



Part 7

Stable Diffusion

生成图像



1. Introduction to Prompt

- 1. What is prompt?**
- 2. Prompt Notes**
- 3. Prompt structures**
- 4. Prompt types**
- 5. Auxiliary generation prompt tools**

1.1 What is prompt?

- 在人工智能领域，Prompt 是一个非常重要的概念。它是一段文本或语句，用于指导机器学习模型生成特定类型、主题或格式的输出。
- 广义的讲，Prompt是指用户输入的文本或图像信息，目的是知道模型根据一些特定的需求生成艺术作品。
- 文生图主要以文字来实现这个沟通过程，图生图还可以依赖图片来传达信息，但图生图中也需要用到文字信息，而且同样重要。



1.1 What is prompt?

- 这边简单说明一下，首先我们的计算机其实是不认识字，为了让计算机认识理解这个字，需要经过一系列的步骤，而这第一步就是**分词，就是将提示词分割为tokens词元的过程**。但是有的时候我们会发现为什么提示词的字数和token数是不一样的呢？
- Stable Diffusion采用的CLIP (Contrastive Language–Image Pre–Training) 模型来进行分词，但所采用的分词方法不是简单地按字符数量来分的。举个例子，对于像"Dreambeach"（蓝梦岛）这样的单词，系统可能之前没有见过，但是呢在之前其中的"dream"和"beach"两个词序列的频率可能都比较高，因为模型不认识这个词，所以模型把这个词分解成它知道的dream和beach，然后算作2个token，而不是将整个单词视为一个token，这也是为什么提示词可能和token数不一样的原因，当然，大多数的时候他们是一样的。

如果想了解stable diffusion原理，可查看李宏毅讲解生成式AI(大模型，文本，图像)

https://www.bilibili.com/video/BV16N411K7aT/?p=16&vd_source=08e56b77b0b1fdc634c3139ecda871e1

如果想了解CLIP模型，[可查看CLIP 论文逐段精读【论文精读】](#)

1.2 Prompt Notes

A. Word

- Prompt 需要英文书写, prompt的书写需要插入分隔符, 基本的分隔符形式是一个**英文里的半角逗号**, Prompt 可以换行, 但每一行的**行末也要有分隔符**。
- 对于在单词tag上特化训练的模型, 建议使用逗号隔开的单词作为提示词。
- 在书写提示 (Prompt) 时要避免拼写错误, 否则模型可能会将拼写错误的单词拆分为字母进行处理。
- 通过提示 (Prompt) 生成图像具有一定的随机性。如果提示词不够准确, AI将只能依靠随机性进行生成。然而, 提示词的编写并非一次性完成, 而是通过逐步优化的过程来建立初始框架, 并逐渐细化。

- Prompt 的主题非常广泛, 可能包括作品主题、画风、形象特点以及一些具体包含的要素
- 不同的 Prompt 分别向AI描绘了画面风格、人物体貌、服饰特点、场景内容和一些额外的修饰性元素

1.2 Prompt Notes

B. Natural language

自然语言，简单地理解，就是我们平常沟通的一些语言。

比如一个拥有金色头发带着花环，穿着白色纱裙的女孩正坐在一颗月桂树下，她的周围是粉色薰衣草草地。如果是用单词tag的方式，可能就是，一个女孩，金色头发，花环，白色纱裙，月桂树，粉色薰衣草。

- 对于在自然语言上特化训练的模型，建议使用描述物体的句子作为提示词。
- 不要使用太过于复杂的语法，语言模型不一定能够理解，导致在解析上出现偏差。

1.2 Prompt Notes

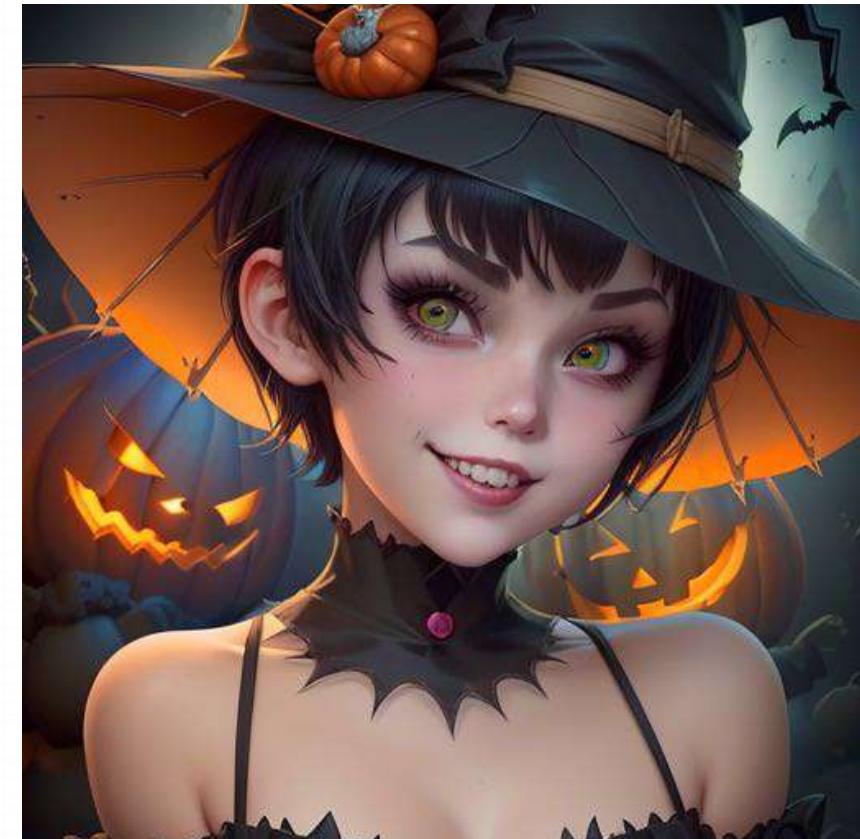
B. Natural language

一般来说越靠前的词汇权重就会越高，比如下面这张图的**Prompt**:

Example: 万圣节假期的拟人化形式是一个留着短发和恶棍笑容的可爱女孩，可爱的帽子，可爱的脸颊，虚幻的引擎，高度详细，艺术种子数字插图，woo tooth，吉卜力工作室，deviantart，锐利的焦点，artstation，由 Alexei Vinogradov 面包店，糖果，绿宝石般的眼睛。

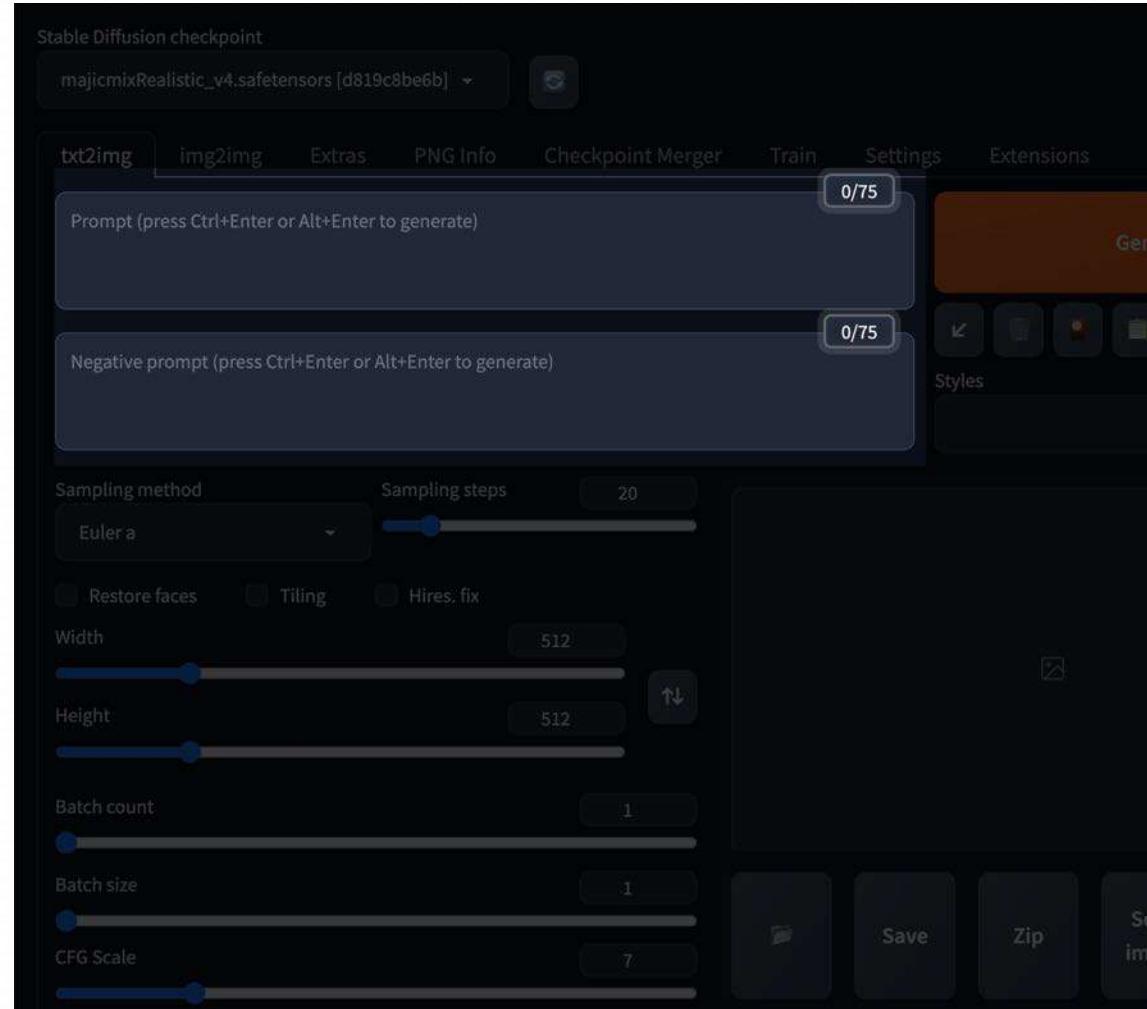
```
The personification of the Halloween holiday in the form of a  
cute girl with short hair and a villain's smile, (((cute  
girl)))cute hats, cute cheeks, unreal engine, highly detailed,  
artgerm digital illustration, woo tooth, studio ghibli,  
deviantart, sharp focus, artstation, by Alexei Vinogradov  
bakery, sweets, emerald eyes
```

第一句关键词词组：万圣节假期的拟人化形式是一个留着短发和恶棍笑容的可爱女孩。那生成的图片主体画面就会是万圣节短发可爱笑容女孩



1.2 Prompt Notes

B. Natural language Length



超过标准75个tokens的输入会导致Stable Diffusion接受更高的prompt 大小限制，从75增加到150。再写更多字的话会进一步增加 prompt 大小。这是通过将 prompt 分成75个tokens的块，独立处理每个块，使用CLIP的Transformers neural network，然后在反给Stable Diffusion 的下一个组件Unet之前将结果连接在一起完成处理。

例如，包含120个tokens的prompt 将被分成两个块：第一个块包含75个tokens，第二个块包含45个tokens。两者都将填充到75个tokens，并使用起始/结束标记扩展到77。通过CLIP传递这两个块后，我们将获得两个形状为 (1, 77, 768) 的张量。将这些结果连接起来形成 (1, 154, 768) 张量，然后传递给Unet。

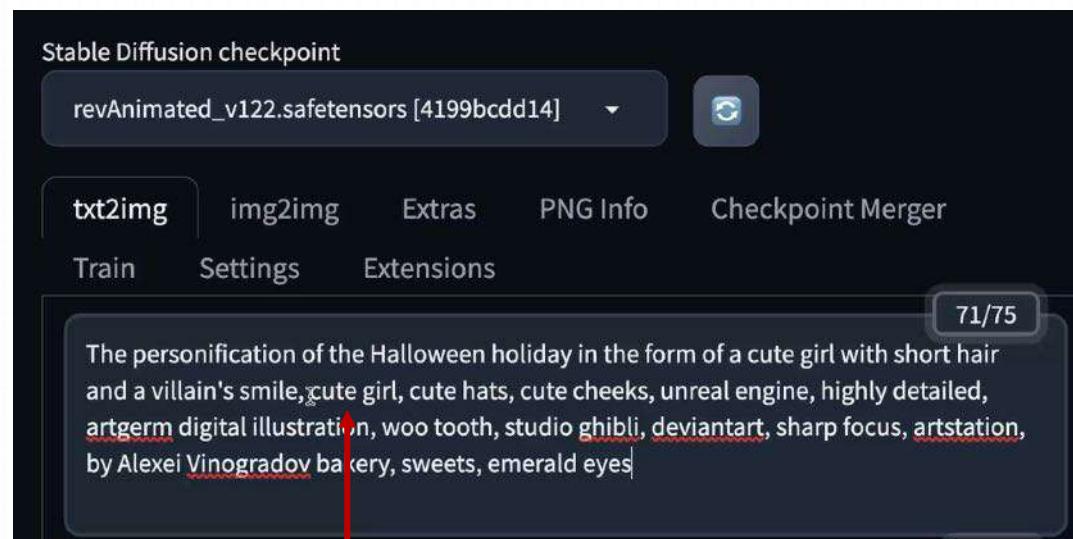
1.2 Prompt Notes

C. 标点符号

The personification of the Halloween holiday in the form of a cute girl with short hair and a villain's smile, **((cute girl))**cute hats, cute cheeks, unreal engine, highly detailed, artgerm digital illustration, woo tooth, studio ghibli, deviantart, sharp focus, artstation, by Alexei Vinogradov bakery, sweets, emerald eyes

((cute girl)) 两侧的括号是权重调节，类似 Midjourney 的 **::**

- ① 最直接的权重调节就是调整词语顺序，越靠前权重越大，越靠后权重越低。
- ② 可以通过下面的语法来对关键词设置权重，一般权重设置在 0.5~2 之间，可以通过选中词汇，按 **ctrl/command+↑↓** 来快速调节权重，每次调节为 0.1，也可以直接输入。



ctrl/command+↑↓

1.2 Prompt Notes

C. 标点符号

特定的标点符号具有一定地作用，这里先简单介绍一些基础的用法，详细内容在进阶部分进行介绍。

- ， 逗号，分割提示词，有一定的权重排序功能，逗号前权重高，逗号后权重低。
- ： 冒号，自定义数值，比如`(blue hair:1.5)`，这个就是权重1.5的意思。
- () 圆括号，`(blue hair)` 没有数值，**增加权重，默认为权重1.1, 两个圆括号是 $1.1 * 1.1$** 。
- [] **方括号，减少权重**，方括号不支持自定义数值，`[blue hair]`在新的版本是0.9，由于NovelAI使用的是在22年9月29号之前的版本，方括号的权重是减弱0.05，就是0.95。**请注意方括号是不能自定义权重的。**
- { } 花括号，在NovelAI日版本，他的权重加强就是花括号，花括号的权重为加强0.05，就是原1.05倍。

1.2 Prompt Notes

C. 标点符号

Example

- `a (word)` 将提高“word”的关注度1.1倍
- `a ((word))` 将“word”的关注度提高1.21倍 ($= 1.1 * 1.1$)
- `a [word]` 将“word”的关注度降低1.1倍
- `a (word:1.5)` 将“word”的关注度增加1.5倍。
- `a (word:0.25)` 将“word”的关注度降低四倍 (即 $1/0.25$)

1.2 Prompt Notes

C. 标点符号

Example

括号加数字

例: (white shoes:2),

含义: 调节白鞋的权重为原来的2倍 (增强)

例: (White shoes:0.8),

含义: 调节白鞋的权重为原来的0.8倍 (减弱)

套括号

圆括号: (((White shoes))),

每套一层, 额外 X 1.1倍

含义: 调节白鞋的权重为原来的
 $1.1 \times 1.1 \times 1.1 = 1.331$ 倍 (增强)

大括号: {{ {White shoes } }},

每套一层, 额外 X 1.05倍

含义: 调节白鞋的权重为原来的
1.15倍 (增强)

方括号: [[[White shoes]]],

每套一层, 额外 X 0.9倍

含义: 调节白鞋的权重为原来的
0.729倍 (减弱)

1.2 Prompt Notes

C. 标点符号

让某些词语更加突出的方法

混合: black / White shoes ,

混合两个描述同一对象的提示词要素

含义: 生成黑色和白色混合的鞋子

迁移: [black/White/red] shoes ,

**连续生成具有多个不同特征的对象,
不断迁移**

含义: 先生成黑色鞋子, 再生成白鞋, 在生成
红色鞋子

迭代: (White shoes:socks:0.8) ,

**与采样进程关联, 一定阶段以后再
生成特定对象**

含义: 进程到达80% (0.8) 之前生成白鞋,
80%之后再生成袜子

1.2 Prompt Notes

D. Emoji 和颜文字

Emoji(表情符号)由于其单个字符的特点，在语义理解方面展现出出色的准确性，因此在需要的情况下可以采用。例如，在比“✌”这种动作上，使用表情符号可以取得非凡的效果。对于某些模型来说，利用颜文字可以在一定程度上控制表情。目前模型对西方颜文字支持情况比较好。

UN Emoji Charts

Emoji List, v15.0

[Index & Help](#) | [Images & Rights](#) | [Spec](#) | [Proposing Additions](#)

[Adopt a Character](#)

Indian Type Foundry

This chart provides a list of the Unicode emoji characters and sequences, with single image and annotations. Clicking on a Sample goes to the emoji in the full list. The ordering of the emoji and the annotations are based on [Unicode CLDR data](#). Emoji sequences have more than one code point in the **Code** column. Recently-added emoji are marked by a @ in the name and outlined images.

Emoji with skin-tones are not listed here: see [Full Skin Tone List](#).

For counts of emoji, see [Emoji Counts](#).

While these charts use a particular version of the Unicode Emoji data files, the images and format may be updated at any time. For any production usage, consult those data files. For information about the contents of each column, such as the [CLDR Short Name](#), click on the column header. For further information, see [Index & Help](#).

Smileys & Emotion

face-smiling

No	Code	Sample	CLDR Short Name	Other Keywords
1	U+1F600		grinning face	face grin grinning face
2	U+1F603		grinning face with big eyes	face grinning face with big eyes mouth open smile
3	U+1F604		grinning face with smiling eyes	eye face grinning face with smiling eyes mouth open smile
4	U+1F601		beaming face with smiling eyes	beaming face with smiling eyes eye face grin smile
5	U+1F606		grinning squinting face	face grinning squinting face laugh mouth satisfied smile
6	U+1F605		grinning face with sweat	cold face grinning face with sweat open smile sweat

西方 [编辑]

在西方，颜文字大多是由左到右看的，并且需要顺时针旋转90度才能看到。因此，西方最常用的颜文字都是左边为眼睛、中间为鼻子（有时候不被包括在内）、而右边则为嘴巴。^[3]通常情况下，冒号会被用作代表眼睛，除非该表情表情是在眨眼。在这个情况下，分号就会被用作代表眼睛。然而，等号、8、和字母B大写都能代表其他种类的眼睛，如“8”能代表戴着眼镜的人。另外，人们有时候会用大括号“[”代表胡子，并且会放在嘴巴“(或)"上。^[4]

颜文字	意思
: -) :) : o) :] : 3 : c) : > =] 8) =) :) : ^) : -)	微笑或快乐。[5][6][7]
: - D : D 8 - D 8 D x - D x D X - D X D = - D = D = - 3 = 3 B ^ D	笑[5]、咧嘴[6][7]、戴着眼镜笑[4]
: -))	非常高兴或有双下巴[4]
> : [: - (: (: - c : c : - < : > c : < : - [: [: {	皱着眉头[5][6][7]、伤心[8]
: - : @ > : (愤怒[4]
: ' - (: ' (哭泣[8]
: ' -) : ')	喜极而泣[8]
D : < D : D 8 D ; D = D X v . v D - ' :	恐怖、厌恶、悲伤、恐慌[6][7]
> : O : - O : O ° O ° ° O : O o _ O o _ O o . O 8 - O	惊讶、震撼[5][9]、打哈欠[10]
: * : * ^ (') { ')	亲吻，情侣接吻[4]
: -) ;) * -) *) ; -] ;] ; D ; ^) : - ,	眨眼[5][6][7]、傻笑[9][10]
> : P : - P : P X - P x - p x p X P : - p : p = p : - b : b : b : - b : - b : b	伸出舌头、厚脸皮、俏皮[5]

表情符号参考: <https://unicode.org/emoji/charts/emoji-list.html>

颜文字参考: <https://zh.wikipedia.org/w/index.php?title=%E8%9B%8A%E6%88%AA&oldformat=true>

1.3 Prompt structures

文生图模型的精髓在于 Prompt 提示词，如何写好 Prompt 将直接影响图像的生成质量。



vectorartz, (masterpiece!), (best quality), a beautiful illustration of
a cute female doctor standing among the colorful flowers, black hair, white outfit, stethoscope,
full shot, half-length, looking at viewer, cinematic lighting, soft lights, simple background,
by Michael Whelan, trending on dribbble

1.4 Prompt types

内容型 Prompt

多数时候是因你想创作的内容而异的，每次都要改成不同的东西

人物及主体特征

服饰穿搭 white dress

发型发色 blonde hair, long hair

五官特点 small eyes, big mouth

面部表情 smile

肢体动作 stretching arms

场景特征

室内、室外 indoor / outdoor

大场景 forest, city, street

小细节 tree, bush, white flower

环境光照

白天黑夜 day / night

特定时段 morning, sunset

光环境 sunlight, bright, dark

天空 blue sky, starry sky

补充：画幅视角

距离 close-up, distant

人物比例 full body, upper body

观察视角 from above, view of back

镜头类型 wide angle, Sony A7

1.4 Prompt types

标准化 Prompt

相对固定可以直接替换使用的 Prompt

画质提示词

通用高画质

`best quality, ultra-detailed, masterpiece, hires, 8k`

特定高分辨率类型

`extremely detailed CG unity 8k wallpaper` (超精细的8K Unity游戏CG)

