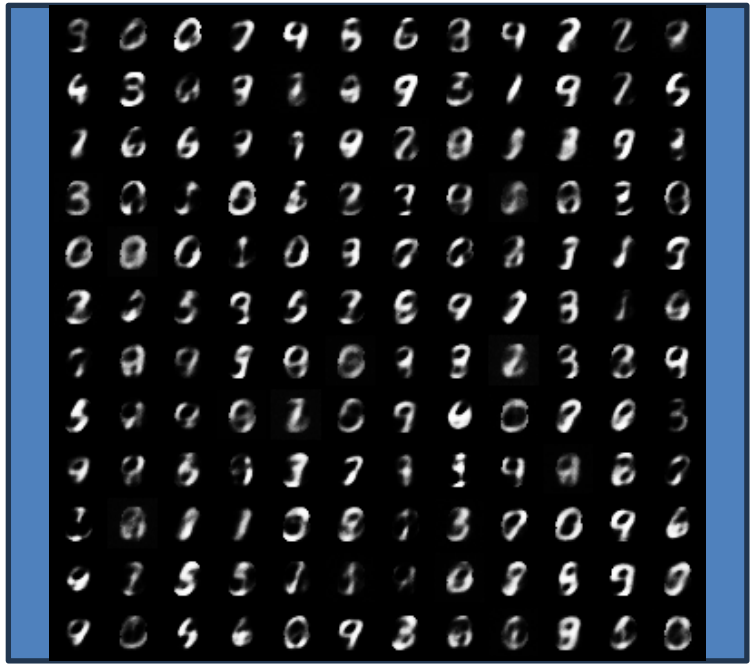
Assignment 4 Writeup

DO NOT TAG

Variational Autoencoder (VAE) Q1 - MNIST

1. Tune beta value with L2 reconstruction loss.
2. Keeping beta fixed switch reconstruction loss to L1.
3. Show examples of L1 and L2 recon loss: (L1 on the left, L2 on the right)



Variational Autoencoder (VAE) Q1 cont.

Explain your observations for VAE with L1 vs L2?

Runing using the default config and only tuning beta. My final beta is 0.02. The valuation loss with L1 is 74 while loss with L2 is 39.

The Recon loss for L1 and L2 is 68 and 31, respectively.

The KL loss (Divergence) for L1 and L2 is 449 and 420 respectively

Due to low beta, both L1, L2 have very high KL divergence, meaning both are not heavily penalizing deviation from the prior. They also learn similar latent space structure.

The recon loss of L1 is higher because L1 loss grow linearly while L2 is square error and grow quadratic, so L2 push values harder toward the mean hence lower loss.

Visually, L1 loss grid show better edge and structure, the sharpness is better, less blurry. L2 loss grid looks smoother but blurry and less sharp. L2 have less error or more accurate.

Variational Autoencoder (VAE) Q2 - MNIST

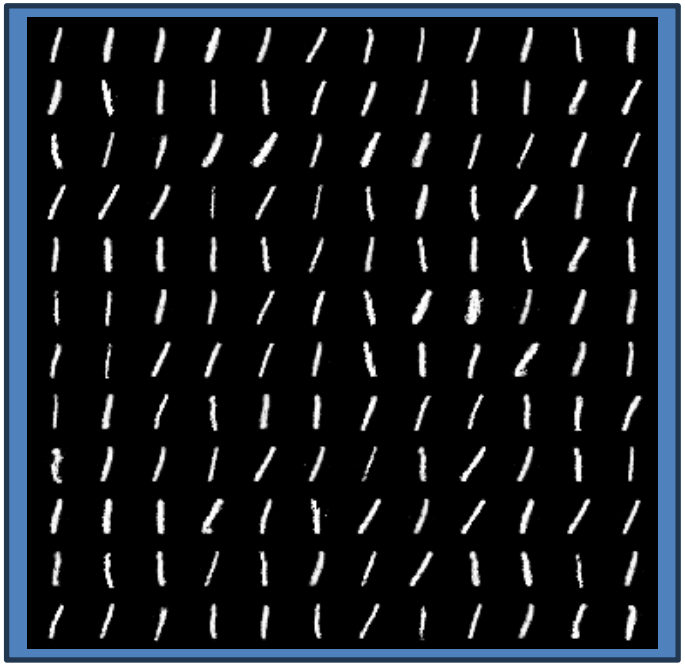
As we optimize for our few-shot setting (KNN classifier) what happens to the quality of generated samples? Share an example and explain your observations.

We tune both L1 and L2 but opt to use L2 due to less error and more correct. Visually, the digits are readable , diverse and correct. It is good quality but not perfect/overfit. The final result has low recon -loss of 18 and high loss_kl of 600. Hence, image look like real digits Latent space is not tightly following prior. As we trained the VAE, the -model gradually improved at structuring and its latent encodings. Its -classification prediction improve. The digits look readable but some -time blurry or distorted. In my case, large KL loss mean latent space Deviates from the prior, which hurt generation. The encoder help with - classification but decoder quality had to compromise.
But overall, the results are really good and should meet requirement

Generative Adversarial Network - MNIST

There is a common failure mode of GAN training termed *mode collapse*. that is when the generator learns to fool the discriminator producing only a certain subset of P(x). Find hyperparameters that exhibit this failure mode and show generated examples.

