Report

CSI 5138 Homework Exercise 4

Pavan Balaji Kumar

Student number: 300169572

In this assignment, an unsupervised model for image generation Variational Autoencoder, a GAN and WGAN, were implemented and used to generate new images based on the MNIST dataset and the CIFAR10 dataset. The results obtained from the models are discussed in this report.

All the terminologies used in the assignment is listed below:

- mnist_vae_x: VAE models trained on mnist dataset where x denotes model number
- cifar_10_vae_x: VAE models trained on cifar_10 dataset where x denotes model number
- Discriminator_x: Discriminator of GAN_x
- Generator_x: Generator of GAN_x
- GAN_x: GAN trained on mnist and cifar_10 dataset where x denotes model number
- input_shape: Shape of the training dataset
- mnist_train, mnist_trainlabel: training data of MNIST dataset
- mnist_test, mnist_testlabel: testing data of MNIST dataset
- cifar_10_train, cifar_10_trainlabel: training data for CIFAR10 dataset
- cifar_10_test, cifar_10_testlabel: testing data for CIFAR10 dataset

Both MNIST and CIFAR10 are image datasets where MNIST contains handwritten images, while CIFAR10 contains various object images. A Variation Autoencoder, a Generative adversarial network and a Wassermann Generative adversarial network were built to generate new images using The TensorFlow framework.

Different loss function was used to optimize the three models. KL divergence loss was used to optimize the VAE. JS divergence loss was used to optimize GAN, and EMD loss was used to optimize WGAN. The loss against each epoch is plotted and discussed in the report.

Let us study the performance of the classifiers on the two datasets.

1. Performance of Variational Autoencoder model on the two datasets.

Ten different VAE were trained on the MNIST dataset. First 5 with different complexity, and the last 5 with different latent space. The models were optimized using Adam optimizer, and the loss was calculated using KL divergence loss. All the VAE models were run for ten epochs on the MNIST dataset. The results obtained from models can be seen in figure 1.

	Hidden Layers	Latent Dimesnion	KLD loss
1	2.0	2.0	0.177
2	4.0	2.0	0.179
3	8.0	2.0	0.186
4	10.0	2.0	0.217
5	12.0	2.0	0.217
6	4.0	10.0	0.108
7	4.0	20.0	0.099
8	4.0	30.0	0.102
9	4.0	50.0	0.101
10	4.0	75.0	0.102

Figure 1. VAE results

Figure 1. shows the loss obtained for different configurations of the hidden layers and latent space. As seen in the figure, it is clear from the results that when the number of the hidden layer is increased without change the latent space of the model, the loss increases gradually. Hence having a deep VAE will not result in better performance. However, it is not the case with latent space increasing the latent space decreases the loss gradually. Hence the more the latent dimension, the better performance.

The gradual increase and decrease of loss associated with the increase in no hidden layers and latent space plotted for each epoch of the training in the below figure. It can be seen that after a specific latent space, there is not a drastic increase in the performance of the models. The images generated by the model can be seen in figure 3. The images are generated from random latent points, and the results are good as the digits are recognizable.