`unreal engine rendered` (虚幻引擎渲染)

画风提示词

插画风 `illustration, painting, paintbrush`

二次元 `anime, comic, game CG`

写实系 `photorealistic, realistic, photograph`

1.5. Auxiliary generation prompt tools

A. ChatGPT

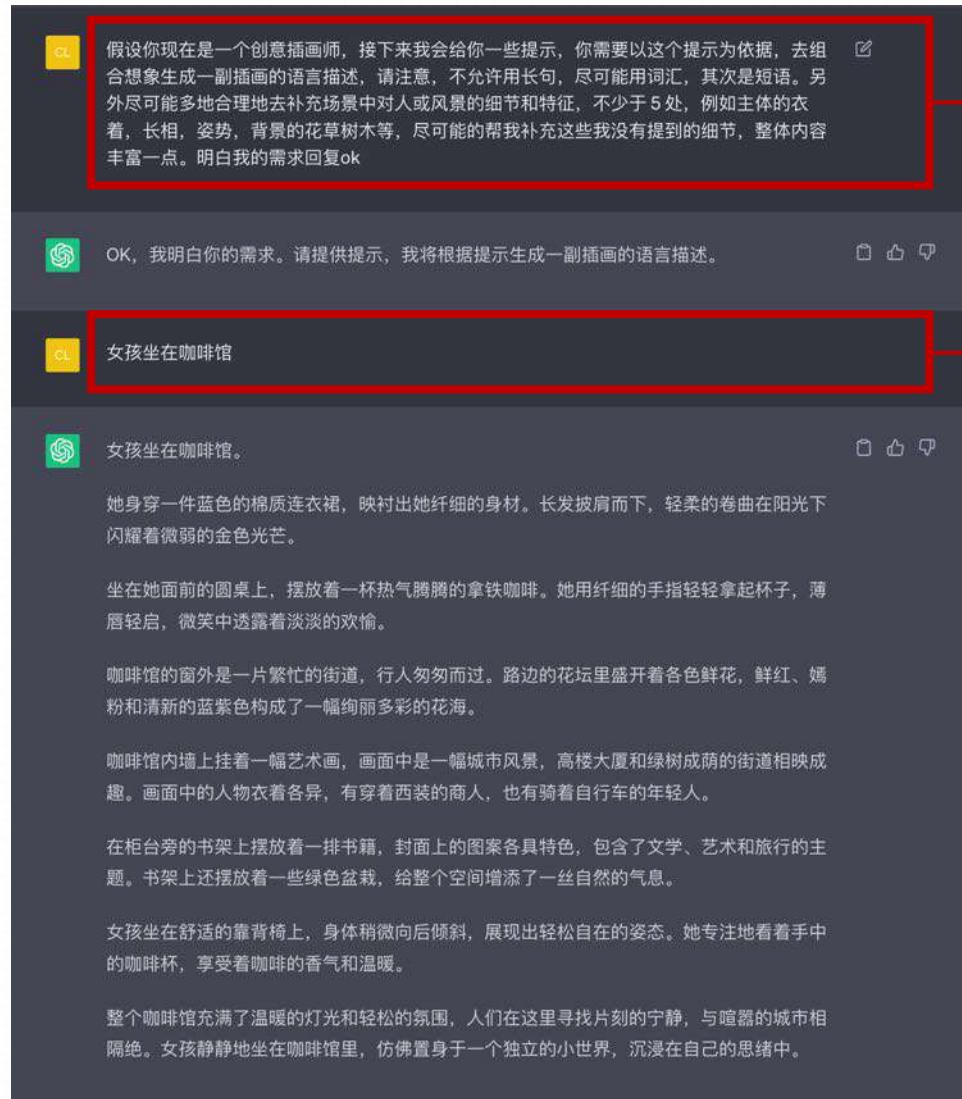
在编写提示词时，我们可能会遇到不知道如何描述一个场景的情况。例如，我们想要绘制一副女孩在咖啡馆里看书的画面，但不知道如何描述这个场景。这时，我们可以借助 ChatGPT 来提供灵感，打开思路。

具体操作如下：首先给 ChatGPT 一个基本的提示，让它帮助我们产生画面描述。

“假设你现在是一个创意插画师，接下来我会给你一些提示，你需要以这个提示为依据，去组合想象生成一副插画的语言描述，请注意，不允许用长句，尽可能用词汇，其次是短语。另外尽可能多地合理地去补充场景中对人或风景的细节和特征，不少于 5 处，例如主体的衣着，长相，姿势，背景的花草树木等，尽可能的帮我补充这些我没有提到的细节，整体内容丰富一点。”

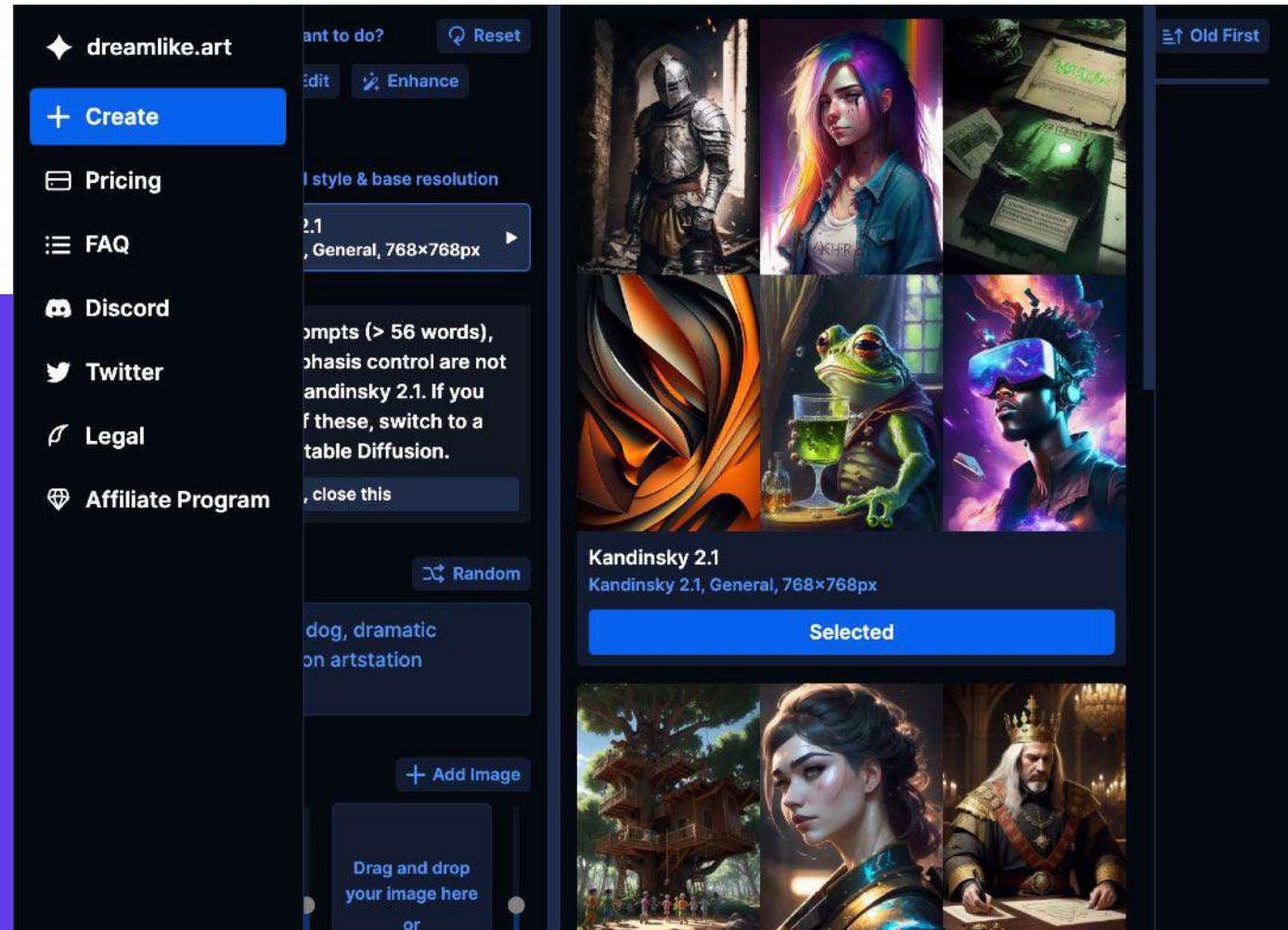
之后，我们输入一个简单的场景描述，例如 “女孩坐在咖啡馆” ，ChatGPT 就会给出对应的画面描述，接下来，我们从中提取出重要的提示词即可。

1.5. Auxiliary generation prompt tools



1.5. Auxiliary generation prompt tools

B. dreamlike.art



我们可以借助 ChatGPT 帮我们生成提示词参考:

1. 给 ChatGPT 一段示例参考:
<https://dreamlike.art/guides/using-openai-chat-gpt-to-write-stable-diffusion-prompts>
2. 根据参考生成 Prompts，再添加细节润色

1.5. Auxiliary generation prompt tools

C. Moonvy

The screenshot shows the OpenPromptStudio interface with two workspaces and a central style library.

Top Workspace:

- Title:** untitled
- Prompt:** apple, forest ::-1 big bad wolf, wood ::2 unreal engine, cinematic lighting, UHD, super detail --aspect 2:3
- Style Tags:** apple 苹果, forest ::-1 big bad wolf 森林:: -1大灰狼, wood ::2 unreal engine 伍德: : 2个虚幻引擎, super detail --aspect 2:3, cinematic lighting 电影光效
- Quality:** UHD 超高清
- Buttons:** 复制, 上, 下, 导出, 稳定扩散模型
- Model Selection:** stable-diffusion-webui

Bottom Workspace:

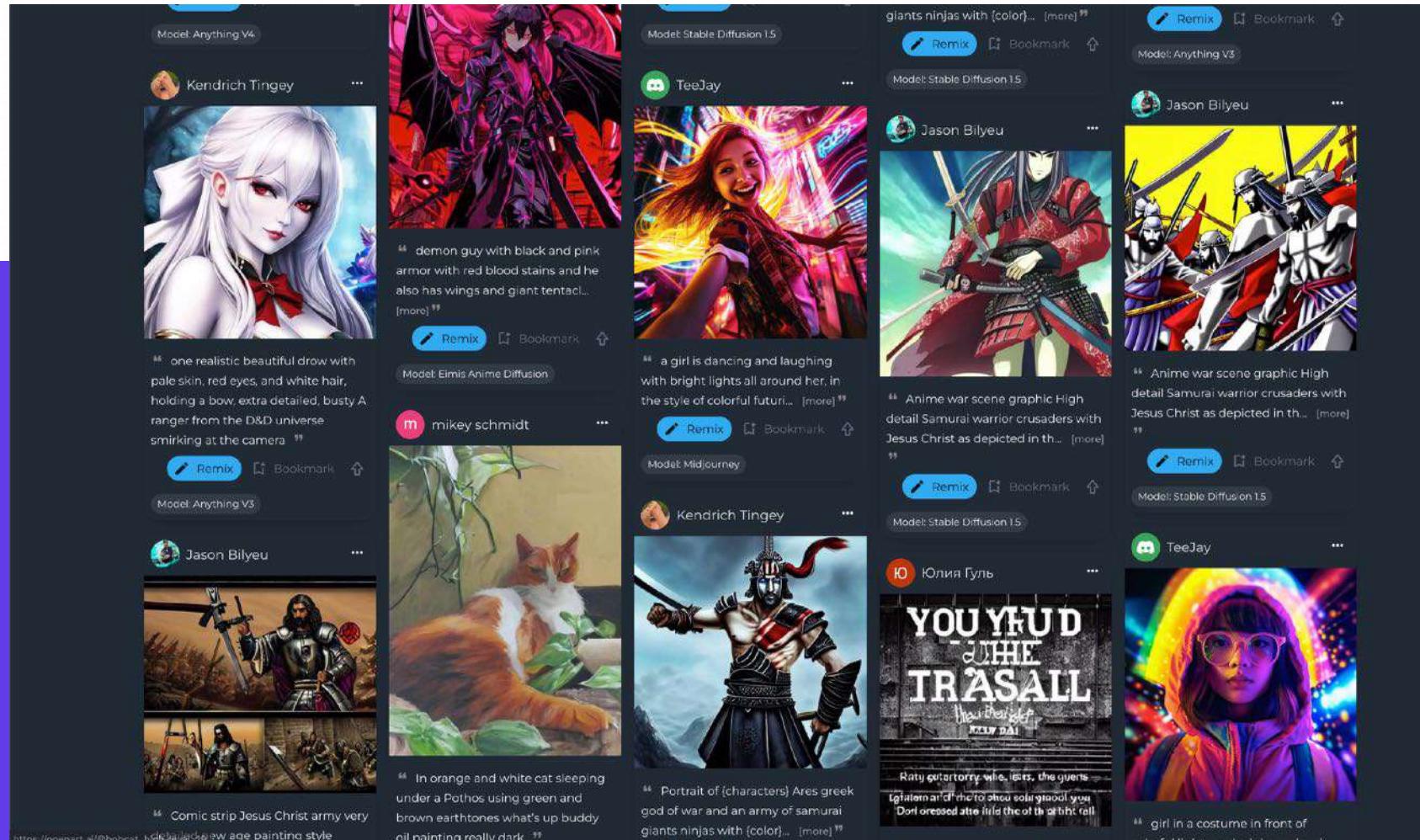
- Title:** untitled
- Prompt:** 输入提示词
- Output:** 输出与输入相同
- Style Tags:** Midjourney
- Buttons:** 复制, 上, 下, 导出, 稳定扩散模型
- Model Selection:** Midjourney
- Tools:** 添加工作区, 复制链接, 导出 PNG 时固定尺寸, 导出 PNG 到剪贴板
- Logos:** 广告, N, 次世代账号合租平台

Style Library (右侧):

- Quality:** 提示词词典, 连接我的 Notion
- Painting Styles:** 质量, 绘画, 画面效果, 容貌, 构图, 命令
- Styles:**
 - 普通:** apple 苹果, forest ::-1 big bad wolf 森林:: -1大灰狼, wood ::2 unreal engine 伍德: : 2个虚幻引擎, super detail --aspect 2:3, cinematic lighting 电影光效
 - 风格:** Surrealism 超现实主义, Color Field painting 色块画, Art Deco 装饰艺术, rococo style 洛可可风格, high detail 高细节, blind box toy style 盲盒玩具风格, Cubist Futurism 立体派未来主义, Conceptual art 概念艺术, Futurism 未来主义, Social realism 社会现实主义, interior architecture 室内建筑, Renaissance 文艺复兴, Neoclassicism 新古典主义, modern 现代, anime 动漫, Minimalism 极简主义, Romanticism 浪漫主义, Gothic art 哥特式艺术, American propaganda poster 美国宣传海报, Tonalism 色调主义, Baroque 巴洛克, Fauvism 野兽派, Expressionism 表现主义, Carl Larsson 卡尔·拉尔松风格的画作, Op art 视错觉艺术, Realism 现实主义, Contemporary art 当代艺术, Genre painting 风俗画, Constructivism 构成主义, Mannerism 后期文艺复兴, Bauhaus 包豪斯, Action painting 行动绘画, by Alfonso Mucha 由阿尔方斯·穆卡制作, Dutch Golden Age painting 荷兰黄金时期绘画, Pop art 波普艺术, Monet 莫奈, Northern Renaissance 北方文艺复兴, Dadaism 达达主义, Ukiyo-e 浮世绘, Luminism 明亮主义, Abstract expressionism 抽象表现主义, Impressionism 印象派, Pre-Raphaelite Brotherhood 前拉斐尔派兄弟会

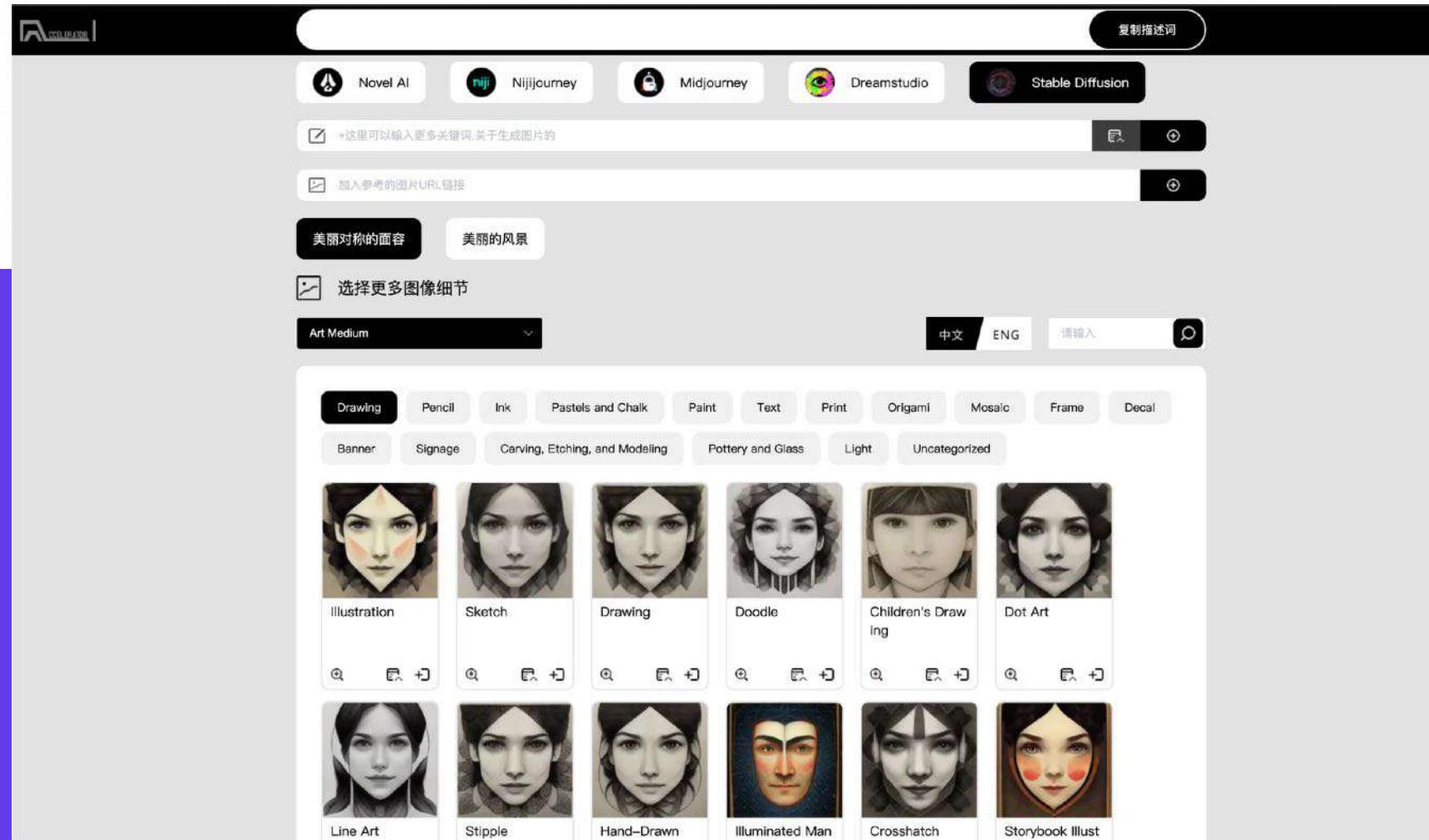
1.5. Auxiliary generation prompt tools

D. Openart.ai



1.5. Auxiliary generation prompt tools

E. 辅助生成 prompt 工具



2. Stable Diffusion Webui

1. **txt2img**

2. **img2img**

3. **Extras**

4. **PNG Info**

5. **Checkpoint Merger**

6. **Train**

7. **Settings**

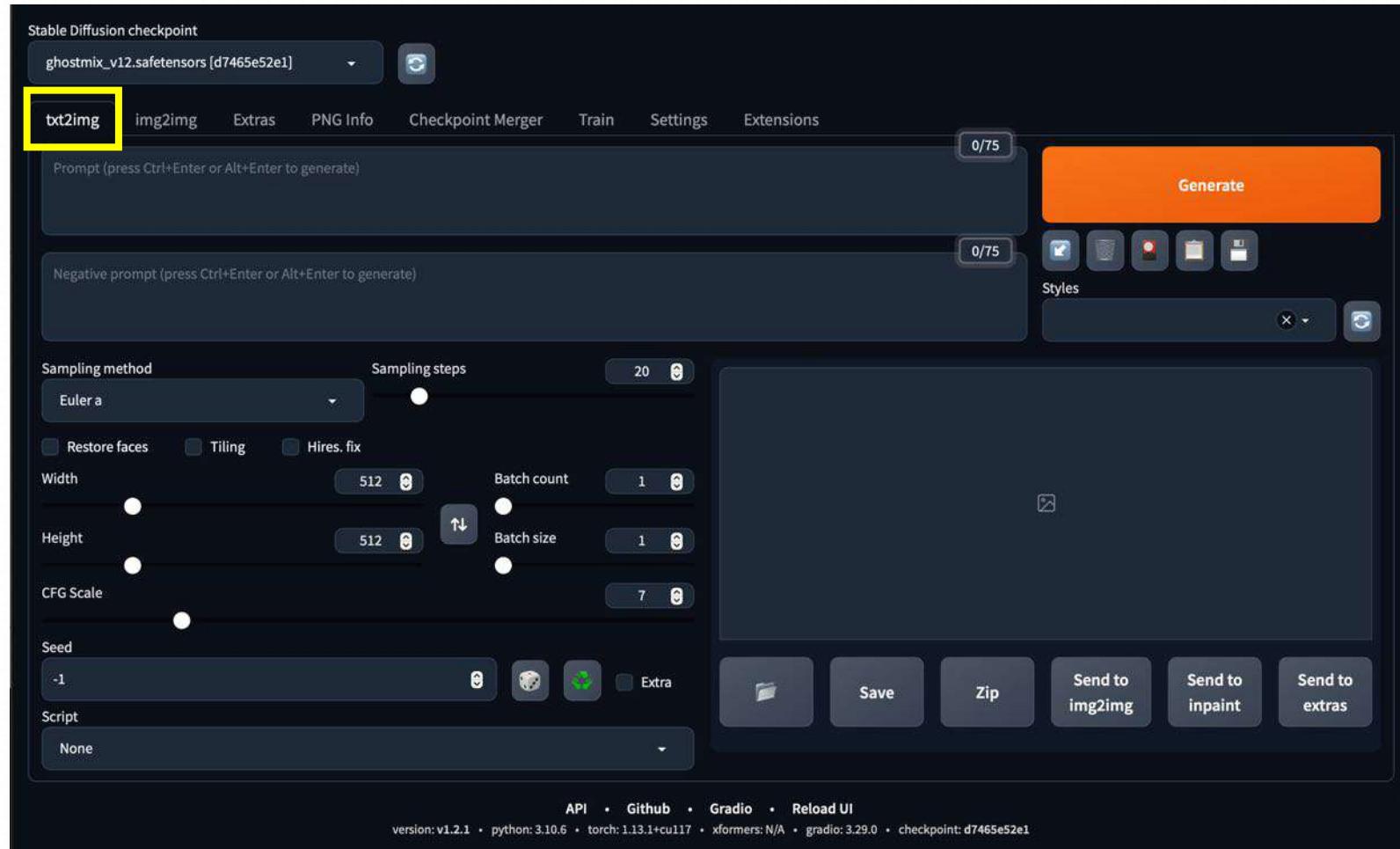
8. **Extensions**

2. Stable Diffusion Webui

The image shows the Stable Diffusion WebUI interface with several annotations:

- 基础模型选择**: Basic model selection dropdown.
- 提示词输入**: Prompt input field with a "Generate" button.
- 一些参数输入的快捷操作**: Quick operation buttons for clearing keywords and reusing previous generation parameters.
- 采样器**: Sampling method dropdown (e.g., Euler, dpm).
- 采样步数**: Sampling steps slider (set to 25).
- 图片宽高**: Resolution settings (Width: 512, Height: 512).
- CFG比例**: CFG scale slider.
- 种子**: Seed input field (28367815).
- 批次**: Batch size slider (set to 1).
- 插件操作区**: Additional Networks section (ControlNet, Script, None).
- 生成图片**: Generated image of a girl with black hair and red eyes.
- 输出文件夹**: Output folder options: Open output folder, Save image and parameters, Save as zip, Send to img2img, Send to inpaint, Send to extras.
- 生成信息**: Generation information: Step: 25, Sampler: Euler, CFG scale: 7, Seed: 3807815, Size: 512x512, Model hash: 30953ab0de, Model: minniu/mi...
- 说明** (Annotations):
 - 负面提示词: Input that you don't want SD to generate.
 - 图片宽高: Change picture resolution.
 - 批次: Control how many batches to generate, each containing how many images.
 - 采样步数: Sampling steps, higher is better but usually 25 is enough to reduce GPU load.
 - 采样器: Sampling algorithm, Euler series is more realistic, dpm series is more dreamlike.
 - CFG比例: Image and prompt consistency ratio, lower values often lead to more creative results.
 - 种子: Same seeds and prompts will result in nearly identical images.

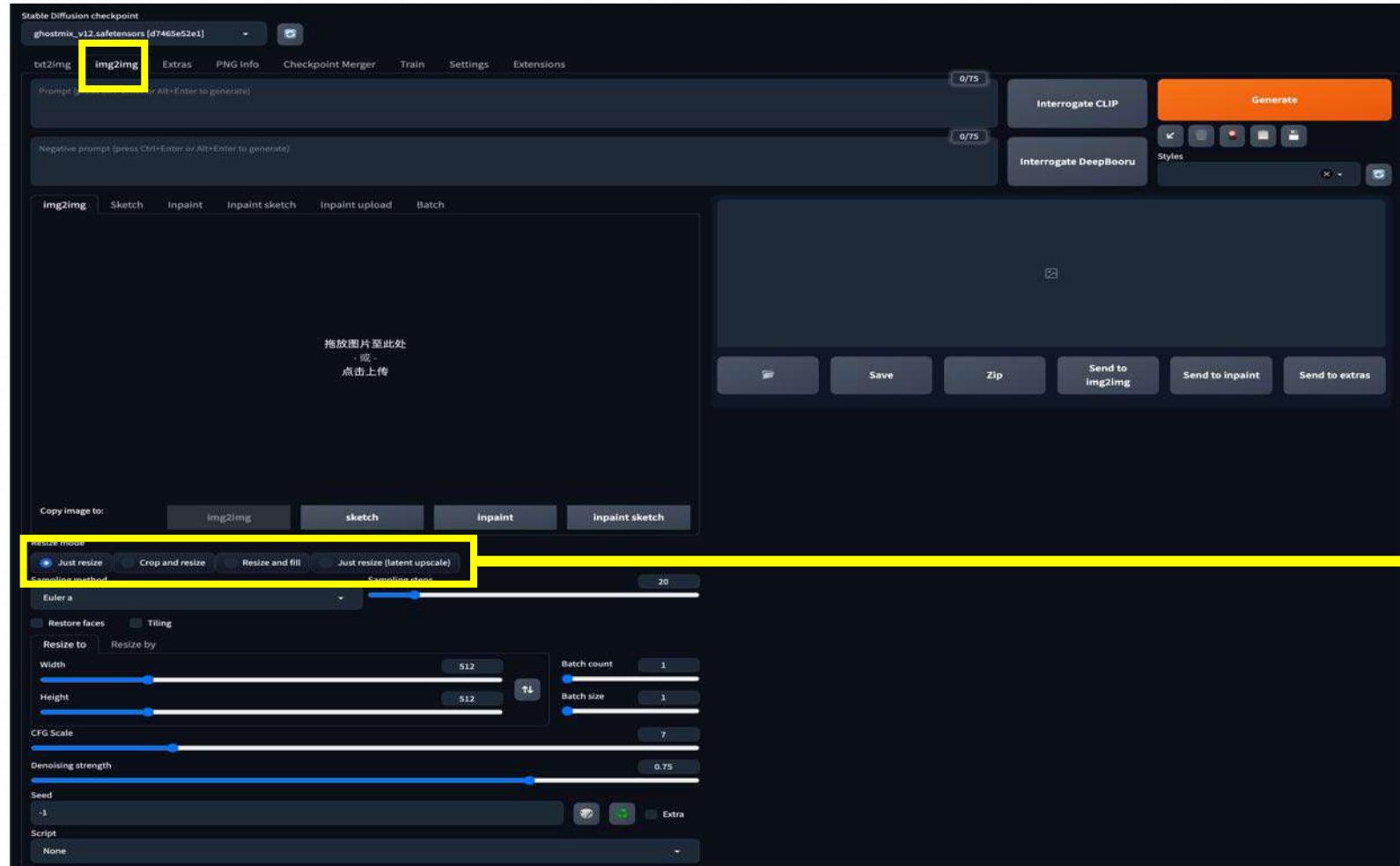
2.1 txt2img



- **txt2img** : 标准的文字生成图像
- **img2img** : 根据图像成文范本、结合文字生成图像
- **Extras** : 优化(清晰、扩展)图像
- **PNG Info** : 图像基本信息
- **Checkpoint Merger** : 模型合并
- **Textual inversion** : 训练模型对于某种图像风格
- **Settings** : 默认参数修改
- **Extensions** : 插件

2.2 img2img

Img2img 与 txt2img 的大部分参数一致，这里主要说明不一致的部分



内容输入部分：

这里主要增加的是要模仿的图片，可以是手绘的、也可以是类似的；

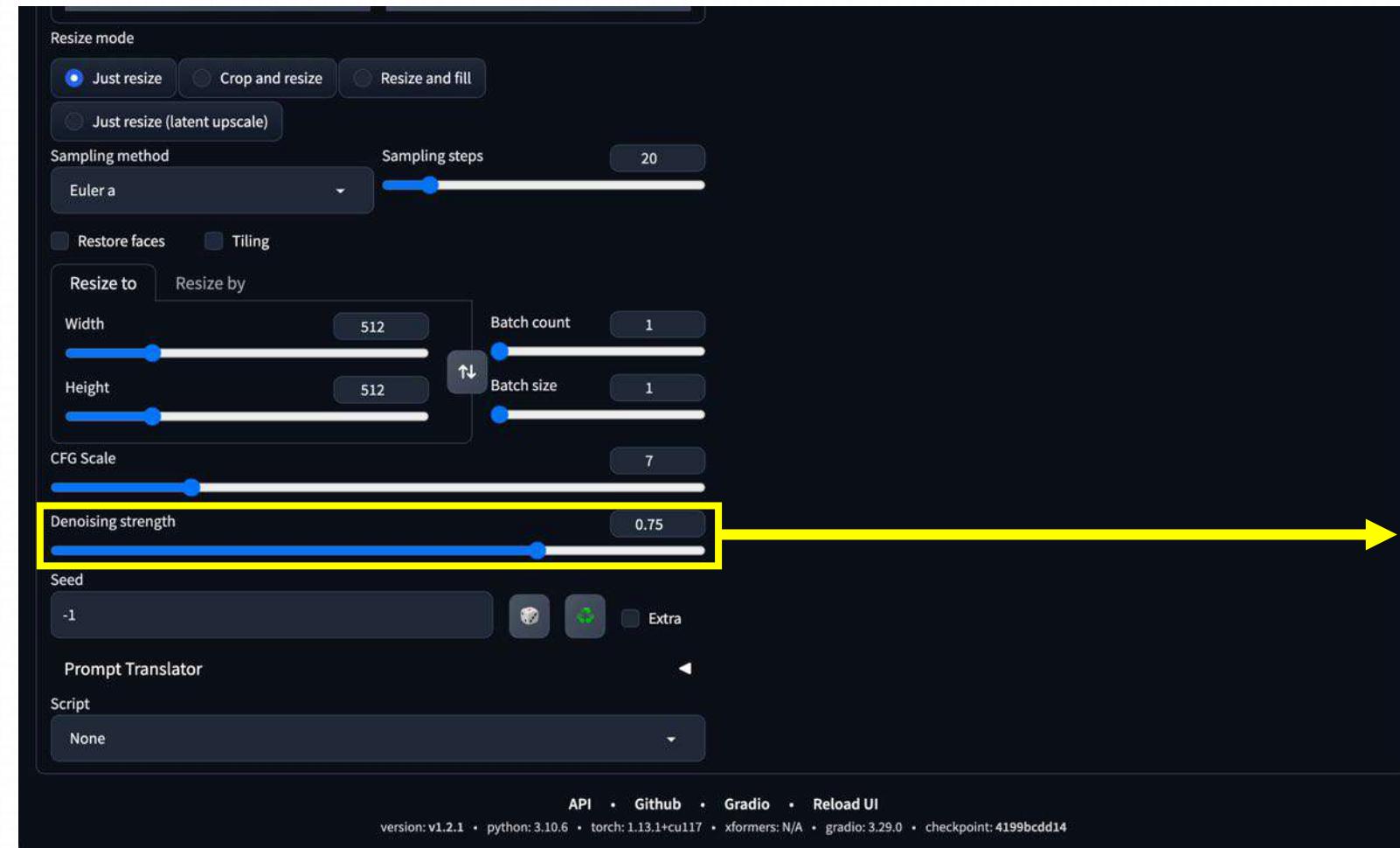
其他文字信息类似，这里依然是描述越准确越好；

对于其中参数主要是图像是否要保存相同尺寸：

Just resize、Crop and resize、Resize and fill 这三种模式保证图输出效果，因为下面会有新的尺寸，是填充还是对应缩放；

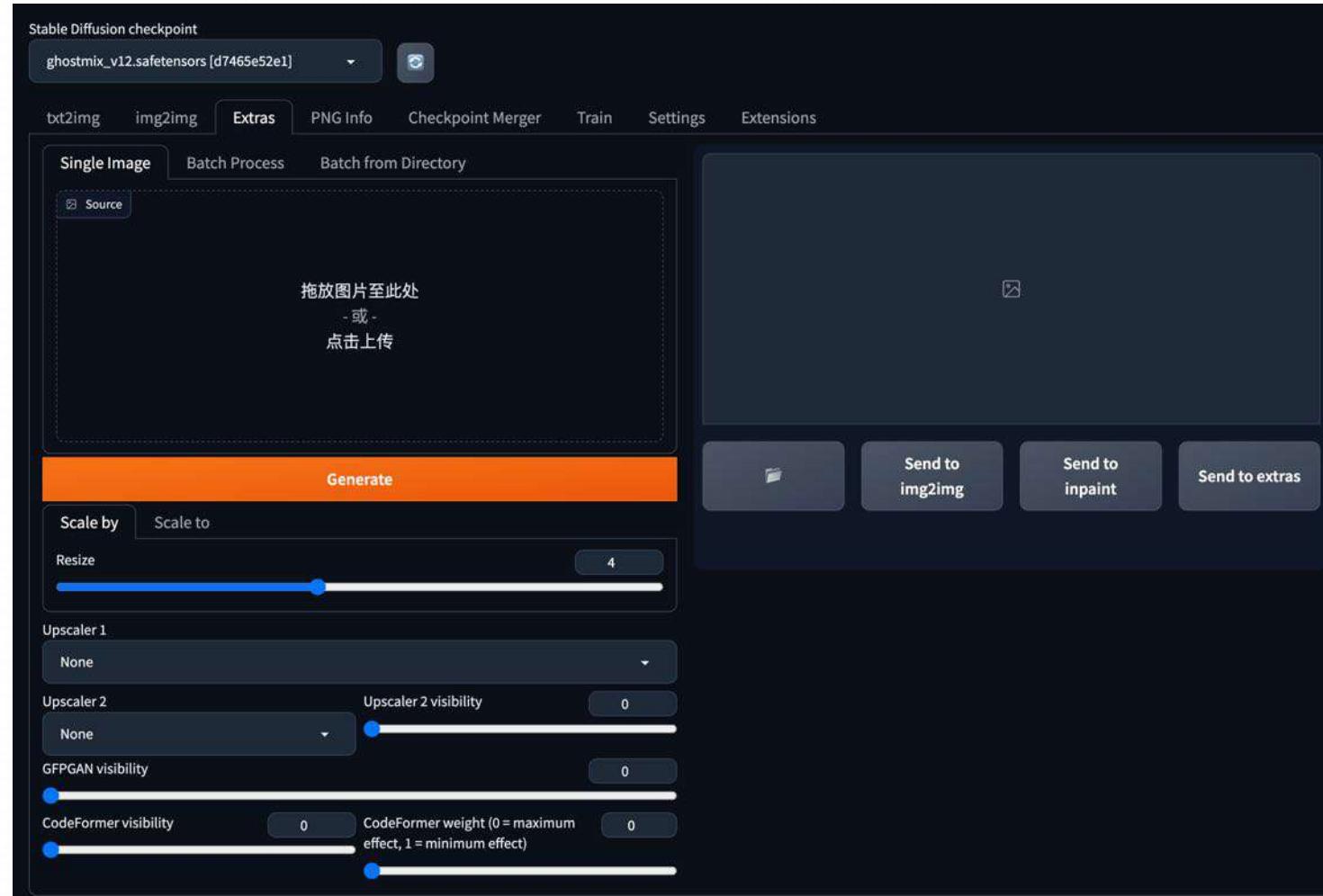
2.2 img2img

2. img2img



这部分的信息与txt2img参数几乎一致，主要新增加的是：
Denoising strength 与原图一致性的程度，一般大于0.7出来的都是新效果，小于0.3基本就是在原图的基础上进行小的修改调整

2.3 Extras



该部分主要将图像进行优化，其中很多方法的模型使用的时候会自动下载，很容易下载失败导致报错。

图片导入，也可以通过其他模块中的 send to extras 直接使用

2.4 PNG Info

Stable Diffusion checkpoint
ghostmix_v12.safetensors [d7465e52e1]

txt2img img2img Extras **PNG Info** Checkpoint Merger Train Settings Extensions

Source 

parameters

(masterpiece, top quality, best quality, official art, beautiful and aesthetic:1.2), (1girl), extreme detailed,(fractal art:1.3),colorful,highest detailed
Negative prompt: (worst quality, low quality:2), monochrome, zombie,overexposure, watermark,text,bad anatomy,bad hand,extra hands,extra fingers,too many fingers,fused fingers,bad arm,distorted arm,extra arms,fused arms,extra legs,missing leg,disembodied leg,extra nipples, detached arm, liquid hand,inverted hand,disembodied limb, small breasts, loli, oversized head,extra body,completely nude, extra navel,easynegative,(hair between eyes),sketch, duplicate, ugly, huge eyes, text, logo, worst face, (bad and mutated hands:1.3), (blurry:2.0), horror, geometry, bad_prompt, (bad hands), (missing fingers), multiple limbs, bad anatomy, (interlocked fingers:1.2), Ugly Fingers, (extra digit and hands and fingers and legs and arms:1.4), ((2girl)), (deformed fingers:1.2), (long fingers:1.2),(bad-artist-anime), bad-artist, bad hand, extra legs ,(ng_deepnegative_v1_75t)

Steps: 30, Sampler: DPM++ 2M Karras, CFG scale: 6, Seed: 1015468391, Size: 500x500, Model hash: d7465e52e1, Model: ghostmix_v12, Version: v1.2.1

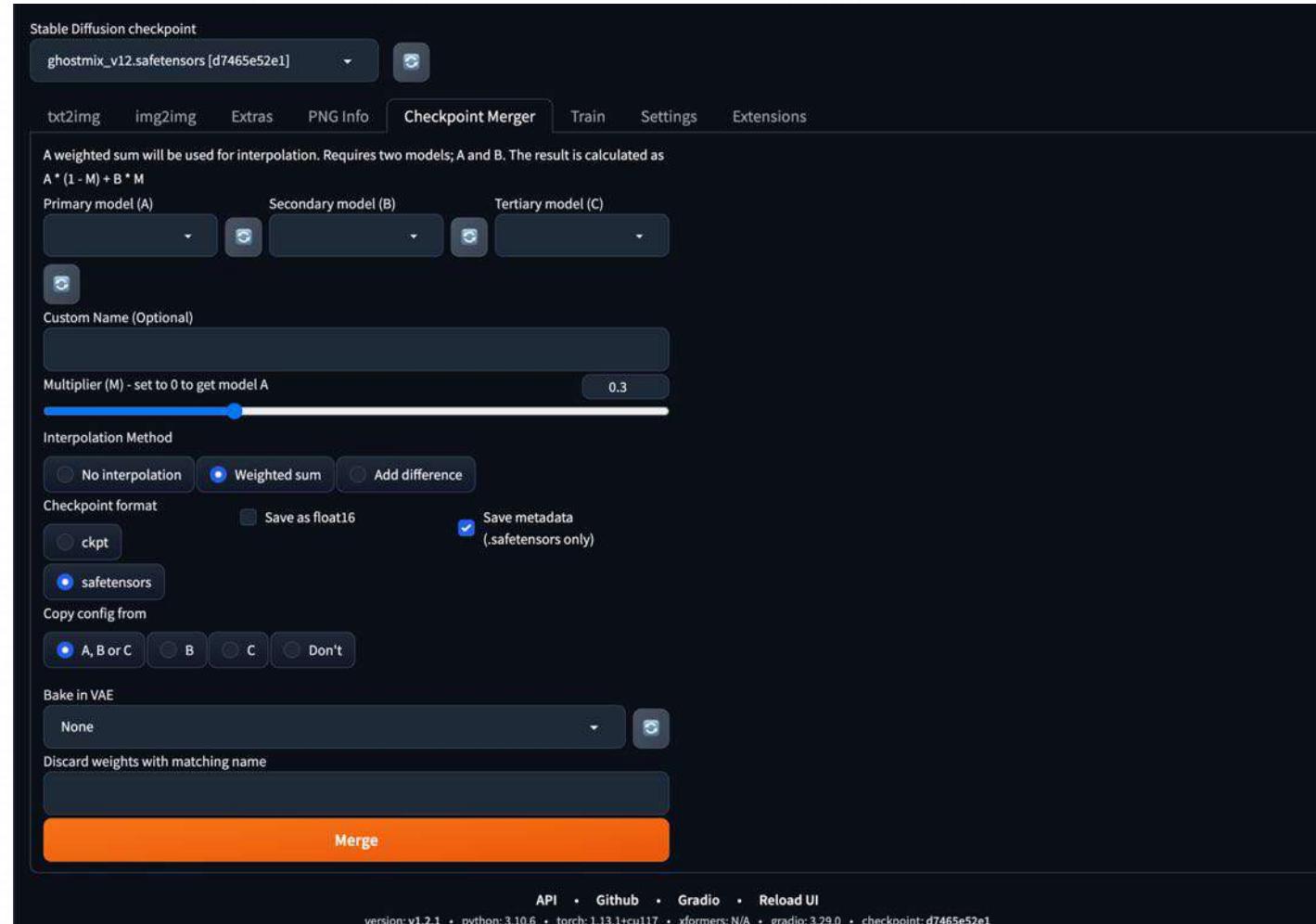
Send to txt2img **Send to img2img** **Send to inpaint**

