Image2Tags 项目简介
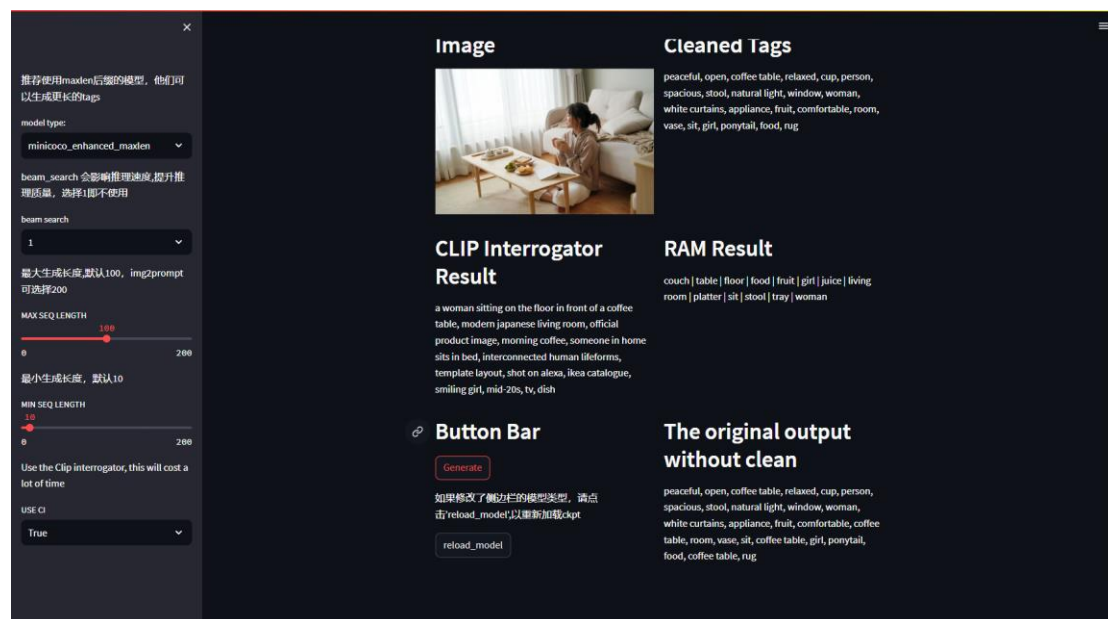
1. 部署 stable diffusion+webui+control net，测试 Stable Diffusionv1.5 与 Stable Diffusion v2.0 的区别。比较 control net v1.0 与 control net v1.1xx 的各个模型的效果。
2. 调研、测试开源文生 prompt 的工作，prompt 包括 caption 和 tag，并完成一个主观的 benchmark
3. 调研现有的 prompt 评价指标。
4. 完成从 AI image 到 prompt 的生成工作
   a) 制作 prompt 扩充数据集。采集线上用户输入，送入 Chat GPT，将 GPT 的输出作为 GT。将 GT 输入 stable diffusion 得到的图片作为输入图片。
   b) 尝试了 Mini GPT4（微调第二阶段，效果不错，但是计算开销大，需要 A100）和 BLIP（形式满足，但是长度较短）。
5. 制作室内装修image 到prompt的数据集。从互联网爬取室内装修的效果图。部署 Mini GPT 4，并根据图片生成描述，作为 GT。模型选择 BLIP。经过实验，发现使用 tag 形式更容易预测。
6. 在 5 走通的基础上增大数据量，同时使用 Streamlit 接口搭建一个前端 demo。
7. 在 56 的基础上，使用 coco 数据集的一部分，实现通用域的 img2prompt 功能。与其他方法进行横向比较（danbrooru、原版 blip、clip Interrogator、RAM），在实体识别上稍逊于 RAM，但是具有更多的风格描述能力。
8. 在 7 的基础上，添加 RAM 预测的 tags，与 minigpt4 的 tags 融合，兼顾风格与实体。最终模型推理的 tags 更加丰富。
9. 整合所有的训练模型以及比较模型 RAM、CLIP Interrogator，集成到一个前端 UI 中，并添加长度控制、数据后处理等功能，成为一个完整的工具。

链接：caption · Streamlit (matpool.com)

运行实例：

开放域效果实例（在 3000 张 coco 图片上训练）

Image

Cleaned Tags

peaceful, open, coffee table, relaxed, cup, person, spacious, stool, natural light, window, woman, white curtains, appliance, fruit, comfortable, room, vase, sit, girl, ponytail, food, rug

