TU/e

# 5LSL0 Assignment 3: Variational Autoencoders


| **Name** | **Student ID** |
|---|---|
| Detian Guo | 1662341 |
| Yicheng Zang | 1621114 |


Eindhoven, June 3, 2022

# I. Deterministic autoencoder

*Exercise 1*

The first 10 images of 'example_clean' and corresponding output is shown in Fig1.

Fig. 1: 'example_clean' and corresponding output images

The output images of the autoencoder is less clear than the input one. And many of the digits can not be recognised, only 0,1 and 9 preserve the original structure.

The plot of the loss is shown in Fig2.

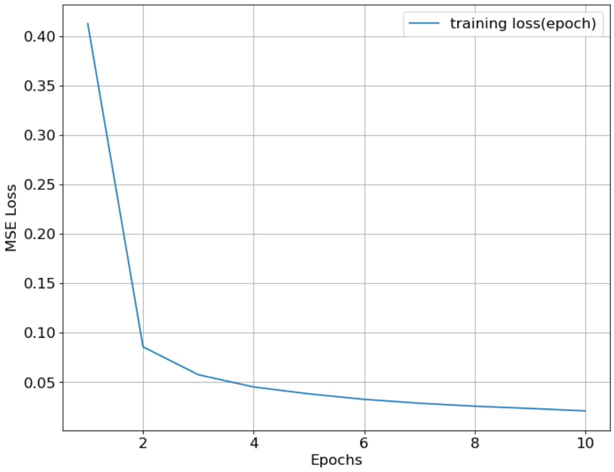


Fig. 2: Training Loss - Epochs

*Exercise 2*

*(a):* The scatter plot is shown in Fig 3.

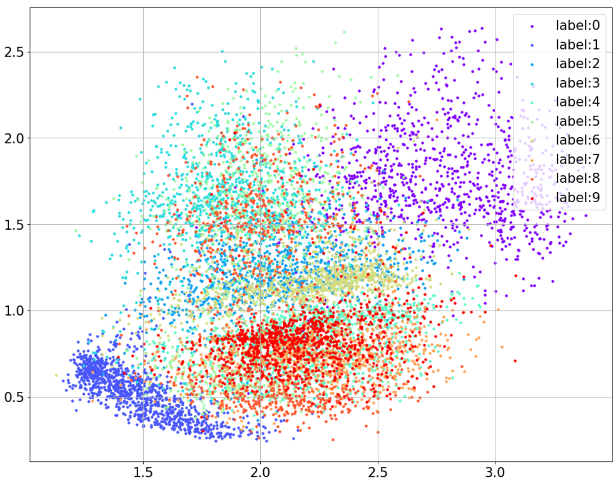


Fig. 3: The Visualized Latent Space Plot of the Autoencoder

| | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Accuracy | 81% | 82% | 34% | 45% | 41% |
| | 5 | 6 | 7 | 8 | 9 |
| Accuracy | 26% | 36% | 35% | 20% | 28% |

TABLE I: Classification accuracy for 10 digits with only convolution layers

*(b):* In 10 digits, only 0, 1 provide clear clusters, and the other digits are both have a large overlap with each other. The reason is semantic vector in latent space cannot fully represent the information of the whole sequence. However, 0 and 1 have a unique structure, and the feature information can be saved in the semantic vector. Thus 0 and 1 provide clear clusters.

*(c):* Because the weight is zero, thus there is no input to the latent space. In this way, some of the clusters are clipped at weight of zero.

*Exercise 3*

*(a):* The accuracy as percentages of the total number of zeros, ones, twos, etc. is shown in TableI. The average accuracy over all 10 digits is only 42.7%.

*(b):* We expect 430 correct classifications. The correct rate is 43.5%.

1 and 0 will be classified significantly better than the other. However, 5 and 8 have the lowest accuracy of the 10 numbers, so they seems to be classified worse.

Combined with the plot of the latent (Fig 3), 1 and 0 shows clear clusters. And 5 and 8 both overlap with others. For other numbers, the classification accuracy is negative correlated with the cluster overlap area in the plot. The more overlap in the plot, the less accuracy the number has.

