

第 1 章

走进 Stable Diffusion 的世界



Stable Diffusion 是一款在 2022 年发布的深度学习模型，它专注于将文本描述转化为详细的图像。其功能多样，不仅限于图像生成，还能应用于内补绘制、外补绘制等任务，甚至在特定提示词（英语）的引导下实现图生图的转换与翻译。据维基百科介绍，Stable Diffusion 主要聚焦于“文本到图像”的深度学习应用，即广为人知的“文生图”领域。用户只需要输入文本提示词（text prompt），该模型便能生成与之相匹配的图像，展现了强大的图像生成能力。

Stable Diffusion 的魅力在于其作为一种强大的文本到图像生成模型，为用户提供了前所未有的创意表达途径和视觉创作的诸多可能性。

1. 高度的灵活性与创意性

Stable Diffusion 允许用户通过简单的文本描述生成与之相对应的图像。这种从抽象语言到具体视觉内容的转换过程极大地激发了用户的创造力和想象力，使得每个人都可以成为自己视觉故事的讲述者。图 1-1 为 Stable Diffusion 生成的图像。



图 1-1

2. 高质量的图像输出

经过训练的 **Stable Diffusion** 模型能够生成细节丰富、风格多样的高质量图像。这些图像在细节处理、色彩搭配以及整体构图上都达到了令人惊叹的水平，为用户提供了接近甚至超越专业摄影师和画家的作品体验，如图 1-2 所示。

3. 广泛的应用场景

Stable Diffusion 的应用领域极为广泛，从艺术创作、广告设计到游戏开发、虚拟现实等，都能见到它的身影。它不仅可以为专业人士提供高效的创作工具，也能让普通用户轻松实现自己的创意想法，降低了视觉创作的门槛，如图 1-3 所示。



图 1-2



图 1-3

4. 持续的技术进步

随着人工智能技术的不断发展，**Stable Diffusion** 模型也在不断优化和升级。新的算法、更强大的计算能力以及更丰富的数据集使得 **Stable Diffusion** 能够生成更加逼真、多样化的图像，从而满足用户日益增长的需求。图 1-4 为 **Stable Diffusion** 模型生成的图像。



图 1-4

5. 促进艺术交流与融合

Stable Diffusion 打破了传统艺术创作方式的界限,使得不同文化、不同背景的人都能通过共同的语言——文字来交流和理解彼此的艺术理念。这种跨界的艺术交流不仅促进了艺术的多元化发展,也推动了文化的融合与创新。图 1-5 为 Stable Diffusion 生成的建筑图像。

总之,Stable Diffusion 以其独特的魅力吸引了全球范围内用户的关注和喜爱,成为当下热门的文本到图像生成技术之一。

本章深入剖析了 Midjourney 与 Stable Diffusion 这两款前沿软件的优势与局限,旨在引领读者全面理解它们在不同应用场景下的独特价值。本章不仅详尽地阐述了这两款软件的核心特性与差异,还贴心地附上了详尽的安装指南与注册流程教学,旨在帮助读者根据个人需求与实际情况做出明智的软件选择,并顺利踏上后续的学习与实践之旅。



图 1-5

1.1 Midjourney 和 Stable Diffusion 的优缺点对比

当前市场上存在两款备受人们推崇且广泛应用于工作的主流 AI 绘画软件,一款是 Midjourney,另一款是 Stable Diffusion。它们虽然都是基于文本提示生成 AI 图像的工具,但是在功能上各有千秋,各自具有独特的优势与局限。这两款软件为用户提供了不同的创作途径,以适应用户多样化的需求与偏好。

1.1.1 Midjourney 的优势

Midjourney 是一款基于 Disco Diffusion 平台构建的创新 AI 绘画工具,能够将文本描述转化为生动的图像。它不仅擅长创作视觉冲击力强的艺术作品,还作为一个国际化的 AI 绘画平台。特别值得一提的是,它能识别并响应中文输入,确保了广泛的用户适用性。Midjourney 生成的图像的版权直接归属于创作者,为艺术创作提供了坚实的法律保障。Midjourney 公司已全面优化模型开发、训练、调整及用户界面设计,让用户享受“开箱即用”的便捷体验,无须高端计算机配置即可轻松上手。此外,Midjourney 的图像生成速度极快,极大地提升了用户的创作效率,如图 1-6 所示。

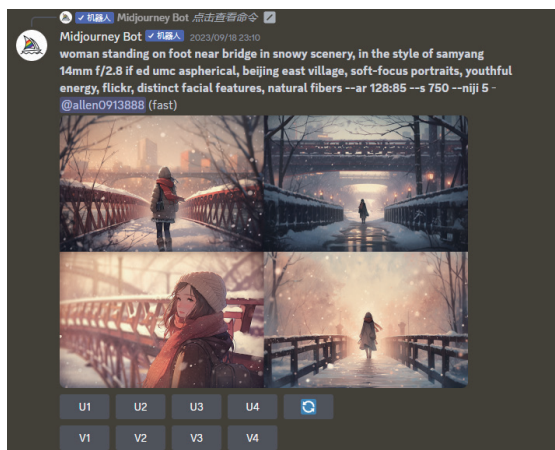


图 1-6

1.1.2 Midjourney 的劣势

Midjourney 以其高度的随机性著称，但在可控性方面相对较弱，仅提供数量有限的模型变体供用户选择。尽管用户能够调整如纵横比等参数，并选择不同的算法生成版本，但与 Stable Diffusion 相比，Midjourney 在变化和选项的丰富度上略显不足。此外，Midjourney 设置了一定的内容限制，包括政治、血腥、敏感人体部位、毒品、侮辱性词汇等，违反这些规定可能导致账号被封禁。作为一个开放的社区平台，Midjourney 上的图像一旦生成即可被他人访问，除非用户选择开通每月 60 美元（费用可能因促销、折扣等活动有所变动）的会员服务并激活隐身模式，以保护自己的作品不被公开浏览。鉴于 Midjourney 是一款成熟的商业产品，它采用付费模式，更适合那些追求快速上手且不介意支付费用的用户群体，如图 1-7 所示。

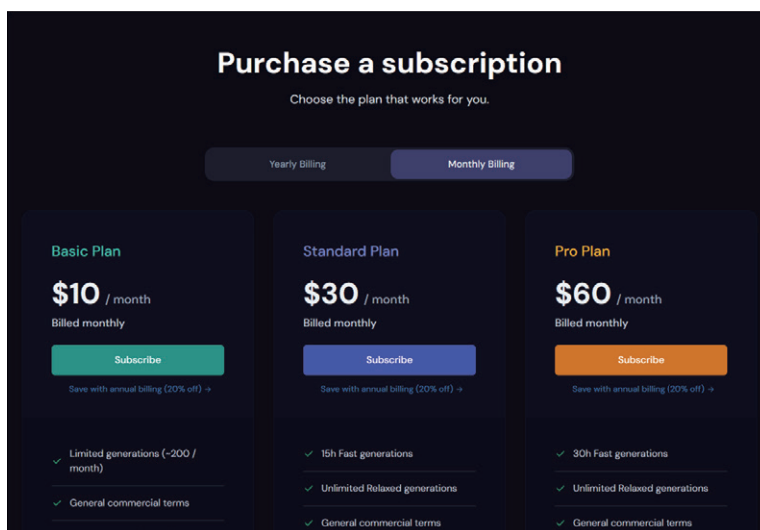


图 1-7

1.1.3 Stable Diffusion 的优势

Stable Diffusion 作为一款开源模型，鼓励全球用户共同参与其创新与发展，因此存在众多在线及离线版本供大家选择。离线版本完全免费，且用户可根据个人需求自由挑选模型，展现出极强的可扩展性。Stable Diffusion 在图像定制方面尤为出色，它允许用户精细地调整图像至每个像素级别，同时创作者对 AI 遵循提示的严格程度拥有完全控制权，包括设置种子值、挑选采样器等，以引导 AI 引擎生成预期效果。此外，Stable Diffusion 平台提供了数千种艺术模型，能够基于用户提示生成多样化的艺术风格，极大地丰富了创作空间。

Stable Diffusion 的可定制化程度较高，得益于丰富的插件生态，这些插件为用户提供了更多控制图像构图、姿势等细节的可能性。对于注重隐私保护的用户而言，Stable Diffusion 的本地部署模式尤为吸引人，因为它无须联网即可使用，且默认情况下不对外公开图像内容，除非创作者主动分享，否则作品将安全地保存在个人计算机上。

1.1.4 Stable Diffusion 的劣势

Stable Diffusion 在数据模型的应用上较为灵活，但这也意味着许多模型需要用户自行训

练（或从网络下载），且参数调整过程相对复杂，对于初学者而言，其上手难度相对较高。此外，Stable Diffusion 对运行环境配置有一定的要求，特别是显存资源，若显存不足可能导致图像渲染出错。在用户选择显卡时，显存容量成为优先考虑的因素，因为它直接影响到图像能否成功生成，而显卡的算力决定了运行的效率。为了保证软件流畅运行，推荐大家至少配备英伟达 3060 系列或更高规格的显卡，且显存达到 12GB 及以上。因此，Stable Diffusion 更适合那些对图像隐私有较高要求、计算机配置优良且具备较强学习能力的用户群体。

