvTranspose2D Layer 2: 32 filters, kernel size (4x4), stride 2, ReLU activation.
- ConvTranspose2D Layer 3: 1 filter, kernel size (4x4), stride 2, Sigmoid activation (final 28x28 reconstruction).

2. Generated Images from Both Models

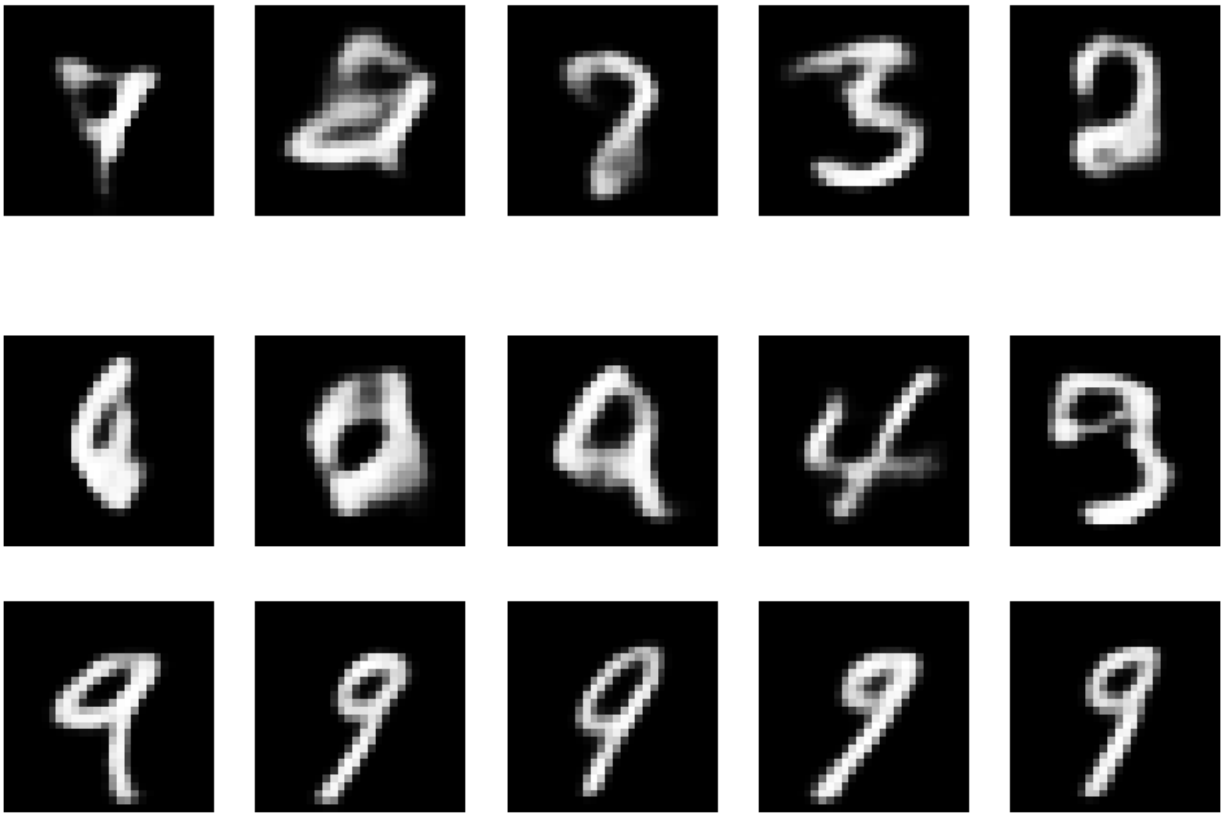
Below is the comparison of the generated images from GAN and VAE models;

