Notes & Conclusions

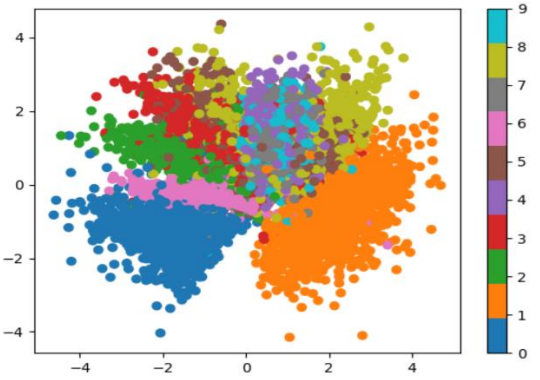
For the continuous VAE I used 2 layers, one for the encoder(class) and the other for the decoder(class), the encoder layers have 2 fully connected layers which return a latent vector consisting of μ & σ. During the forward, I apply relu and the reparameterization trick on μ & σ. The decoder has 2 fully connected layers and applies relu and sigmoid during the forward, the decoder returns a new generated image. In the training the loss has 2 expressions, binary cross entropy of the original image and the generated one and the KL divergence of the latent vector.

The discrete VAE has an encoder which has 3 fully connected layers applying relu on each one, and a decoder which has 3 fully connected layers applying relu on each one and sigmoid at the last one, it uses the Gumbel-Softmax reparameterization trick in the forward. The forward returns a generated image and a predicted category. In the training the loss has 2 expressions, binary cross entropy of the original image and the generated one and the KL divergence of the latent vector of the categories.

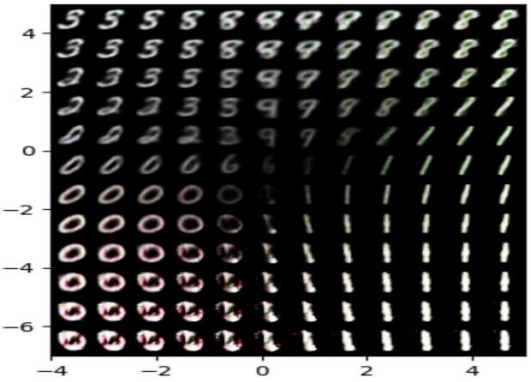
The both VAE which has both continuous & discrete latent spaces, consists of both the previous continuous VAE & discrete VAE and is similar to them except that when it comes to training, the loss is the sum of the previous 2 losses. The final generated image is the mean of the 2 generated images of the continuous VAE & discrete VAE.

All the models were used with the following hyper-parameters:
learning_rate=0.001, num_epochs=20, batch_size=28.
N=3, K=20, temp=1.0 for the models that use discrete.

Continuous VAE performance:
Latent space:



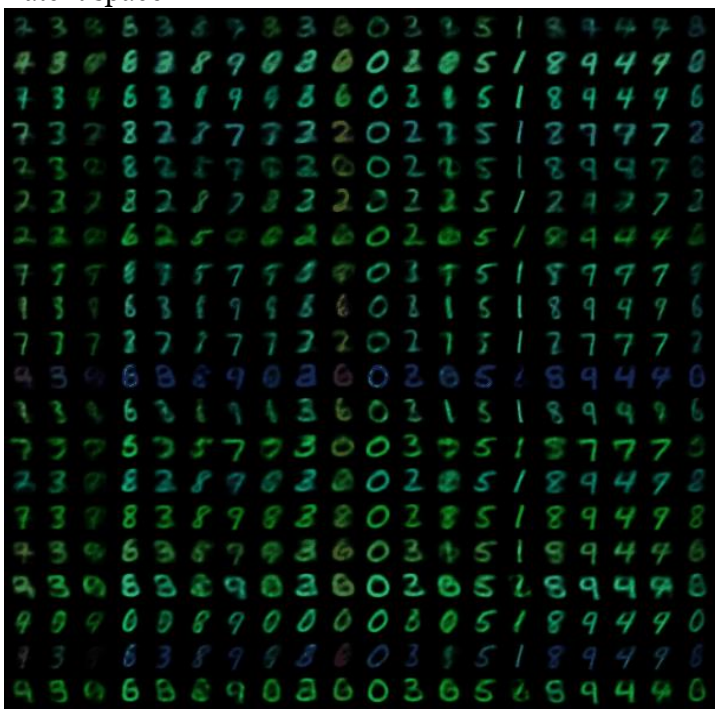
Reconstructed latent space:



Original test batch images:

Generated test batch images:

Discrete VAE performance:
Latent space:

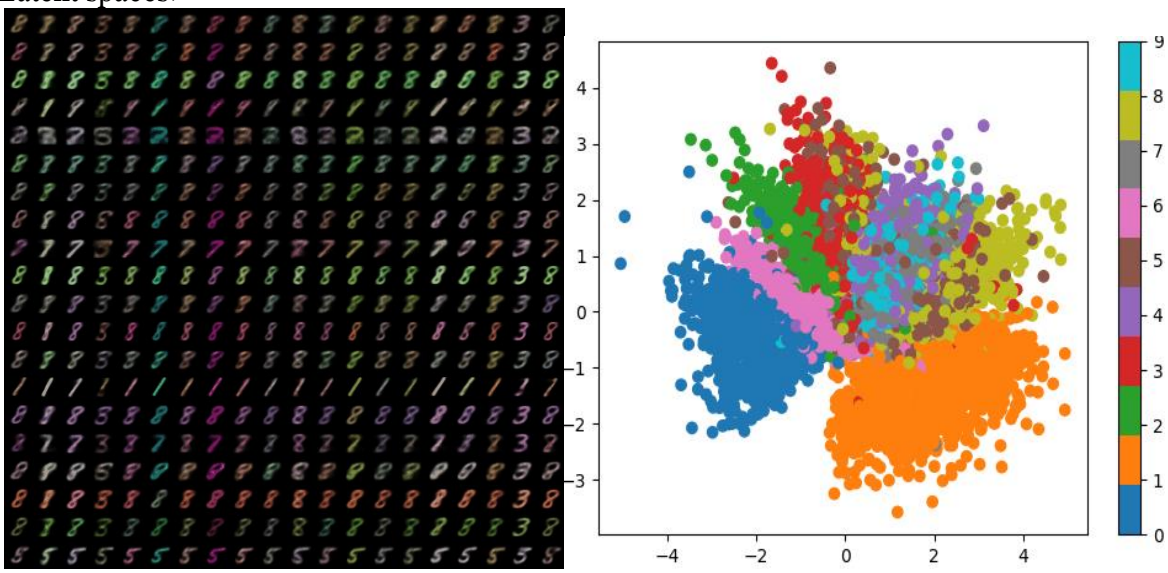


Original test batch images:

Generated test batch images:

Both VAE performance:

Latent spaces:



Original test batch images:

Generated test batch images:

Conclusions:
By looking on the test batches visualizations it seems that the best performing VAE is the discrete VAE, the discrete VAE seems to capture colors in a better way than the continuous VAE and the "both VAE". In the "both VAE" the continuous latent varaibles seem to hinder the discrete latent variables ability. As we can see in the reconstructed latent space visualization of the Continuous VAE , it seems that the VAE doesn't capture the colors well enough, we can understand that by the whiteness of the digits which is a result of the accumlation of all the colors in the same space. Interestingly though the latent space of the "both VAE" has more color variation in it than the other two VAEs. We can choose a color and a digit by picking a specific coordinate in the latent space.