

Assignment2 StableDiffussion GPT Documentation Report

Yeyang Liu (S2000549)

Find an open source code for LLM and/or Text to Image Generation

- LLM Open Source Model: NanoGPT structure
- Text to Image Generation Model: Stable Diffusion 2.0

Play around, explore the codes, be creative.

Objectives:

- To generate Van Gogh style “Starry Sky”
- To train a nano-GPT model to generate prompts related to “starry night”
- To fine-tune the Stable-Diffusion model to generate Van Gogh style images

Documentation

Topic: Generating Van Gogh style “Starry Sky” images using nangGPT and Stable Diffusion

1. Constructing Dataset of “star”, “sky” and “night”

Dataset: <https://github.com/krea-ai/open-prompts>

Open Prompts is a dataset constructed by krea-ai. It is made up of 9,637,287 records of prompts with raw data, it is published by Turquoit#3987 from the Stable Diffusion Discord server (<https://github.com/paperdave/stable-diffusion-sqlite>) which contains links to every image generated by the bot from August 6th (its inception) to August 20 (when they closed the bot off).

Methodology:

1. Download the large dataset, and keep only the “prompt” column
2. Filter the prompts whose words is less than or equal to 6
3. Keep the prompts that contains the word “star”, “sky” and “night”
4. Save the filtered prompts as starry sky prompts dataset to fine-tune the GPT model

Examples from the constructed starry sky prompts dataset:

```

0
beautiful lush magic mana forest, night sky with dazzling stars, fairies, fireflies, bokeh, octane render, unreal engine, raytracing, crystallized, intricate, hyper detailed, light rays

1
The collage features a woman with wings made of stars, surrounded by a blue and white night sky. The woman is holding a staff in one hand, and a star in the other. She is wearing a billowing white dress, and her hair is blowing in the wind. leather, koke dama by Mark Briscoe terrifying

2
low angle wide shot, the elder scrolls vi hammerfell, capital city, marketsquare, mesa, oasis, sand dunes, ruined temple covered with moss, meadows, sub tropical, illustration, night time, starry sky, perfectly shaded, soft painting, art by krenz cushart and wenjun lin

3
Gediminas_Pranckevicius a dramatic, epic, ethereal painting of shirtless Ari Shaffir dressed as clown with a beer belly wearing a large belt and bandana offering a balloon animal | he is relaxing by a campfire | background is a late night with food and jugs of whisky | stars, tarot card, art deco, art nouveau, mosaic, intricate | by Mark Maggiori | trending on artstation

4
wide shot panorama 20mm lens of a nostalgic megalopolis skyline, at night, detailed, 8K, architecture photography, city, skyscrapers, starry sky, glass reflections, melancholy, lens flares, city lights, neons, raining, fog, silhouettes

```

2. Train GPT model to generate Stable Diffusion prompts

Model Backbone: nanoGPT

Dataset: Starry Sky Prompts Dataset generated in the first step

(<https://www.kaggle.com/datasets/yyleaves/starry-sky-prompt-dataset>)

Methodology:

1. Start from Andrej Karpathy's nanoGPT structure as backbone
2. Train the model on the constructed Starry Sky Prompts Dataset
3. Monitor the loss in training every 100 iterations and generate example prompt every 1k epochs
4. Generate prompts using the model after the training is done. The model could generate several prompts separated by '\n\n', the same format as the training dataset.

Example output after 1k epoch

```

iter 1000, train loss 0.740, val loss 0.740

at fironealies cryshinempe, bok behifulri starry with reaving a big mediell, digne and ciental reactorst, octane, dreamic lightin g, trending on artstation

poittine grave, beautiful eagu a wings thing down large over caphing town ohe dress on the sall ancies at night skyinewovs cowbo y!!, mivichous, ats, michinene weishot, shigh starp, pootorto poce, almen coutyosyai night sky! with a in very scentry mollindig crostaition, cinemationc in on handst painting other in a of a ireld bockgry scus mile

```

Example output after 3k epoch

iter 3000, train loss 0.170, val loss 0.187

enmetric god ration, 4 k Mignola, trending on artstation

a group of the girl, reatio, at night sky, stars in the windows, in he she indlobowing the stars and, in the sky

a realistic portrait of a medieval print, red etrender, - he, soothe night sky is black and full of stars, in the intrificial mak euve reading down, digital art, warm long hour, wildres, glowing s, a leavy, nightting sky stars, light fog, huge resolution, 4k, masterpiece, by greg rutkowski, trending on artstation,

light irides

Example final output

Ak Flom B Thooder Brucolossoft Dürer Koaking lade at the stars in the night sky, midnight, spectacular milky way, shining meteor, official media, anime key visual, detailed, artwork by makoto shinkai. - h 5 7 6

harmony of stea young woman with micron pen ink face by mimedium that night sky with stars, scading aremon. high detailed digital painting by van AlexelOP, symmetr moon rubberlit moon and rimac light beeples, star wars ilm, a a compurking, digital art, octane render, palm cinematic color

portrait of a beautiful cinematic view of a merging with and sirl, in the sky at night, hazy detailed, digital purThe sky night sky, dark stars, aurora borale, nebula artstation, architecture, anime scene, details, concept art, 4 k