Midjourney 与 Stable Diffusion 各有千秋，将两者结合使用，能够充分发挥它们各自的优势，从而胜任更多元化的商业化任务，例如电商服装模特图的制作、AI 插画与动画的创作等。这种互补性使得这两款软件在创意产业中展现出强大的合作潜力。

1.2 Stable Diffusion 的部署和安装

Stable Diffusion 本质上并非传统意义上的软件，其初始形态仅为一组源代码数据。直到 GitHub 平台上的开发者 Automatic1111 将这些代码转化为可通过浏览器网页运行的程序，才生成了如今人们所看到的可视化操作界面——Stable Diffusion Web UI（简称 SD Web UI）。随着技术的不断演进，还出现了如 Comfy UI 这样的节点式操作页面，与 Web UI 相比，它赋予了用户创建稳定扩散工作流程的能力，不仅更复杂而且更灵活，进一步丰富了用户的使用体验。

1.2.1 Stable Diffusion 的安装方法

鉴于手动配置过程存在较高门槛，且安装过程中常会遇到各类问题，市场上涌现出众多集成了 Stable Diffusion 两种操作界面（包括 Web UI）的一键部署包，这些部署包均基于 GitHub 上的相关开源项目复制而来。下面将以广受欢迎 Web UI 界面为例，详细说明如何从 GitHub 部署 Stable Diffusion。

（1）访问 GitHub 并找到 Stable Diffusion 项目的页面，如图 1-8 所示。

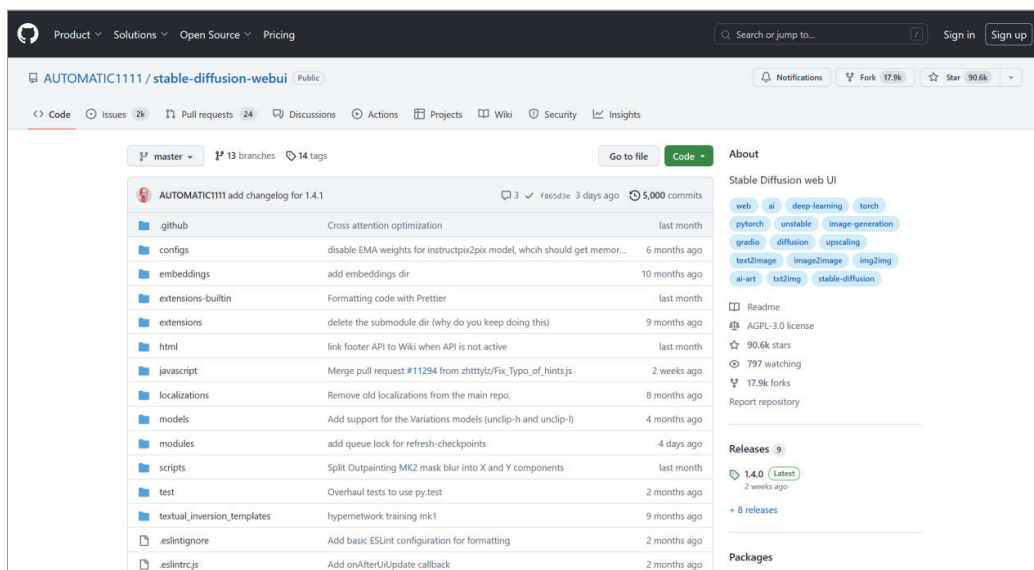


图 1-8

(2) 滚动页面至“安装和运行”部分，此时可以启用浏览器的翻译功能以便理解。这里提供了多种安装方式，选择第一种进行安装，单击 **v1.0.0-pre** 选项来到官方下载页面，如图 1-9 所示。



图 1-9

(3) 单击 `sd.webui.zip` 选项下载压缩包，如图 1-10 所示。



图 1-10

(4) 当下载并解压完成后得到文件，双击运行 `update.bat` 文件，将自动更新 Web UI 到最新版本，等待完成，然后关闭窗口，如图 1-11 所示。

名称	修改日期	类型
今天		
update.bat	2023/10/15 15:39	Windows 批处理文件
environment.bat	2023/10/15 15:39	Windows 批处理文件
run.bat	2023/10/15 15:39	Windows 批处理文件
webui	2023/10/15 16:37	文件夹
system	2023/10/15 15:39	文件夹

图 1-11

(5) 双击以启动 run.bat，在第一次启动时它将自动下载大量文件，如图 1-12 所示。

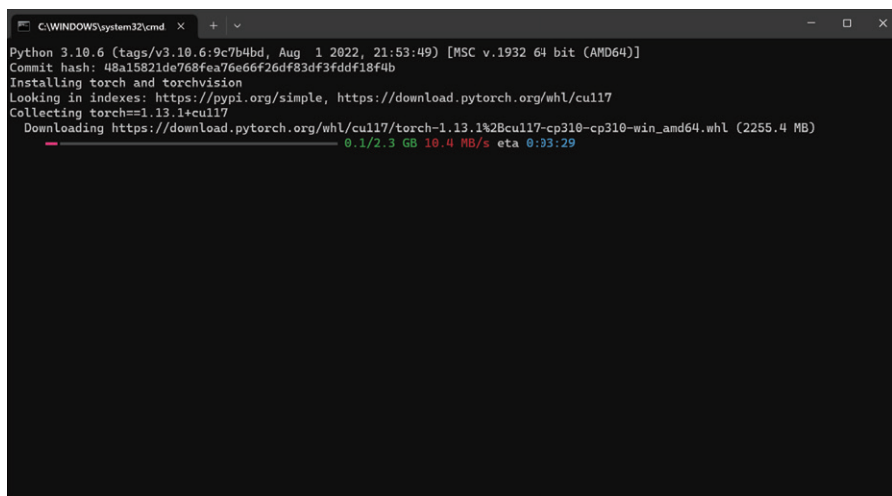


图 1-12

(6) 正确下载并安装所有内容后，将会看到消息“Running on local URL: http://127.0.0.1:7860”，如图 1-13 所示。

```

LatentDiffusion: Running in eps-prediction mode
DiffusionWrapper has 859.52 M params.
Running on local URL:  http://127.0.0.1:7860
  
```

图 1-13

(7) 在浏览器中打开链接，将显示 Web UI 界面，如图 1-14 所示。注意在下载过程中要全程保持网络畅通。

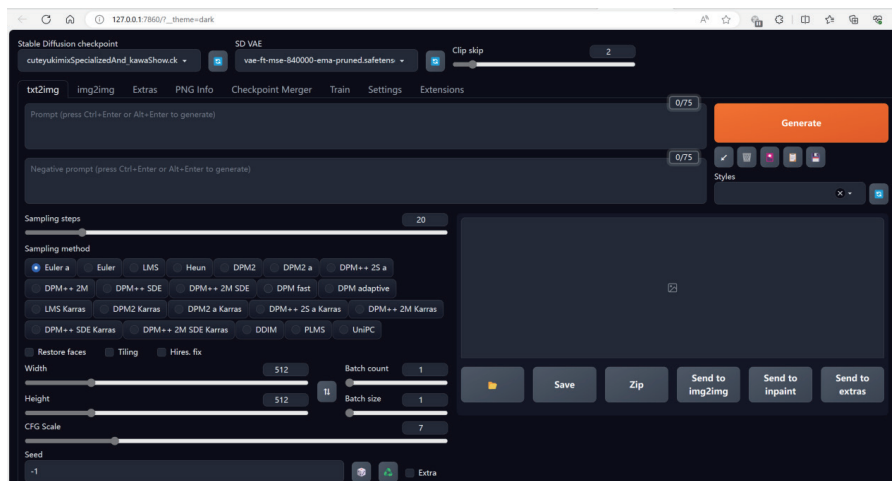


图 1-14

1.2.2 添加所需的 AI 模型

在完成 Stable Diffusion 项目的本地部署后，为了确保图像正常生成，用户至少需要添加一

个大模型。在模型资源方面,国际上知名的平台如 Civitai 和 Hugging Face,因起步较早,汇聚了大量高质量的模型及高水平的创作者,但访问这些平台可能需要借助 VPN 以克服网络限制。相比之下,国内也有如哩布哩布 AI、吐司 AI 等模型网站,访问它们无须担心网络问题,且它们往往涵盖了国外热门模型的本土化版本。国内网站的一个潜在的缺点是模型质量可能参差不齐,需要用户自行甄别。下面以国内模型网站为例说明添加模型的方法。

(1) 进入哩布哩布 AI 网站,地址为 <https://www.liblib.art/>。进入网站后可以看到首页中有各种类型模型训练的图像预览,在左上角有不同的标签,这些模型主要分为 LoRA 和 Checkpoint 模型,如图 1-15 所示。

