Assignment 3

1 Overview

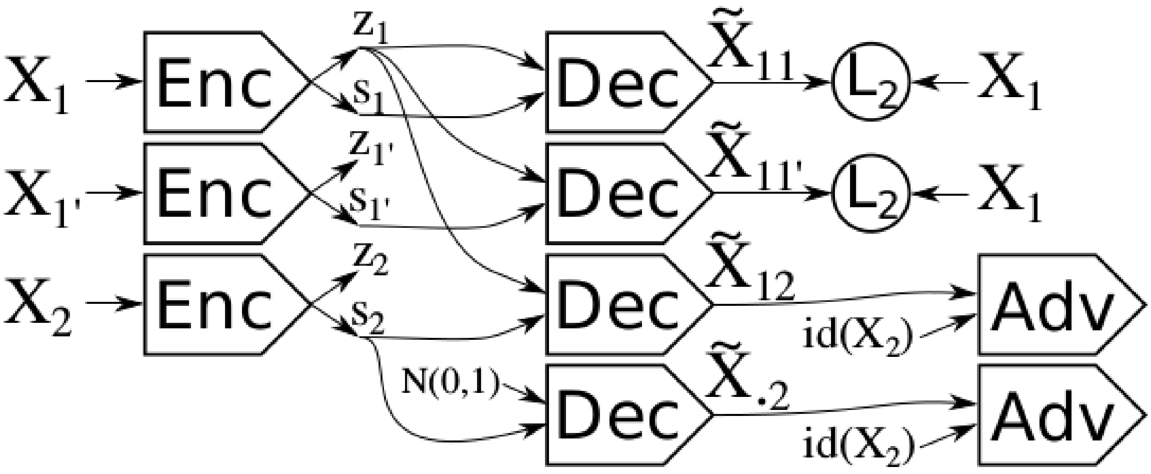

Figure 1: Mathieu et al.