Send to extras

API · Github · Gradio · Reload UI
version: v1.2.1 · python: 3.10.6 · torch: 1.13.1+cu117 · xformers: N/A · gradio: 3.29.0 · checkpoint: d7465e52e1

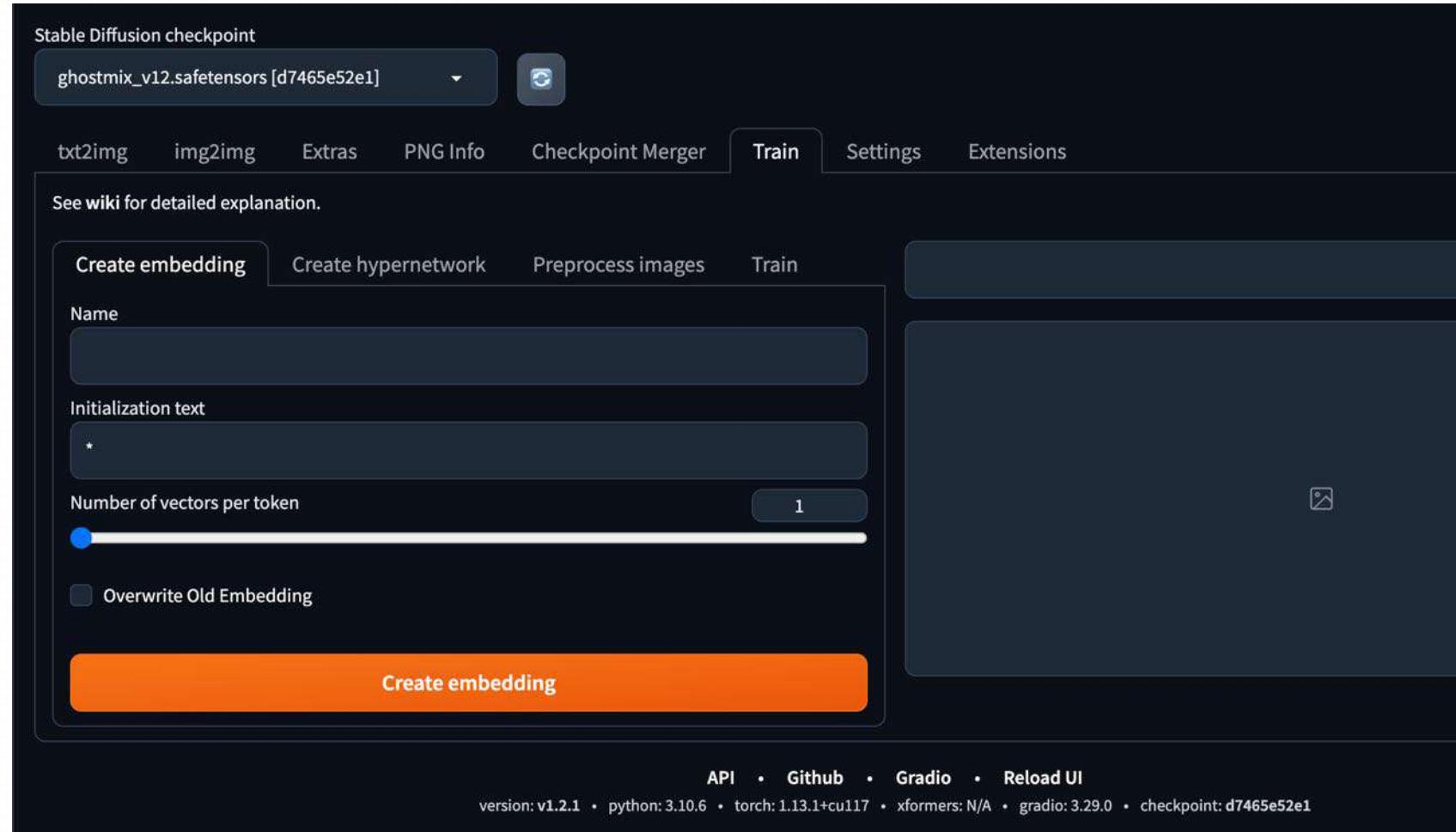
图片信息，意思是使用 SD 绘制的图，放在这个页面中，在右侧会显示这个图片的一些参数，包括关键词

2.5 Checkpoint Merger



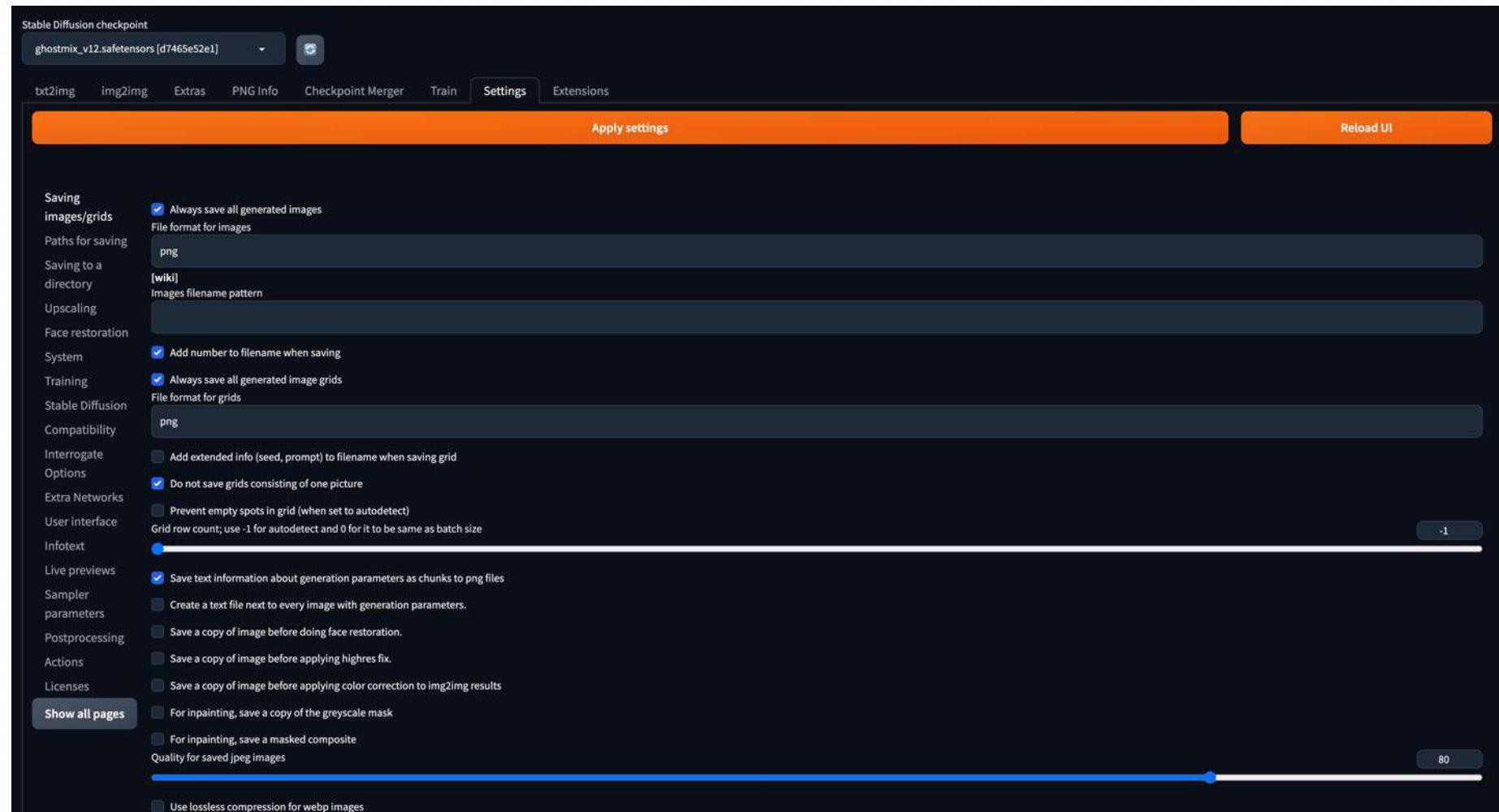
是把多个模型混合起来

2.6 Train



Train, 这个就是训练自己的模型，新手用的比较少，后面可用作工作流的部署。

2.7 Settings



2.8 Extensions

Stable Diffusion checkpoint
ghostmix_v12.safetensors [d7465e52e1]

txt2img img2img Extras PNG Info Checkpoint Merger Train Settings Extensions

Installed Available Install from URL Backup/Restore

Apply and restart UI **Check for updates** Disable all extensions
 none extra all

Extension	URL	Version	Update
<input checked="" type="checkbox"/> LDSR	built-in		
<input checked="" type="checkbox"/> Lora	built-in		
<input checked="" type="checkbox"/> ScuNET	built-in		
<input checked="" type="checkbox"/> SwinIR	built-in		
<input checked="" type="checkbox"/> prompt-bracket-checker	built-in		

API • Github • Gradio • Reload UI
version: v1.2.1 • python: 3.10.6 • torch: 1.13.1+cu117 • xformers: N/A • gradio: 3.29.0 • checkpoint: d7465e52e1

3. txt2img Interface

1. Negative Prompts

2. Samplers

3. Sampling steps

4. Restore faces & Tiling

5. Hires.fix

6. Aspect Ratios

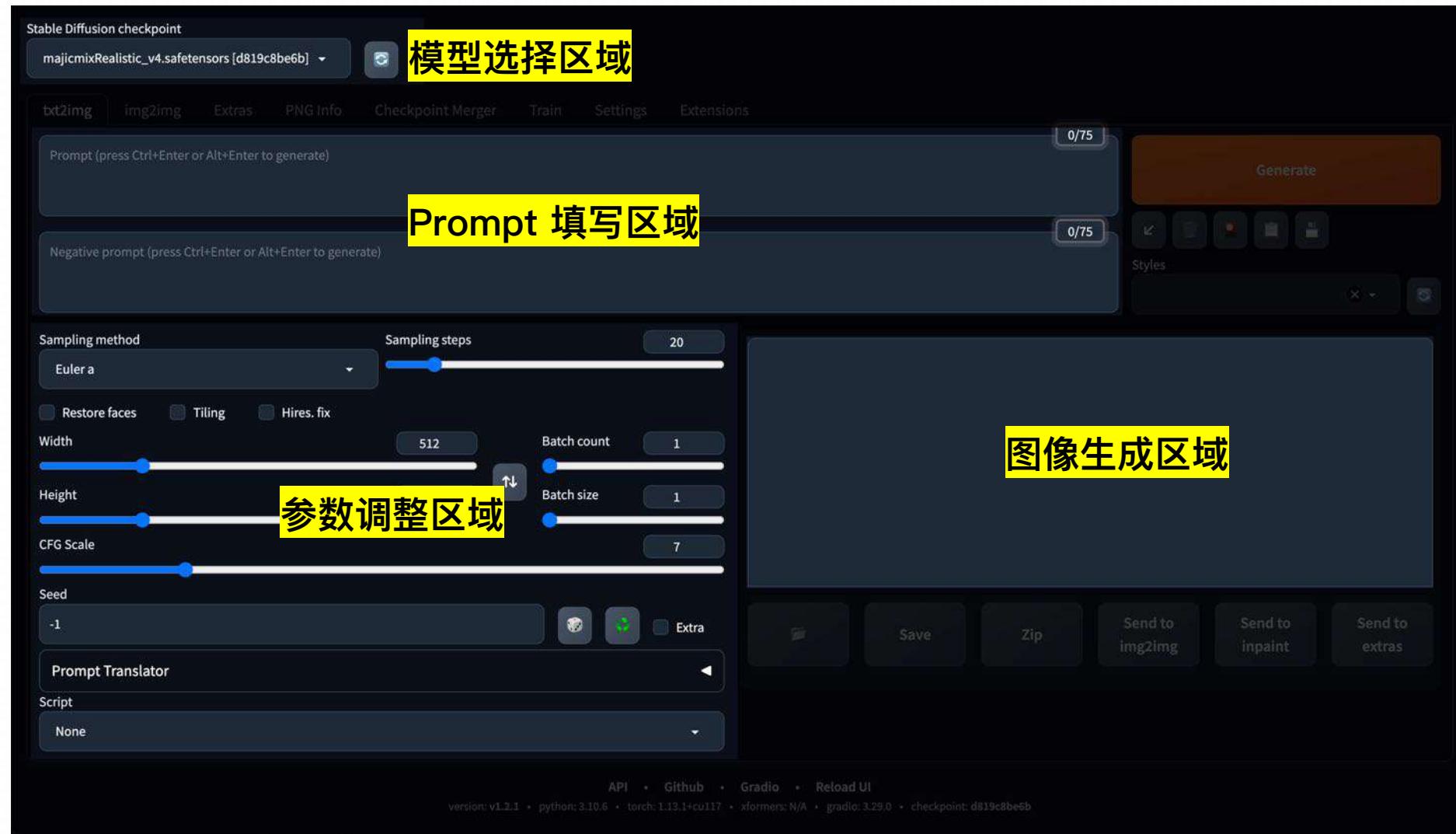
7. Batch Count & Size

8. CFG Guidance Scale

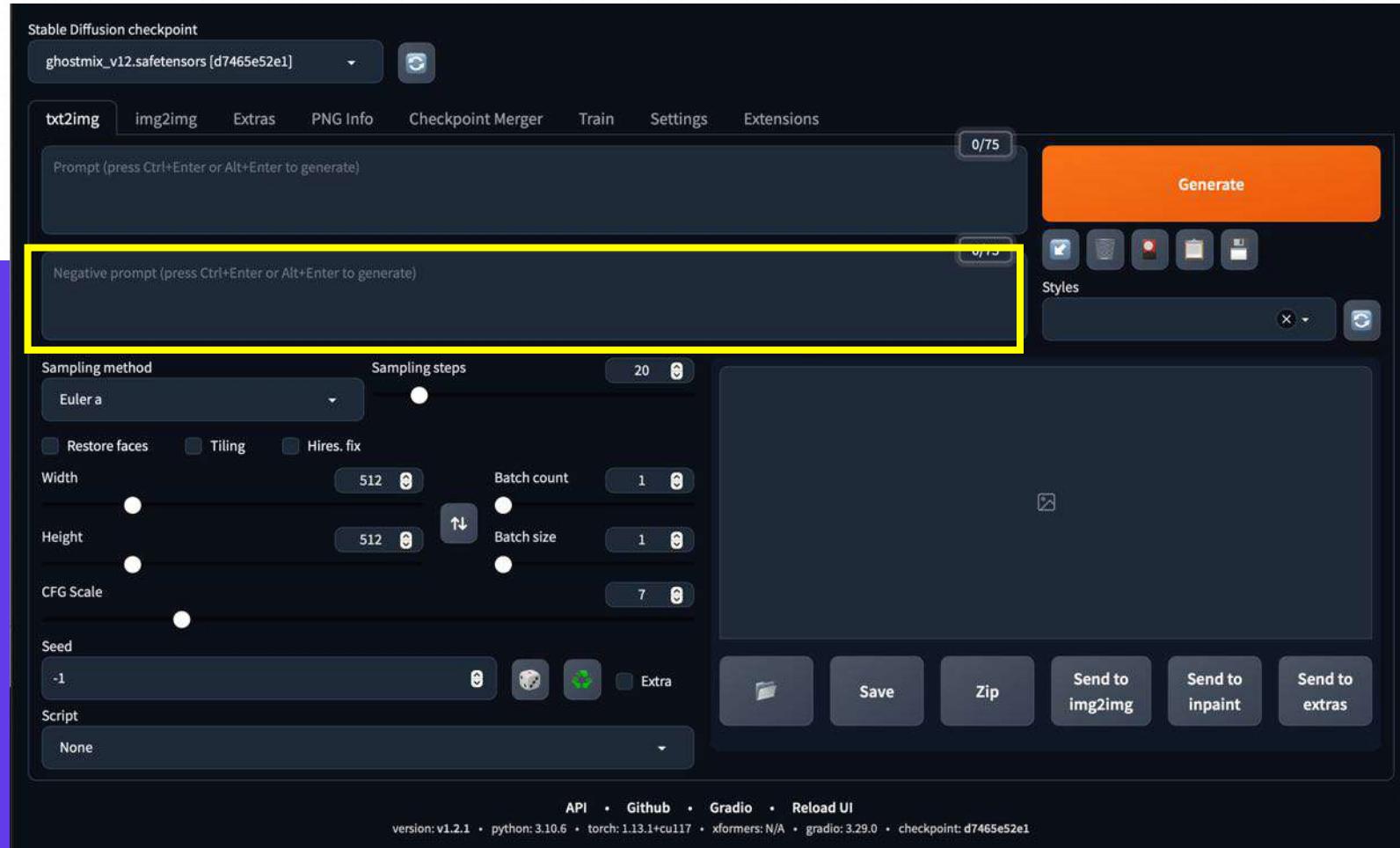
9. Seed

10. Others buttons

3. txt2img Interface



3.1 Negative Prompts



在SD中，提示词分为prompt正向提示词和Negative prompt反向提示词。

正向提示词内输入的东西是想要生成的作品，反向提示词内输入的是不希望画面生成的内容。

鉴于我们所使用的多种模型包含大量训练图片，其中可能存在大量低质量图像信息，因此我们需要通过输入反向提示词来排除这些内容。

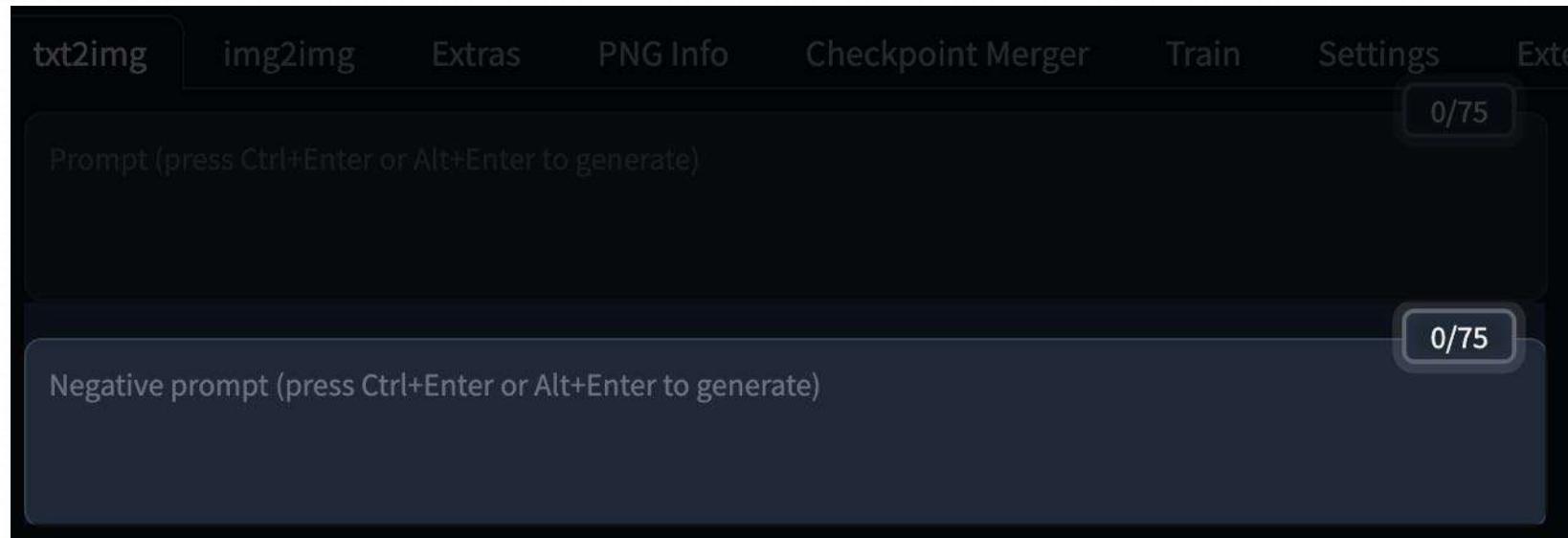
值得注意的一点是，提示词输入框中显示为0/75。许多初学者可能误解为该处可以输入75个字符，但实际上这里的75指的是75个令牌tokens。

3.1 Negative Prompts

Negative prompt 与 prompt 相反， Negative prompt 代表的是用户不希望 Stable Diffusion 生成的内容。这是 Stable Diffusion 的一个非常强大但未被充分利用的功能，它可以通过微调 positive prompt，帮助用户实现那些通常需要花费更多时间才能达到的结果。

General negative prompts :

lowers, error, cropped, worst quality, low quality, jpeg artifacts, out of frame, watermark, signature