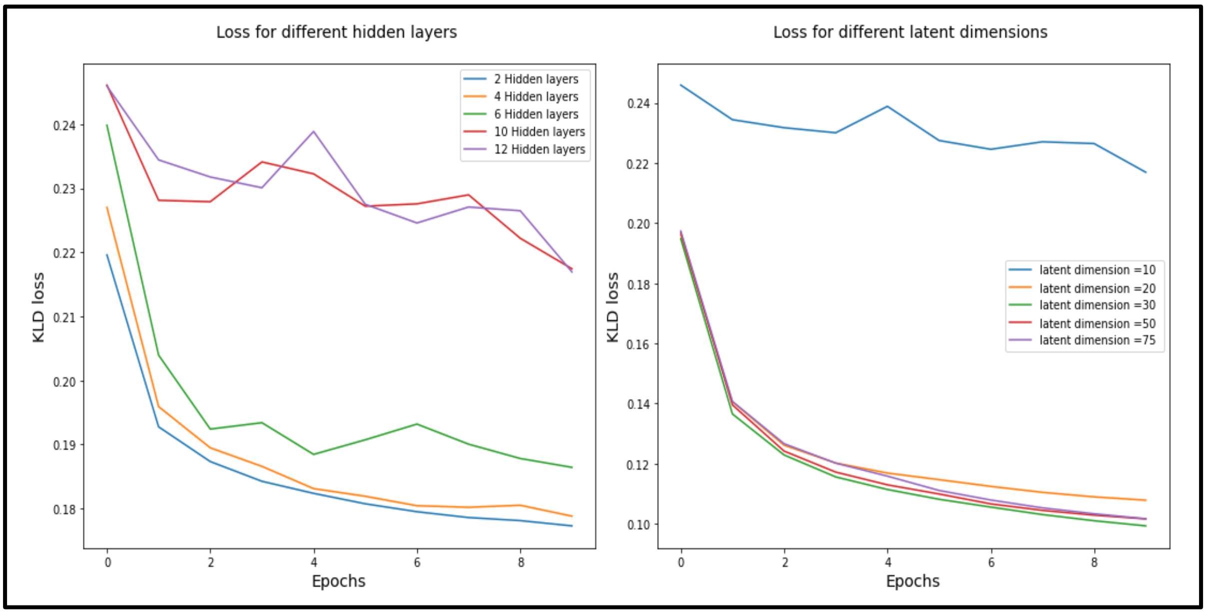

Figure 2. The plot of KLD loss for different hidden layers and latent space

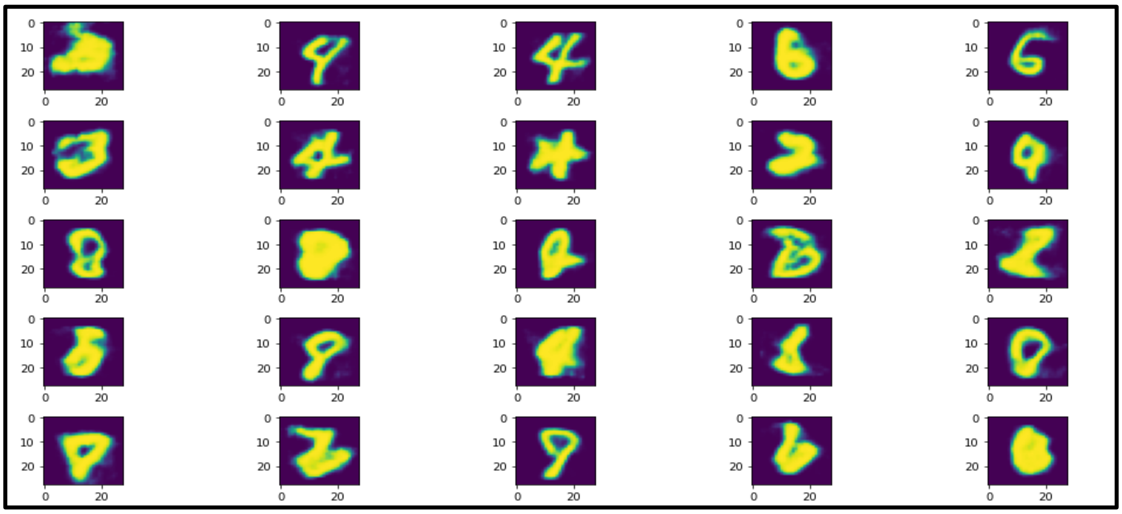

Figure 3 Images generated by VAE based on MNIST

Just like for the MNIST dataset, ten different VAE were trained on the CIFAR_10 dataset. First 5 with different complexity, and the last 5 with different latent space. The models were optimized, and the loss was calculated using the same optimizer and loss as for the MNIST dataset. All the VAE models were run for 30 epochs on the CIFAR_10 dataset. The results obtained from models are plotted, as seen in figure 4.