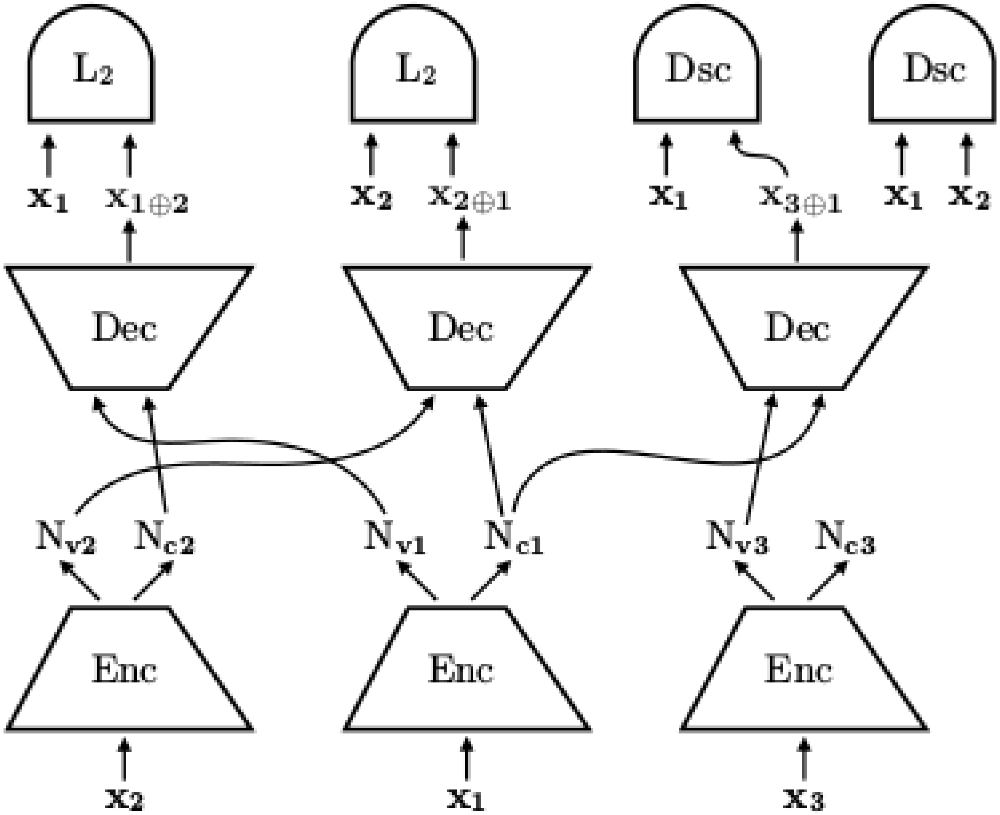

Figure 2: Szab et al.

Here, we keep the training procedure same as Mathieu et al., but change the discriminator according to Szab et al. This is done as appending lookup tables or embedding of identity of image is bit unstable, so complete image is sent instead of embedding.

Shortcut Problem: According to Mathieu et al. implementation, to avoid shortcut problem when all information is passed through style space itself, one do adversarial training such that specified space has information of the class.

2 Implementation

Algorithm for model training

for number of training iterations do
Training generative model
1. take three sample images from two different classes s_1,s_2.
2. pass all three sample images through the encoder to get (μ_i,σ_i,s_i) where (μ_i,σ_i) is denoted by z_i which is called as unspecified latent space and s_i as specified latent space, i ∈ sample number.
3. output of encoder is passed through decoder for image reconstruction for images with the

same class.
4. loss between original image and decoder output is backpropagated.
5. now pass the z_1 with second class label to the decoder for computing adversarial losss and backpropagte the gradients, and keep the Adv frozen.
6. now take a normalized sample (0,1) and repeat step 5.
Now for training adversary
1. take two sample images from two classes.
2. compute (z_i,s_i) for both the sample through encoder.
3. through these encoder output reconstruct the images through decoder.
4. now compute the adversarial loss for both negative and positive sample and backpropagate the gradients keeping encoder and decoder frozen.

end

3 Results

3.0.1 Celeba Disentangled

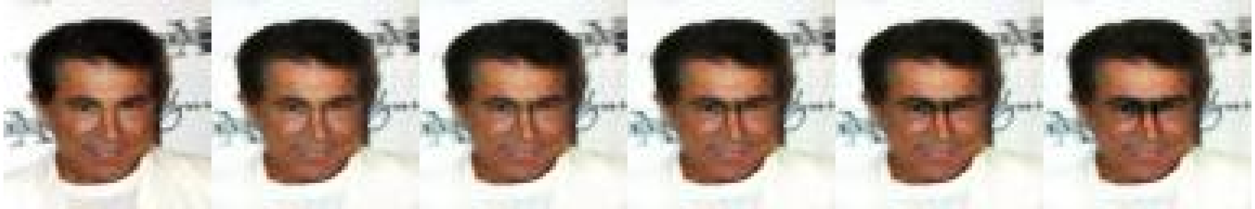

Figure 3: Interpolation (Eyeglasses)

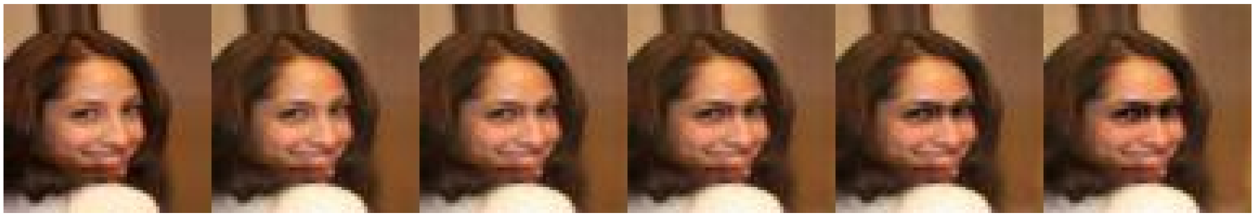

Figure 4: Interpolation (Eyeglasses)