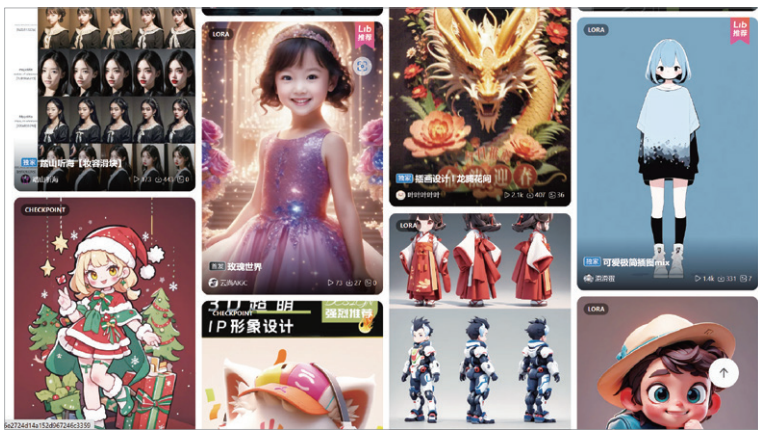


图 1-15

(2) 目前使用得最为广泛的就是这两个模型,使用 Stable Diffusion 至少需要搭载一款 Checkpoint 模型,LoRA 模型则是一种微调模型,在之后的章节会详细介绍。选择 Checkpoint 模型后,便可看到该模型的信息,也能看到模型作者对该模型的使用方法的讲解,在右侧单击“下载”按钮即可将模型下载到本地使用,如图 1-16 所示。

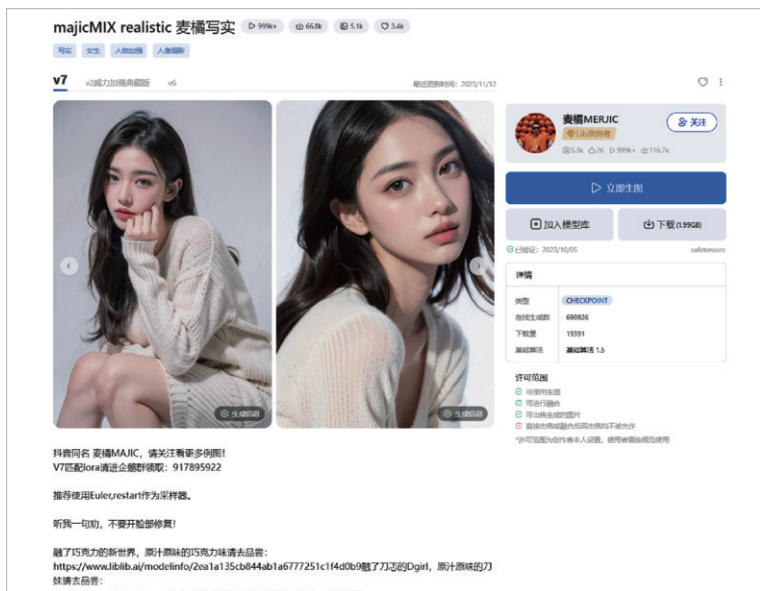


图 1-16

(3) 将下载好的 Checkpoint 模型放到 Stable Diffusion 的 models\Stable-diffusion 目录中，如图 1-17 所示。

名称	修改日期	类型
Lora	2023/11/15 20:02	文件夹
LyCORIS	2023/8/23 22:10	文件夹
RealESRGAN	2023/5/25 20:11	文件夹
roop	2023/6/28 12:58	文件夹
ScuNET	2023/5/25 20:11	文件夹
Stable-diffusion	2023/11/15 20:06	文件夹
SwinIR	2023/5/25 20:11	文件夹
TaggerOnnx	2023/11/7 20:06	文件夹
torch_deepdanbooru	2023/5/25 20:11	文件夹
VAE	2023/10/22 21:26	文件夹
VAE-approx	2023/5/25 20:11	文件夹

图 1-17

目前最新的大模型通常都内置了 VAE 模型，用户无须再通过外挂 VAE 模型的方式来改善生成图像偏灰的问题。但一些较早的大模型仍需用户手动下载，如果使用，需要将下载好的 VAE 模型放入 Stable Diffusion 的 models\VAE 中，在使用时也可以通过外挂 VAE 模型的方式覆盖原本的 VAE 模型，如图 1-18 所示。

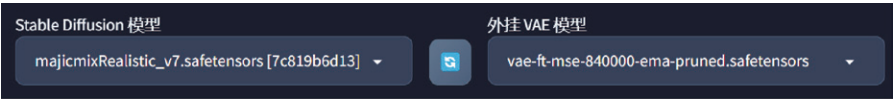


图 1-18

第 2 章

Stable Diffusion 绘图基本参数



本章为 AI 绘画入门的必学内容，通过学习 Stable Diffusion Web UI 的相关界面知识，掌握基础描述语参数和反推功能，创作者便可以在本章的实操中生成属于自己的第一张 AI 图像。

2.1 认识 Stable Diffusion Web UI

Stable Diffusion 的功能十分强大，但掌握 AI 绘画需要从认识最基础的界面开始，因为对界面的掌握程度直接影响了使用体验和后续灵活运用效果。在学完本节之后，读者会发现上手操作其实并不难。

在完成 Stable Diffusion 的本地部署，并且至少配置了一个大模型 Dreamboot 之后，便可以开始对 Web UI 的界面展开基础层面的认识了。

1. 启动界面

启动界面可以被划分为 4 个主要区域，即模型区、功能区、参数区和出图区，每个区域都有其特定的用途和功能，以满足用户的不同需求，如图 2-1 所示。

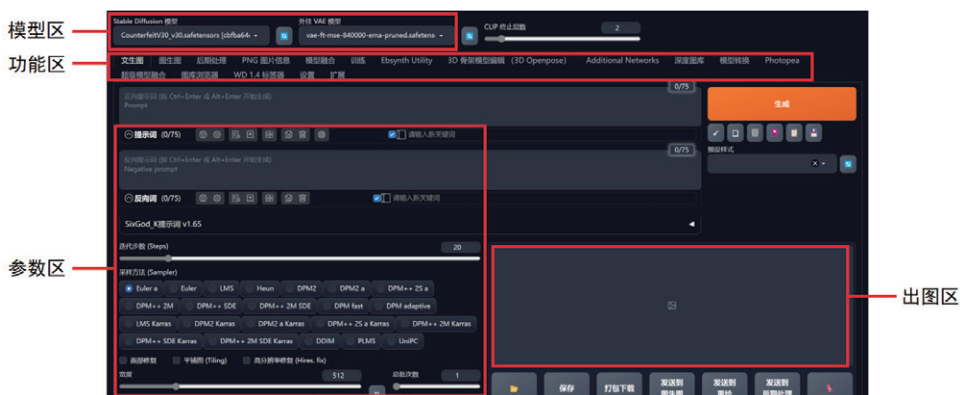


图 2-1

- 模型区：模型区的主要功能是让用户能够切换所需的模型。用户可以从网络下载所需的 Safetensors、CKPT、PT 模型文件，并将其放置在 \modes\Stable-diffusion 目录下。单击模型区的刷新箭头后，用户可以在此选择并加载新的模型。
- 功能区：功能区提供了一系列的功能选项，用户可以根据需要进行选择。在安装完对应的插件后，重新加载 UI 界面将会功能区添加对应插件的快捷入口。

- 参数区：参数区提供了一系列可调整的参数设置，这些设置会根据用户选择的功能模块变化。例如，在使用文生图模块时，用户可以指定要使用的迭代次数、掩膜概率和图像尺寸等参数。
- 出图区：出图区是供用户查看 AI 绘制最终结果的地方。在这个区域，用户还可以看到用于绘制图像的相关参数的各类信息。

2. 文生图页面

在文生图页面，用户可以输入文本、选择模型，并配置一些其他参数，以此来生成图像。文本是生成图像的基础，必须提供。用户可以选择预定义的模型，或者上传自己的模型。此外，用户还可以选择一些其他参数，如批处理大小、生成的图像尺寸等。接下来针对图 2-2 中的一些参数进行详细说明。

