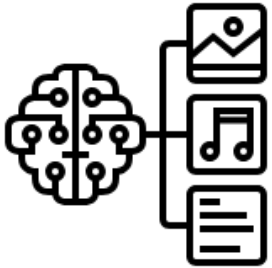
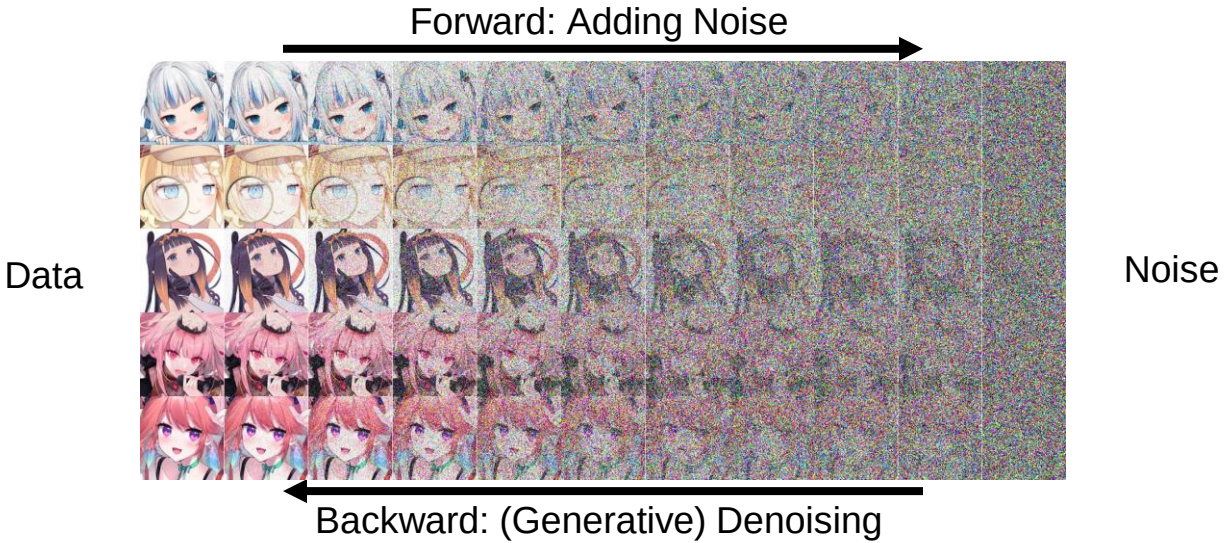
Homework 5 – Generative AI

Due Date: To Be Determined



Introduction

In this assignment, the focus is on diffusion models in the field of generative AI, which generate images from a batch of noise. The process consists of two stages: the forward process and the backward process.

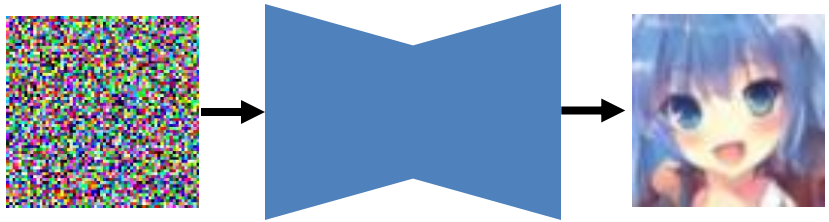


Introduction

Part 1 – Anime Face Generation

In this part, you will learn how to train a diffusion model from scratch. The given dataset contains 71,314 anime face images.

- Given: dataset
- Todo: diffusion model
- Target: 1000 anime face images

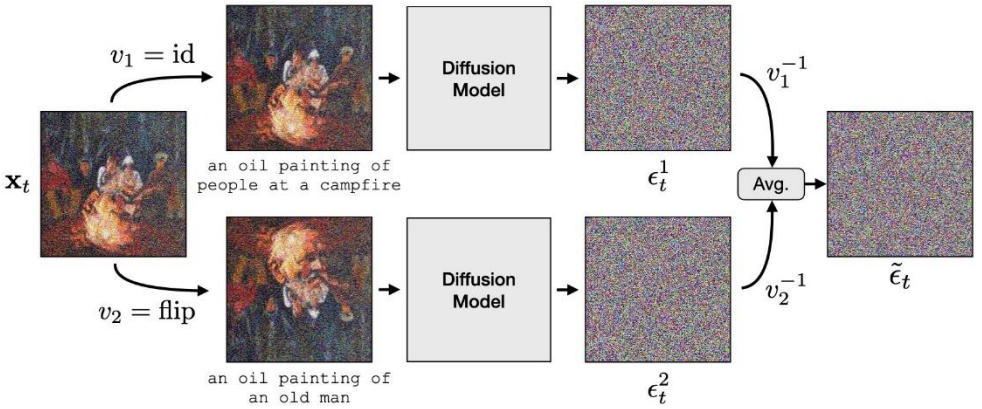


Introduction

Part 2 – Optical Illusion Generation

In this part, you will learn how to use existed text-to-image diffusion model. In addition, you need to change the denoising operation to realize the effect of optical illusion.

