Image Generation with Limited Dataset

Cheng Qian
University of California, San Diego
La Jolla, U.S.A.
chq001@ucsd.edu

Xin Zhao
University of California, San Diego
La Jolla, U.S.A.
xiz225@ucsd.edu

Jinbo Ma
University of California, San Diego
La Jolla, U.S.A.
jim031@ucsd.edu

1 Proposal

1.1 Problem to solve

In recent years, deep generative models have demonstrated remarkable capabilities in image generation. Denoising Diffusion Probabilistic Models (DDPMs) were first introduced by Ho et al. [1]. However, for some datasets that contain a large number of images but a small number of samples per class, it is challenging for the models to learn the fine-grained features associated with each category. Figure 1 illustrates how diffusion models are doing well in image generation. We used the `huggan/smithsonian_butterflies_subset` dataset from Hugging Face [2] as our training data. The butterfly dataset contains 1000 images. We set 1000 steps to add and remove noise, batchsize 16 and 80 epoches.

Figure 1: Butterfly image generated with a good dataset.

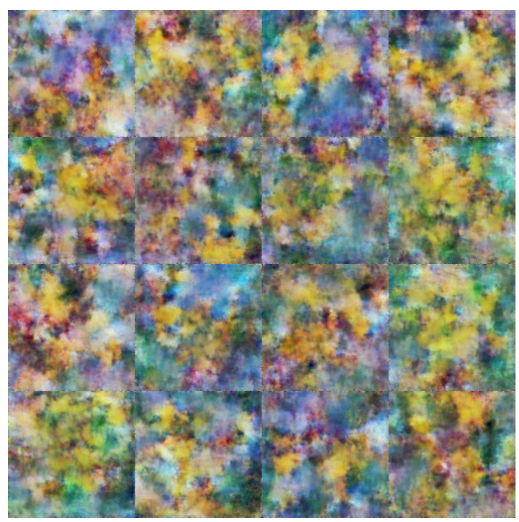

Figure 2: Pikachu image generated with a limited dataset.

However, with the same model, we have experimented with the Kaggle Pikachu Classification Dataset [3], which only contains 387 low-quality pikachu images. Then, with the same training parameters, we get Figure 2. It is significantly worse than what the denoising diffusion probabilistic model does in Figure 1. In Figure 2, it is impossible to tell the shape of pikachu.

1.2 Why important

Having a high-quality and comprehensive training dataset is both time-consuming and resource-intensive. These challenges in data collection motivate the use of data-efficient models. Still, denoising diffusion probabilistic model needs to be modified for more limited datasets to improve their efficiency.

1.3 existing works that aim to solve this problem and the limitations of existing works

Figure 1–2 are the results from the models we trained actually. To find the problem, we have trained a denoising diffusion probabilistic model. The code and results are publicly available at `https://github.com/prayerERROR/ECE285_SP25_Project.git`.

1.4 proposed solution for solving this problem

Latent Diffusion Models (LDMs) were introduced by Rombach et al. [4] to enable efficient high-resolution image synthesis in latent space. We can implement and modify this model to the Kaggle Pikachu Classification Dataset [3]. We can also perform semantic segmentation.

1.5 Tentative timeline

It is week 3. We are supposed to have a project presentation in week 10. So we plan to give a comprehensive proposal in week 3. Then we are going to explore and collect the available data sets in week 4. Encoder and decoder modules should be implemented in week 5. After that, we are scheduled to build Unet in week 6. Having good models, we are about to try to train our models in week 7. A week should be enough for us to tune the parameters in week 8. In week 9, we are supposed to evaluate out models with different benchmarks and prepare for the final presentation. We are going to give presentation and write final project report during week 10 and week 11.

References

[1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. arXiv preprint arXiv:2006.11239, 2020.

[2] Hugging Face. *Smithsonian Butterflies Subset*. `https://huggingface.co/datasets/huggan/smithsonian_butterflies_subset`, 2023. Accessed April 2025.

[3] Hal0samuel. *Pikachu Classification Dataset*. `https://www.kaggle.com/datasets/hal0samuel/pikachu-classification-dataset`, 2022. Accessed April 2025.

[4] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. *High-Resolution Image Synthesis with Latent Diffusion Models*. arXiv preprint arXiv:2112.10752, 2022.