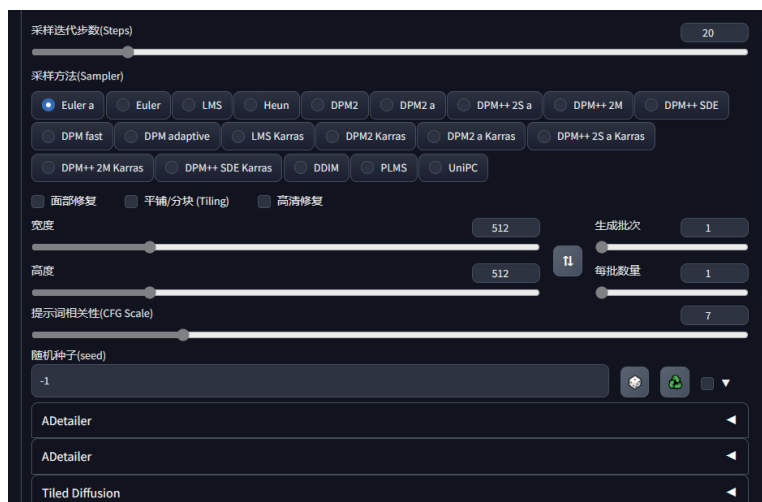


图 2-2

- 采样迭代步数：此参数用于指定图像生成的迭代次数。较多的迭代次数可能会让生成的图像质量更好，但也需要更长的时间来完成。
- 采样方法：此参数选择用于生成图像的采样方法。在默认情况下，该参数设置为 Euler a，但也可以选择 DPM++ 这些新加入的系列选项，这将使所生成图像的细节更丰富。
- 面部修复：如果绘制面部图像，可以选择此选项。当头像是近角时，选择此选项可能会导致过度拟合和图像虚化的现象。相较而言，当头像是远角时选择此选项更为适合。
- 平铺 / 分块：用于生成一个可以平铺的图像。
- 高清修复：此选项使用一种两步式的过程生成图像，首先以较小的分辨率创建图像，然后在不改变构图的情况下改进其中的细节。
- 宽度、高度：这两个参数用于指定所生成图像的宽度和高度，较大的宽度和高度需要更多的显存以及计算资源。
- 生成批次：此参数用于指定模型针对每一幅要生成的图像所能运行的最大迭代次数，增加其值，模型便可以多次生成图像，但生成的时间也会更长。
- 每批数量：此参数用于指定一次可以生成的最大图像数量。
- 提示词相关性：此参数可以调整图像与提示符的一致程度。增大其值，将使图像更接近提示内容，但过高会使图像的色彩过于饱和。此参数的值越小，AI 绘图的自主发挥

空间越大，越有可能产生有创意的结果（默认为 7）。

- 随机种子：此参数可以指定一个随机种子，用于初始化图像生成过程。相同的种子值每次都会产生相同的图像集，这对于保障图像生成的再现性和一致性很有用。如果将种子值保留为 -1，则每次运行“文生图”时将生成一个随机种子。

3. 图生图页面

在图生图页面，允许用户使用 Stable Diffusion 生成与原图像的构图色彩相似的图像，或者指定一部分内容进行变换。与文生图功能相比，图生图功能新增了图像放置区域和“重绘幅度”参数设置，如图 2-3 所示。

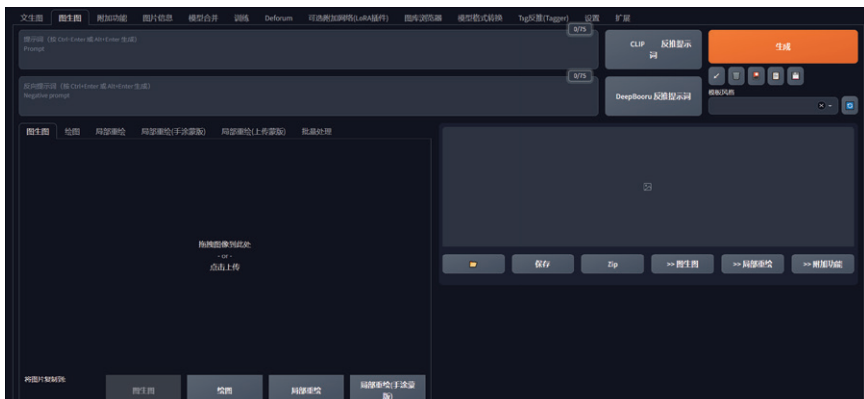


图 2-3

接下来针对图 2-4 中的相关参数进行说明。

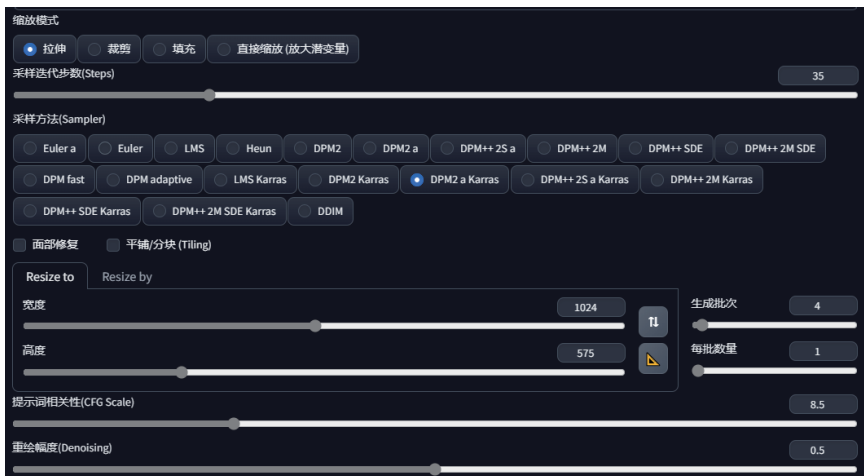


图 2-4

- 缩放模式：此参数主要用于设置在调整图像尺寸后以何种模式保证图像的输出效果，可选项包括拉伸、裁剪、填充和直接缩放。
- 重绘幅度：此参数决定图像模仿的自由度，数值越高，越能自由发挥；数值越低，生成的图像与参考图像越接近。通常，当数值小于 0.3 时，基本上就是在原图上加一个滤镜效果。
- 绘图：允许使用 Stable Diffusion 中的画笔在图像中进行绘制。如果想在图像的某

个具体位置增添物体，可以先通过涂鸦的方式画出其大概形状，再配合提示词辅助生成。

- 局部重绘：允许使用 Stable Diffusion 对图像中被手工遮罩的部分进行重新绘制。如果用户找到了一张整体尚可、细节较差的图像，可以单击这个按钮开始局部重绘。Stable Diffusion 会自动生成一个遮罩层，此时可以用鼠标在图像上涂抹需要修复的区域。单击“生成”按钮，Stable Diffusion 会根据遮罩层和原图生成一个新的图像，并显示在右侧。值得一提的是，官方提供了基于 1.5 版本的专门的 In-paint 修复模型。
- 局部重绘（手涂蒙版）：允许使用 Stable Diffusion 中的画笔在图像上进行重新绘制。该功能结合了局部重绘以及涂鸦功能，可以通过调整画笔的颜色以及形状更好地对图像进行局部更改。与局部重绘相比，该功能会参考画出的蒙版形状进行相应操作；与涂鸦功能相比，该功能会对原图进行重绘更改，而涂鸦只会原图上增添物品。
- 局部重绘（上传蒙版）：允许在 Stable Diffusion 中手动上传图像以及蒙版。该功能会对蒙版内或蒙版外的图像进行重绘，常用于较为精细的修改。

2.2 Stable Diffusion 描述语参数详解

在大概了解了 Stable Diffusion 的基础界面后，便可以开始尝试生成第一张 AI 绘画作品。接下来将通过实例从文生图、图生图以及如何规范地写 prompt 来控制图像，提升图像的细节，从而帮助读者迅速了解 Stable Diffusion 的作图流程。

2.2.1 尝试通过文本操控图像

首先通过最基础的文生图来制作图像，这里使用一款二次元赛璐璐风格的大模型 Counterfeit-V3.0 以及对应的 VAE 进行演示，大家可以根据已有的下载模型进行选择，具体操作步骤如下。

(1) 将提示词输入文生图（text-to-image generation）的正向提示词框中，此时为了强调所生成图像的质量，通常会先列出一些与质量相关的关键词，如 best quality、masterpiece、Highly detailed、absurdres 等，这些词有助于引导模型生成品质更高的作品。接着根据想要描述的主体内容进一步完善提示词，例如想生成一个身着水手服的粉色长发女孩的图像，则需要提示词中添加 1girl、sailor_shirt、long hair、pink hair 这类具体描述。与 Midjourney 等采用较为连续自然语言文本的方式不同，Stable Diffusion 模型更倾向于将自然语言拆解为独立的词组或短语。遵循这样的规则，正向提示词就可以按照上述方式编写完成，如图 2-5 所示。这样的输入方式有助于模型更精确地理解并生成符合预期的图像。



图 2-5

- (2) 单击“生成”按钮，在生成图像的区域会根据输入的提示词不断演化，进行去噪操

作。生成图像的时间与显卡的配置有关，显卡的性能越高，出图的速度越快。当进度条完成后，

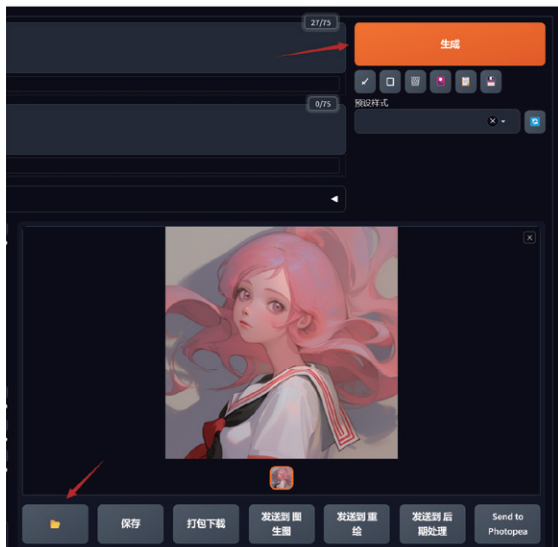


图 2-6

后，图像会显示在出图区域，用户可以通过单击进行放大查看，同时图像会自动保存在本地。单击界面下方的黄色文件夹图标，即可查看以前生成的所有图像，如图 2-6 所示。

(3) 加入负向提示词去除画面中不想要的物体或改善画面的质量，例如加入 badhandv4、EasyNegative、ng_deepnegative_v1_75t、rev2-badprompt、verybadimagenegative_v1.3、negative_hand-neg、bad-picture-chill-75v 等，如图 2-7 所示。这些负向提示词有助于精确地控制图像的生成过程，避免生成不必要的元素以提高图像的整体质量。

