Dr Muhammad Rana and Dr Anjan Dutta
TAs: Anindya, Iqra, Silpa, Srinivasa, Ubaid, Umar
Institute for People-Centred AI
University of Surrey

Spring Term 2025

EEEM068: Applied Machine Learning
Project: Human Faces Generation with Diffusion Models

- Submission deadline: **TBC**

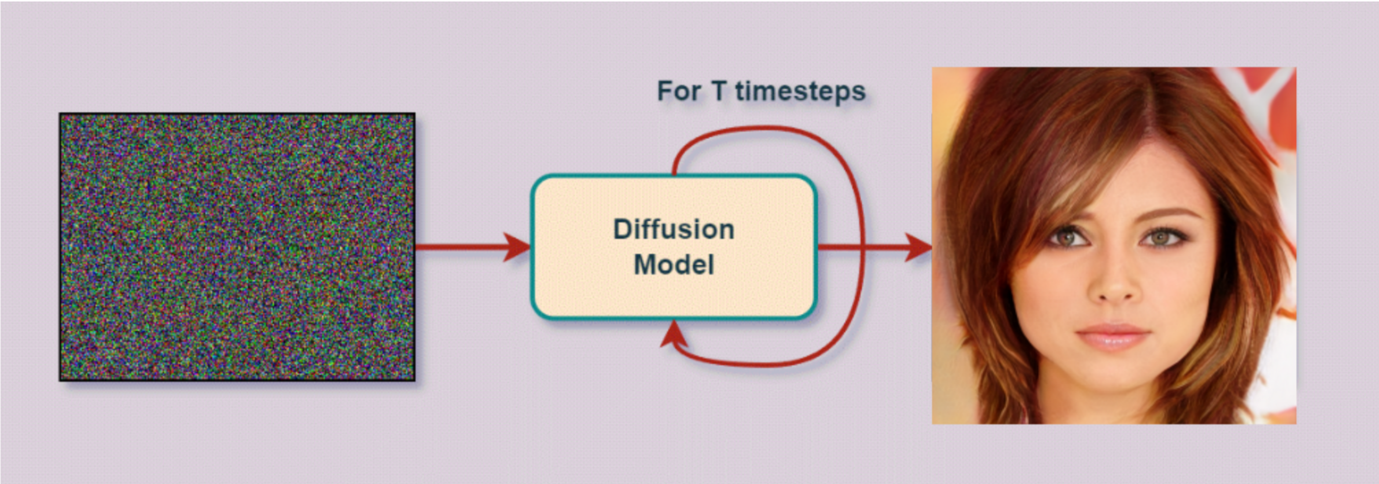

Figure 1: A diffusion model iteratively transforms random noise into a detailed realistic image over 'T' timesteps.

Project specification

Overview

This project requires you to apply the knowledge gained from the tutorials of the EEEM068: Applied Machine Learning module. It accounts for 100% of the total module marks, distributed as follows: 60% for the working code and project report, and 40% for the oral exam and viva. The objective of this project is to train a diffusion model network on the 'CelebA-HQ 256x256' dataset to perform unconditional image generation of human faces.

Submission

You should submit a zip file containing the following:

- Your complete code;
- A **4-page report** (IEEE double-column format) explaining your code, visualizing the results you obtained, and discussing your observations. If you want, you can include additional pages only with visualizations (this is optional and you won't lose any marks if you do not include additional pages). However, the main text of your report should fit in the first 5 pages and the additional pages (if any) should only include visualizations with short captions.

Please note that for all the visualizations and tables you include in your report, it is important to include a reference in the main text (typically using a Figure or Table number).

Background

In this project, you will use diffusion models to generate realistic images of human faces, leveraging the cutting-edge capabilities of these models to produce high-fidelity and diverse portraits. Diffusion models [1] [2], a novel class of generative models, iteratively refine noise into detailed images through a reverse Markov process, closely mimicking the distribution of the training data. Unlike traditional GANs (Generative Adversarial Networks) which directly learn to generate images from a latent space, diffusion models gradually transform a random noise distribution into a coherent image structure,

[1] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. Advances in neural information processing systems, 34:8780–8794, 2021.