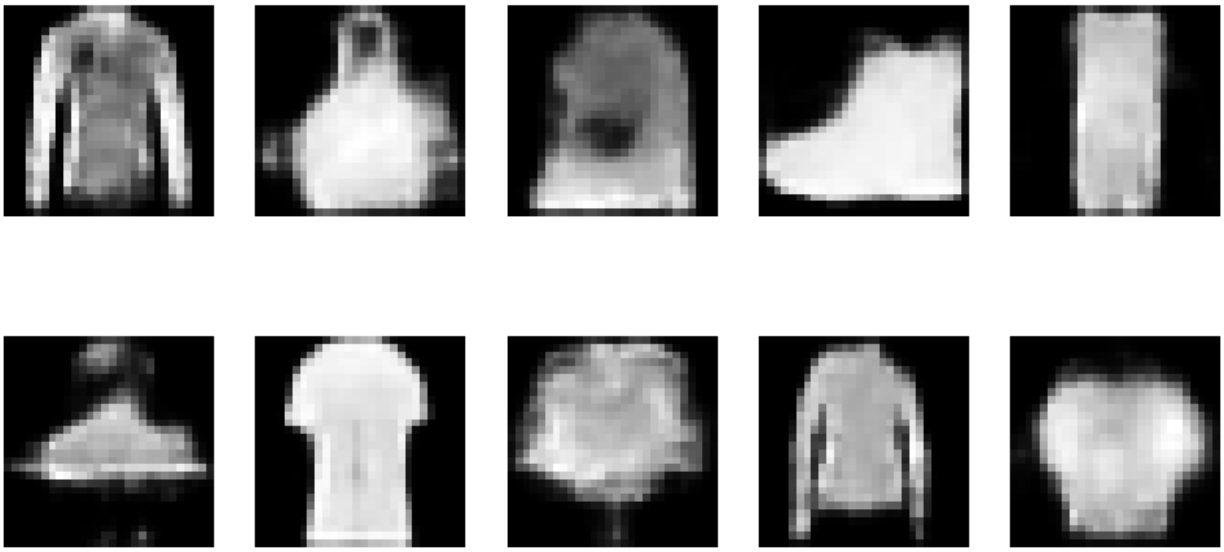
- GAN: This is due to the training instability hence the blurriness evident from the generated images. There are some situations when the generator creates similar samples of text and the term mode collapse is used.





- VAE: The reconstructed images are of better quality than the original images and have less distortion, which implies that encoder-decoder based model is efficient in capturing features of input images.