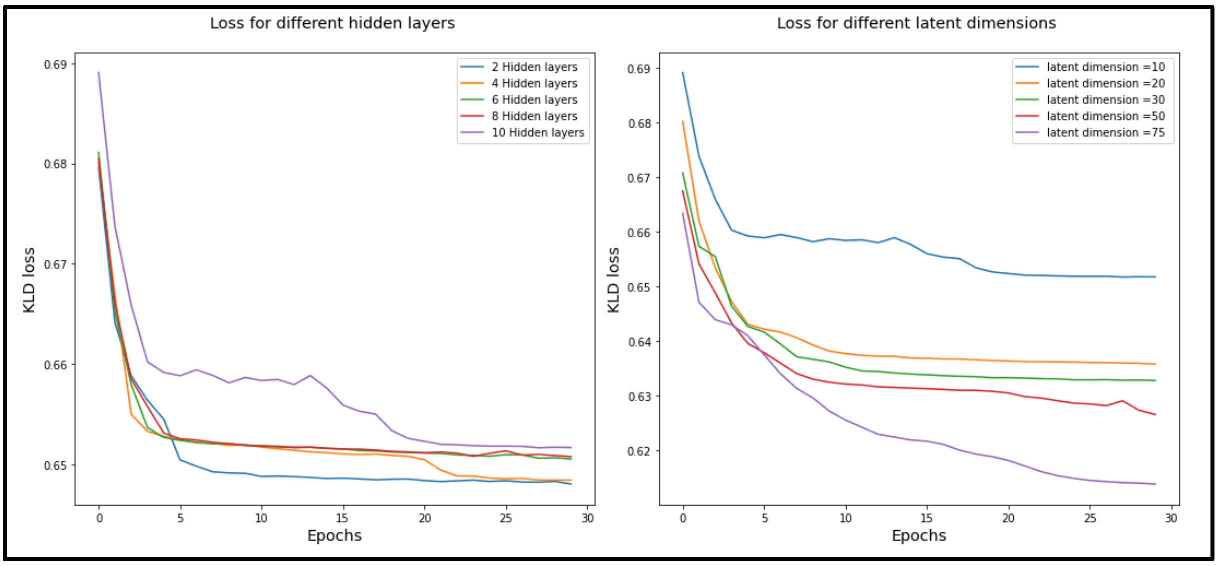

Figure 4 Plot of KLD loss for different hidden layers and latent space CIFAR_10

The figure above shows that the model exhibits the same behaviour as it did with the MNIST dataset. When the VAE is trained with the CIFAR_10 dataset, the change in loss is similar to MNIST when the number of the hidden layer increases or the latent space is increased. The images generated by the VAE based on the CIFAR_10 dataset is given in the figure below.

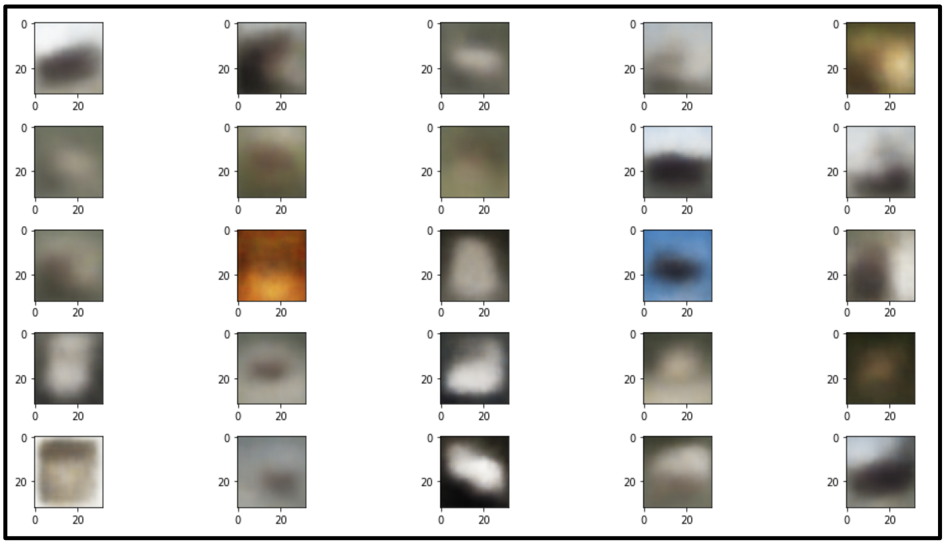

Figure 5 Images generated by VAE based on CIFAR_10

The images plotted in the figure are based on the images from the cifar_10 data. Though the images are not apparent to the naked eyes, it can be considered a good result because the models were not trained using HD images, but only images of the 32-bit size and hence the generated resolution is also less. The generated images are somewhat recognizable.