[2] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456, 2020.

allowing for an unprecedented level of control and quality in the generation process. By training on a dataset of human faces, the model learns to capture the details of facial features, expressions, and lighting conditions, resulting in images that are remarkably realistic. Through this project, you will explore the capabilities of diffusion models, understand their working principles, and apply them to the generation of realistic human faces.

CelebA-HQ 256x256 Dataset

CelebA-HQ 256x256 is frequently used for evaluating unconditional generation ability of generative models, such as GANs, denoising diffusion, and score-based models. The dataset folder contains 30,000 images of human faces in JPG format.

Figure 2: Samples from CelebA-HQ 256x256 Dataset

Train on a Butterfly Dataset

You will first use a pre-defined butterfly dataset to train a toy diffusion model. Following the provided notebook:

- Load and preprocess the butterfly dataset to fit the model's input requirements.
- Visualize samples from the butterfly dataset.
- Configure the diffusion model's parameters, such as the number of diffusion steps, the noise schedule, and any specific parameters related to the reverse diffusion process. These parameters significantly impact the quality of generated images.
- Define the U-Net architecture. Adjust the model parameters based on the size and complexity of your dataset.
- Set up your model with the appropriate loss function and optimizer. For diffusion models, a common choice is to use a Mean Squared Error (MSE) loss and Adam optimizer.
- Train your model using the prepared dataset and visualize the training loss.
- After training, use your trained diffusion model to generate butterfly images. This can be done by feeding noise into the model and performing the reverse diffusion process to produce images.
- Examine the generated images for quality, diversity, and realism. Assess how well the model has captured the characteristics of butterflies, including colors, patterns, and shapes.

Human Faces Generation

Instead of using butterfly images, you will now train on CelebA-HQ human faces dataset.

- Create a dataset class to fetch images from 'celeba_hq_256' folder.

- Create a dataloader for train and test sets. Use 2700 images for train and 300 images for test.
- Follow previous steps to preprocess the images, set-up the diffusion model and train your model on the new human faces dataset.
- After training, use the trained model to generate 300 images.
- Calculate the FID score between your generated images and the test set images.
- Try different values of hyperparameters to improve the training behavior and the FID measure. Observe how the choice of hyperparameters affects the results.

Discuss your Observations

- In your report, you should discuss your observations with reference to the fine-tuning of hyper-parameters, training behavior, FID measure, and generated images visualizations. Whenever appropriate, please refer to the corresponding figures/tables to make your observations more concrete.

Extra credit

Extra credit will be awarded if one could potentially perform additional tasks related to the main image generation task. These additional tasks might include but are not limited to text conditioning, visualization, etc.

Marking Criteria

The assessment for this project will be divided into two main components:

- **Group Coursework Submission (60%)**: Includes the technical report and working code.
- **Group Presentation and Oral Exam/Viva (40%)**: The presentation will be 5 minutes long, followed by a 15-minute Q&A session with questions directed at each group member.

Group Requirements

- Each group should consist of **four students**.
- **Equal Contribution**: All members must contribute equally to:
  - Technical/coding tasks.
  - Experiments.
  - Report writing.
  - Presentation.
- The project should utilize **PyTorch**, **Python notebooks**, and **Google Colab**. Access to the **Surrey AI Supercomputer** is also available.

Assessments: Marking Criterion

The group submission aspect of the project will be assessed as follows:

- **Technical Report (50% - 30 marks)**: Clarity, structure, results presentation, and discussion.
- **Functionality (30% - 18 marks)**: Completeness of implementation and performance of the model.
- **Code Quality (20% - 12 marks)**: Efficiency, maintainability, and thorough documentation.

Detailed Report Assessment (30 marks)

- Abstract: 5 marks
- Introduction: 5 marks
- Literature Review (minimum 5 papers): 5 marks
- Methodology: 7 marks
- Experiments: 5 marks
- Conclusion and Future Work: 3 marks

Functionality (18 marks)

- Successful implementation of all required steps, with thorough validation of results.

Code Quality (12 marks)

- Efficient, maintainable, and well-documented code adhering to best practices.