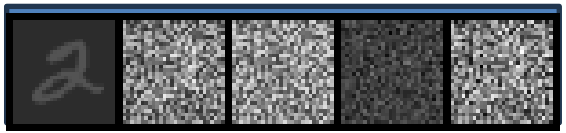
The generator collapse to only a few modes, it generate only one digit repeatedly and fool the discriminator. Here it look like number “1”. The generator stuck producing only 1 digit. To achieve this, I lower latent dim to 4, low dimension mean limited variety, generator cannot explore enough. Increase learning rate, and remove leakyReLU mean weak Discriminator that can be exploited by G, and get stuck in narrow solution.



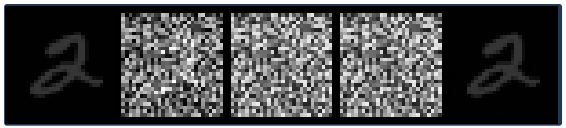
Simple Diffusion

Insert your simple diffusion visualizations here:

iteration 0

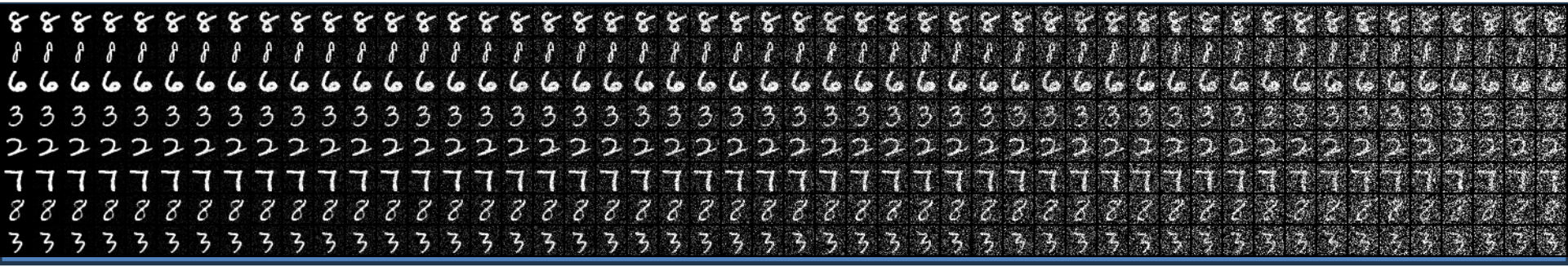


iteration 99



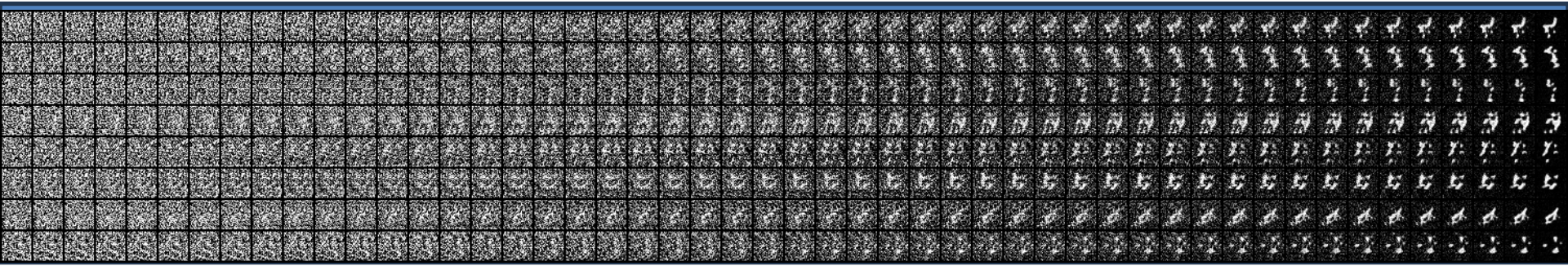
DDPM - MNIST

Insert your forward diffusion visual here (any epoch is fine):

