CS 4782 Coding Assignment 4 Written Responses

Due: 4/24/25 11:59 PM on Gradescope

Late submissions accepted until 4/26/25 11:59 PM

Note: For homework, you can work in teams of up to 2 people. Please include your teammates' NetIDs and names on the front page and form a group on Gradescope. (Please answer all questions within each paragraph.) Please show all the relevant steps in your solutions.

Problem 1:

Do you see all 10 digits? Do you notice that some digits are better quality than others? How would you try to improve the quality of the digits?

Yes, we can see most of the 10 digits in the generated samples, but not all of them are equally clear. Simpler digits like 1, 0, and 7 tend to look better, while more complex ones like 5, 8, and 9 sometimes come out blurry or distorted. To improve the quality, we would either increase the model's size (for example, a larger latent space or deeper network), train for more epochs, or modify the decoder to use convolutional layers instead of only linear layers.

Problem 2:

Do you notice that clusters in the center are smaller than the peripheral clusters on average? If so, why do you think is the case?

Yes, the clusters in the center of the t-SNE plot are generally more compact than the ones on the edges. This is due to the TNSE algorithm, which tries to draw similar points together, and different points apart. Since the center clusters are surrounded by all sides, they are 'pushed' from all directions, forming a more compact shape. The peripheral clusters on the other hand, have no 'pressure' coming from outside, which allows them to have wider stretched shapes.

Problem 3:

What are the sources of stochasticity in the DDIM sampler? How do different samples from the same initial draw of noise relate to each other, if at all? Write 2-3 sentences below.

The only source of randomness in DDIM is the initial noise sample, the starting point $\mathbf{z}_1$. Once that's drawn, the DDIM steps themselves are fully deterministic there's no extra noise added during the denoising steps. This means if we start from the same $\mathbf{z}_1$, we will always get the exact same final sample.

Problem 4:

In section 2.3, the diffusion architecture accepts the noisy latent and the timestep and predicts the added noise. Briefly describe the diffusion architecture being used for this problem? How is the timestep information being incorporated? Write your answer in 3-4 sentences

This diffusion architecture uses a simple fully connected score network to predict noise at each step. The input is first projected into a higher-dimensional space, then passed through several residual feedforward layers with SwiGLU activations and RMS normalization. Timestep information is incorporated by embedding the noise level α^2

using sinusoidal positional embeddings and a small MLP; this allows the network adapt its predictions depending on the current stage of the diffusion process.

Problem 5:

Q: What do you observe about the diffusion sampling process? Does there appear to be any relationship between the initial Gaussian latent variable z_1 and the final sample from the data distribution z_0. Write your answer 3-4 sentences.

Watching the sampling process, we can see that the points gradually become more structured over time. At the beginning, it looks like random noise, but by the end, the points have clearly formed into organized rings. There is a weak relationship between $\mathbf{z}_1$ and $\mathbf{z}_0$. Nearby points at the start tend to stay somewhat close to each other at the end, However, the transformation is quite large so the points of the same color do tend to spread out.