*(c):* The cross entropy loss will be used for a multi-class classification task.

*Exercise 4*

*(a):* The average accuracy overall all 10 digits now comes to 97.8%, which is much higher than the previous.

*(b):* Fig 4 shows the training and test loss. After the 2nd epoch, The number of the cross entropy loss is optimized to 0.06 and 0.08 for test dataset and train dataset respectively. This means that the distance between the classification distribution generated from our network and the ground truth is close.

*(c):* It can be found that the network is overfitting from the Fig 4. It is possible to end the training early by monitoring the test-loss. We can also reduce the complexity of our network, e.g. reduce the neurons in fully-connected layer, to avoid overfitting.

*Exercise 5*

*(a):* With the inputs were samples from the latent space, the picture was created, i.e., the visualised latent space of the autoencoder. The picture is shown as Fig. 5

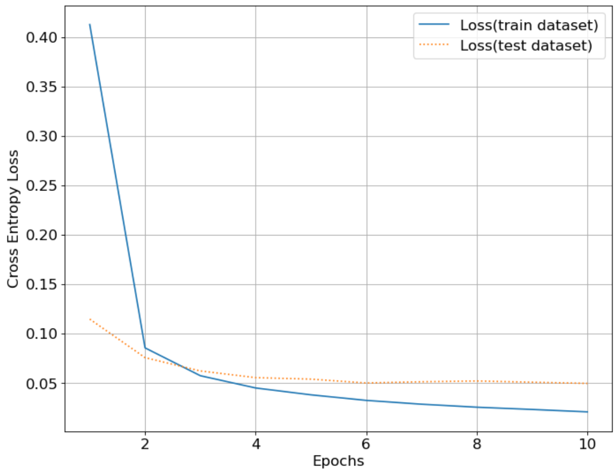


Fig. 4: Training and Test Loss - Epochs (Autoencoder)

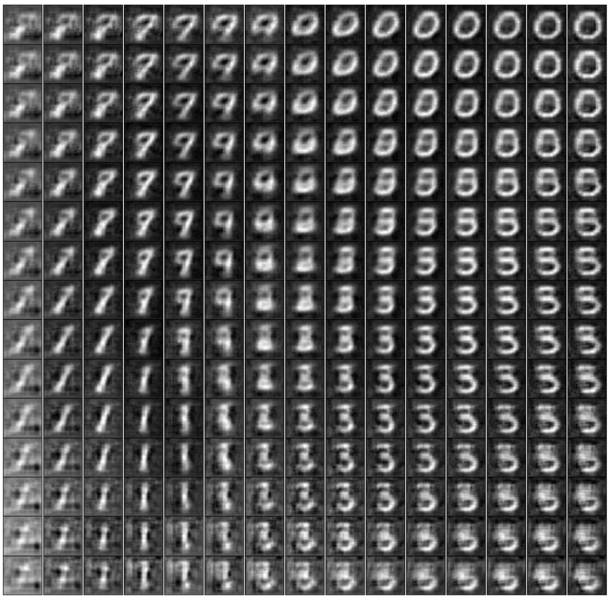

Fig. 5: Visualised Latent Space of the Autoencoder

*(b):* The digits 1, 3, 8, 9, 0 can be recognized. However, others not. 1 and 0 have unique structure, thus their feature information can be saved in latent space. For 3, 8 and 9, they have the same feature with other numbers. And the first entered information in latent space will be diluted by what is entered later. Thus all images have trends to mix and the result is 3, 8 and 9.

*Exercise 6*

*(a):* Use the ‘x_noisy_example‘ as input to the autoencoder. We generated an image, Fig. 6 Row 1 shows the noisy input, row 2 shows the denoising output, and row 3 shows the corresponding clean image ‘x_clean_example‘.