2. Performance of GAN model on the two datasets.

Five GAN models with different latent dimensions were trained for each MNIST and Cifar_10 dataset. Adam optimizer with learning rate 0.0002 and beta_1 0.5 as hyperparameter was used to optimize the model. Jenson-Shannon Divergence was used as the loss function. Each model was trained for 30 epochs with a batch size of 256. The results obtained for the model training can be in Figure 6.

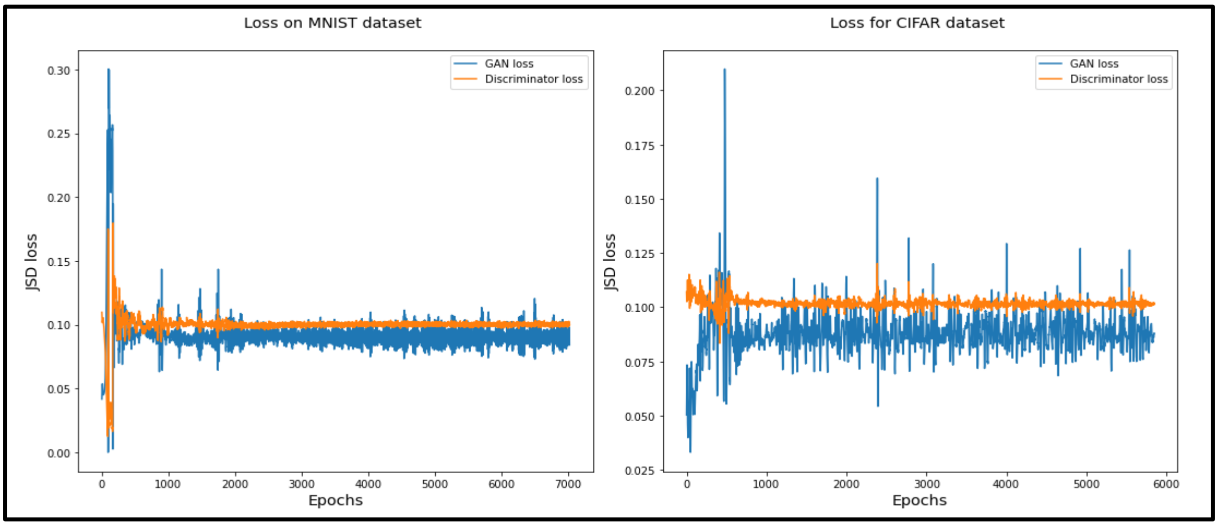

Figure 6 Discriminator and GAN loss against epochs on different datasets

Figure 6 clearly shows that there is no clear trend in the optimization of the loss function throughout the training epochs. Though the GAN loss is very high at the beginning of the interval between both, the loss function closes, and both travel around the same path. As seen in the figure above, this is the case for both the MNIST dataset and CIFAR_10 dataset. The images generated by the model based on the MNIST dataset is plotted in the figure below.