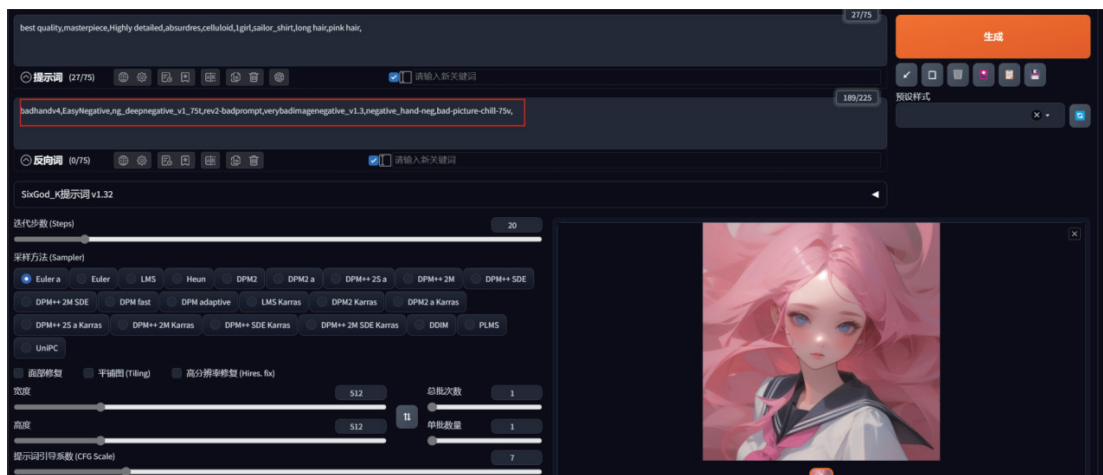


图 2-7



图 2-8

(4) 此时可以观察到画面中人物的脸部细节有了显著提升，但仍存在一些不合理之处，例如头发有多处未能自然连接，如图 2-8 所示。由于 Stable Diffusion 本质上是基于扩散模型的，它并不直接理解图像的内容或绘制方式，而是通过反向扩散过程直接生成图像，所以对于大多数模型而言，这些不合理之处是难以避免的。当然，对于这些细节问题，大家不必过于纠结。

(5) 如果想在保持同一张图的基础上获得更高精度的图像，可以先单击绿色小图标保存之前图像的种子，以便复现上一张图像。然后单击“高分辨率修复 (Hires. fix)”

选项，并在放大算法中选择 R-ESRGAN 4x + Anime6B（该算法对动漫图像的修复效果较佳）。接着设置“重绘幅度”为 0.2，注意幅度越大，所生成图像与原图的区别也会越大。在完成设置后，单击“生成”按钮等待图像生成。需要注意的是，高分辨率修复会显著减慢生成速度，并对显卡的算力提出更高要求。这一步骤的操作示意如图 2-9 所示，生成的效果如图 2-10 所示。

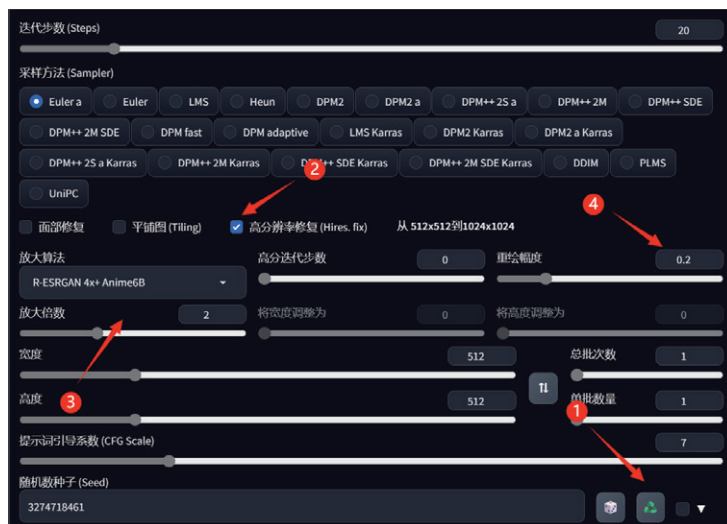


图 2-9



图 2-10

此时可以直观地观察到，图像的整体细节以及分辨率都有了显著提升，图像的像素也增加到了原先的两倍。在图像的下方还会展示当前图像的正向提示词、负向提示词、尺寸、种子等附加信息，这些信息会一并包含在图像的信息中，如图 2-11 所示。

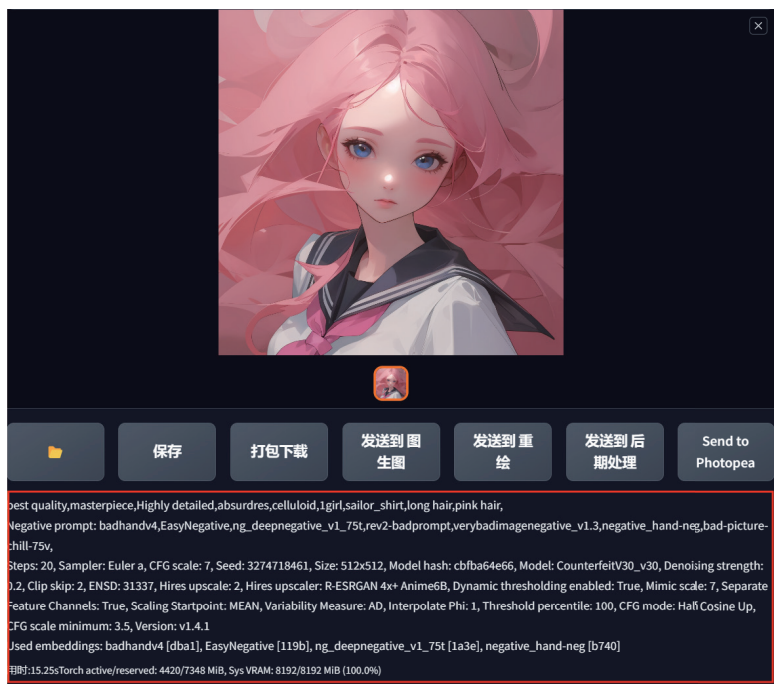


图 2-11

创作者同样可以使用 PNG 图像信息来展示图像详情。通过单击“发送到文生图”或“发送到图生图”按钮，可以快速复现并生成该图像，如图 2-12 所示。



图 2-12



图 2-13

通过高清放大技术，可以在保留原图画风格的基础上为图像增添更多细节，而不仅仅是简单地提升分辨率。若希望原图保持原有风貌，避免大幅度改动，则应谨慎设置重绘幅度，不宜过大。最终生成的动漫女孩图像效果如图 2-13 所示。

2.2.2 应用图生图功能

下面尝试一下图生图的功能，具体操作步骤如下。

(1) 与文生图界面相比，图生图界面增加了一个专门用于放置图像的区域。这一界面在生成图像时，除了会参考输入的文本内容外，还会受到所放入参照图像的影响，使得生成的图像更加贴合指定的主题或风格，如图 2-14 所示。

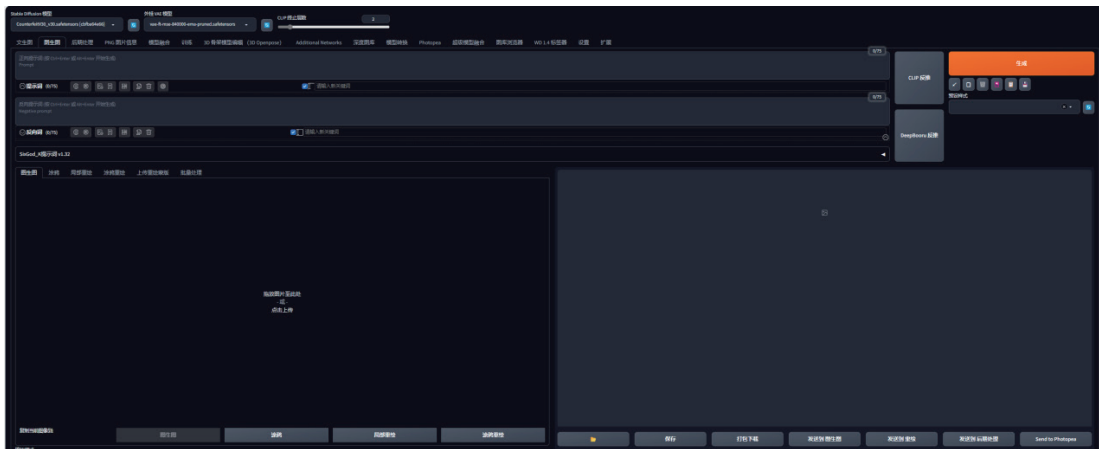


图 2-14

(2) 尝试将刚才生成的动漫女孩的眼睛从蓝色变为黑色。首先将刚才生成的图像放入选框中，保持其他参数不变，仅在正向提示词框中添加 **black eyes**，然后单击“生成”按钮，如图 2-15 所示。