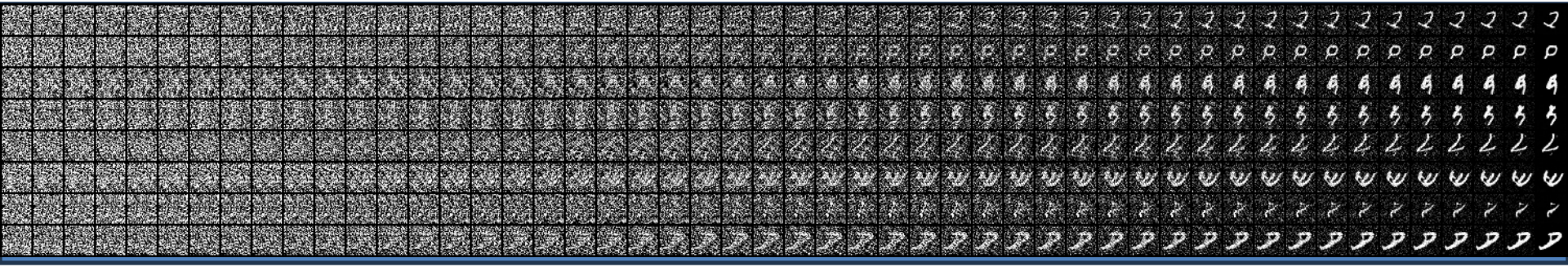


DDPM - MNIST

After tuning only noise schedule and timesteps:
Insert your reverse diffusion visual here **(epoch 0)**:



Insert your reverse diffusion visual here **(epoch 9)**:

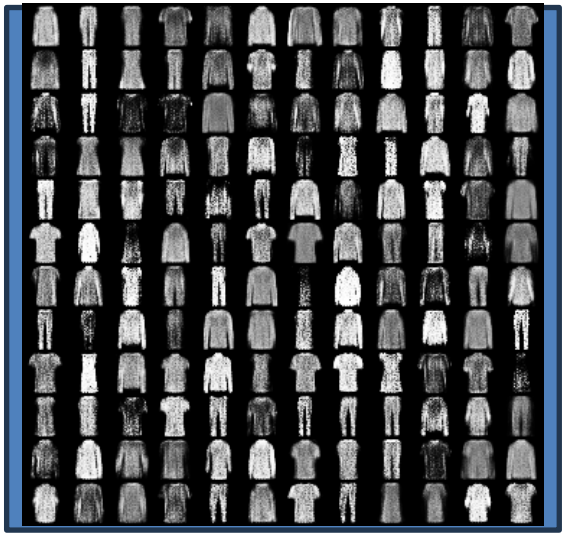


Tune each model (fashionMNIST) to achieve the best images and show your results:

VAE

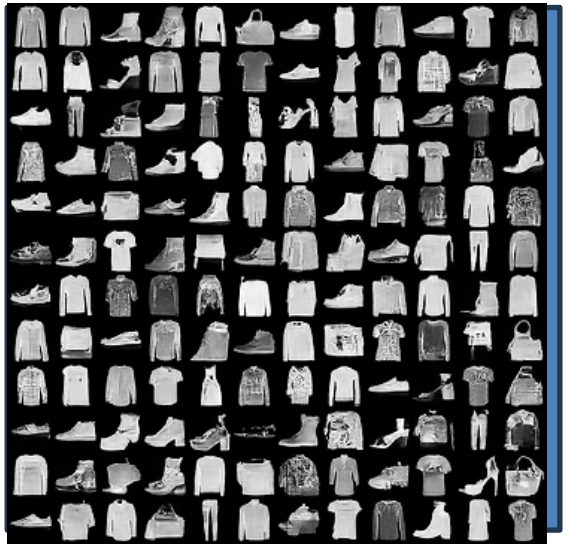
FID: 91.5

GAN



FID:128.3

Diffusion



FID: 35.4

Compare and contrast the results from the different models. Why might we see the results that we see? what role does the target dataset play?

So Diffusion model have best FID score and GAN has worst. In term of visual traits, both diffusion and VAE model look coherent and diverse, with diffusion is a bit better. Diffusion provide clothing picture with good quality and details, it also took longest time to train indicating a lot of computational power.
VAE model (L2) looks smooth but lack sharp detail. GAN model can generate sharp image look closely, it has a lot of noise. GAN however is more likely to collapse so require more time tuning to find optimal hyperparameters.
The reason we see this result is because Diffusion model learn denoise step by step, and benefit from gradual refinement. It work well with FashionMNIST clothing picture that have fine details. Diffusion work better than GAN who is single step.
VAE with L2 loss is also a good model here. It is robust and stable, but sometime blurry and lack sharpness. Look closely, the VAE picture capture the general shape but lack textures.
GAN can generate sharp image but it is unstable and sometime suffer from mode collapse and more likely overfit.
FashionMNIST target dataset play some role here. It has high variability of same class, low resolution 28x28. Its characteristic fit certain model than others. Like VAE L2 struggle with model sharpness, GAN can generate sharp picture but lack stability. Diffusion with gradual refinement perform best with this target dataset but require high computational power.

FID is commonly used to benchmark generative models. Is it a good metric? Why or why not?

FID is a good metric. It measure both how close the generative images to real data and how diverse they are. Lower FID mean better perceptual quality because FID correlates with human perception of visual quality. FID is also a robust benchmark that has been studied.

FID however have limitation. FID cannot be compare across different dataset unless all implementation conditions are the same. FID is also unreliable on small datasets so score can vary widely.