We notice that the noise is canceled. However, the autoencoder failed to recover some digits.
We think the trained autoencoder learned the distribution of clean inputs. So this network can perform the denoise task well. As for the digit inpainting task, the information entropy

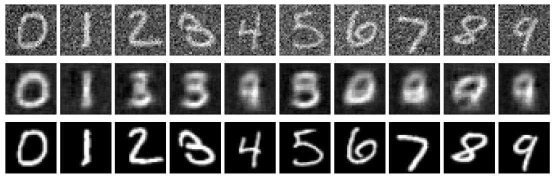

Fig. 6: Noisy Input, Denoising Output by Autoencoder and Corresponding Clean image

in the latent space is limited by only 2 dimensions in the latent space. In other words, the result will be better if there are more dimensions in the latent space.

## II. Variational Autoencoder

*Exercise 7*

*(a):* Fig. 7 shows the visualized images in ‘example_clean‘ and the respective output images when they mapped through the VAE.

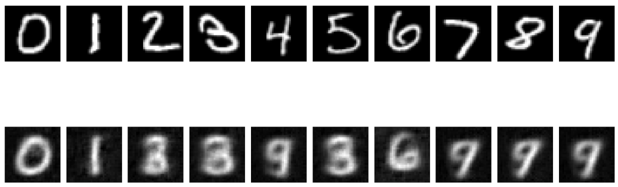

Fig. 7: Visualized Input and Output of ‘example_clean‘ Mapped Through the VAE

Moreover, the loss plot as a function of the number of epochs is shown in Fig. 8.

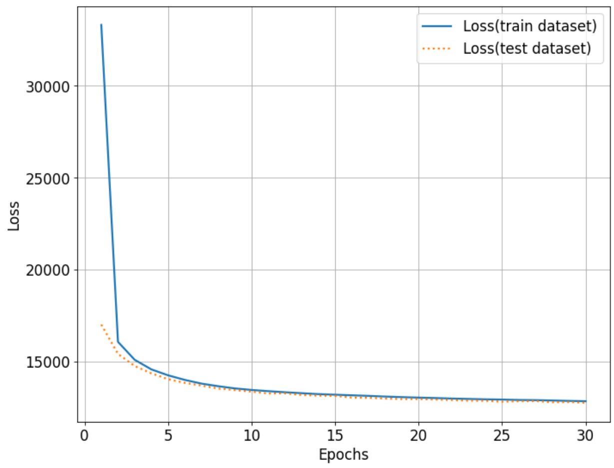


Fig. 8: Training and Test Loss - Epochs Plot (VAE)

*(b):* The visualized latent space of the variational encoder is shown in Fig. 9

*(c):* The vcalues that are adjacent to each other in the Latent Space correspond to very similar reconfigurations. Such as 3, 8, 9 have a large overlap.