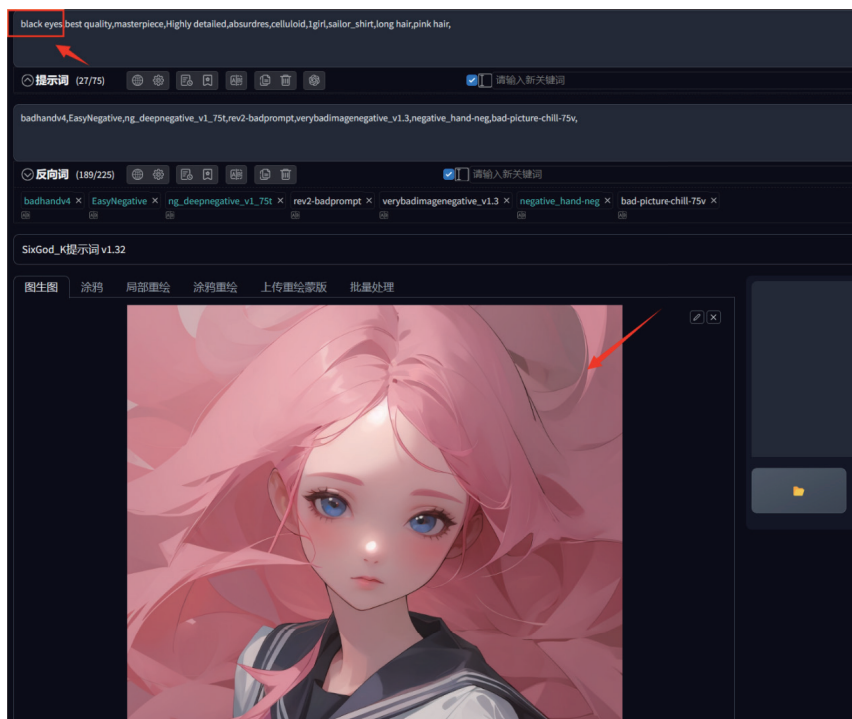


图 2-15

(3) 生成的图像与原图在构图上保持相似，同时也体现了在正向提示词框中新加入的 **black eyes**。然而，在许多细节上，新图像与原图相比有了显著变化。这是因为下方默认的重绘幅度设置为 0.75，重绘幅度越大，与原图的差异就越明显，AI 自由创作的空间也就越大。如图 2-16 所示，新图像在保留原图基本特征的同时融入了更多的创意和变化。

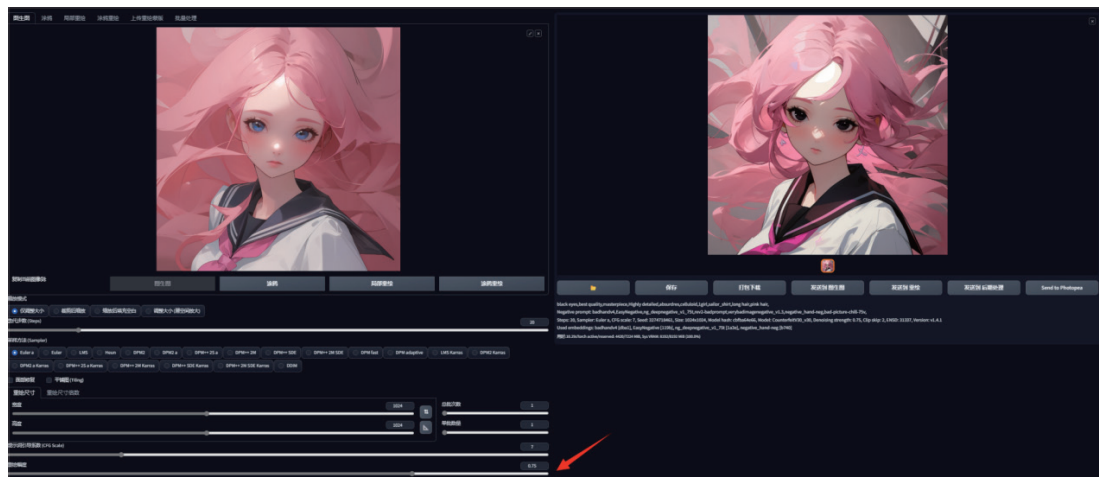


图 2-16

(4) 如果只想改变眼睛的颜色，希望图像的其他部分保持原样，那么可以使用图生图的一个功能——局部重绘。这个功能允许用户针对图像的特定区域进行编辑，而不会影响其他部分，如图 2-17 所示。

(5) 若想让刚才图像中的人物闭上眼睛，可以先将图像放入选框中，然后使用右侧的画笔工具调整其大小，接着在需要更改的部分（眼睛区域）绘制一个蒙版，这样之后生成的内容将仅针对所绘制蒙版的内部进行更改。随后在提示词中加入 **closed eyes**，并再次单击“生成”按钮，如图 2-18 和图 2-19 所示。



图 2-17



图 2-18



图 2-19

(6) 最终生成的图像中，人物仅闭上了双眼，其他区域并未发生任何改变，如图 2-20 所示。

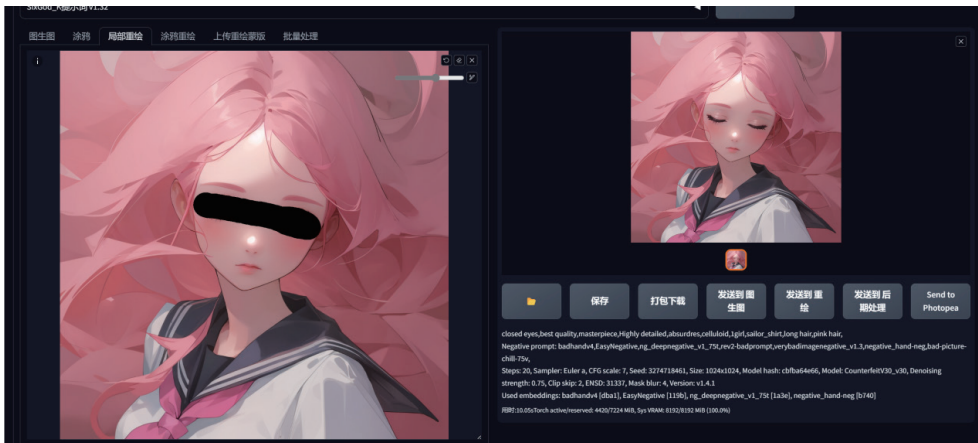


图 2-20

(7) 通过刚才的操作，读者已经大致了解了局部重绘的用法。如果想给这张图像再添加一些元素，如 gold earrings（金色的耳坠），可以先将这个描述加入正向提示词中，然后使用图生图的另一个功能区域——涂鸦重绘来实现更具体的创作意图。

(8) 进入“涂鸦重绘”页面，单击右上角的调色板图标，然后吸取或选择想要的颜色。接下来就像使用局部重绘一样，用画笔画出希望更改的区域。在完成绘制后单击“生成”按钮，即可看到效果，如图 2-21 和图 2-22 所示。

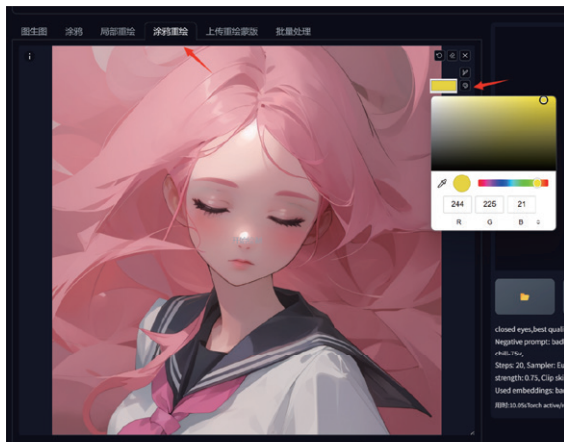


图 2-21

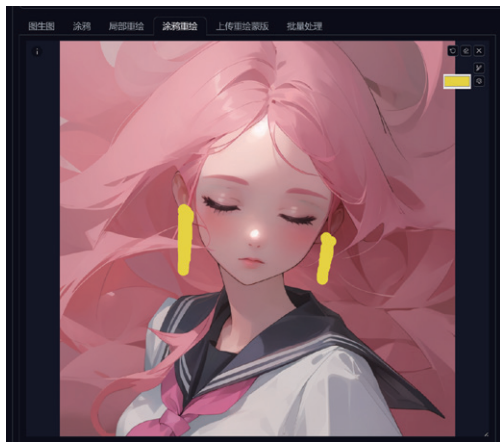


图 2-22

(9) 与单纯的局部重绘相比，涂鸦重绘功能加强了对颜色的控制。在使用局部重绘时，如果颜色效果始终无法达到理想状态，可以尝试使用涂鸦重绘进行更精确的调整。最终的效果如图 2-23 所示。

