AI (爱) Model: Development Documentation

In creating this model, I had three principles that I wanted to follow: I did not want to violate copyright/harm graphic artists, I did not want to contribute to a closed-source for-profit AI company, and I wanted to use only open-source tools. I used Python, Jupyter Notebook, Stable Diffusion, and the GNU Image Manipulation Program (GIMP), the kohyass LoRA native training script, and the AUTOMATIC1111 Stable Diffusion graphical user interface to complete this project. Unfortunately, due to issues with a security protocol on my computer that could not be resolved through actions in Windows Powershell, I was not able to train the model locally on my own hardware, and had to run the scripts through Google CoLab.

To create this AI model of love, I used scans from three Chinese comic books dating from 1965-1973. Each page was hand-scanned at 300 DPI, cropped, and adjusted for brightness and contrast using the open-source image processing software, GIMP. This was an arduous process, because after the first round of processing, I found that there were traces of the image on the other side of the page showing through the background. This would have been a major problem for training in Stable Diffusion. As a diffusion model, Stable Diffusion starts with a “latent space,” a randomly generated pattern of noise that it resolves through statistical calculations run on a model that approximates a neural network. If all of the images had a trace of the other page showing through, it would have caused the model to “assume” that these traces were desired parts of the image and repeated them. Though this could invoke some interesting creative possibilities, I decided that for the sake of a general-use model, it would be better to remove these. Image examples and a processing workflow are in the appendix to this documentation.

After scanning the images, I was left with 425 individual images, each roughly 550x1100 pixels. To increase the size of the dataset, I used a basic script in Python to mirror each image, resulting in 850 individual images. I then used an open-source AI ecosystem built around Stable Diffusion, the popular open-source image model, to build the training data set. Using the built-in data processing tools, I chose settings that would break each oversized image into two separate, discrete images, resulting in a set of 1,700 total training images. Popular training guides, such as the Rentry guide, suggest that 100-2,000 images would be appropriate for training an artistic style; 1,700 was more than good enough.

After processing the data, I uploaded it to Google Drive so that it would be accessible for CoLab. I then entered Google Colab and used the Khoyass training notebooks to complete the process. As part of the data preparation, I tagged the images with BLIP captioning, which uses machine vision to tag general-purpose images. Though there are captioning tools designed for illustrations and comics, they are intended for Anime and I did not want to pollute the distinctive style of these books with the Anime style. This is both an aesthetic and political choice; from conversations with other Asian game studies academics, Anime has been identified as a point of contention—it is a source of exotification, its origins in making culture “palatable” to a western audience are problematic, and it is often used to center a western-centric understanding of “Japan” as a representative for East Asia held in opposition to “China” due to geopolitical rivalries.[[1]](#footnote-1)

I chose to natively train the model instead of using LoRA training. The process of training a model checkpoint is more computationally intensive, but the deeper training may create a more consistent style with additional layers of training. This allows for me to develop stronger versions of this model in the future. I chose to train using the Stable Diffusion AnyLoRA model, because it was designed to be more amenable to stylistic training, and I thought that success would be more likely given that “style” was the most important factor here. My training configuration can be found as an appendix at the end of this document.

I trained the model, downloaded it, and placed it into the “checkpoints” section of my Stable Diffusion install. At first, I was worried that it would not work because the training data showed relatively high amounts of loss that did not resolve as well as they should have. I am very happy to say that the results were quite successful, though I believe the quality and consistency of this model would be improved with additional training data and more training epochs.





“a drawing of a woman riding on a horse, black and white”

Left: image generated using the trained model and calling the keyword (token) for the dataset

Middle: image generated using the trained model without calling the token

Right: image generated using standard Stable Diffusion 1.5

One issue I had with creating these images is that “anime pollution” is still very prevalent in AnyLoRa. This may be because many of the current DIY LoRas are created by anime fans, resulting in one of the most popular Stable Diffusion offshoots, “Waifudiffusion.” For a LoRA designed to retain styles well, it makes a lot of sense that removing “anime pollution” would be an important step. This “anime pollution” can be seen in the following image, created using the trained model:



“a drawing of a woman speaking at a podium in a crowded room, black and white”

While the aesthetic quality of the scanned artwork is preserved, the automatic generation sees a kimono as the most likely garment for these women to be wearing. Adding negative keywords “japan” and “kimono” resulted in images with characters wearing traditional Chinese peasant clothing:



“a drawing of a woman speaking at a podium in a crowded room, black and white,” negative prompt “kimono, japan”

The images used in the written piece were generated using AI, with universal additions to the prompt as “a drawing, 70年式, 70年小人书” to ensure the concepts were properly called. I found that this model had difficulty with drawing urban locations; it excelled in creating dynamic images of movement, but was less successful at creating images of stationary people. To generate the futuristic prompt, I prompted it to include “a drawing of a person standing in front of a city with tulou-style skyscrapers, Chinese traditional architectures, in the foreground is a lake with lotus leaves, lotus flowers,” (I did not record the full prompt because this is still a prototype) because I wanted to use the traditional “tulou” style communal living architecture along with symbols of regeneration (lotus) and the cycles of life. Some interesting images resulted! They can be found in the third appendix to this documentation.

Some of the images are derived from edited versions of my own photographs of the location. The image of the abandoned building, for instance, is derived from an original image altered to be black-and-white and run through a basic “cartoon” filter in GIMP. I did this and used the “image-to-image” function of the AUTOMATIC1111 interface to create images that recalled the original while maintaining the style.





Left: altered photograph, Right: generated image using image-to-image function

For the rest of the images accompanying the writing, I just decided to experiment and see what I liked. That is not the best way of documenting it, but the process of creation and generation is very quick and exciting, and it can be prohibitive to the creative process to record all aspects of each prompt. A few extras can be found in the final appendix to this documentation project.

**SOURCES:**

Stable Diffusion 1.5 Base Model: <https://github.com/Stability-AI/StableDiffusion>

Automatic1111 Stable Diffusion GUI: <https://github.com/AUTOMATIC1111/stable-diffusion-webui>

Python: <https://www.python.org/>

Jupyter Notebook: https://jupyter.org/

GNU Image Manipulation Program: <https://www.gimp.org/>

Kohya-ss Scripts: https://github.com/kohya-ss/sd-scripts

Kohya-ss Notebooks: <https://github.com/Linaqruf/kohya-trainer>

DistroAnon, EZScriptsAnon, and Anonymous, “/hdg/ LoRA Training Guide.” <https://rentry.org/lora_train>

Yashang She, “Transnational Galatea: Raciaizing Anime-esque Character Design in Genshin Impact.” Conference Presentation. Denver, Colorado. Society for Cinema and Media Studies Conference, April 12, 2023.

1. Yashang She, “Transnational Galatea: Raciaizing Anime-esque Character Design in Genshin Impact.” Conference Presentation, Society for Cinema and Media Studies Conference. April 12, 2023. [↑](#footnote-ref-1)