*(d):* The result of repeat the exercise 5a is shown in Fig. 10

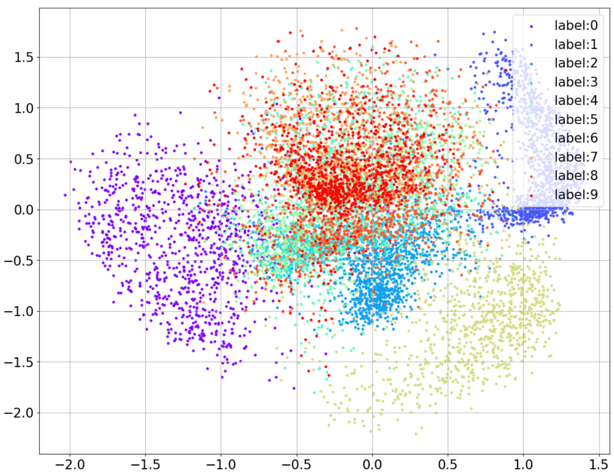


Fig. 9: The Visualized Latent Space Plot of the VAE

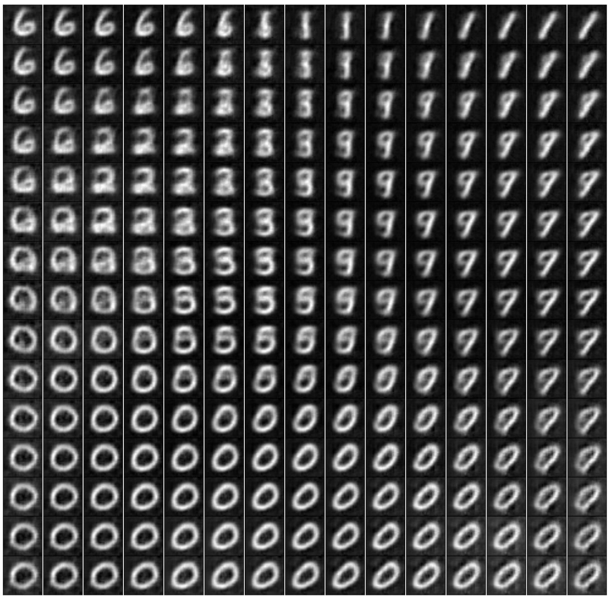

Fig. 10: Visualised Latent Space of the VAE

*(e):* The images of VAE latent space is clearer than the images of Autoencoder.

Autoencoder uses a single value to describe the performance of the input image in terms of latent features. VAE is a model that uses a "probability distribution of values" to describe the observation of features instead of the original single values.

## III. Denoising with VAEs

*Exercise 8*

*(a):* The output of using the new VAE to repeated exercise 6a is shown in Fig. 11

*(b):* The noisy measurements that we used as input, is shown together with the reconstructions after implement equation 4 and the ground truth for the example images in Fig. 12. The MAP loss is shown in Fig.13

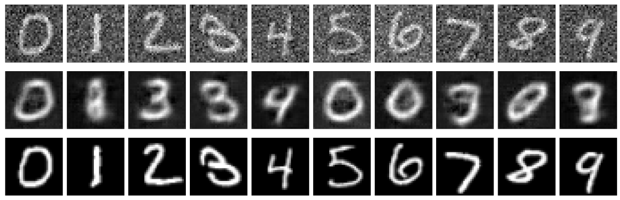

Fig. 11: Noisy Input, Denoising Output by New VAE and Corresponding Clean image

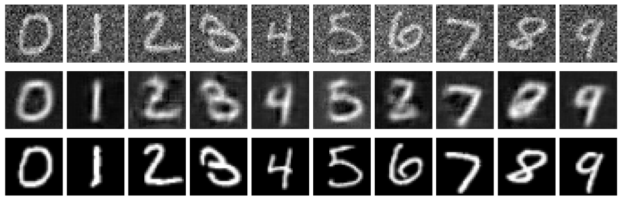

Fig. 12: Noisy Input, Reconstructions after Implement Equation 4 and the Ground Truth

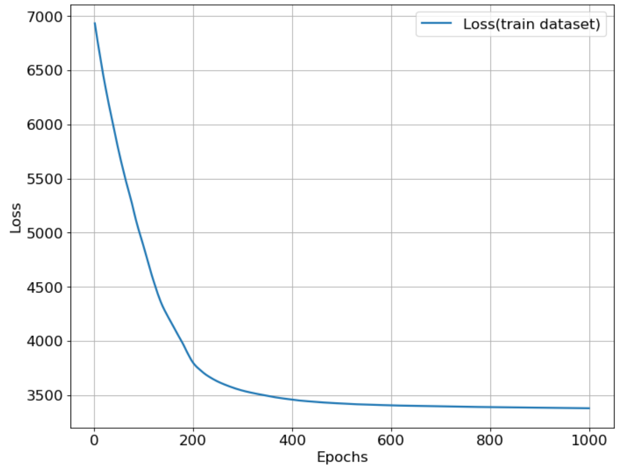


Fig. 13: MAP training loss with 1000 iterations

*(c):* The second method is better. The reason for that are as follow,

1) It is simpler to optimize the values in the latent space.
2) The first method is trained on train dataset. The second method is trained directly on the test dataset.
3) MAP uses a trained decoder and trained for another 1000 iterations.