Midterm Report

Team 13 – Chengkai Men (cmen), Rui Shao (ruishao), Tianlin Sun (tianlins)

1. Overview

For this part, we have implemented the DDPM and DDIM Schedulers. We have successfully trained the initial models. Although their performance needs further improvement, the overall architecture is well-functioning.

2. Implementation Scope

- ddpm.py
- scheduling_ddpm.py
- scheduling_ddim.py
- train.py
- inference.py

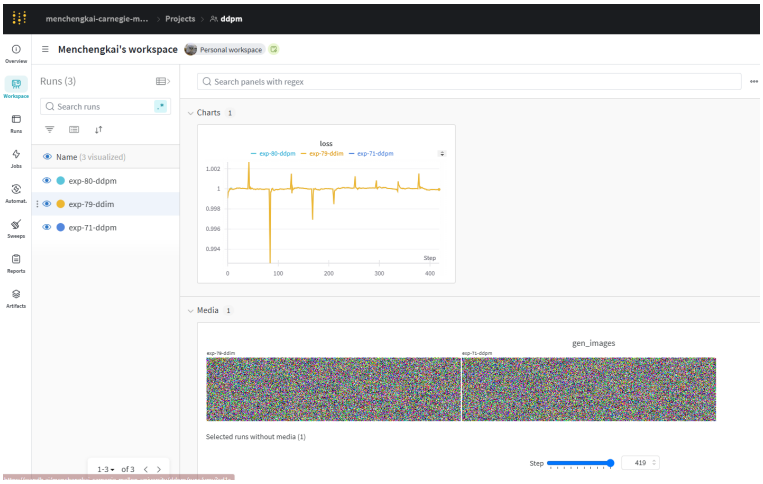
3. Details

We implemented exactly according to the formulae in the original DDPM and DDIM papers. The underlying network is a Unet with a bottleneck design. This structure allows it to extract high-level abstract features and efficiently generate images.

During training we have encountered and solved many issues.

- The dataset took more than 100 hours to train which was unfeasible. We reduced the image resolution using torch.transforms.Resize
- We downloaded the CIFAR-10 dataset for initial training
- We encountered memory overflow when computing the Frechet Inception Distance (FID) as well as Inception Score (IS). They exceeded my computer's memory limit. So we generated fewer number of images at first.

4. Results



Wandb run snapshot

```
FID: 584.018798828125
```

```
IS: (tensor(1.1013), tensor(0.0046))
```

5. Evaluation, Next Steps

Overall, our results are *not* satisfactory. The generated image cannot be distinguished from pure noise. Both the FID and IS scores were high. We will improve our model by trying out other hyperparameters. We plan to use the smaller CIFAR-10 dataset to achieve a small-scale success before training on the given large dataset. Furthermore, we will implement VAE and CFG to see their performances.