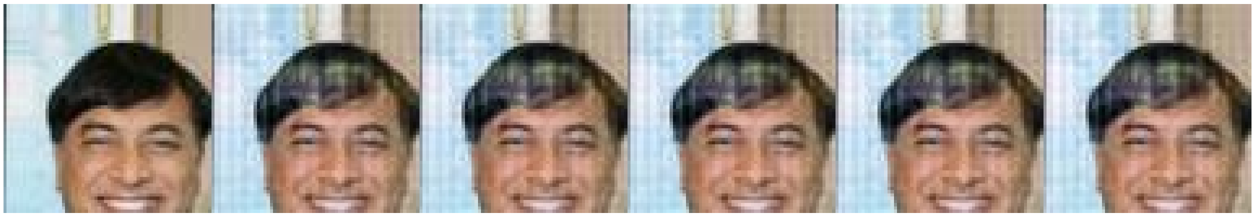

Figure 5: Interpolation (Blonde Hair)

3.0.2 MNIST Disentangled

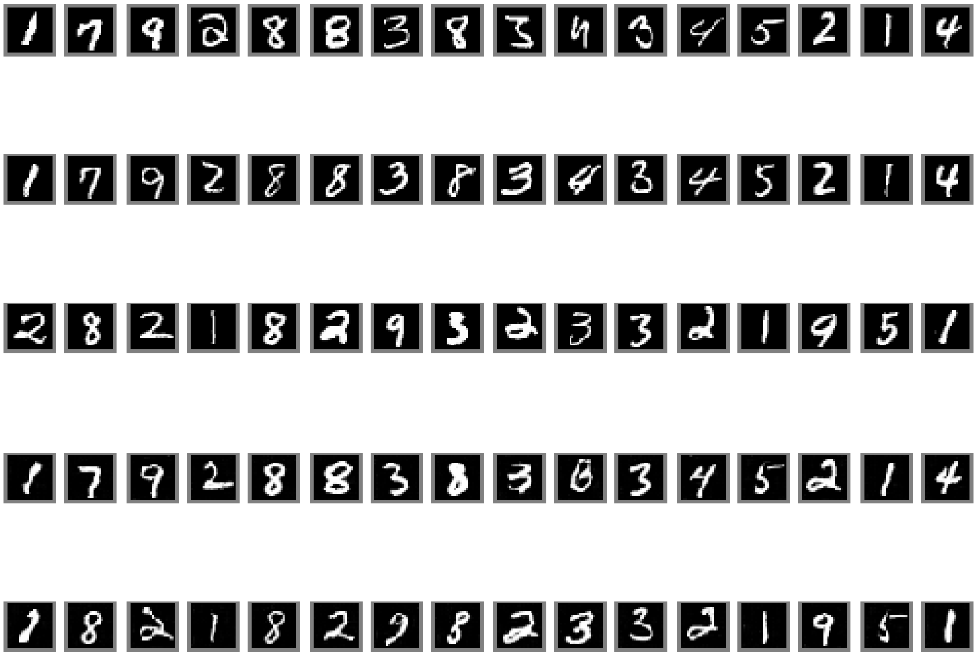

Figure 6: Recreated image

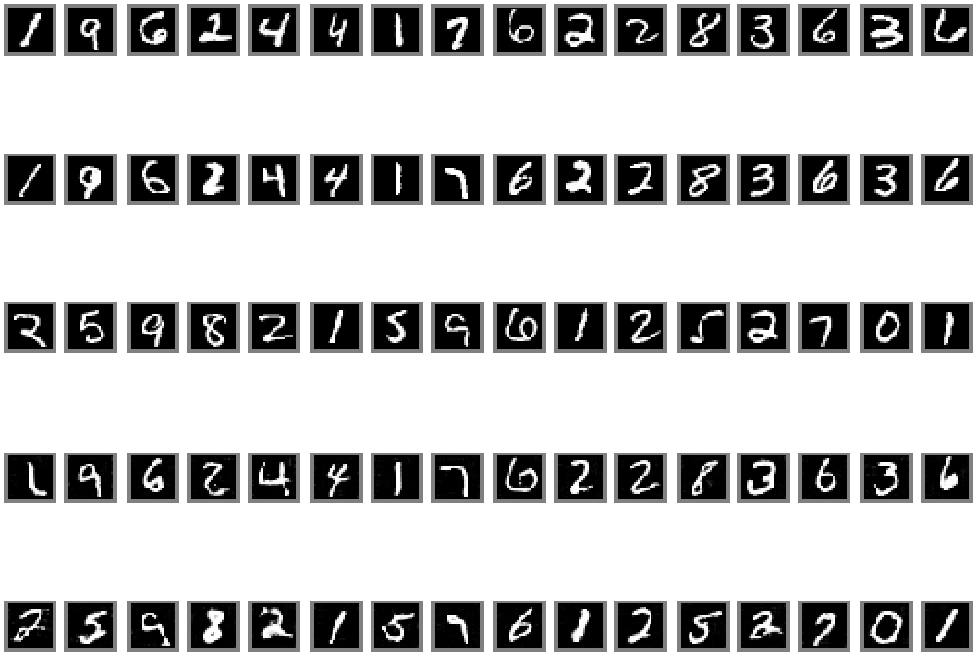

Figure 7: Recreated image

Here 1st 3 are input, while last 2 is swapped image.