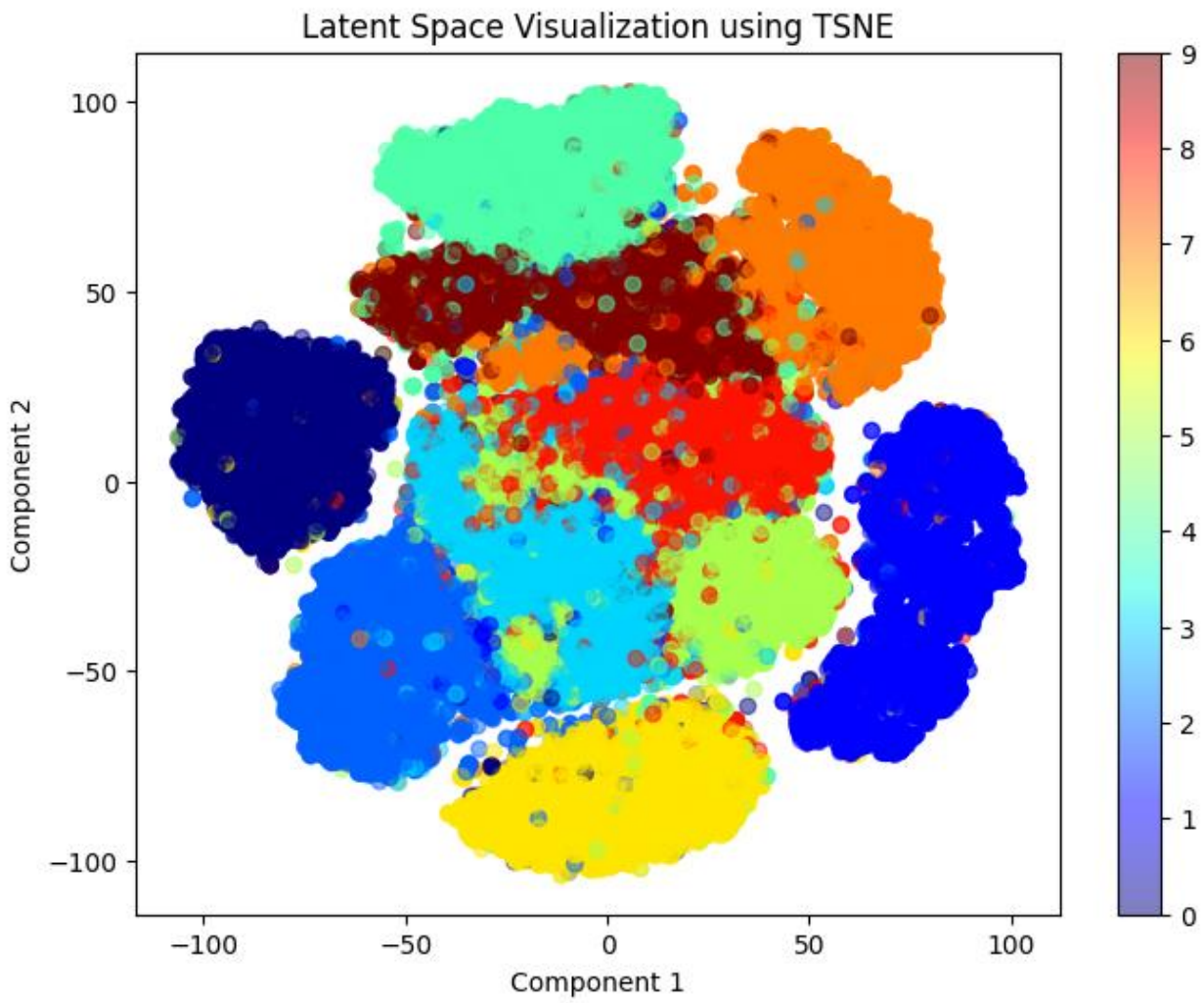


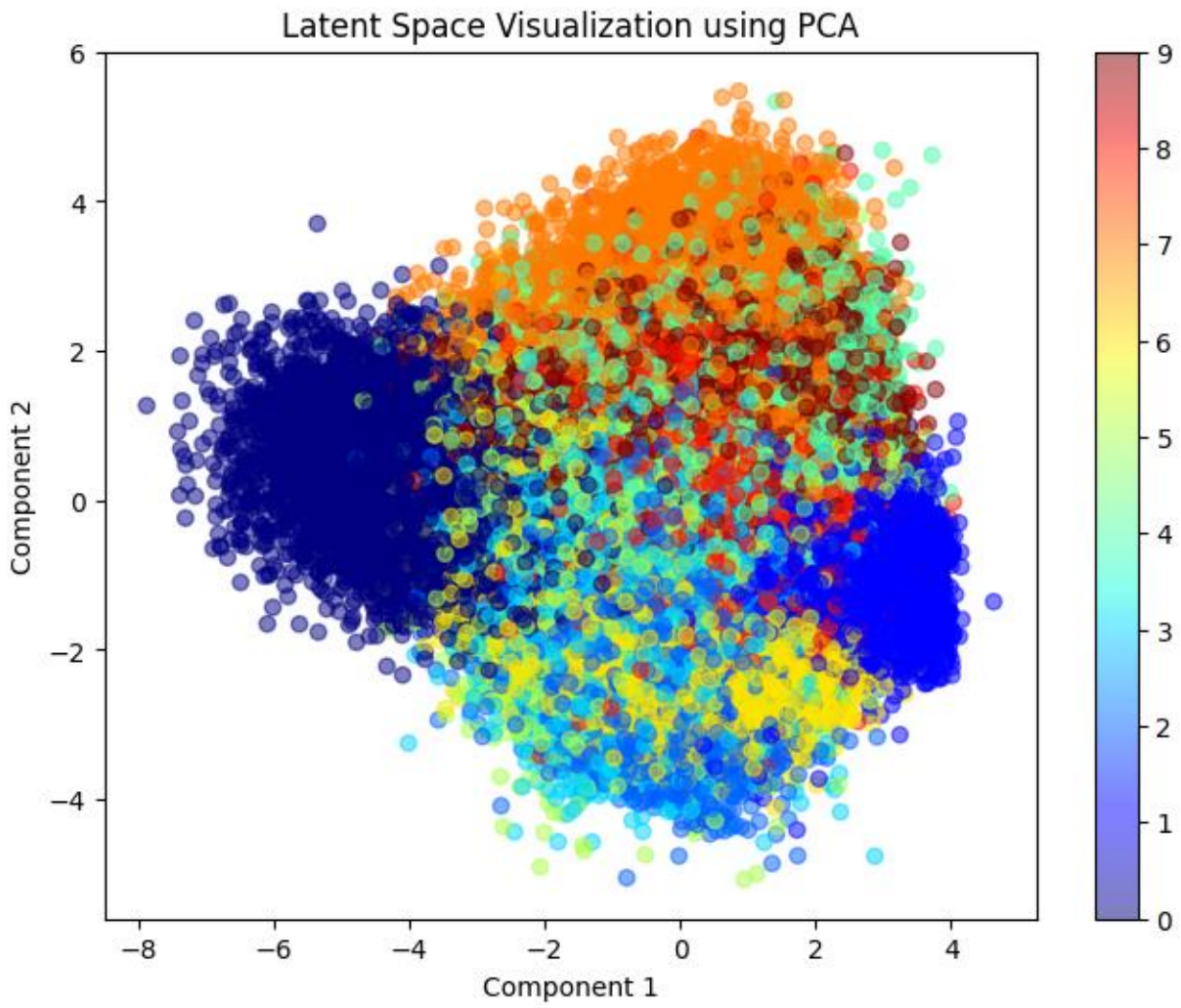


3. Plots for Latent Space Representation

The plotted figures 1 and 2 below illustrate the t-SNE and PCA results of latent space representation of VAE.

- t-SNE Visualization: Organized clusters imply that the VAE is preserving class information and successfully keeps various digits or objects segregated.
- PCA Visualization: Illustrates the separability of the input data along the first two principal components which show the linear separability in the latent space meaning that the encoded representation features could be easily identified for reconstruction.


Latent Space Visualization using TSNE
Component 2
Component 1
100
50
0
−50
−100
−100
−50
0
50
100
9
8
7
6
5
4
3
2
1
0



4. Discussion

Image Quality

- GAN-generated images: Usually they are distorted at the edges or less defined especially due to adversarial instability. If the generator component overcomes the discriminator component, then the latter will create low-quality images.
- Of images generated by VAEs: More likely to be regular and to exhibit high interpretability due to vectors of the latent space.

Issues with GAN Training

- Mode Collapse: Instead of providing unique images, and Figures after each iteration, the generator continues to generate similar pictures.
- Discriminator Overpowering: If too intense, the generator is deprived of learning a proper set of features.
- Less Generative Structure: Addition of layers or architectural changes to the network may contribute to this problem of generating higher quality images.

Why VAE Performed Better

- Segmented Latent Space: Contributes to the enhanced feature representation in this study as it helps to achieve better quality reconstructions of the image.
- This prevented overfitting and ensures that the latent space M is continuous with its neighbors.
- Stable Training: VAEs are relatively easier to train as compared to GANs since it does not involve the alternate minimization of adversarial loss.