那些通常需要花费更多时间才能达到的结果输入框内输入的标签内容就是你画面中不想要出现的东西，如：低质量的，缺手指，五官不齐等等

3.1 Negative Prompts

以人物肖像的negative prompt为例：

Negative prompts for people portraits:

deformed, ugly, mutilated, disfigured, text, extra limbs, face cut, head cut, extra fingers, extra arms, poorly drawn face, mutation, bad proportions, cropped head, malformed limbs, mutated hands, fused fingers, long neck

畸形的，丑陋的，残缺的，毁坏的，不好的比例

Prompt: Photo of a woman smiling



No negative prompt

With People negative prompt on OpenArt

<https://blog.openart.ai/2023/02/13/the-most-complete-guide-to-stable-diffusion-parameters/>

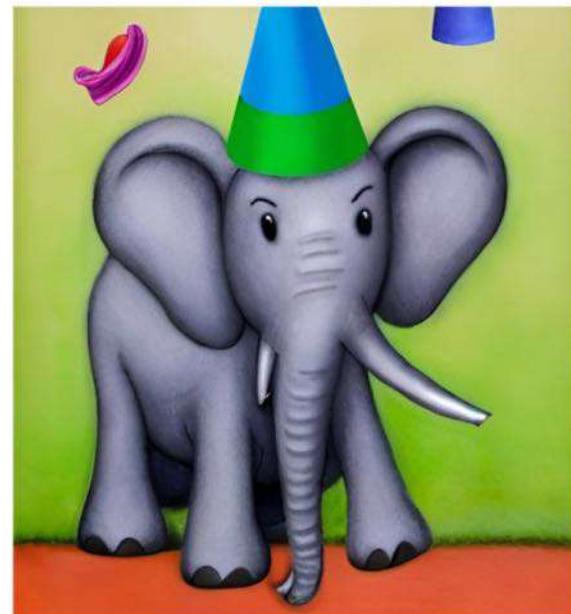
3.1 Negative Prompts

以照片类图片的negative prompt为例：

Negative prompts for photorealistic images:

illustration, painting, drawing, art, sketch

插图 绘画 素描 艺术 草图



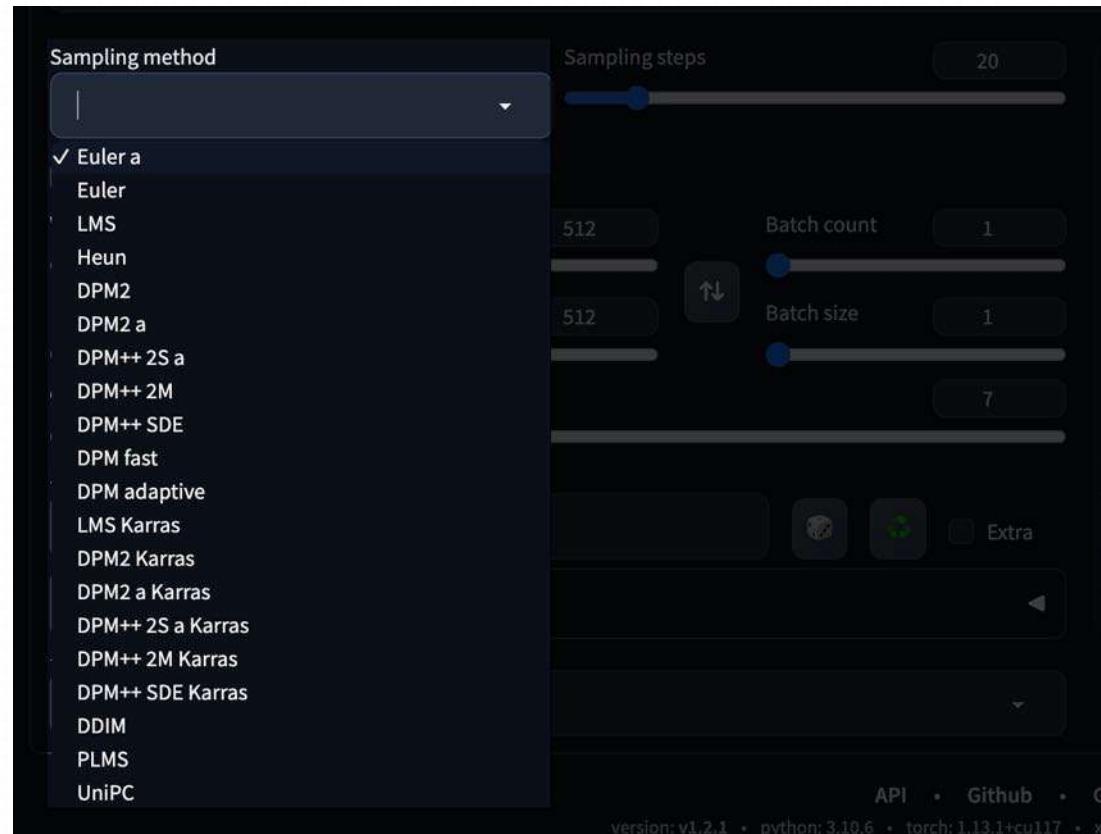
No negative prompt



Photorealistic negative prompt

3.2 Samplers (采样器)

Diffusion模型的工作原理是通过对初始的噪声画布进行去噪。这就是sampler的工作所在。简单来说，sampler是一种算法，它们在每一步之后取出生成的图像，与文本提示进行比较，然后对噪声进行一些修改，直到它逐渐达到与文本描述相匹配的图像



在SD上，我们最常用的采样器有DPM系列和Euler系列，其中DPM系列会有更加接近真实世界的效果，Euler系列会让真实感减弱同时更加富有艺术感。

常用的Sampler有三种：

- Euler a
- DPM++2S a Karras
- DDIM

3.2 Samplers (采样器)

A. Euler a



适当的 Euler a 值能够**捕捉到细节和纹理**, 但如果值太大会导致过度拟合, 生成图像出现噪点等不良效果

一句话概括: 采样生成速度最快, 但是如果说在高细节图增加采样步数时, 会产生不可控突变 (如人物脸扭曲, 细节扭曲等)

适用情况:
插画风格, ICON, 二次元图
像, 小场景

<https://zhuanlan.zhihu.com/p/612572004>

3.2 Samplers (采样器)

B. DPM++2S a Karras



采用 DPM++2S a Karras 采样方法生成高质量图像，该方法在每个时间步长中执行多次操作，同等分辨率下细节会更多，比如可以在小图下塞进全身，代价是采样速度更慢

适用情况：
写实人像，复杂场景刻画

3.2 Samplers (采样器)

C. DDIM



DDIM 采样方法可以快速生成高质量的图像，相比其他采样方法具有更高的效率，想尝试超高步数时可以使用，随着步数增加可以叠加细节

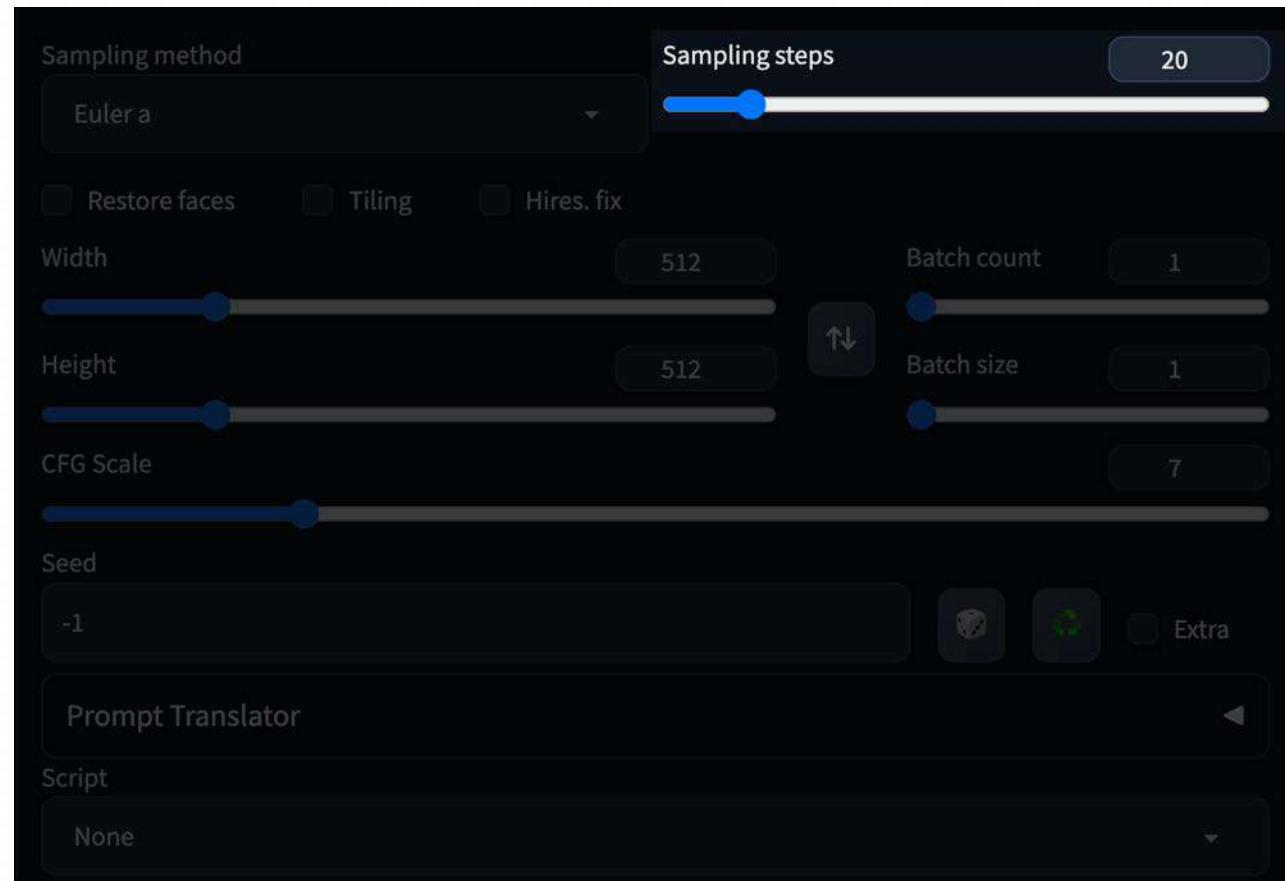
适用情况：
写实人像，复杂场景刻画

3.2 Samplers (采样器)



3.3 Sampling steps

Stable Diffusion 是通过从一个充满噪声的画布开始，然后逐步进行去噪，最终达到最终输出的图片。这个参数控制这些去噪步骤的数量。通常来说，步骤越多越好，但也有一定的限度。我们使用的默认值是25步，这应该足够生成任何类型的图像



一般来说采样迭代步数保持在 18–30 左右即可。低的采样步数会导致画面计算不完整，高的采样步数仅在细节处进行优化，对比输出速度得不偿失

3.3 Sampling steps



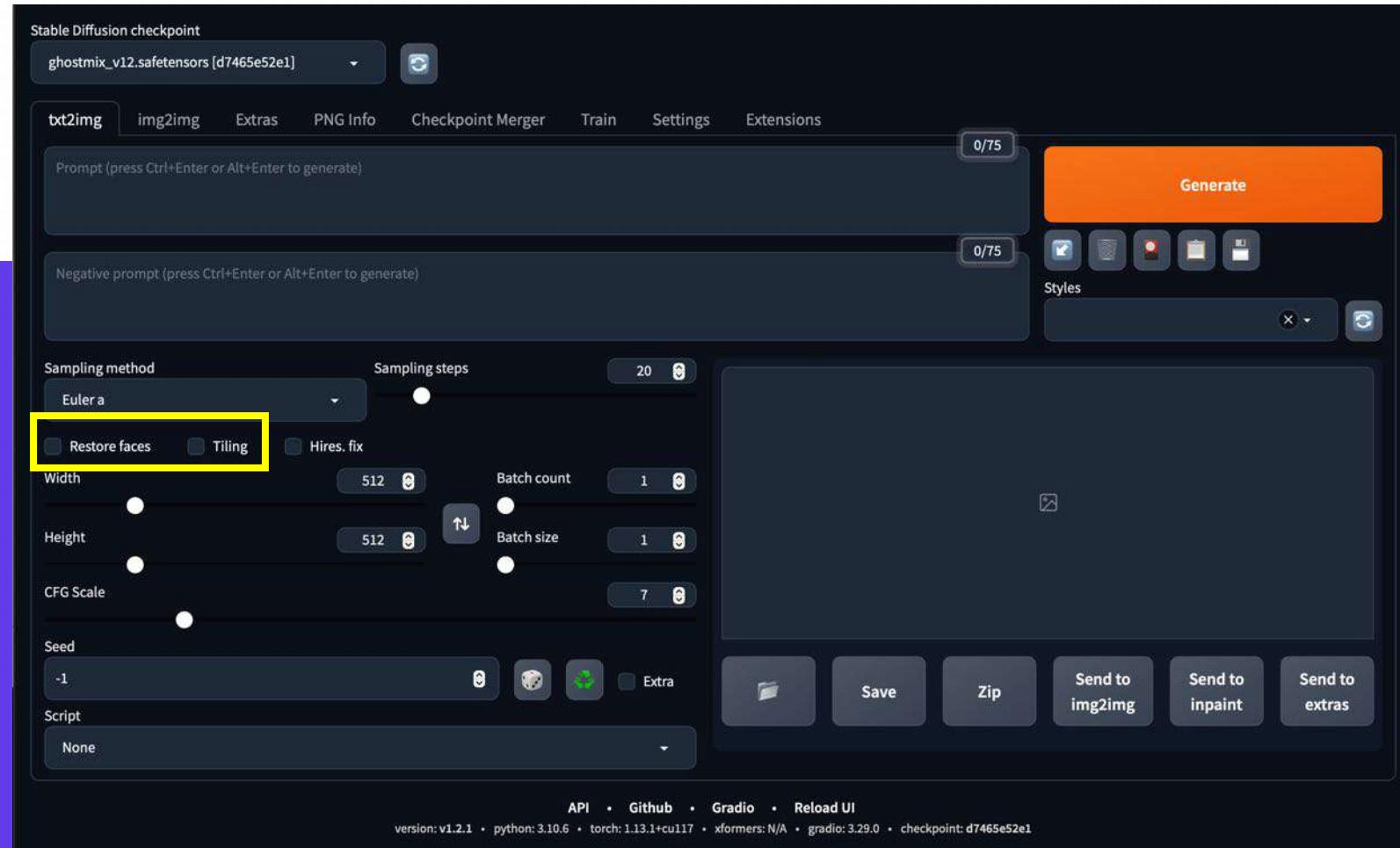
具有不同step的相同prompt

1girl, reading book,

在不同情况下应使用的step一般指南：

- 如果正在测试新的提示，并希望快速得到结果以便调整你的输入，使用10–15步
- 当找到喜欢的prompt后，将step增加到25
- 如果正在创建一个人脸、有毛皮的动物或任何具有详细纹理的主题，而你觉得生成的图像缺少一些这些细节，试着将其提高到40

3.4 Restore faces & Tiling



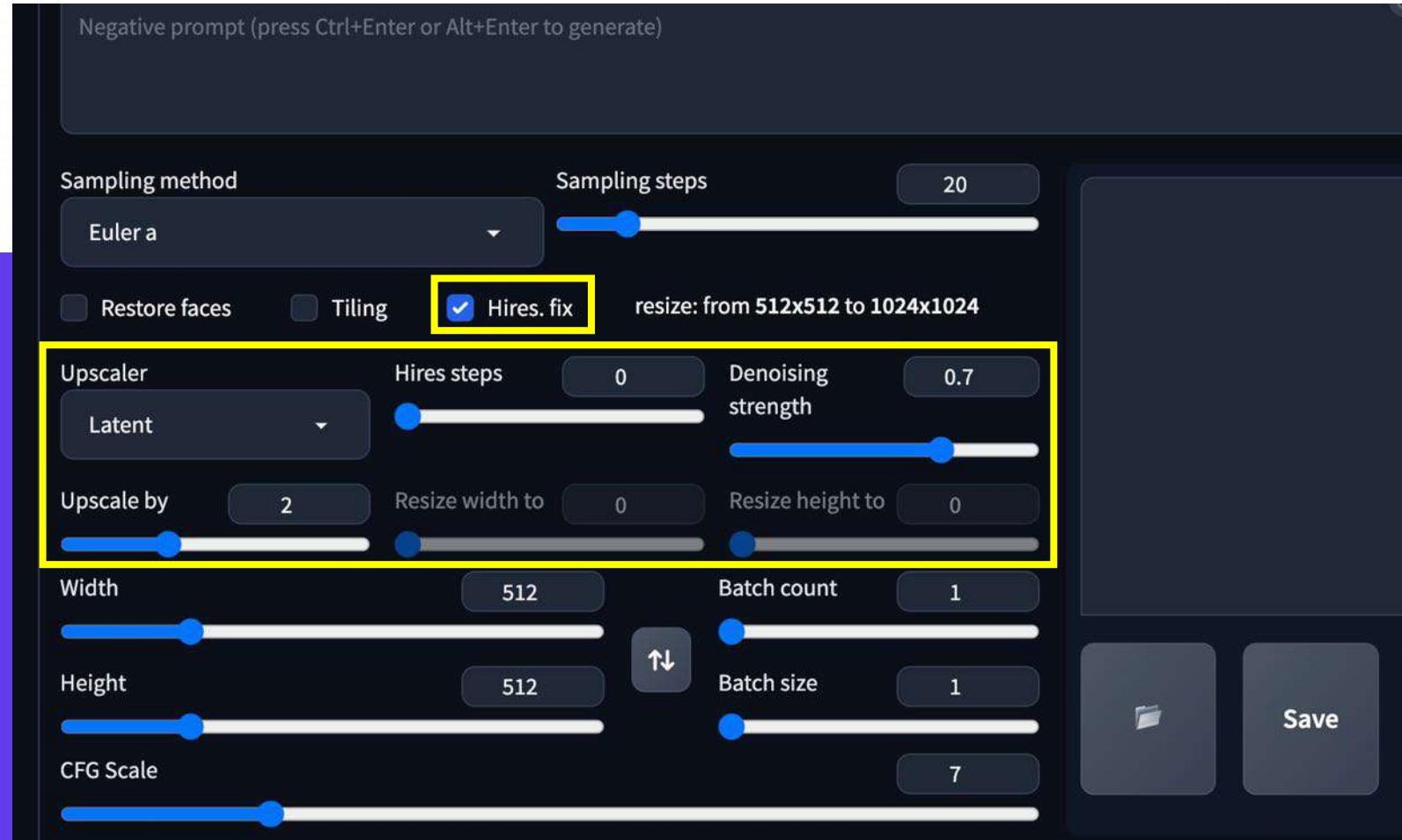
Restore faces(面部修复)

为恢复脸部缺陷而训练的附加模型。

Tiling(平铺图):

将生成的图片拼接在一起，类似于 Midjourney 中的 `-tile` 的指令

3.5 Hires.fix

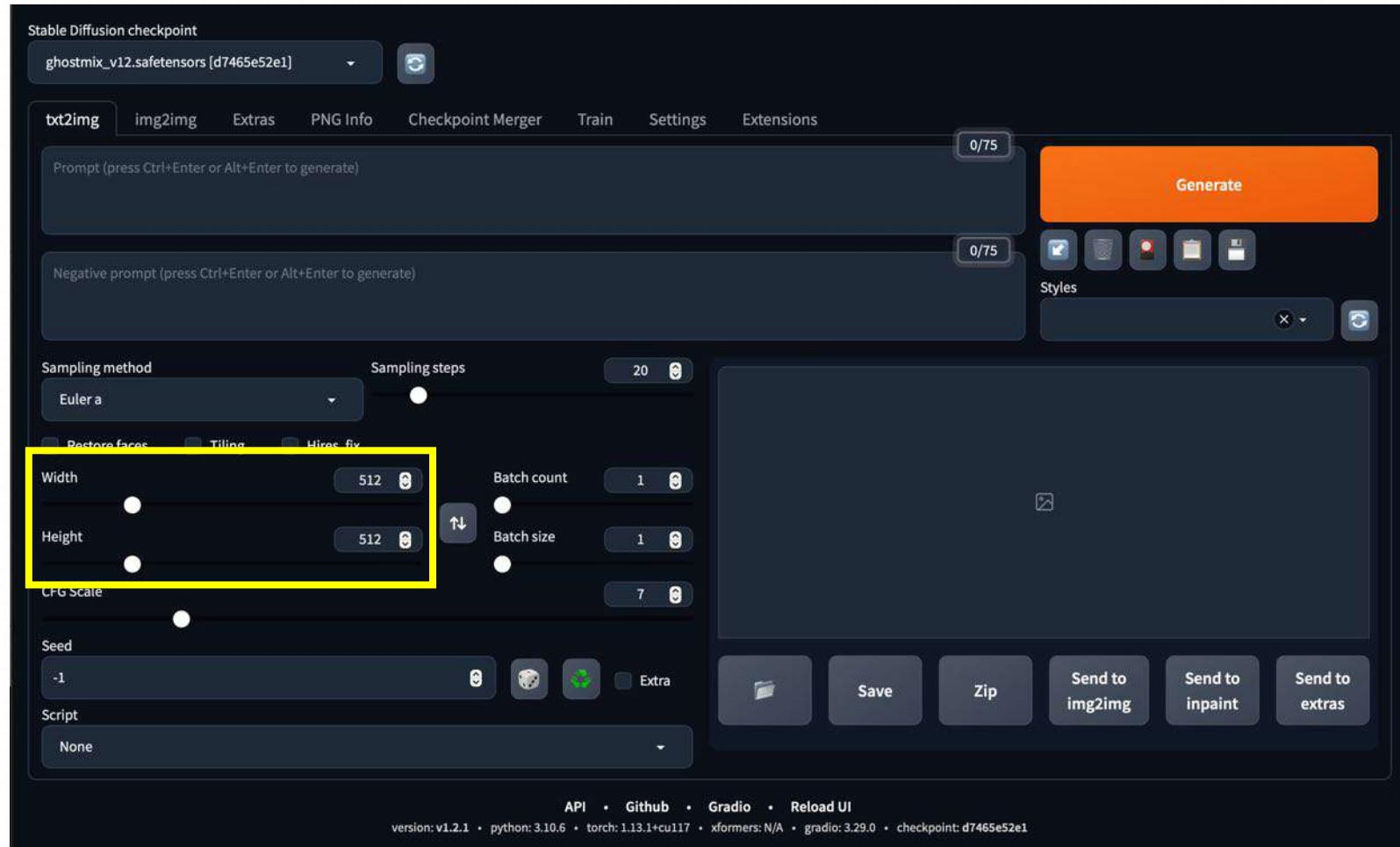


Hires.fix(高分辨率修复)

通俗来说，就是以重新绘制的方式对图像进行放大，并且在放大的同时补充一些细节。

打开后，这里会出现一些子参数。
放大算法：用默认值即可。
高分迭代步数：一般选在 10~20 即可。
一般超过20步之后会发生一些变形。

3.6 Aspect Ratios

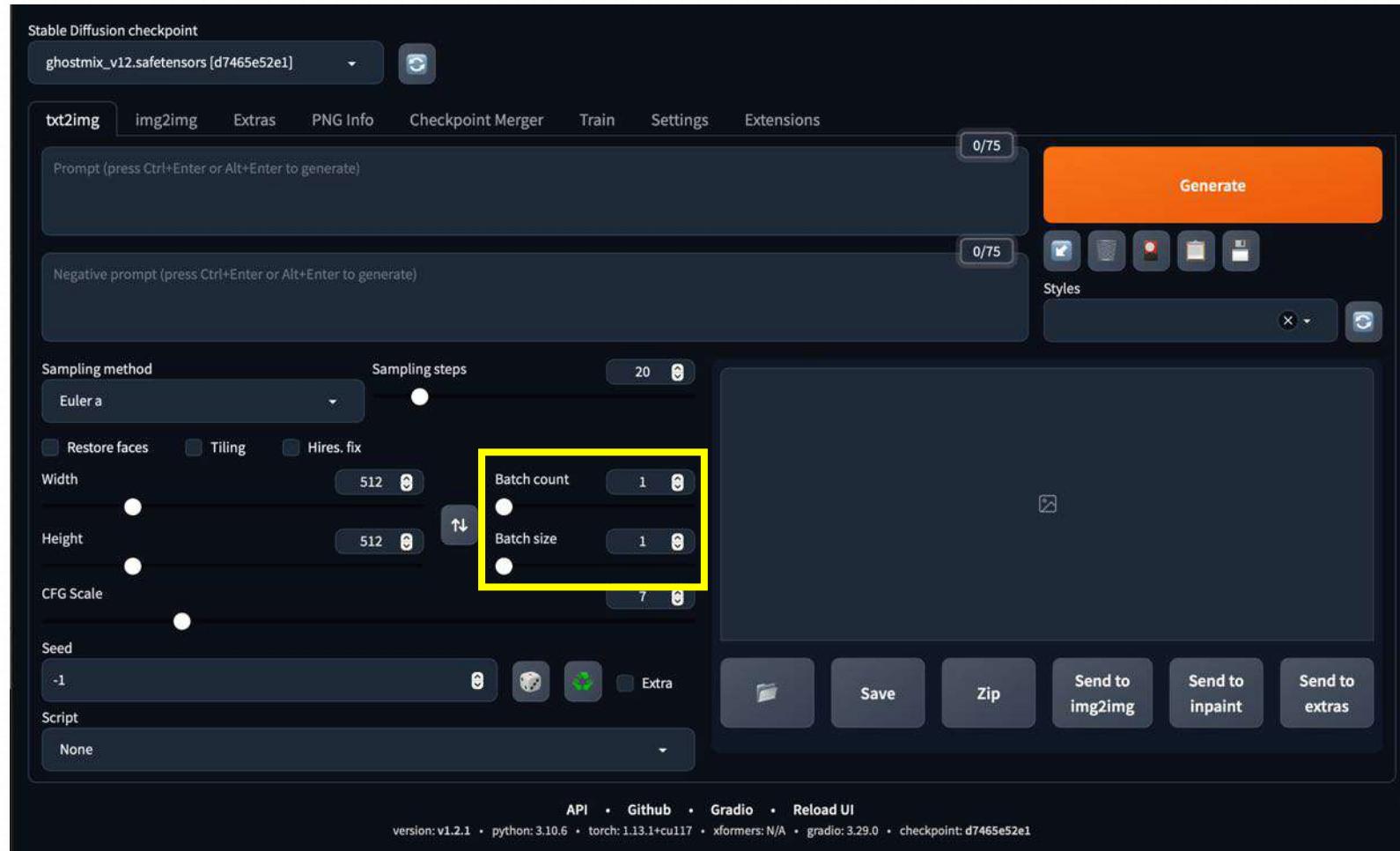


默认尺寸是 $512 * 512$ ，生成的图像会有些模型，画面会有一些涂抹感。想同 Prompt 使用不同的尺寸生成会回不同的质感。

我们一般可以使用 $800*800$ 的尺寸。较高的尺寸对显卡的要求也较高，会影响出图的速度。此外，较高的尺寸生成图像会出现多人多头像的照片。

➤ 原因是 AI 在进行模型训练时使用的图片分辨率一般比较小，如果分辨率设置太大，它就会认为你是多张图片拼接而成的，那出现多的人就不奇怪了。

3.7 Batch Count & Size



Batch count 代表生成批次数，
Batch size 代表每一批次生成的
图片个数。如果将 Batch count
和 Batch size 的值分别设置为 2 和
4 时，生成的图片总数就为：
 $2 * 4 = 8$ 张。

- 生成批次的数量可以高一些，
每批数量的数值要小一些，不然会爆显存~

3.8 CFG Guidance Scale (提示词的相关性)

CFG 比例可以看做是提示词与创造力的比例 (creative vs. prompt – 图像与提示符的一致程度) ,该值越低时，产生越有创意的结果；反之如果该参数越高会使得SD尽量去生成与Prompt 一致的结果。



例子中默认使用的 CFG 是7，这为创造性和生成你想要的内容之间提供了最佳的平衡。

- CFG < 5时，图像可能开始看起来更像是 AI 的幻觉
- CFG > 16时，图像可能会开始产生带有难看的人工痕迹的图像

3.8 CFG Guidance Scale (提示词的相关性)

Prompt: 1girl, reading book,

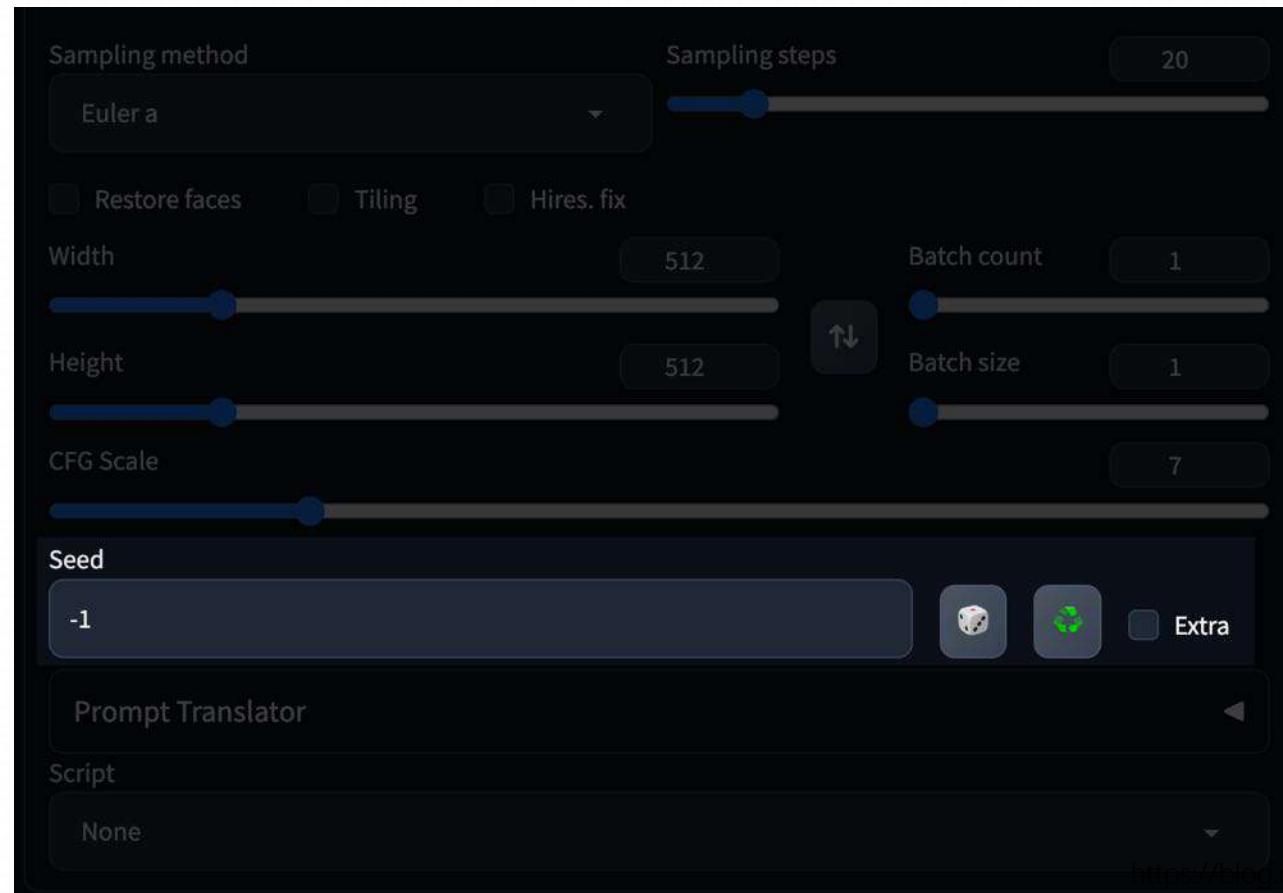


CFG 比例可以分为不同的范围:

- CFG 2 – 6: 富有创意, 但可能过于扭曲且不符合prompt。对于短prompt可能会很有趣且有用
- CFG 7 – 10: 推荐用于大多数提示。创造性和引导生成之间的良好平衡
- CFG 10 – 15: prompt很详细且清晰的描述了希望图像看起来像什么时
- CFG 16 – 20: 通常不推荐, 除非prompt非常详细。可能会影响连贯性和质量
- CFG > 20: 几乎从不适用

3.9 Seed

Seed是一个决定我们之前谈到的初始随机噪声的数字，由于随机噪声决定了最终的图像，所以它是在Stable Diffusion系统上每次运行完全相同的提示都会得到不同图像的原因。也是如果多次使用相同的seed和相同的prompt运行，会得到相同生成图像的原因



- 当输入-1 或点击旁边的骰子按钮时，生成的图像是完全随机的，没有任何规律可言
- 当输入其他随机数值时，就相当于锁定了随机种子对画面的影响，这样每次生成的图像只会有微小的变化
- 因此，使用随机种子可以控制生成图像的变化程度，从而更好地探索模型的性能和参数的影响

3.9 Seed

随机数种子

随机数种子就像在做一道菜时加入的特定调料，它可以影响整道菜的味道。在这个图像生成方法中，随机数种子就是一个特定的数值或代码，可以影响最终生成的图像的过程，就像特定的调料会影响整道菜的味道一样。不同的随机数种子会生成不同的图像，就像加入不同的调料会让同一道菜变成不同的味道。每个图都会有它对应的随机数种子，如果想还原这张图，或者绘制一张相似的图，必须保证这个值是相同的。**随机数种子不变的情况下，即使模型发生了改变，最终生成图的大概结构和配色也会有一定相似性。**就像是无论是以土豆还是以茄子作为食材，只要加入的都是甜辣酱，最终的味道是差不多的。我们可以在输入框中输入特定的值，以保证随机数种子固定，否则可以点一下旁边的骰子 icon，变成 -1 (-1 就指的是随机值) 使得模型采用随机的值来生成图像。

变异随机种子

在理解了随机种子之后，我们再去理解变异随机种子就更好理解了。变异随机种子相当于又加入了第二种特定调料。而后面的变异强度就指的是两种调料的占比，数值越大，越接近第二种。**数值为 0 时，就指的是完全不用变异随机种子。一般是用于确定了图后，对图生成其他相似画面的图像。**例如，在做 ip 时，通过加入变异随机种子，可以实现画面内容不变的情况下，生成带有些许差异的画，从中选择更优的。

3.9 Seed

portrait of a beautiful woman with blonde hair and blue eyes



portrait of a beautiful woman with blonde hair and blue eyes, smiling



portrait of a beautiful woman with blonde hair and blue eyes, crying



portrait of a beautiful woman with blonde hair and blue eyes, eyes closed



由于相同的seed和prompt组合每次都会给出相似的图像，我们可以通过多种方式利用seed参数：

控制角色的特定特征：

在这个例子中，我们改变了情绪，但这也适用于其他身体特征，如头发颜色或肤色，但变化越小，它就越有可能起作用

3.9 Seed

由于相同的 seed 和 prompt 组合每次都会给出相似的图像，我们可以通过多种方式利用seed参数：

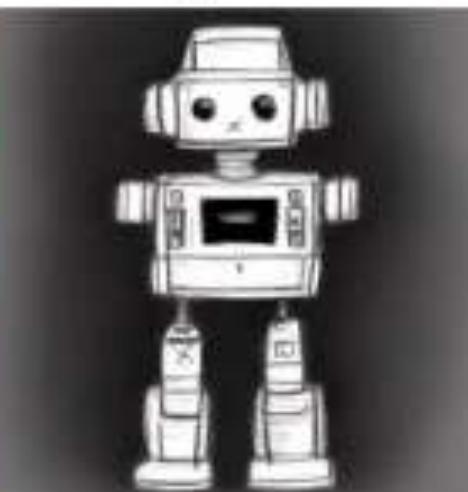
测试特定词的效果：

如果想知道提示中特定词的变化，你可以使用相同的种子和修改后的prompt来测试它，通过改变单个词或短语来测试prompt

A cute robot, standing in a dark room, **synthwave**



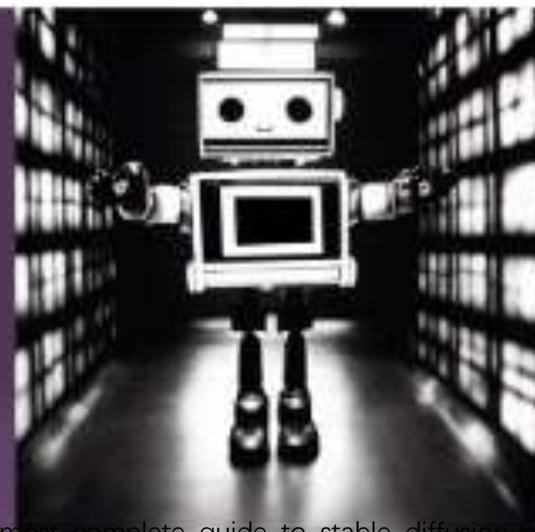
A cute robot, standing in a dark room, **pencil sketch**



A cute robot, standing in a dark room, **low poly**



A cute robot, standing in a dark room, **film noir**



<https://blog.openart.ai/2023/02/13/the-most-complete-guide-to-stable-diffusion-parameters/>

3.9 Seed

由于相同的seed和prompt组合每次都会给出相似的图像，我们可以通过多种方式利用seed参数：

改变风格：

如果您喜欢图像的构图，但想知道它在不同风格下的外观。这可用于肖像、风景或您创建的任何场景

By Van Gogh



By Frida Kahlo



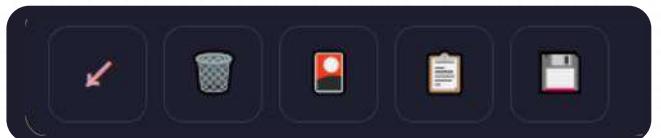
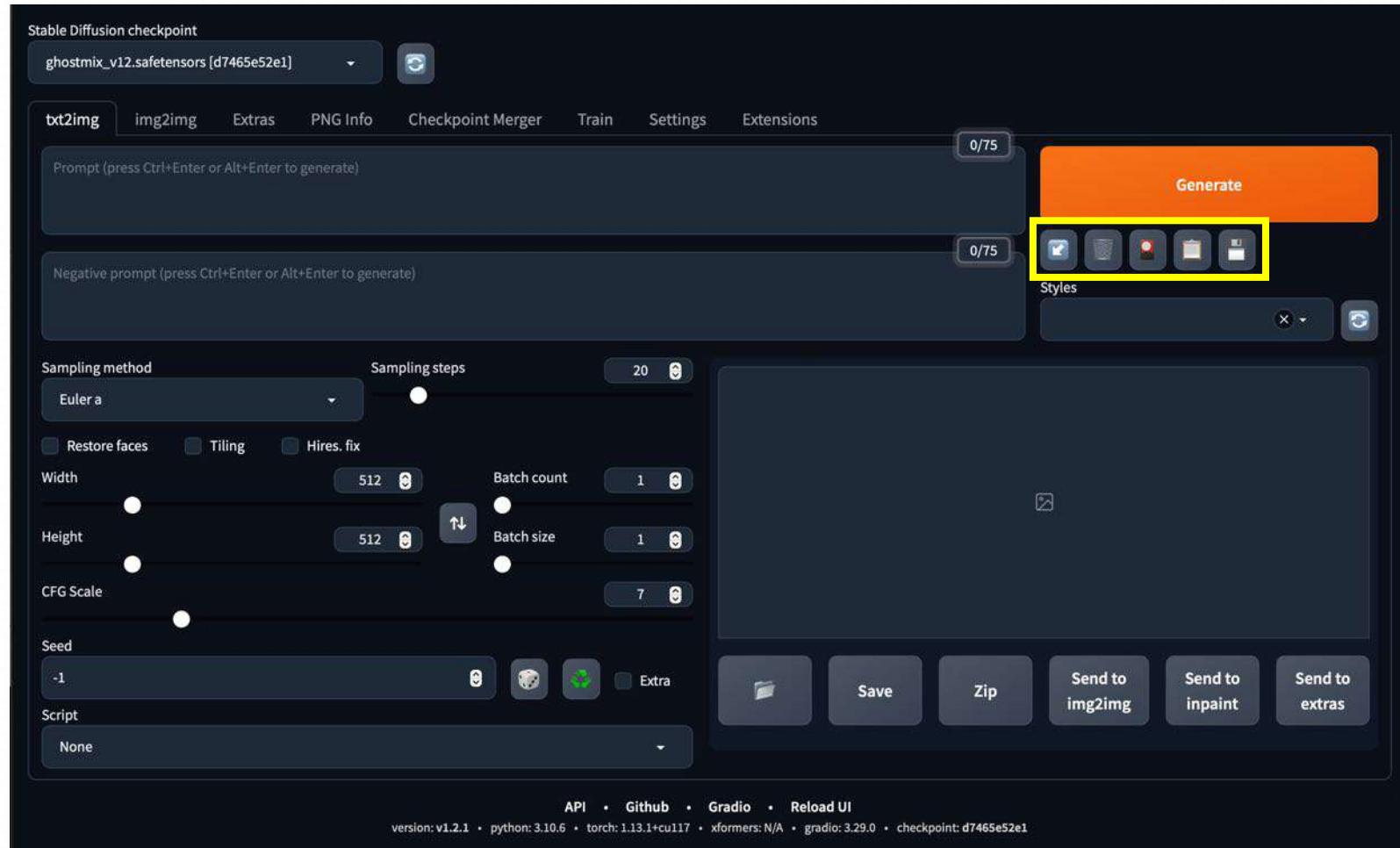
Pixar render



By Greg Rutkowski



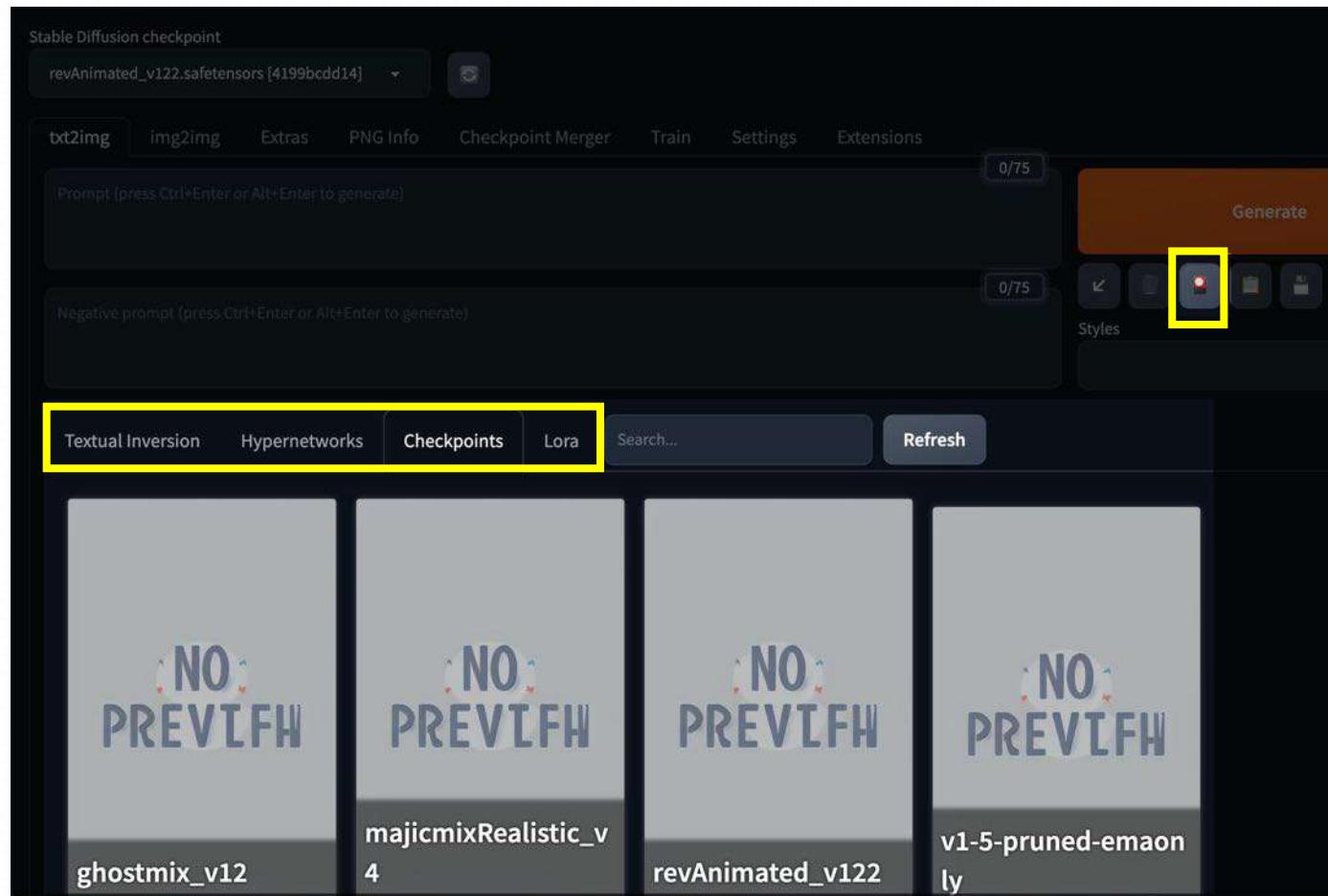
3.10 Other buttons



- 第一按钮的意思是读取上一张图的所有参数信息（包括关键词），比如你画一张图之后，关掉了软件，再次启动点击这个就会把参数复制进来。
- 第二个按钮是删除，清空关键词；这里可以和第一个按钮结合使用
- 第三个按钮是模型选择管理，安装的模型可以在这里进行调整
- 第四和第五个按钮式提示词模版功能：可以先点击第五个按钮，存储并为现有的 Prompt 命名，然后在模版风格这里选中刚刚创建的模版，选中之后按一下第四个按钮，创建的关键词就被填入到文本框中了

3.10 Other buttons

模型管理



Textual Inversion 文本倒置（又称Embedding）

一种从少量示例图像中捕获新颖概念的技术，其方式可以在以后用于控制文本到图像的管道。

该方法只需要3–5张图像，通过定义新的关键词，就能生成和训练图像相似的风格。

Hypernetwork 超网络模型

是一种微调技术，最初由Novel AI开发。它允许动态生成神经网络的参数（权重）。hypernetwork的应用领域较窄，主要是训练画风，训练难度很大，未来很有可能被后出现的lora所替代，可以将hypernetwork理解为低配版的lora。

Checkpoints 检查点文件

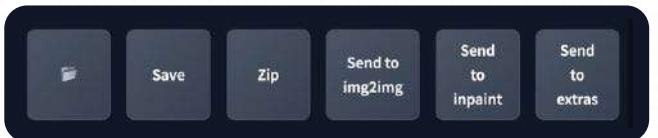
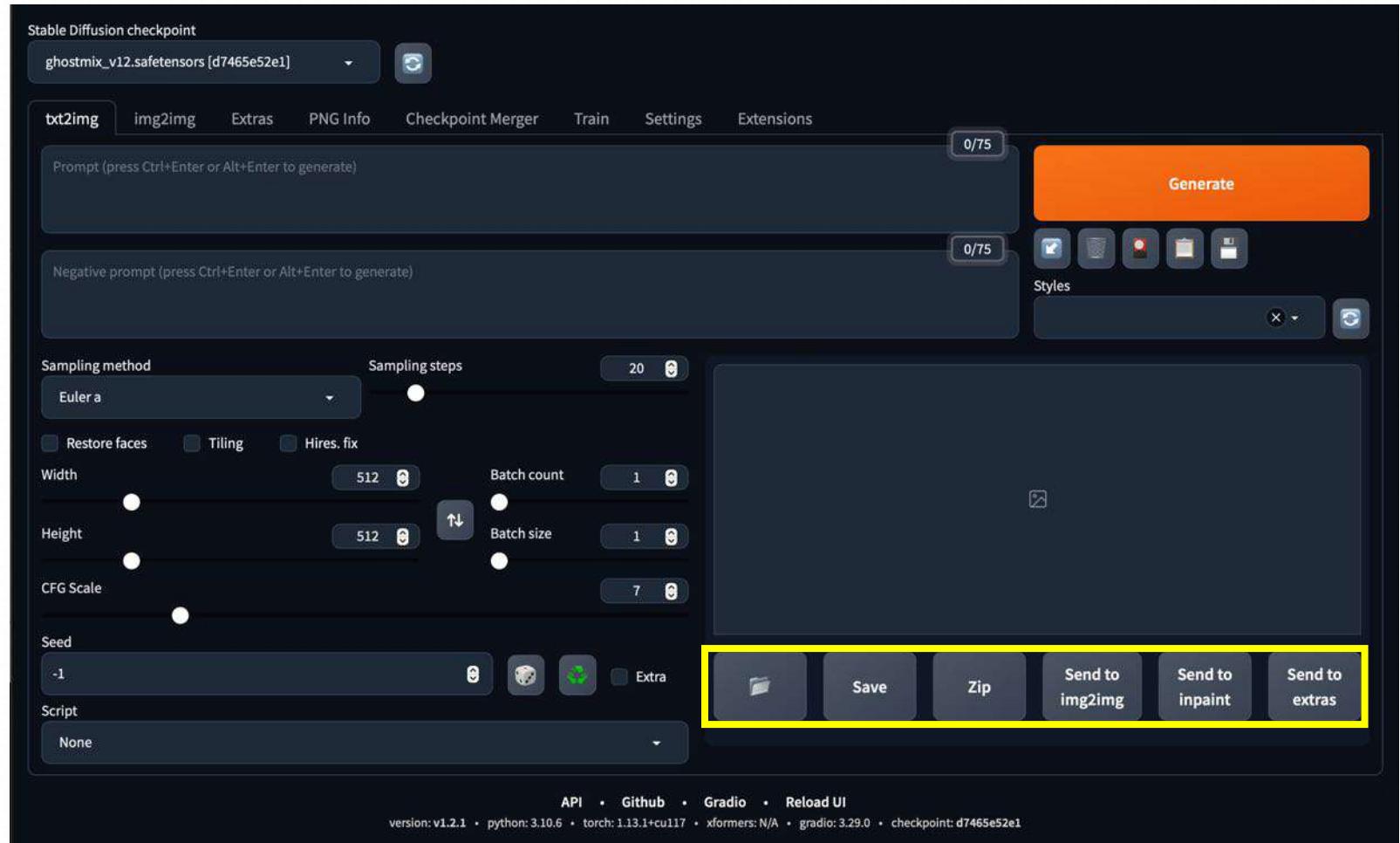
一个二进制文件，其中包含所有权重、偏差、梯度和所有其他保存的变量的值。记录了模型使用的所有参数 tf.Variable) 的确切值。该文件的扩展名为.ckpt

意味着里面存放着模型的权重参数，checkpoint 文件本质是一个变量名-变量值的词典。

Lora

全称是Low-Rank Adaptation，即大型语言模型的低阶自适应。仅需要少量的数据就可以进行训练的一种模型。在生成图片时，LoRA模型会与大模型结合使用，从而实现对输出图片结果的调整。

3.10 Other buttons



第一个按钮: 打开存储输出图片的文件夹，
你生成的图片都在文件夹内

Save: 保存输出图片的同时，将生成参数以
CSV格式的参数保存

Zip: 将图片以压缩包的形式保存，可以整体
打包下载压缩包，也可以单张下载压缩包

Send to img2img: 将当前图片转入到图
生图模式中，转入时会带着tag组合一起

Send to inpaint: 将当前图片转入到局部
重绘模式中，同上，转入时会带着tag组
合一起

Send to extras: 将当前图片转入到更多功
能中，包括放大图像、XXXXX等功能

4. img2img (图生图)

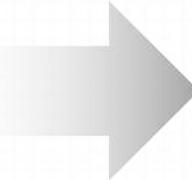
1. 图生图基本原理

2. 图生图基本流程

3. 随机种子作用解析

4. 图生图拓展应用

4.1 图生图基本原理



4.1 图生图基本原理

在 txt2img 的过程里，我们可以借助一些提示词，让AI知道我们想让它画什么。但AI绘画是有随机性的，因此，它画出来的东西也不一定是符合我们的需求。当你觉得 Prompt 不足以表达你的想法，或者希望以一个更为简单清晰的方式传递一些要求的时候，那就再上传一张图片给它，此时，图片的作用和文字的作用是相当的，就是都作为一种信息输送到 AI 那里，让它拿来生成一张新的图片。

简而言之， img2img 是一种生成图像的方法，它可以通过对已有的图像进行修改或变形，生成新的图像。在 Stable Diffusion 中，我们可以将垫图的图像看作是一张“初始图像”，通过多次迭代，对其进行修改和变形，逐步生成与垫图风格类似想要的图片。

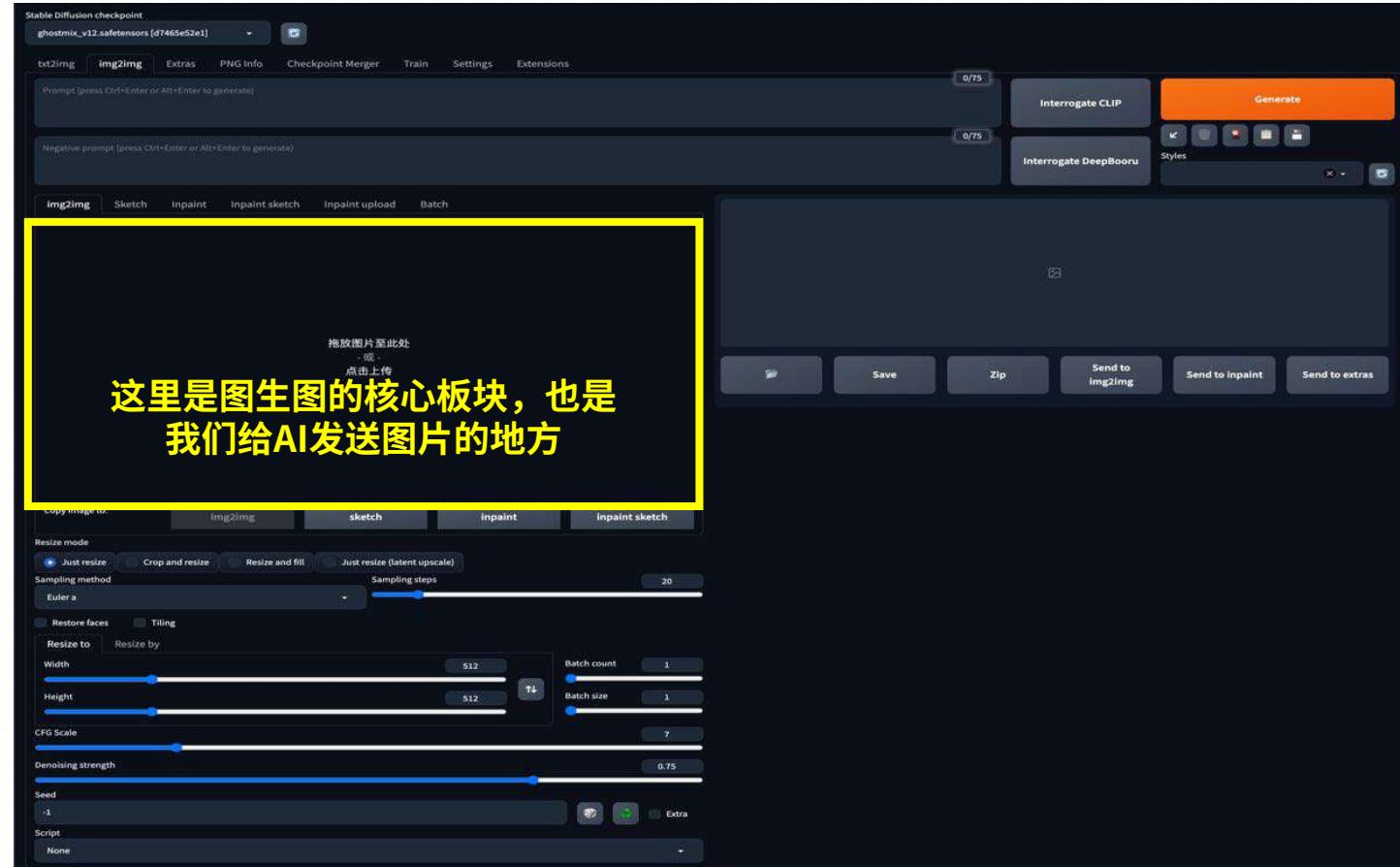
如果只是简单的给它一些文本信息，AI 会根据自己的“经验”和“阅历”去绘画，100次会生成100个不一样的画面。但这个时候你给它一张图片进行参考，那他会从图片上获得更多的信息，原本图片上记录的像素信息，会在这个“加噪”和“去噪”的过程里被作为一种特征反映到最终的成品上，会让最后的成品跟一开始图片足够“像”，就给了我们一种“重新绘制”的表象。

4.2 图生图基本流程



4.2 图生图基本流程

A. 导入图片

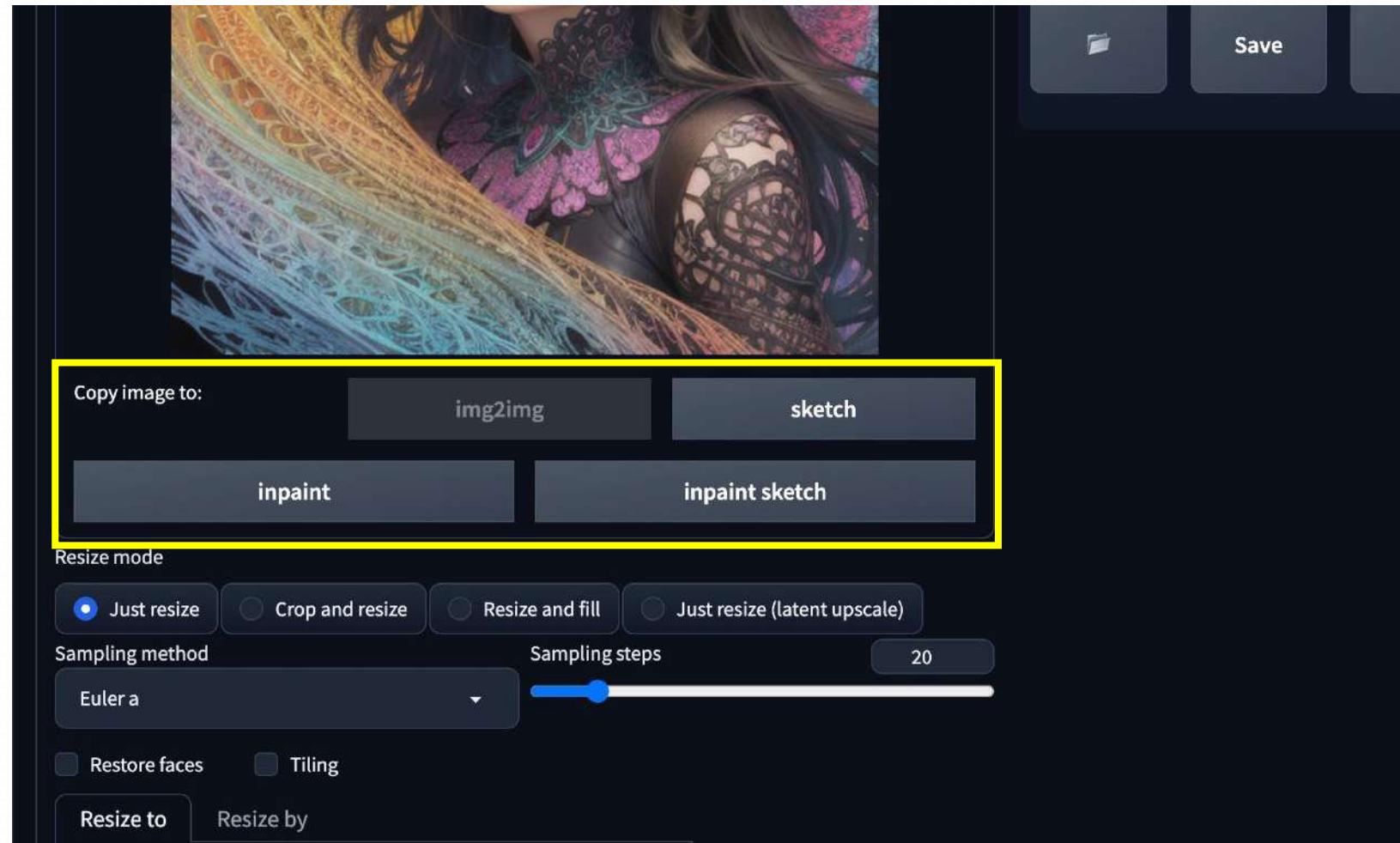


往SD中导入图片的方式有两种：

- 直接将图片拖拽到“图片上传区域”
- 点击“图片上传区域”，再打开我们本地的图片上传即可

4.2 图生图基本流程

A. 导入图片



- Sketch: 绘图
- Inpaint: 局部重绘
- Inpaint sketch: 局部重绘 (手涂蒙版)

4.2 图生图基本流程

B. 书写Prompt

图生图也需要Prompt，而且Prompt与文生图的同样重要，我们书写的Prompt需要具体且准确

类别	说明
提示词排序	越前面的词汇越受 AI 重视，重要事物的提示词放前面
增强/减弱	(提示词:权重数值)，默认 1，大于 1 加强，低于 1 减弱。如 (doctor:1.3)
混合	提示词 提示词，实现多个要素混合，如 [red blue] hair 红蓝色头发混合
+ 和 AND	用于连接短提示词，AND 两端要加空格
分步渲染	[提示词 A:提示词 B:数值]，先按提示词 A 生成，在设定的数值后朝提示词 B 变化。如[dog:cat:30] 前 30 步画狗后面的画猫，[dog:cat:0.9] 前面 90%画狗后面 10%画猫
正向提示词	masterpiece, best quality 等画质词，用于提升画面质量
反向提示词	nsfw, bad hands, missing fingers.....，用于不想在画面中出现的内容
Emoji	支持 emoji，如 😍 形容表情，👉 修饰手