3.1 Unspecified space trained as a VAE

3.1.1 TSNE

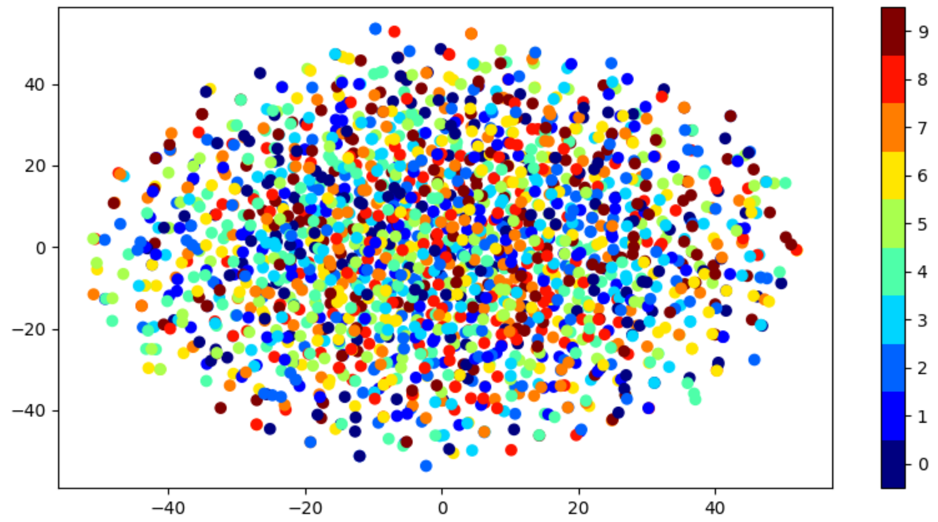

Figure 8: Unspecified Space

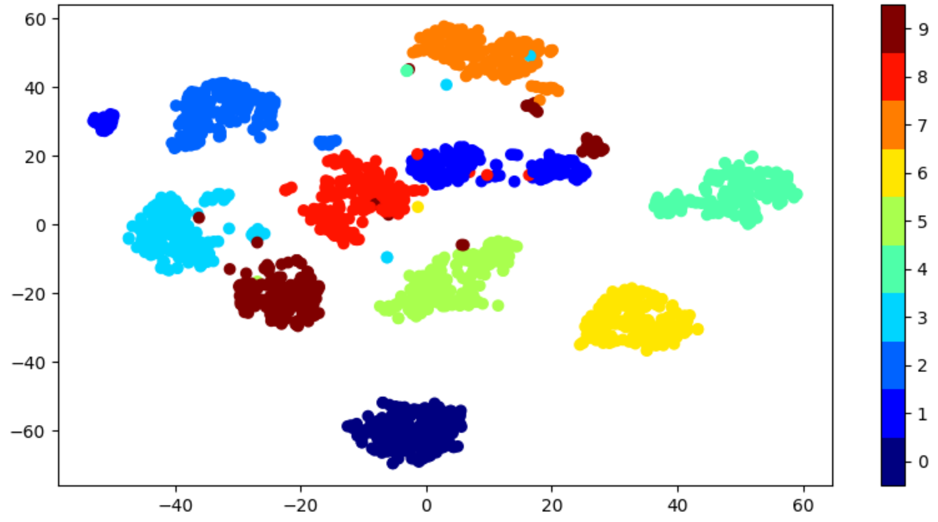

Figure 9: Specified Space

Here, we see that class and latent space is strongly correlated in specified space, however same thing is not true for unspecified space.