CLIP Interrogator Result

a woman sitting on the floor in front of a coffee table, modern japanese living room, official product image, morning coffee, someone in home sits in bed, interconnected human lifeforms, template layout, shot on alexa, ikea catalogue, smiling girl, mid-20s, tv, dish

RAM Result

couch | table | floor | food | fruit | girl | juice | living room | platter | sit | stool | tray | woman

Image

Cleaned Tags

stall, market, outdoor market, women, person, farmer market, sunny day, buildings, umbrella, fruit market, woman, street, dress, man, trees, pose, walk, sunglasses, t - shirts, stand, canopy, produce, girl, fruit stand, vegetable

CLIP Interrogator Result

two women standing next to each other in front of a tent, korean women's fashion model, white font on black canvas, slavic features, street fashion outfit, blackpink, taiwan, images on the sales website, black slime, cheerful atmosphere, sustainability, township, sleeveless tops

RAM Result

canopy | couple | dress | farmer market | fruit stand | girl | goggles | market | pose | produce | shopping bag | stall | stand | sunglasses | tent | vegetable | wear | woman

Image

Cleaned Tags

building, palace, clock, balcony, arches, people, trees, sky, clouds, activity, history, tower, building, town square, plaza, city, government building, town, street, clock tower

CLIP Interrogator Result

a group of people standing in front of a building, including a matador & a bull, clip stadio, pexels contest winner, naboo, inspired by Carles Delclaux Is, color restoration, dirt and unclean, ultra wide angle horizon, exteriors, red bricks, i és complex

RAM Result

brick building | building | car | person | palace | plaza | sky | walk

Image

Cleaned Tags

briefcase, suits, tie, men, formal attire, limo, ties, cloudy sky, building, dress, leader, suit, man, trees, walk, mat, business suit, single file, walking, cars, jackets, carpet, red, car, wear, garment, red carpet

CLIP Interrogator Result

a group of men walking down a red carpet, official government photo, adreas rocha, bomberman, text paragraphs in left, olive skin color, circa 2 0 0 8, jackal, the king of death, with curly black and silver hair, israel, madagascar, pine trees in the background, two young men, podium

RAM Result

briefcase | business suit | car | carpet | dress | garment | leader | limo | man | mat | red | suit | tie | walk | wear

Image

Cleaned Tags

play, group, skateboard, skateboarder, person, skate park, buildings, smiling, enjoyment, building, street, group of people, man, people, ride, road, night, ramp, stand, trick, excitement, street corner

CLIP Interrogator Result

a group of people standing next to each other, wit studio official media, webtoon, set photo, with orange street lights, kawaii playful pose of a dancer, interconnected human lifeforms, ouchh and and innate studio, journalist photo, movie still of a tired, version 3

RAM Result

bus stop | group photo | person | man | pose | stand | youth

Image

Cleaned Tags

eful, cobblestones, alleyway, person, buildings, tranquility, hat, stone, man, gray shirt, trees, walk, road, shadows, cobblestone street, alley, pull, serene, streetlights

CLIP Interrogator Result

a man is walking down a cobblestone street, old asian village, 2 0 2 1 cinematic 4 k framegrab, featured on pexels, by Bian Jingzhao, tonal topstitching, suburbia street, unsplash photo contest winner, guangjian huang, shenzhen, cinemascope, sharp focus on houses, 2 0 5 0 s, by Zheng Xie

RAM Result

alley | bag | carry | path | person | man | pavement | road | shirt | shopping bag | walk | woman

家装专业领域（1200+张三维家设计图训练）

Image

Cleaned Tags

children's room, furniture, yellow and green color scheme, modern and playful style, sofa, coffee table, armchairs, white television, fireplace, green and white tiled surround, green and white striped rug, natural and graphic elements

CLIP Interrogator Result

a living room filled with furniture and a flat screen tv, 2. 5 d cgi anime fantasy artwork, painted pale yellow and green, model trees, in style of disney animation, in game capture 3d render, by Bian Shoumin, inspired by Nassos Daphnis, a room, pre-rendered

RAM Result

armchair | lamp | chair | chandelier | couch | table | furniture | green | hassock | living room | room | stool | television | white | yellow

Image

Cleaned Tags

traditional, blue and white color scheme, wooden bed, upholstered in blue fabric, wooden frame, headboard with ornate carvings, bedside tables, wooden with white marble tops, light blue walls, wooden planks floor, chandelier, natural light, bright and airy space

CLIP Interrogator Result

a bedroom with a blue bed and a chandelier, 3 dcg, light blue colors, ornate declotage, warm and soft and subdued colors, someone in home sits in bed, unbiased render, clean colors, bed on the right, love theme, pre-rendered, other bedrooms can be seen