4.2 图生图基本流程

B. 书写Prompt

如果生成的图片和我们想要的图片差距非常大，那就是AI只读取了我们的图片，却没有了解到画面里具体的内容是什么，那我们就需要非常明确的为它提供内容型 Prompt。

比如如果是人物的话，我们需要加入更多针对性描述外貌和服饰细节的内容型提示词
例如短发、蓝眼睛、络腮胡、穿格子衬衫、带着毛线帽

人物及主体特征

服饰穿搭 white dress

发型发色 blonde hair, long hair

五官特点 small eyes, big mouth

面部表情 smiling

肢体动作 stretching arms

场景特征

室内、室外 indoor / outdoor

大场景 forest, city, street

小细节 tree, bush, white flower

环境光照

白天黑夜 day / night

特定时段 morning, sunset

光环境 sunlight, bright, dark

天空 blue sky, starry sky

补充：画幅视角

距离 close-up, distant

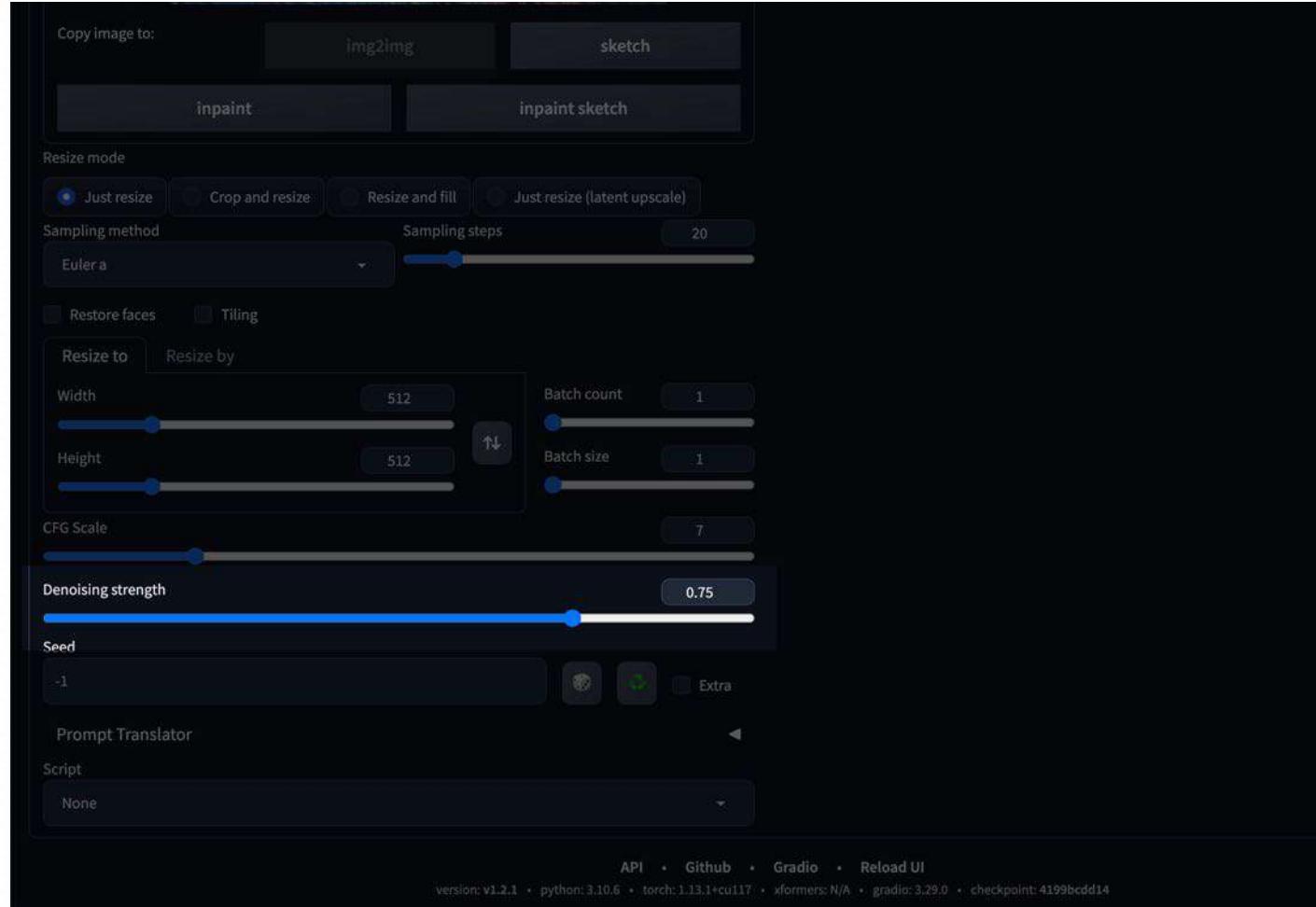
人物比例 full body, upper body

观察视角 from above, view of back

镜头类型 wide angle, Sony A7

4.2 图生图基本流程

C. 参数调整



图生图与文生图参数几乎一致，主要新增加的是：**Denoising strength**。指的是每次迭代中图像被重新绘制的程度或幅度，也就是新生成的图像与上一次迭代生成的图像之间的差异程度。

它的取值范围在 0 到 1 之间。当值等于 0 时，新生成的图像与上一次迭代生成的图像完全一致，即没有重新绘制；当值等于 1 时，新生成的图像与上一次迭代生成的图像完全不同，即完全重新绘制。

一般大于0.7出来的都是新效果，小于0.3基本就是在原图的基础上进行小的修改调整。

4.2 图生图基本流程

C. 参数调整

Denoising strength



原始图像



0 . 3



0 . 4



0 . 5



0 . 6



0 . 7



0 . 8

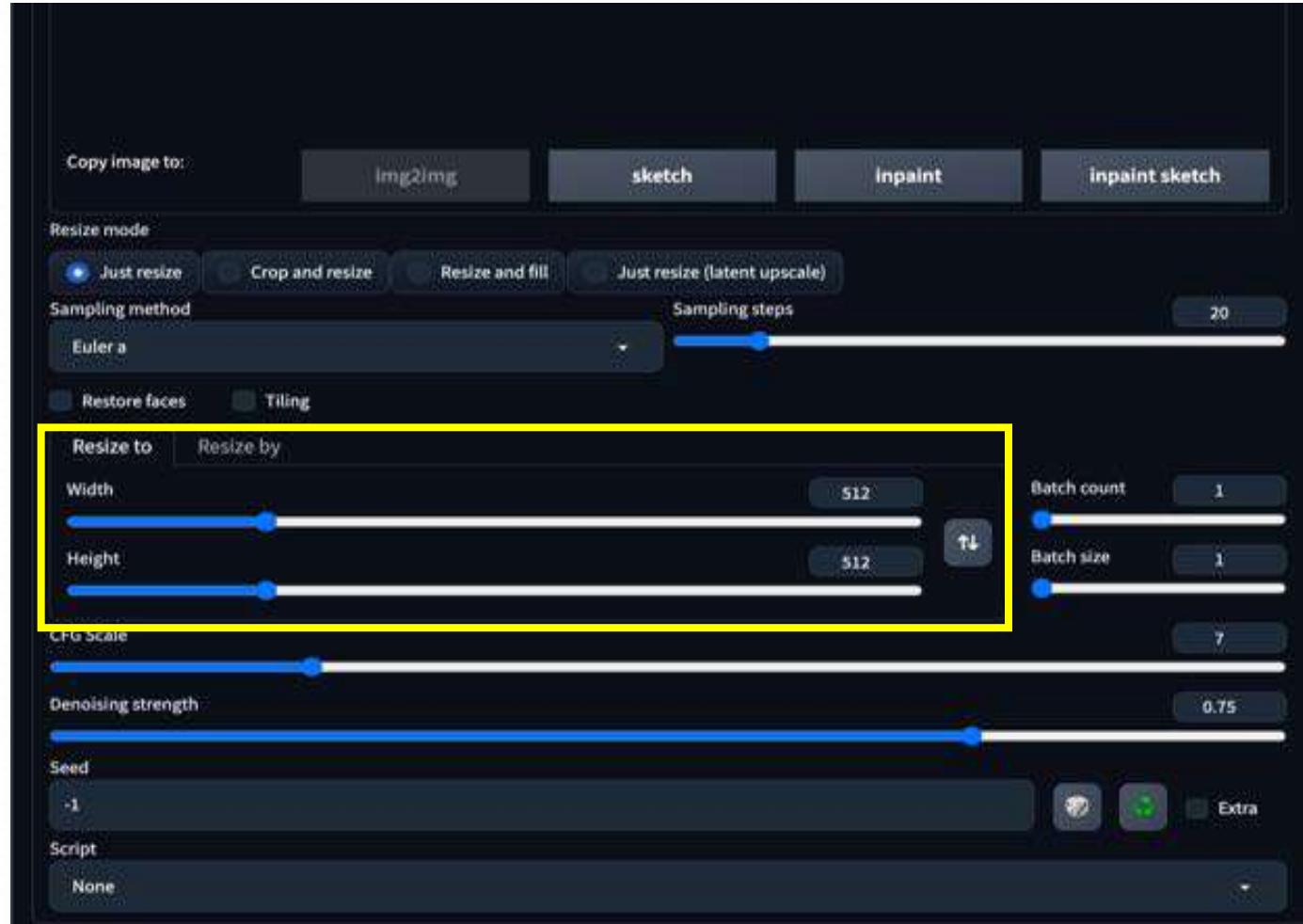


0 . 9



1

4.3 分辨率设置



推荐使用跟上传的这张图片的原始尺寸一模一样的分辨率，但如果我们上传的图片本身分辨率很大，那这个时候我们就按比例往下折算到我们自己设备能画出来的那个安全的区间内。

- 比如我们有一张图片的分辨率是2000*2000，它的长宽比是1:1，那么我们在Stable Diffusion中设置时，设置成1000*1000即可。
- 假如我们有一张图片的长宽比本身是1:1，但是我们想要生成16:9尺寸的图片，那么我们需要先把这张图片在其他软件中裁剪成16:9的尺寸，然后再上传到SD中。如果设置的宽高比例与原始的不一样，那可能会导致画面的变形与拉伸。

4.3 分辨率设置



原始图像

Resize mode

Just resize Crop and resize Resize and fill Just resize (latent upscale)

The interface shows three versions of the same image based on different resize modes. The first version ('Just resize') is stretched and looks distorted. The second version ('Crop and resize') is cropped and looks unnatural. The third version ('Resize and fill') is a smooth, high-quality resize that maintains the original's quality.

Just resize	Crop and resize	Resize and fill
拉伸	填充	裁剪

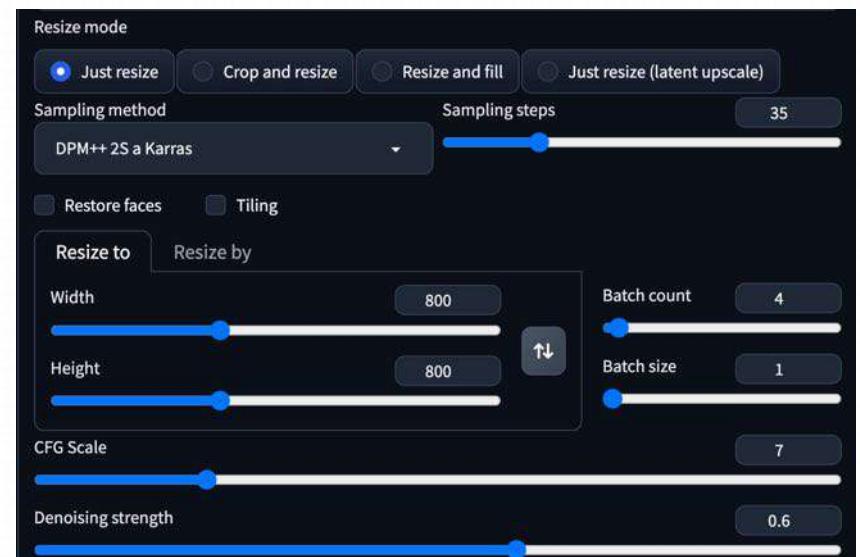
4.4 图生图拓展应用

A. 把真实的人像照片变成二次元的画风



((masterpiece)),((best quality)), ((ultra-detailed)), (best illustration), 1girl, solo, blush, smug, smile, purple eyes, choker, gradient eyes, no pupils, multicolored_hair, short hair, black hair, sangonomiya kokomi, ((Hanfu)), outdoors, sakura trees, sakura, facing towards viewer, front view

EasyNegative,extra fingers, fewer fingers, extreme fingers,wrong hand,wrong tail, missing male, extra legs, extra arms, missing legs, missing arms, weird legs, weird arms, watermark, logo, long hand, (poorly drawn hands:1.331), (bad anatomy:1.21), (bad proportions:1.331), (fused fingers:1.61051), (too many fingers:1.61051), extra digit, fewer digits,(mutated hands and fingers:1.5), fused fingers, one hand with more than 5 fingers, one hand with less than 5 fingers, one hand with morethan 5 digit, one hand with less than 5 digit, extra digit, fewer digits, fused digit, missing digit,text,watermark,



4.4 图生图拓展应用

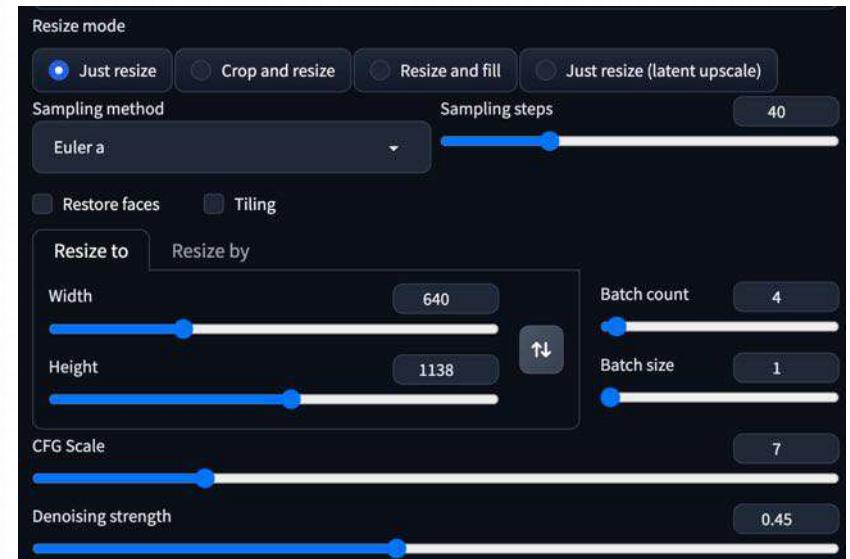
B. 把静物、风景“拟人化”

导入不是人物的图片，用描述人物的提示词进行定义



best quality, masterpiece, ultra high res, photorealistic, 1girl, offshoulder, smile, 1girl, waist up, smile, beautiful lighting, best quality, masterpiece, ((Hanfu)), ultra highres, photorealistic (zentangle abstract background:1.4),(Whole body)

EasyNegative,extra fingers, fewer fingers, extreme fingers,wrong hand,wrong tail, missing male, extra legs, extra arms, missing legs, missing arms, weird legs, weird arms, watermark, logo, long hand, (poorly drawn hands:1.331), (bad anatomy:1.21), (bad proportions:1.331), (fused fingers:1.61051), (too many fingers:1.61051), extra digit, fewer digits,(mutated hands and fingers:1.5), fused fingers, one hand with more than 5 fingers, one hand with less than 5 fingers, one hand with morethan 5 digit, one hand with less than 5 digit, extra digit, fewer digits, fused digit, missing digit,text,watermark,



4.4 图生图拓展应用

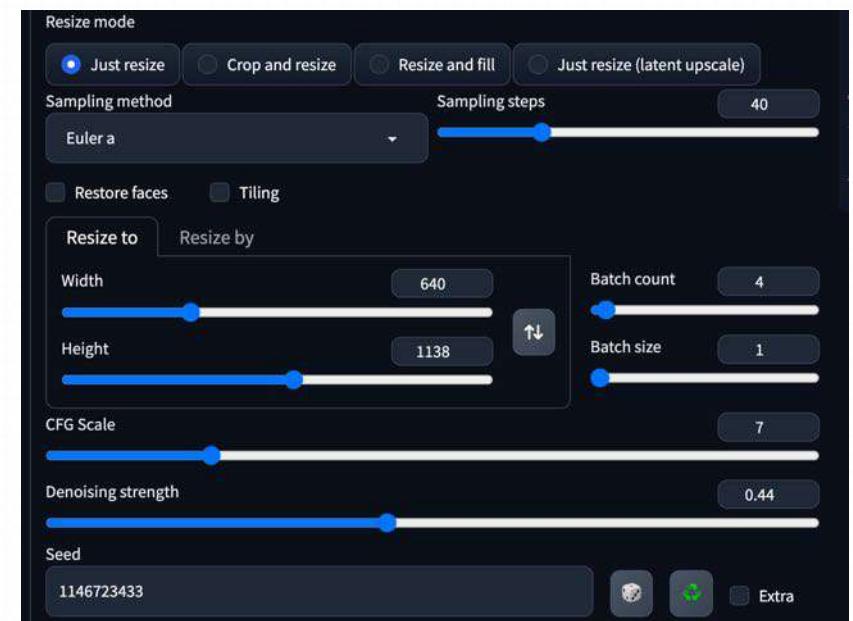
C. 把二次元角色“真人化”

导入动漫、游戏人物的图片，用真实系的模型加上真实系的标准化提示词进行定义



best quality, masterpiece, ultra high res, photorealistic, 1girl, offshoulder, smile, 1girl, waist up, smile, beautiful lighting, best quality, masterpiece, ((Hanfu)), ultra highres, photorealistic (zentangle abstract background:1.4),(Whole body)

ng_deepnegative_v1_75t, (badhandv4), (worst quality:2), (low quality:2), (normal quality:2), lowres, bad anatomy, bad hands, ((monochrome)), ((grayscale)) watermark, (moles:2)



4.4 图生图拓展应用

D. 抽象画法

有时候，图生图都不需要一张具体的图片，即便你只是乱涂乱画，它都能为你做出一些令人惊艳的效果来。



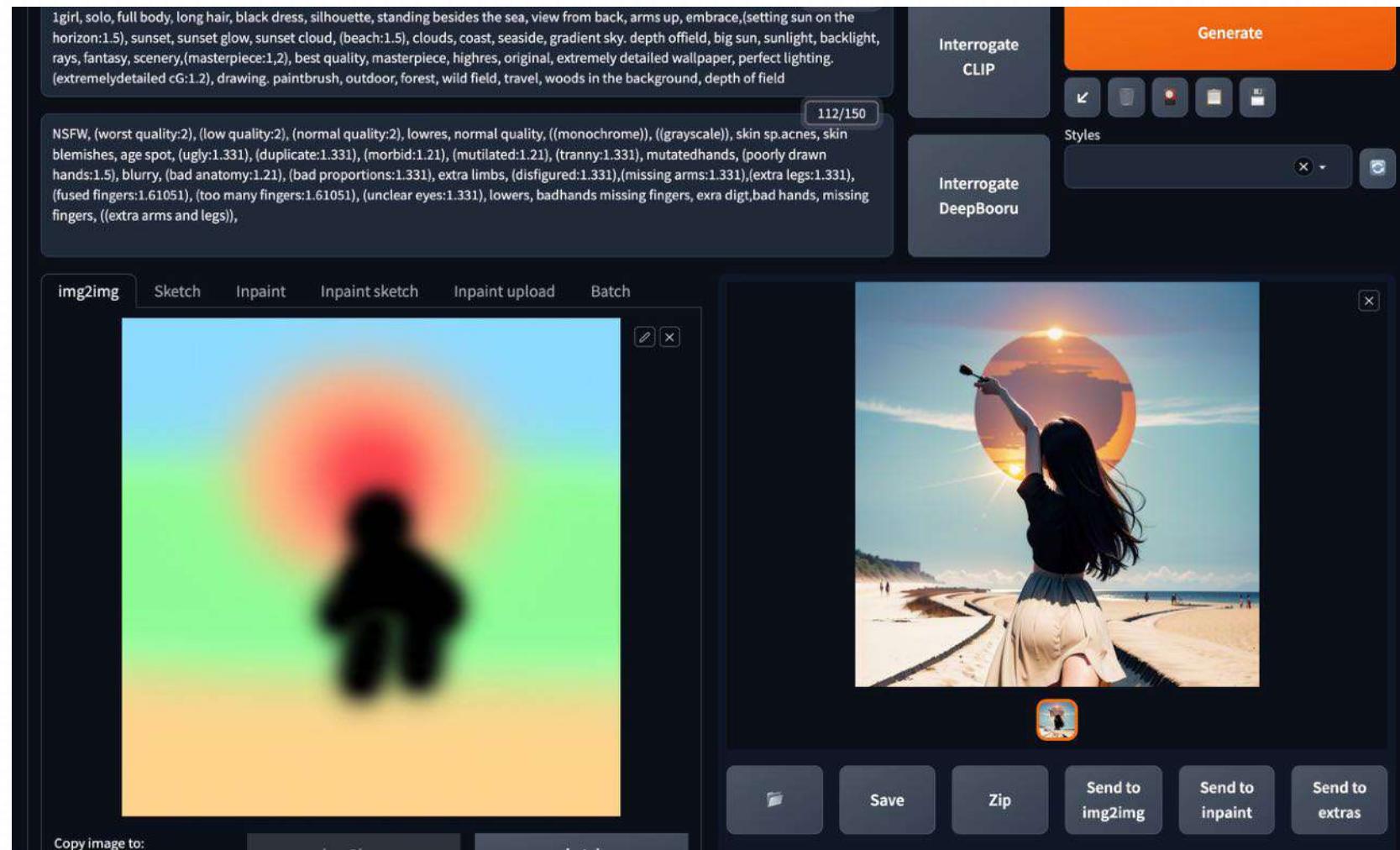
1girl, solo, full body, long hair, black dress, silhouette, standing besides the sea, view from back, arms up, embrace, (setting sun on the horizon:1.5), sunset, sunset glow, sunset cloud, (beach:1.5), clouds, coast, seaside, gradient sky, depth offield, big sun, sunlight, backlight, rays, fantasy, scenery, (masterpiece:1.2), best quality, masterpiece, highres, original, extremely detailed wallpaper, perfect lighting. (extremelydetailed cG:1.2), drawing. paintbrush, outdoor, forest, wild field, travel, woods in the background, depth of field

NSFW, (worst quality:2), (low quality:2), (normal quality:2), lowres, normal quality, ((monochrome)), ((grayscale)), skin sp.acnes, skin blemishes, age spot, (ugly:1.331), (duplicate:1.331), (morbid:1.21), (mutilated:1.21), (tranny:1.331), mutatedhands, (poorly drawn hands:1.5), blurry, (bad anatomy:1.21), (bad proportions:1.331), extra limbs, (disfigured:1.331), (missing arms:1.331), (extra legs:1.331), (fused fingers:1.61051), (too many fingers:1.61051), (unclear eyes:1.331), lowers, badhands missing fingers, exra digit,bad hands, missing fingers, ((extra arms and legs)),

4.4 图生图拓展应用

D. 抽象画法

有时候，图生图都不需要一张具体的图片，即便你只是乱涂乱画，它都能为你做出一些令人惊艳的效果来。





智能大数据可视化实验室
INTELLIGENT BIG DATA VISUALIZATION LAB