**Christian Mpabuka**

**ITAI 2376 Deep Learning in Artificial Intelligence**

**ITAI: 2376**

**Professor: Anna Devarakonda**

**Midterm Creating Images with Diffusion Models**

Imagine you have a clear photo, and you gradually add static to it until it becomes a noisy mess. That’s essentially what the forward diffusion process does. In the notebook, a function called add noise manages this step-by-step noise addition. We don’t add all the noise at once because the goal is to teach the AI how to clean up the image gradually. If we added all the noise in a single step, the model wouldn’t learn how to improve the image bit by bit.

Now, let’s talk about the model we’re using; it’s called U-Net. This model is excellent for image tasks because it can capture both the big picture and the tiny details. One of its cool features is something called “skip connections.” Think of them as shortcuts that help the model remember fine details like edges and lines, so they don’t get lost along the way.

The model also uses something called class conditioning. That simply means it knows what kind of image it’s supposed to generate, like a “5” or a “9,” because we give it a label. It converts that label into a code and feeds it into the model, so the AI understands what it’s aiming to create.

Another important piece is time embedding. This tells the model how noisy the image currently is. It’s like giving the AI a sense of how far along it is in the cleanup process. Without this, the model wouldn’t know how much noise to remove at each step.

Some images are easier for the model to generate than others. Simple, common ones are a breeze because the model saw lots of them during training. But unusual or complex ones, like a funky-looking “3” or a digit in a fancy font, can be trickier. That’s where CLIP scores come in. They help by giving feedback on how close the generated image is to what it’s supposed to be. The model can use that feedback to adjust and get closer to the target.

In the real world, this kind of model has tons of cool uses: generating art, designing logos, creating synthetic data for training other AIs, fixing blurry images, and even helping scientists design new materials. That said, our current model is basic. It’s great for simple tasks like generating handwritten digits, but not quite ready for high-res photos.

If I were to improve this project, I’d do three things:

1. Make the U-Net bigger so it can learn more complex patterns.
2. Use CLIP to pick the best images from multiple attempts.
3. Train the model to generate digits in different styles to make it more versatile and creative.

Image result:

