**LAB2-Denoising Diffusion Implicit Models (DDIM) and LoRA**

**314552036 周子翔**

1. **Task 1-1: [2D Swiss Roll] Reverse Process of DDIMs**
   1. **Complete and Explain of all #TODO code (15pts)**

A computer screen shot of text

AI-generated content may be incorrect.

Predicts the clean data (x0\_pred) and computes the next sample (x\_prev) using DDIM update rule.

eta controls stochasticity (0 for deterministic).

A computer code on a black background

AI-generated content may be incorrect.

Iteratively applies ddim\_p\_sample for each timestep, starting from random noise.

Collects all intermediate samples for visualization or analysis.

A computer screen with text

AI-generated content may be incorrect.

Runs DDIM sampling and computes Chamfer Distance between generated and true data.

* 1. **Fig of Evaluation Result (2.5 pts)**

A diagram of a distribution of samples

AI-generated content may be incorrect.

1. **Task 1-2: DDIM [Image Generation]**
   1. **Complete and Explain all #TODO code (20pts)**
      1. **DDIMScheduler Implementation (image\_diffusion\_todo/scheduler.py):**

A computer screen shot of a program code

AI-generated content may be incorrect.

Implemented set\_inference\_timesteps to select the timesteps for DDIM sampling.

Implemented step to perform the DDIM update, using the noise predictor and eta parameter for controlling stochasticity.

The scheduler computes the reverse process for DDIM, accelerating sampling compared to DDPM.

* + 1. **Sampling Script (image\_diffusion\_todo/sampling.py):**

A screen shot of a computer code

AI-generated content may be incorrect.

Modified to support --sample\_method ddim, --ddim\_steps, and --eta arguments.

The script loads the trained checkpoint, sets up the DDIMScheduler, and generates images using DDIM sampling.

* + 1. **FID Evaluation (image\_diffusion\_todo/fid/measure\_fid.py):**

A screen shot of a computer code

AI-generated content may be incorrect.

Ensured the script resizes images to the correct size and computes FID between generated samples and the AFHQ eval set.

Ran dataset.py once to prepare the evaluation data.

* 1. **Key Code Explanation:**

DDIM sampling loop replaces the DDPM loop, using fewer steps and the eta parameter for deterministic or stochastic sampling.

All TODOs were filled according to the DDIM paper and assignment requirements.

* 1. **Table of Evaluation Results (10pts)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **FID** | | **S** | | | | |
| **10** | **20** | **50** | **100** | **1000** |
| η | **0.0** | 26.7321 | 27.9843 | 30.1247 | 32.4132 | 31.0128 |
| **0.2** | 29.4417 | 29.0312 | 32.7715 | 27.6249 | 33.9913 |
| **0.5** | 26.1215 | 30.7712 | 34.4108 | 36.0217 | 36.8823 |
| **1.0** | 28.4136 | 31.8542 | 33.9217 | 36.7712 | 35.1348 |

* 1. **Discussion and Explanation of Evaluation Results (10pts)**
     1. **Effect of DDIM Steps (S):**
        1. As the number of steps increases from 10 to 1000, FID scores generally increase, which is counterintuitive. Normally, more steps should allow the model to better approximate the reverse diffusion process and yield lower FID (better image quality).
        2. In this result, the lowest FID is observed at lower steps (e.g., S=10, η=0.0: 26.7321), and FID increases as steps go up. This may indicate that the model or scheduler is over-smoothing or introducing artifacts at higher steps, or that the optimal number of steps for this checkpoint is relatively low.
     2. **Effect of η (eta):**
        1. η controls the stochasticity of the DDIM process. η=0.0 is deterministic, while higher η introduces more randomness.
        2. For S=10 and S=20, η=0.0 yields the lowest FID, suggesting deterministic sampling works best for few steps.
        3. For higher steps, the difference between η values becomes less pronounced, but higher η tends to result in higher FID, indicating that too much randomness can degrade image quality.
     3. **General Trends:**
        1. The best FID scores are achieved with low steps and low η.
        2. Increasing η generally increases FID, especially at higher steps.
        3. The results suggest that, for this model and dataset, using fewer DDIM steps and a deterministic process (η=0.0) produces the most realistic images as measured by FID.
     4. **Possible Explanations:**
        1. The model may be overfitting or the DDIM scheduler may not be optimally configured for high step counts.
        2. The checkpoint used may be best suited for fast, low-step sampling.
        3. The evaluation setup (image size, dataset, etc.) may also affect the FID trend.

1. **Task 2-1: Train LoRA on a Specific Style**
   1. **Dataset Description (5 pts)**
      1. The dataset used is lambdalabs/naruto-blip-captions, an open-source image-caption dataset available on Hugging Face.
      2. Source: https://huggingface.co/datasets/lambdalabs/naruto-blip-captions
      3. This dataset contains images of Naruto characters with synthetic captions generated using BLIP-2.
   2. **Visualization of Training Images and Generated Images (15 pts)**
      1. Still training.
2. **Task 2-2: Train DreamBooth with LoRA on a Specific Identity**
   1. **Dataset Description (5 pts)**
      1. The dataset used is sample\_data/dreambooth-cat/, a small custom dataset provided in the assignment.
      2. Source: Local folder in the repository (task2/sample\_data/dreambooth-cat/)
      3. This dataset contains 4 images of a specific cat, used to teach the model a unique identity.
   2. **Visualization of Training Images and Generated Images (15 pts)**
      1. Training images:



* + 1. Generated images:

A person with a beard wearing red sunglasses

AI-generated content may be incorrect.