Universid boy young Ivan Digital painting of a n chard Field postallathe and Engyptian The stars are fisheye body thomas monwit h appy inspired by a claude prear grunk skullhas and shoulders portrait of a emeran jungle city, helepture, walking on the nigh t sky, hair

3. Fine-Tuning Stable Diffusion using Van Gogh paintings

Model Backbone: Stable Diffusion

Dataset: Vincent Van Gogh's painting

Methodology

1. Download “Starry Night” and other 14 similar-styled paintings made by Vincent Van Gogh from <https://www.kaggle.com/datasets/ipythonx/van-gogh-paintings>
2. Process the painting images and reduce the smaller of its length/width to 512 to reduce GPU memory usage.

_sd → 5_VanGoghStarryNight_Processed



在 5_VanGoghStarryNight_Processed...



Cypresses with Two Female Figures.jpg



Cypresses.jpg



Enclosed Field with Ploughman.jpg



Evening Landscape with Rising Moon.jpg



Green Wheat Fields.jpg



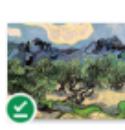
Landscape with Couple Walking ...



Mountainous Landscape Behind Sun...



Mountains at Saint-Remy with Dark...



Olive Trees with the Alpilles in the Back...



Road with Cypress and Star.jpg



Starry Night.jpg



The Mulberry Tree.jpg



Wheat Field with Cypresses at the Ha...



Wheat Field with Cypresses.jpg



Wheat Field with Reaper and Sun...

3. Find and use the code to fine tune the Stable Diffusion model on Van Gogh's paintings from <https://github.com/Akegarasu/lora-scripts>
4. Put processed images to the folder with number of repeats in the folder name as required

█ / ... / train / gpt_sd /

名称

█ 5_VanGoghStarryNight_Processed

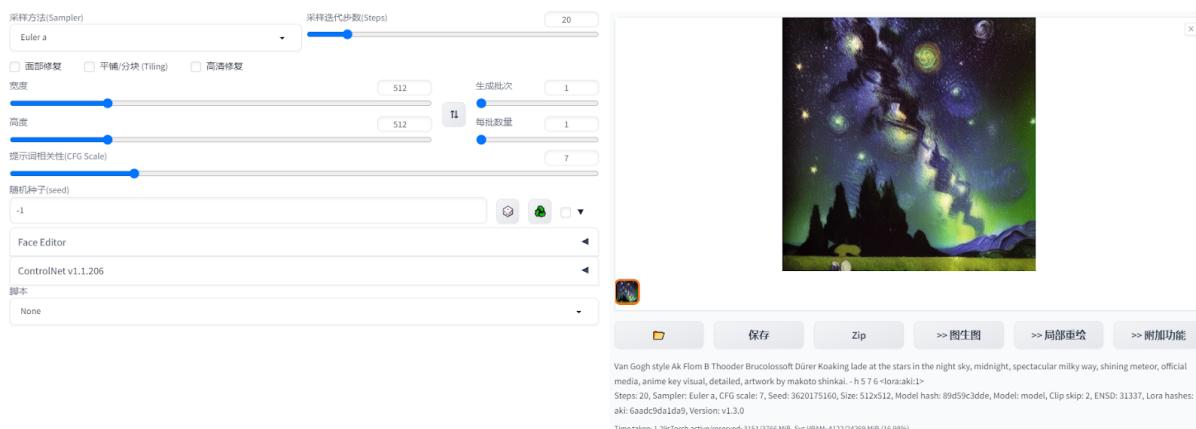
5. Modify the training script train.sh and run the fine-tuning process using LoRA.

```
[Dataset 0]
loading image sizes.
100%|██████████
0<00:00, 2223.99it/s]
make buckets
number of images (including repeats) / 各bucketの画像枚数（繰り返し回数を含む）
bucket 0: resolution (448, 576), count: 20
bucket 1: resolution (576, 448), count: 55
mean ar error (without repeats): 0.026091158365616072
prepare accelerator
Using accelerator 0.15.0 or above.
loading model for process 0/1
load StableDiffusion checkpoint: ./sd-models/model.ckpt
loading u-net: <All keys matched successfully>
loading vae: <All keys matched successfully>
loading text encoder: <All keys matched successfully>
CrossAttention.forward has been replaced to enable xformers.
[Dataset 0]
caching latents.
100%|██████████
:02:00:00, 5.13it/s]
import network module: networks.lora
create LoRA network. base dim (rank): 32, alpha: 32.0
create LoRA for Text Encoder: 72 modules.
create LoRA for U-Net: 192 modules.
enable LoRA for text encoder
enable LoRA for U-Net
prepare optimizer, data loader etc.
```

4. Inferencing

Methodology

1. Inference using webui (<https://github.com/AUTOMATIC1111/stable-diffusion-webui>) using lora checkpoint and set parameters. The original prompted image wasn't ideal even though the model is tuned using Van Gogh Starry Night styled paintings. Maybe the number of paintings and epochs used in the training is not enough. So I add "Van Gogh style" in front of the prompt and the result turns out pretty well and outperforms the original model not being tuned.