RAM Result

bed | bedcover | bedroom | bedding | blue | lamp | bureau | carpet | chandelier | curtain | dresser | floor | furniture | headboard | nightstand | pillow | room | wall lamp | white | wood floor

Image

Cleaned Tags

modern, minimalist, clean lines, neutral color scheme, beige walls, dark hardwood floors, large windows, natural light, white leather sofa, chairs, coffee table, plants, abstract painting, modern, sophisticated, focus on clean lines and neutral colors

CLIP Interrogator Result

a living room filled with furniture and a large window, grey tones, modern city, 8k octan advertising photo, spring light, sleek glass buildings, cinemascope panorama, unreal engine: .2, gray anthropomorphic, muted palette, luxury condo interior, line sleek

RAM Result

armchair | lamp | chair | city | city view | couch | furniture | hassock | living room | modern | panorama | room | stool | white | window

Image

Cleaned Tags

modern, minimalist, dining room, wooden table, chairs, light beige walls, white tiles, natural light, plant, neutral color palette, calm, tranquility

CLIP Interrogator Result

a dining room with a table and chairs, inspired by Niccolò dell' Abbate, beautiful shapes, 8k", inspired by Nassos Daphnis, placed in a living room, official product image, two models in the frame, two skinny figures, welcoming attitude

RAM Result

bureau | carpet | chair | table | dining room | dinning table | drawer | dresser | furniture | green | modern | oval | plant | room | round table | stool

Img2prompt（线上 9w+图文对训练）

Image

Cleaned Tags

a detailed photorealistic computer rendering of a cinematic hyperrealism highly detailed photograph of a beautiful detailed landscape matte painting of a cinematic hyperrealism highly detailed photograph of a masterpiece of art, close - up, a beautiful girl, white dress, long hair flowing, delicate facial features, bright eyes, light smile, soft light,

CLIP Interrogator Result

a woman sitting on a beach next to the ocean, embrace porcelain doll, very very pale white skin, she is about 2 0 years old, asian female water elemental, long luxurious light blond hair, 5 years old, veil, child hybrid, looking into the distance, adopt, 1 8

RAM Result

beach | beautiful | dress | girl | hair | mermaid | sea | stone | sit | veil | white | woman

Image

Cleaned Tags

8k resolution, depth of field, macro lens, telephoto lens, dreamlike, rebecca cyberpunk, ok, here are the cue words you need highest quality, close - up, a cute kitten, cute expression, wimage _ ide eyes, light blue fur, soft light, blurred background, in a meadow, natural tones, cheerful atmosphere,

CLIP Interrogator Result

a girl with long green hair and cat ears, cg artist, (((rainbow))), trustworthy eyes, webtoon, the artist has used bright, from league of legends, long wild black curly hair, aliased, dating app icon, sprite art, in a painting of a cat, fur covering her chest, antialiased, emeralds

RAM Result

bangs | cat | dress | dye | eye | girl | green | hair | neckband | stare | woman

Image

Cleaned Tags

in the style of ink wash painting, there is a beautiful woman in hanfu, with long hair fluttering in the wind, holding a sword in her hand, standing under the blooming cherry blossom tree the branches of the tree are swirling, and the petals fall from the branches, forming a

CLIP Interrogator Result

a woman with long white hair in a red dress, artwork in the style of guweiz, cherry blossom rain everywhere, detailed clothes, very long flowing dark hair, fancy silver dress, nightcore, happy appearance, 中国鬼节, cute:2, discord pfp, beautiful costume, long straight silver hair

RAM Result

anime | art | cherry blossom | dress | girl | stand | woman

Image

Cleaned Tags

8k resolution, panorama, studio ghibli, dreamlike, by hayao miyazaki, create a painting of a pair of sisters and brother, rain scene, close - up, two sisters are hugging each other, the girl is wearing a pink dress, the boy is wearing a green shirt, the girl is wearing a blue short - sleeved shirt, the boy

CLIP Interrogator Result

two little girls standing next to each other under an umbrella, trend in art station, realistic cosplay, embrace porcelain doll, photo style of shawn paul tan, 2 0 2 1 anime, little nightmares, hyper realistic eyes, very sweet, grown together, asian

RAM Result

boy | catch | child | girl | park | park bench | stand | sunshade | umbrella

经验：
数据的重要性
模型围绕数据，数据围绕业务场景。搞定数据就搞定了大部分问题。

接触的模型：
SD、controlNet、Mini GPT4、BLIP、BLIP2、SAM、chatGLM、LORA、RAM
接触的技术栈：
Scrapy, gradio, streamlit

自我评价：
工作没那么困难
工作没那么累