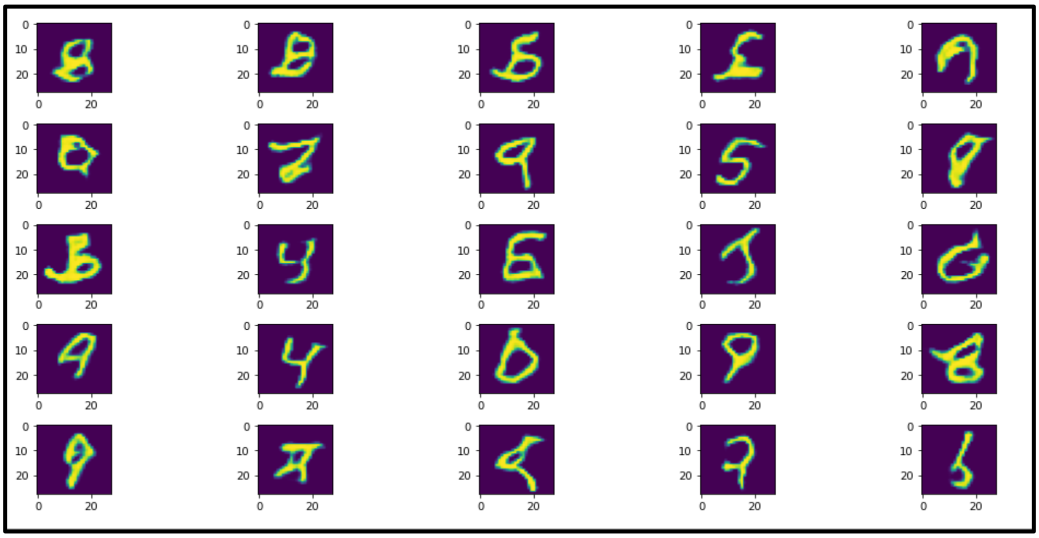

Figure 7 Images generated by GAN based on MNIST

The images generated by GAN based on the MNIST dataset are generated without any distortion. The digits are visible to the naked eye, and most of the generated resemble the actual handwritten digits in the dataset. It is an excellent output considering only noise was fed into the GAN generator based on the latent dimension. However, This is not the case with GAN models trained on the CIFAR_10 dataset. The models trained on the CIFAR_10 dataset produces images that are so blurred that it is difficult to grasp from them what it represents. The image generated can be seen in the figure below.

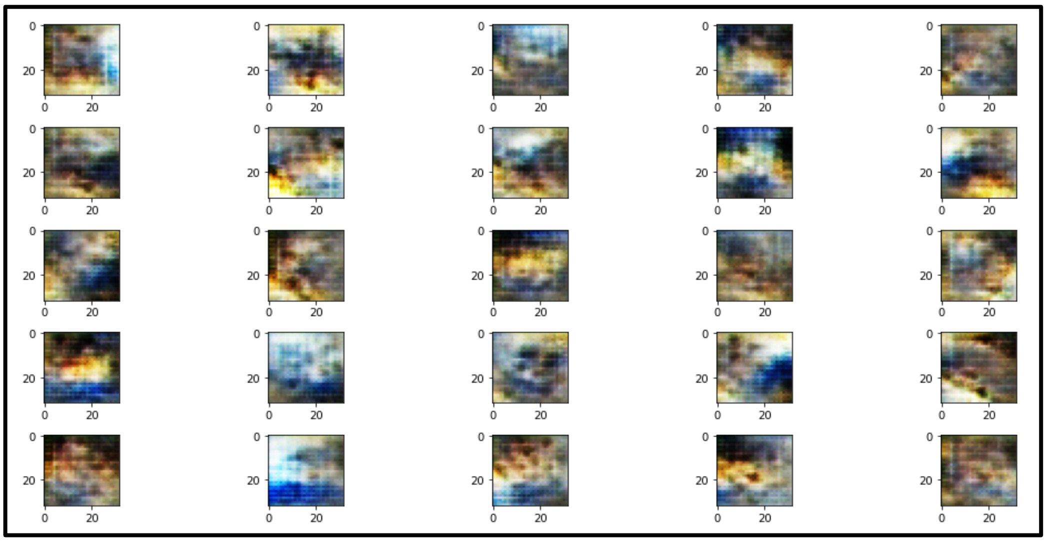

Figure 8 Images generated by GAN based on CIFAR_10

It was observed that for higher latent space, like 750 or 1000, the MNIST dataset was not able to produce a clear image as it produced for the latent space like 350 or 500.

3. Performance of WGAN model on the two datasets.

Five WGAN models with different latent dimensions were trained for each MNIST and Cifar_10 dataset. RMSprop optimizer, with a learning rate of 0.0005 was used to optimize the model. EMD loss was used as the loss function. Each model was trained for 1000 epochs. The results obtained for the model training can be Figure 9.