3.1.2 Classifier

Specified Space

Classifier	**Accuracy**	**F1 score**
MLP	0.866	0.865
SVM	0.885	0.886

Table 1: Performance of Various Classifiers

Unspecified Space

Classifier	**Accuracy**	**F1 score**
MLP	0.103	0.031
SVM	0.125	0.095

Table 2: Performance of Various Classifiers

3.2 Specified space trained as a VAE

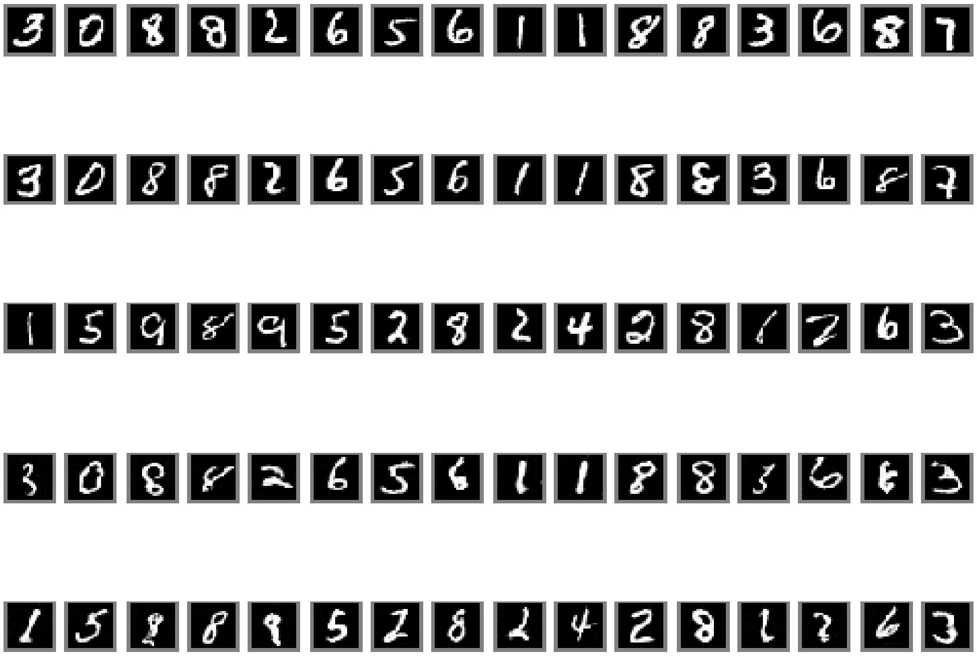

Figure 10: Recreated image

3.2.1 TSNE

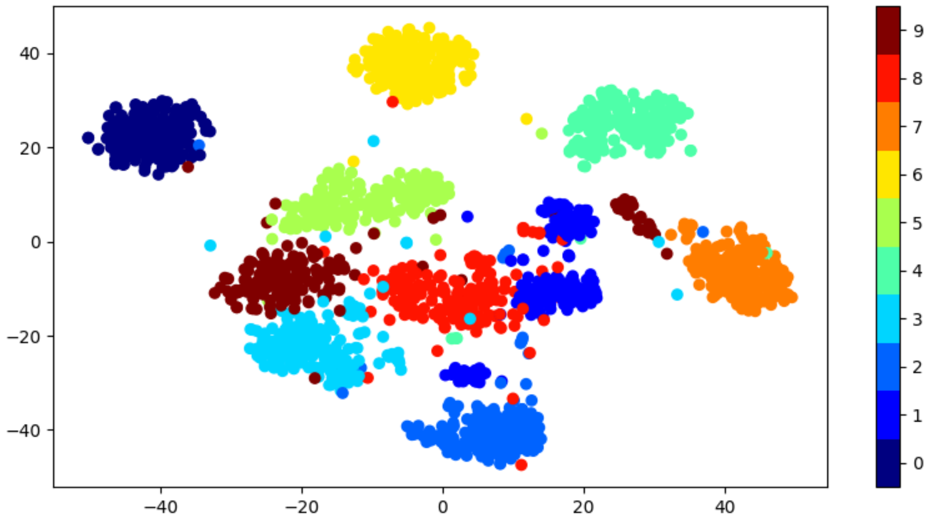

Figure 11: Specified Space

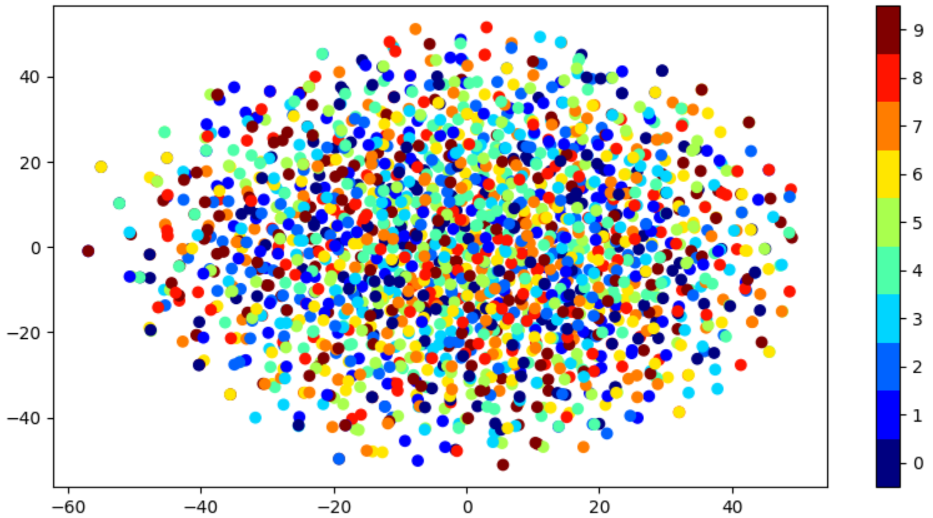

Figure 12: Unpecified Space

We observe that as space is modelled as a Gaussian, there is bit information loss and hence bit less class performance on specified space.

3.2.2 Classifier

Specified Space

Classifier	Accuracy	F1 score
MLP	0.791	0.789
SVM	0.816	0.814

Table 3: Performance of Various Classifiers

Unspecified Space

Classifier	Accuracy	F1 score
MLP	0.116	0.051
SVM	0.108	0.053

Table 4: Performance of Various Classifiers

4 Reference

https://github.com/MichaelMathieu/factors-variation

https://github.com/carpedm20/DCGAN-tensorflow

https://github.com/hwalsuklee/tensorflow-mnist-VAE

https://github.com/Prinsphield/GeneGAN