codeway

# AI Academy - Research Engineer Case

Kaan Oktay
kaan.oktay@codeway.co

## Objective

The goal of this case is to train an unconditional generative model on a large-scale face dataset and then, finetuning it on a much smaller face dataset of a single subject. We strongly encourage you to use `Diffusers` library from `HuggingFace` to complete the case.

## Task 1 - Training Latent Diffusion Model on Face Dataset

For this task, you are required to implement a latent diffusion model which first projects input images to a latent space and then, train the diffusion model on the latent space. To train your model, you will use FFHQ dataset in $256 \times 256$ resolution which can be found here. Please follow the steps below to accomplish this task:

- First of all, download the FFHQ images provided in the link above to a folder named `data/` in your main code folder. You should construct your dataloader with relative paths so that we can effortlessly run it on our own machines for evaluation.
- Project $256 \times 256$ images to a latent space of $32 \times 32$ resolution using `AutoencoderKL` class from `Diffusers` library. You can use the state-of-the-art autoencoder model here.
- You should then train the diffusion model on this latent space using the standard UNet backbone (`UNet2DModel` class from `Diffusers` library). For training, you can use the example here which utilizes the code here. The code is for unconditional image synthesis in pixel space and so, all you need to do is adjusting the example for latent space.
- Lastly, after training your model, generate new face samples and visualize them. You can use DDIM sampler instead of DDPM sampler for fast and deterministic sampling.

After successfully learning the space of general face images, you will finetune the model for subject-specific face generation in the next step.

## Task 2 - Finetuning the Model on a Single Subject

For this task, you will finetune the model on a face dataset of a single subject. In the literature, finetuning diffusion models in this way is called *Dreambooth* which was first introduced in this paper. Please follow the steps below to finetune your model:

- First, you should prepare a small face dataset of a single subject. For this, you can use 10-15 photos of any person on which you would like to train your model. Your favorite athlete, football player, politician, singer or friend, any person works.
- Finetune your model from the previous task on this dataset using the *Dreambooth* method. You can use the example here which utilizes the code here. The code is written for text-to-image conditional image synthesis models and thus, you should slightly adapt it for unconditional image synthesis.
- There are a few points you should be careful about while finetuning your model. For example, you should preserve the prior face information learned in the previous step in order to prevent overfitting. There are a couple of tricks and best practices for *Dreambooth* training that you can find fairly easily on the internet as it is an extremely popular method.
- Lastly, after training your model, generate new face samples of your subject and visualize them. As in the previous case, you can use DDIM sampler instead of DDPM sampler for fast and deterministic sampling.

## Submission

Please write a concise 2-3 pages report including sampled images, loss curves and your short comments. You can submit everything in a single zip file including your code folder and the report. The code folder includes all your files related to model training, inference scripts and the environment file as well as the subject dataset for the second task. Regarding the dataset in the first task, just as the same way instructed in the task, you can assume that we will have it in the folder `data/` and thus, you don't need to include it in your submission.

## Evaluation

Your solution will be evaluated based on the following points:

- First of all, your code should work without any extra effort. To facilitate it, please include all necessary libraries in the environment file and always use relative paths.
- In order to sample from your trained models easily, provide an inference script in addition to your main training script.
- In the report, describe how to run your code for both training and inference with their corresponding arguments.
- The evaluation of the first task is simple, sampled faces should be realistic and diverse.
- For the second task, we will focus on how much generated faces look like the dataset subject e.g. the identity of the subject person should be clearly recognizable in the samples.

## Further Remarks

To complete the above tasks, you can use free GPUs from Google Colab or free $300 credits from Google Cloud if you sign up the first time. If you don't have access to any of these options, contact us and we can try to find a solution for you. Furthermore, if there is anything unclear or confusing in the case, please don't hesitate to contact us.