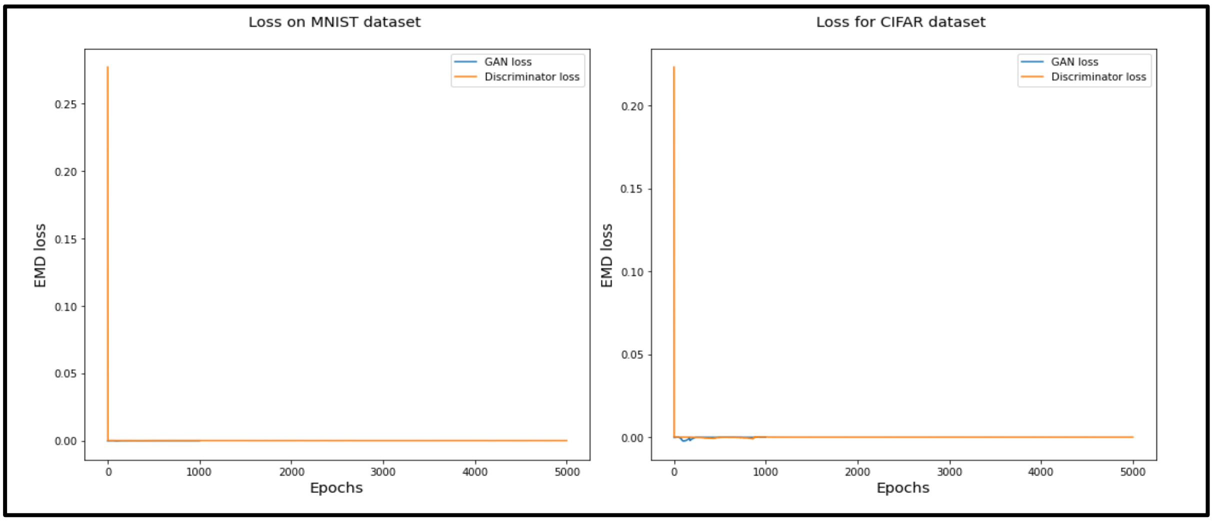

Figure 9 Critic and WGAN loss against epochs on different datasets

It can be seen from the figure that the loss function for both the Critic and the WGAN are almost similar and remain the same throughout the training of the model. The loss trend is the same on both the dataset even though the WGAN is going below the critic loss for a few epochs. The images generated for the MNIST dataset are somewhat clear, and some images were recognizable and similar to that of the images in the dataset. The generated images are shown in figure 10.

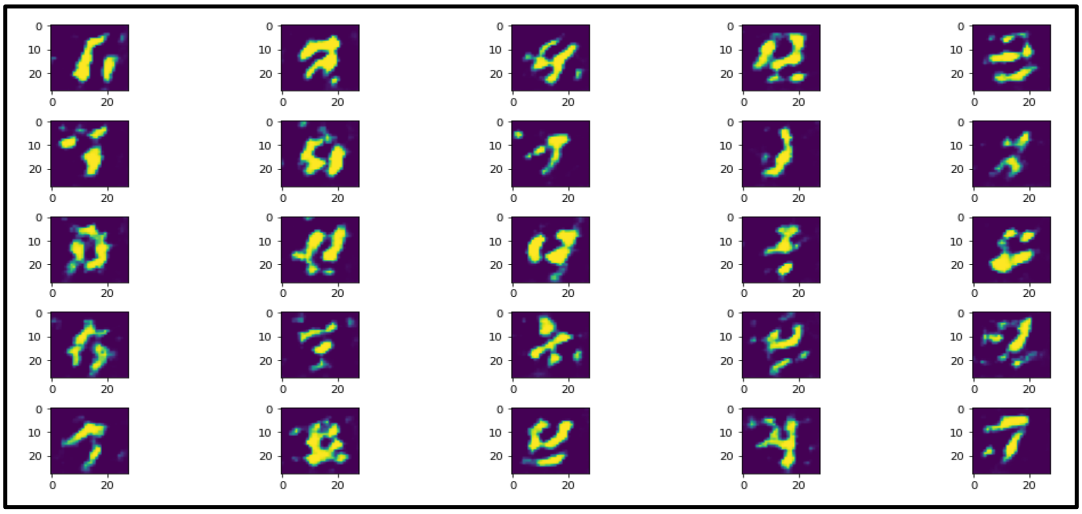

Figure 10 Images generated by WGAN based on MNIST

The WGAN model did not perform well on the CIFAR_10. The various model that was constructed failed to learn to generate images from the dataset. The generated images did not contain any information in them and were not recognizable.

4. Conclusion:

In this assignment, new images were generated using all existing images in the MNSIT and CIFAR_10 dataset. Out of all the model trained, VAE with latent space 1000 and 4 hidden layers produced the best-generated images for the CIFAR_10 dataset. For the MNIST dataset, the best images were produced by Gan with latent space 350. Generating new images maybe be helpful in data augmentation and help where there is a low amount of data available. With more compute power and advanced algorithms, images with better quality can be generated.