2.2.3 基本提示词写法

目前，AI 制图主要依赖于文本来决定生成的图像，故提示词的恰当性对最终图像的质量和效果具有显著影响，因此编写优质、规范的提示词显得尤为重要。这里介绍一种有效的语序结构，即采用“引导词 + 焦点 + 环境 (+ 修饰词)”的方式来书写提示词。

引导词在 AI 制图中扮演着指南针的角色，它引领着创作者的创作方向。引导词可以细分为三个子部分，即基础引导词、风格词和效果词。基础引导词如同创作者的目标，它明确了创作者所追求的质量标准，如“顶级作品”或“最高品质”。风格词则是创作者手中的工具，帮助创作者选择适合的艺术风格，如“素描”“油画”或“浮雕”。效果词则如同创作者的调色板，允许创作者挑选合适的光效，如“优秀照明”“镜头光晕”或“景深”，以增添图像的视觉效果。

焦点在 AI 制图中如同画布上的主角，可以是人物、建筑或景物等。为了让主角更加栩栩如生，创作者需要对其进行详尽的描述。例如，若主角为人物，则可能需要描绘其面部特征、发型、身材、穿着及姿态等细节。在没有明确角色时，创作者可以转而描述环境中的关键元



图 2-23

素,如壮观的瀑布 (waterfall)、盛开的向日葵 (sunflower)、古老的时钟 (clock) 等,以构建丰富的场景氛围。

环境在 AI 制图中扮演着舞台的角色,它是主角展现魅力的必要背景。若缺少环境描述,则容易生成单调的纯色背景或与效果标签相关的简单背景,同时焦点元素也会显得突兀而庞大。环境词汇能够构建出围绕焦点、充斥整个画面的生动场景,如繁茂葱郁的森林 (forest)、绚烂多彩的彩虹 (rainbow)、温暖明媚的阳光 (sunlight)、清澈宁静的湖泊 (lake),以及色彩斑斓的玻璃 (colored glass) 等,为图像增添丰富的层次和氛围。

修饰词在功能上常与效果词相似,用于增添场景的细节和丰富度,例如彩虹 (rainbow)、闪电 (lightning)、流星 (meteor) 等自然元素。需要注意的是,如果焦点描述过于简略,而修饰词被放置在引导词的末尾作为效果词使用,可能会导致场景的描述过于突出,从而削弱了焦点的存在感,使得整个图像显得场景权重过大、焦点不够突出。

在生成艺术图像的过程中,权重控制扮演着至关重要的角色,它决定了 AI 是否会对期望的元素给予足够的重视。调整提示语在咒语(即输入指令)中的位置是最基本的权重控制方法之一,通常位置越靠前的词汇越受重视。此外,还可以通过给提示语添加括号并指定权重值来进一步调整权重,如 (castle:1.5) 就是直接给 castle 这个词汇赋予权重,其中的数字即为权重大小。权重值越大,表示该元素在生成图像时的重要性越高。在默认情况下,权重值为 1,而在实际使用中,权重值通常在 0 到 2 之间调整。

在 Web UI 界面中,使用 () 和 [] 进行权重调整时,前者会使其中内容的权重乘以 1 (实际无变化),后者则尝试通过除以 1 (同样无实质影响) 来调整权重,但这种方式并不直观且效率不高。尽管多重括号能够增加权重,如 (((((castle)))))) 通过 4 个括号将 castle 的权重提升至 1.4641,但这种做法既低效也不优雅。

因此,推荐采用 (提示语: 权重数) 的方式更直接、高效地调整权重。在具体操作时,先选中一组词汇,然后通过方向键来设置或调整其权重值。具体操作如下。

(1) 以下是一个例子,展示如何输入正向提示词“最佳质量,杰作,基于物理的渲染,生动的色彩,墨水,水彩,一个女孩,身材修长,汉白玉发光皮肤,中式连衣裙,汉服,唐装,波波头,腰带,长发,刘海辫子,闭着一只眼睛,花海,日落,中国风格,绽放的效果,细致美丽的草原上有花瓣、花朵、蝴蝶、项链、微笑(花瓣重复描述以增强效果),周围飘浮着沉重的花瓣流动”。对应的英文提示词为“best quality, masterpiece, physically-based rendering, Vivid Colors, ink, water color, 1girl, slender, white marble glowing skin, china_dress, hanfu, tang style, pibo, waistband, long hair, braided bangs, one eye closed, flower ocean, sunset, chinese_style_loft, Bloom effect, detailed beautiful grassland with petal, flower, butterfly, necklace, smile, petal, surrounded by heavy floating petal flow”。

同时加入基础的负向提示词“nsfw, logo, text, badhandv4, EasyNegative, ng_deepnegative_v1_75t, rev2-badprompt, verybadimagenegative_v1.3, negative_hand-neg, muted hands and fingers, poorly drawn face, extra limb, missing limb, disconnected limbs, malformed hands, ugly”。

其中, badhandv4、EasyNegative、ng_deepnegative_v1_75t、verybadimagenegative_v1.3、negative_hand-neg 是 Embedding 模型,需要读者从模型网站自行下载,它们有助于提升生成图像的质量,如图 2-24 所示。



图 2-24

(2) 创作者可以在初期选择生成多张低分辨率的图像以进行筛选，然后使用之前介绍的保存种子的方法挑选出最满意的几张图进行高清重绘与放大。此处示例为一次生成 4 张图，单击“生成”按钮，如图 2-25 所示。

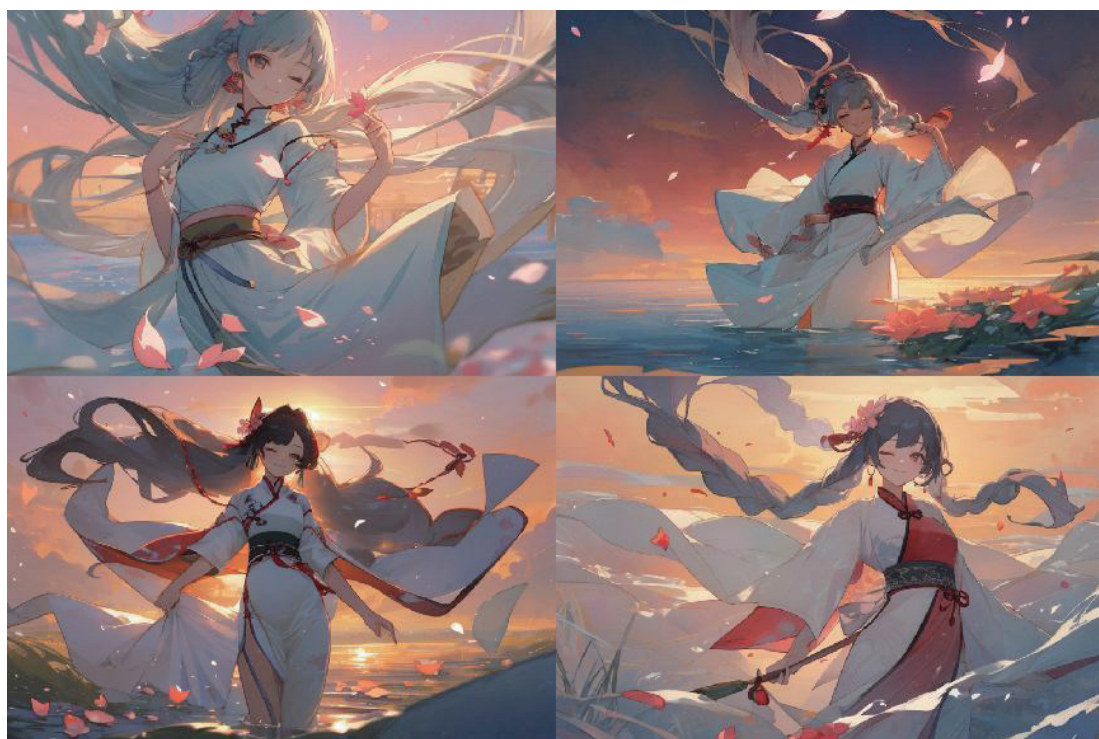


图 2-25

(3) 从刚才生成的图中可以观察到既有满意的也有不满意的。其中，第一张和第四张图像较为符合先前的提示词描述，而第二张和第三张则错误地将“花海”解读为“花和海洋”，且未能体现出蝴蝶元素。基于此，选择图 2-25 中的第四张图保存其种子，并特别增强了蝴蝶的权重至 1.3，然后再次单击“生成”按钮，以期获得更符合预期的图像，如图 2-26 所示。