Suggested Improvements

- Find ways to improve GAN generator to produce higher quality images and one of the ways is to add more convolutions to the network.
- Try to improve the learning rate and make further modifications to the batch normalization of GAN for the balance of training.
- Introduce the shallow implementation of WGAN to enhance the training stability.
- Increase the dimension of the VAE's latent space for reconstruction in order to retain fine details.

5. Fraud Detection Using VAE

Another model used in this work is called the Variational Autoencoder, where fraud was detected in credit card transactions dataset.

Process:

- Data preparation: Data scaling was done to complete standardization; the dataset was made up of normal and fraudulent transactions.
- Training: VAE was trained on normal transactions with an aim of familiarizing with normal or typical transaction patten.

A reconstruction error was used to detect fraud as it is evident that fraudulent patterns have high variations compared to the normal patterns.

Results:

These assumptions are consistent with the results obtained in the evaluation of the experiment, that the model has successfully detected 388 out of 492 fraudulent transactions.

IT Thresholds: Based on the analysis of the distribution of the score of normal transactions, the anomaly threshold was set to the 95th percentile of the reconstruction loss.

Conclusion

GANs and VAEs are two forms of generative models that should be differentiated based on their utility:

Conventional GANs have very good abilities to synthesize realistic images whereas they suffer from the issue of instability. The encoders and decoders that are used in VAEs generate coherent outputs and can be used in anomaly detection because of their probabilistic nature. Future work include improvement of structural and numerical stability of GAN and its comparison with other hybrid model like VAE-GAN could give better results.