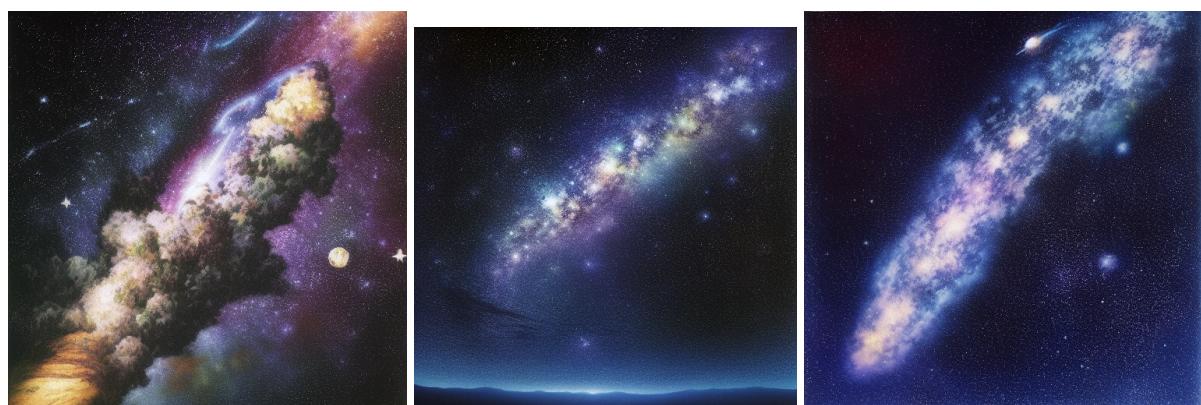
Appendices: Stable Diffusion output using nanoGPT generated prompts

"Ak Flom B Thooder Brucolossoft Dürer Koaking lade at the stars in the night sky, midnight, spectacular milky way, shining meteor, official media, anime key visual, detailed, artwork by makoto shinkai. - h 5 7 6"

Tuned with LoRA, add “Van Gogh Style” in the prompt



Add “Van Gogh Style” in the prompt, not tuned



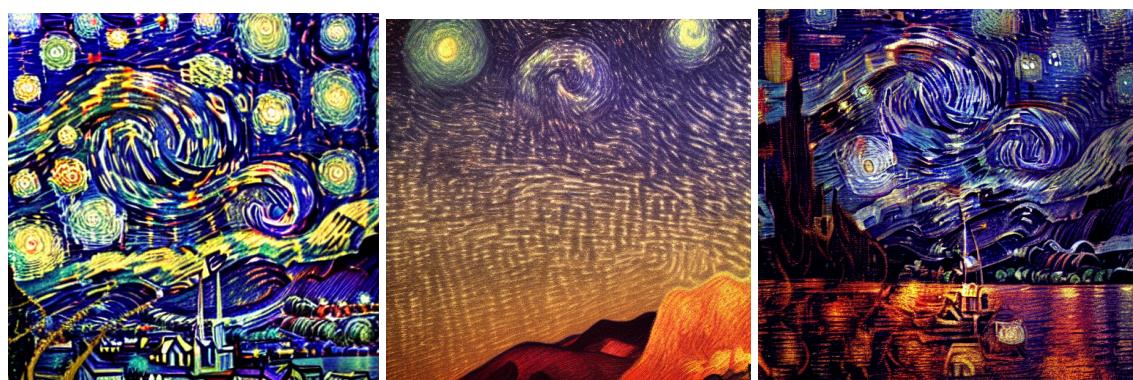
Tuned with LoRA, did not add “Van Gogh Style” in the prompt



"Van Gogh style harmony of stea young woman with micron pen ink face by mimedium that night sky with stars, scading aremon. high detailed digital painting by van AlexeIOP, symmetr moon rabberlit moon and rimac light beadle, star wars ilm, a a compurking, digital art, octane render, palm cinematic color <lora:aki:1>"



"Van Gogh style portrait of a beautiful cinematic view of a merging with and sirl, in the sky at night, hazy detailed, digital purThe sky <lora:aki:1>"



"Van Gogh style night sky, dark stars, aurora borale, nebula artstation, architecture, anime scene, details, concept art, 4 k <lora:aki:1>"



"Van Gogh style Universid boy young Ivan Digital painting of a n chard Field postallatthe and Engyptian The stars are fisheye body thomas monwith appy inspired by a claude prear grunk skullhas and shoulders portrait of a emeran jungle city, helepture, walking on the night sky, hair <lora:aki:1>"

