AIST4010 Project Proposal

YUEN Yu Ching 1155143580

February 20, 2023

1 Project Goal

In this project, we will be investigating the problem of anime posture transfer. Specifically, we will attempt to fine-tune a diffusion model such that given reference image of an anime character A and an image with an arbiturary character in pose B, the model will generate an image of character A in pose B.

2 Significance of the project

The anime industry in East Asian countries are having in increasingly large influence across the globe. Considering the heavy workload for desinging anime-esque characters, automating the generation of posing pictures of such characters could not only represent a reduction in cost, but also help creators get a better feel of their characters.

3 Dataset

We will be using a subset of the Danbooru2021 [1] dataset. It is an extensive anime image dataset collected from the anime image aggregator Danbooru.

The dataset in its totality is around 4.5 terabytes in size. However, as part of the project, we will be choosing a small subset of suitable images for training, and we expect it to be around 100 gigabytes after filtering. However, we may face disk quota issues even with this reduced size. In such case, we will further refine the filter parameters until the working dataset is reduced to a reasonable size.

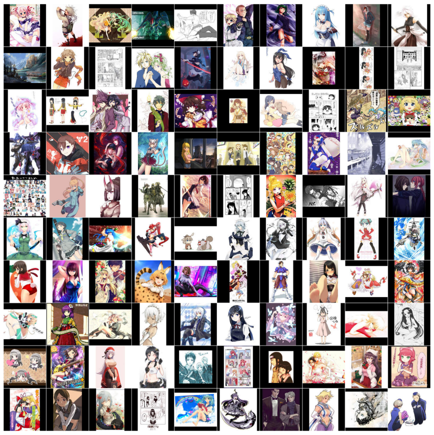

Figure 1: 100 sample images from the danbooru2021 dataset

4 Proposed method

Given that we will be primarily working with anime images, we plan to fine-tune the `waifu-diffusion` diffusion model provided by hakurei on huggingface [2]. We plan to fine-tune it using ControlNet to ensure the validity of the final model.

5 Related works

There are similar models on feature transferral such as [3], which is based on GAN. Should there be enough time, we will attempt to compare the performance of the two architectures by separately training a GAN-based pose transerral model based on the findings in [3].
How will you evaluate your results? Qualitatively, what kind of results do you expect (e.g.,plots or figures)? Quantitatively, what kind of analysis will you use to evaluate and/or compare your results (e.g., what performance metrics or statistical tests)? What is your hypothesis regarding your results compared to baselines?

6 Expected results

Qualitatively, we will evaluate it using plots of its training loss and testing loss, as well as periodically generating testing samples to ensure that the training is progressing towards the objective.
Quantitavily, we will test various loss functions such as L1 loss and L2 loss for their efficicency in training, as proposed by Rogge and Rasul [4].
We hypothesise that the results will be more detailed and accurate compared to baseline, as the training process will enable the model to recognize and utilize more features from the input images.

References

[1] Anonymous, D. community, and G. Branwen, “Danbooru2021: A large-scale crowdsourced and tagged anime illustration dataset,” https://gwern.net/danbooru2021, January 2022, accessed: 2023/02/20. [Online]. Available: https://gwern.net/danbooru2021

[2] hakurei, “waifu-diffusion v1.4 - diffusion for weebs,” Huggingface, accessed 2023/02/20. [Online]. Available: https://huggingface.co/hakurei/waifu-diffusion

[3] M. Mobini and F. Ghaderi, “Stargan based facial expression transfer for anime characters,” in *2020 25th International Computer Conference, Computer Society of Iran (CSICC)*, 2020, pp. 1–5.

[4] N. Rogge and K. Rasul, “The annotated diffusion model,” Huggingface, June 7 2022, accessed 2023/02/20. [Online]. Available: https://huggingface.co/blog/annotated-diffusion