- Given: pretrained diffusion model
- Todo: denoising operation
- Target: an optical illusion image



Requirements

1. Please modify the codes in source code between # Begin your code and # End your code.

2. This assignment requires high-level hardware resource (GPU), it's recommended to run your code on Colab or Kaggle.

3. It's forbidden to use additional data, pretrained weights, other packages in your implementation.

For more detail requirements, please check the spec!

Implementation (70%)

- Part 1-1: Denoising Process (15%)
  - Implement **initialization, forward and backward function** in **diffusion.ipynb**

- Part 1-2: Result Visualization (10%)
  - Visualize the **denoising progress** and plot the **loss curve** in **diffusion.ipynb**

- Part 1-3: Evaluation Baseline (20%)
  - In this section, your generated images must meet a certain standard, including **FID Score** and **AFD rate**.

Implementation (70%)

Standard	FID Score	AFD Rate	Grade
Default	≤ 160.0	≥ 0.60	0%
Simple	≤ 120.0	≥ 0.70	4%
Normal	≤ 100.0	≥ 0.80	4%
Medium	≤ 90.0	≥ 0.85	4%
Hard	≤ 80.0	≥ 0.90	4%
Boss	≤ 70.0	≥ 0.95	4%

Implementation (70%)

- Part 2-1: Prompt Design (5%)
  - Please design **at least 2 different prompts** for different views in **multi-view.ipynb**
- Part 2-2: Viewing Transformation (5%)
  - Implement **IdentityView()** and **Rotate180View()** in **multi-view.ipynb**
- Part 2-3: Denoising Operation (10%)
  - Apply viewing transformation on **denoising process** in **multi-view.ipynb**
- Part 2-4: Evaluation Baseline (5%)
  - In this section, your generated images must meet a certain **CLIP Score (>0.3)** before and after image transformation. More details please refer to our spec.

Report (30%)

- You should write your report following the report template.
- The report should be written in **English**.
- Please save the report as a **.pdf** file. (font size: 12)
- For every parts, please take some screenshots of your code and explain how you implement codes **in detail**.

Submission

Due Date: To Be Determined

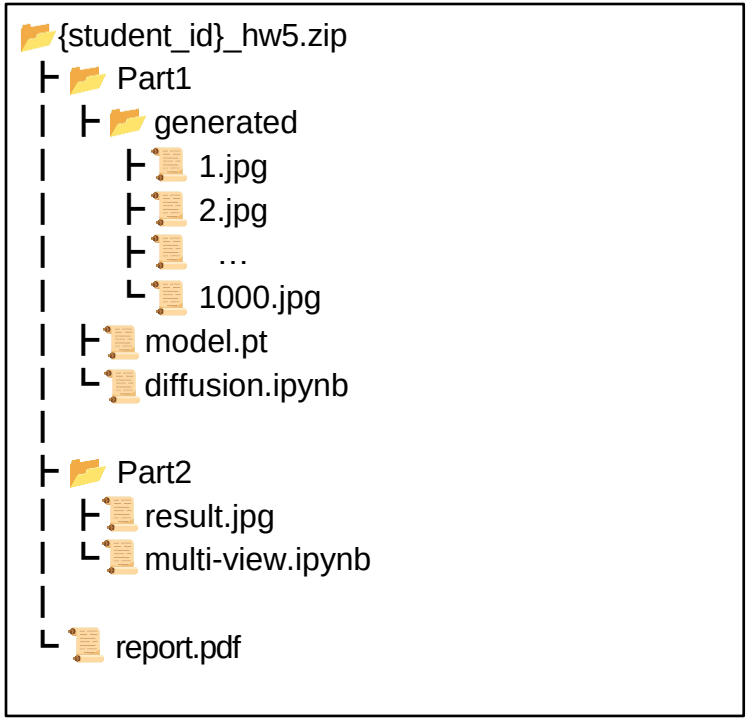
Please compress your source code, results, and report (.pdf) into

{STUDENTID}_hw5.zip. The file structure should look like:

Please zip the contents, instead of the {student_id}_hw5 folder, there should NOT be a folder named {student_id}_hw5 in the zip file.
Wrong submission format leads to -10 point.

Late Submission Policy

-20 points per late day (rounds up the day to the next whole number).



QA Page

If you have any questions about this homework, please ask them on the following Notion page. We will answer them as soon as possible. Additionally, we encourage you to answer other students' questions if you have any idea 😄

Link: (preparing…)

Reminder

When you begin to work on this assignment, it’s recommended to use Ctrl+F+“Todo" to find each part for implementation.

Please note that training a diffusion model is highly time-consuming! Each training round may take over an hour.

It’s best to start this assignment as soon as possible. Good Luck!

Reference

[1] Machine Learning Material from Hung-Yi Lee
https://reurl.cc/RLp2xg

[2] Denoising Diffusion Probabilistic Models (NeurIPS 2020)
https://arxiv.org/pdf/2006.11239

[3] Denoising diffusion implicit models (ICLR 2021)
https://arxiv.org/pdf/2010.02502

[4] DeepFloyd IF on Hugging Face
https://huggingface.co/docs/diffusers/api/pipelines/deepfloyd_if

[5] Visual Anagrams (CVPR 2024 oral)
https://arxiv.org/pdf/2311.17919

Please check out the spec for more details!