(4) 保存种子后多次生成图像，实际上是基于原图种子数据的相邻种子数据进行取样。在图 2-26 中，除了第一张图像与先前的图像较为相似外，其余图像均存在显著差异。当创作者调整权重后，虽然成功地引入了蝴蝶元素，但也出现了权重溢出现象。这是由于权重设置过大，导致原本应作用于特定修饰词的权重影响范围扩大，甚至“溢出”到了其他元素上。在

此例中表现为人物的发饰被影响，呈现出类似蝴蝶的形状。当权重调整至 1.5 或更高时，这种“画面污染”现象会更为严重，可能导致图像整体崩坏。因此不建议创作者将权重设置得过高，以避免出现此类问题，如图 2-27 所示。

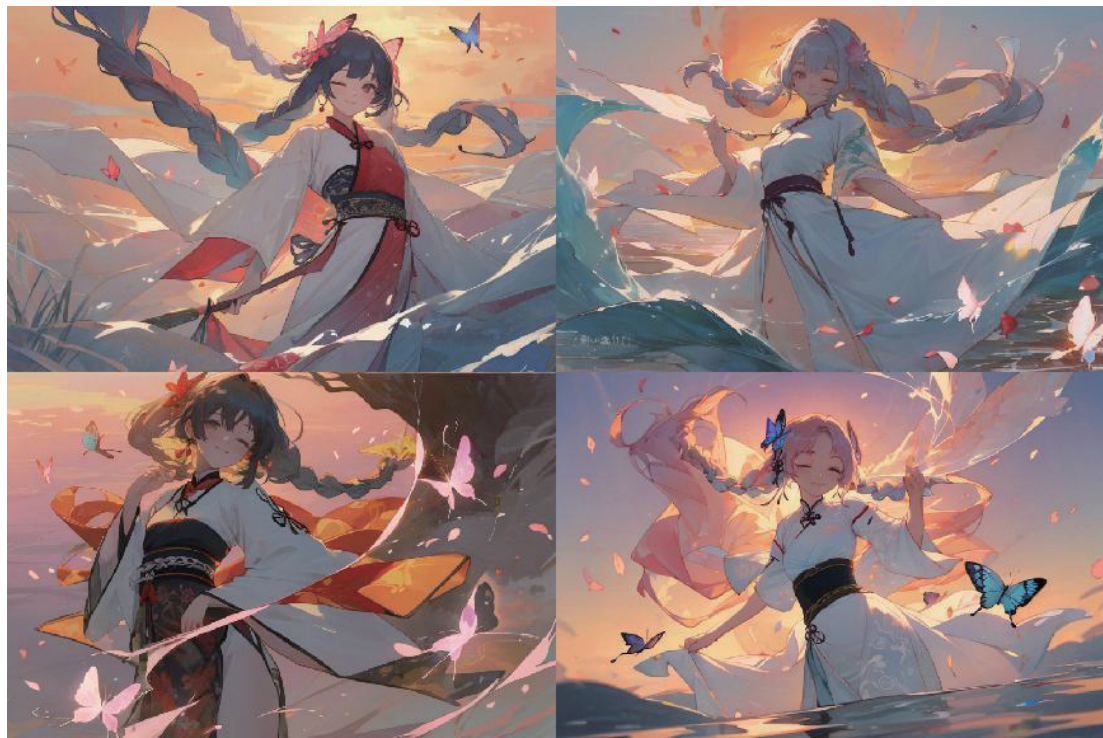


图 2-26

(5) 除此之外，创作者还可以尝试仅通过调整提示词的先后顺序来观察图像的变化。例如，将“蝴蝶”放置到最前面，相比直接提升蝴蝶的权重，调整顺序对整个图像的影响较小，主要表现为对人物发饰的轻微影响，而非直接出现蝴蝶元素。然而，这种调整也间接地提升了蝴蝶在图像中的“感知权重”，同时可能降低了其他部分的权重，对人物的衣服、姿势、背景等也产生了一定程度的影响，如图 2-28 所示。



图 2-27



图 2-28

在 Web UI 的文本理解机制中，确实存在着一种逻辑性的运作方式。这一逻辑性在蝴蝶实

例中得到了明确的验证,表明 AI 在解析描述时会遵循特定的顺序,即词组的排列顺序会直接影响其权重的分配。鉴于此,创作者在编写标签时可以将其视为撰写一篇小作文的过程,运用作文的逻辑来构思和编排词组。通过借鉴作文的行文逻辑,创作者可以更有效地确定关键词的排列顺序,从而提升 AI 在生成图像时的准确性和符合度。

在语言学中,描述一个事物时,常遵循“目标、定义、细节”的递进方式。以描述为例,在描述一幅画时,首先指明“这是一幅画”,接着阐述其类型或特点,最后细致地描绘画的细节;在描述一个人时,同样先确立“这是一个人”的前提,然后描述其性格、身份等特征,最后详细刻画其外貌、服饰及行为;在描述一个背景时,也是先明确“这是一个背景”,再概述其氛围或特点,最后具体描述背景中的元素及独特之处;对于背景中的物体,同样遵循先指出“这是一个物体”,再说明其属性或种类,最后详细描绘其形状、颜色等具体特征步骤。

运用这种逻辑思维方式,创作者能够有条不紊地对画面中的每个元素进行由整体到局部的详细描述。在描绘一幅画时,创作者可以先概括整幅画的氛围和主题,然后深入阐述画面中的主要对象,包括其数量、种类、具体形态、特效及装饰细节等。接着转向次要对象,同样细致入微地描述其各方面的特征。在细化到每一个对象时,创作者还可以进一步采用三段法来剖析其各个组成部分,如描述人物时,先总览其外貌与服饰特色,再逐一刻画五官、发型等细节。通过这样的层层分解与详尽阐述,创作者能够全面而深刻地分析并描述出整幅画中的所有元素。

通过以上方法不仅能够详尽地分析和描述整幅画中的所有元素,而且可以保证每个元素都能通过三段法进一步深挖细节。这一流程确保了创作者对整幅画有一个既全面又深入的理解。下面将通过具体例子来进一步展示这种文本逻辑的妙用。

masterpiece, top quality, ultra HD 8k wallpaper, vivid watercolor painting, 1 boy located on a beach (一幅画,这是一幅生动的水彩画,画中展现了一个男孩在海滩上的场景)。

1 casually dressed surfer boy, alone, full body, the boy standing next to a rock on the beach (男孩是一位身着休闲装的冲浪者,独自站立在海滩的岩石旁,身姿挺拔)。

The boy has a sunny smile, healthy skin tone and bright blue eyes, brown short hair tousled by the sea breeze, wearing a baseball cap and a fishbone necklace (他脸上洋溢着阳光般的笑容,拥有健康的肤色和明亮的蓝眼睛,棕色短发在海风中略显凌乱。他头戴棒球帽,颈间挂着一条鱼骨项链)。

the boy is wearing a yellow T-shirt and blue shorts, both adorned with vibrant patterns (他身着一件黄色T恤和蓝色短裤,衣物上均饰有鲜艳的图案,为整体增添了几分活力)。

sand and shells on the beach, various shells scattered on the beach, waves gently lapping the shore, sunlight, dazzling reflections, shimmer, golden glow (海滩上,细软的沙粒与五彩斑斓的贝壳交相辉映,海浪轻柔地拍打着岸边,阳光洒在海面上,波光粼粼,整个海滩沐浴在一片金色的光辉之中)。

最终生成的效果如图 2-29 所示。



图 2-29

2.2.4 分布渲染

在 Web UI 中存在一种独特的语法功能, 允许创作者在同一幅画中分阶段地绘制不同的提示元素, 这被称为分步渲染。其语法格式为 [A:B:step], 其中 A 和 B 代表不同的内容描述, step 则指定了渲染的步骤或比例。当 step 大于 1 时, 它表示具体的步骤数; 当 step 小于 1 时, 则表示占总步数的百分比。

例如, [a boy with a blue shirt:red cap:0.3] 这条指令意味着在前 30% 的渲染过程中将绘制一个穿着蓝色衬衫的男孩, 而在剩余的 70% 中, 在这个男孩的头上添加一顶红色帽子。这可以理解为 AI 首先绘制出蓝色衬衫男孩的初步形象, 然后在此基础上增添红色帽子的细节。

此外还有两种灵活的变体语法, 即 [:B:step] 和 [A::step]。前者通过将 A 留空, 实现仅在指定 step 后绘制 B 的内容; 后者则将 B 留空, 意味着仅根据 A 的描述进行绘制, 直至达到指定的 step。

这种分步渲染的方法特别适用于在图像中嵌套图像的场景。例如, 可以首先绘制一个穿着 T 恤的男孩, 然后在 T 恤的特定位置绘制一个苹果图案。如果 T 恤和苹果的绘制顺序颠倒, 那么苹果可能会被绘制在画面中的任意位置, 而非预期的 T 恤上。

在使用这种语法时, 务必要仔细核对括号的使用, 以确保准确无误。同时, 分步渲染可能会带来一定的渲染延迟。例如, 即使设定男孩渲染前 50 步、背景渲染后 50 步, 男孩的完整渲染可能要在第 60 步左右才完成。这可能是因为 AI 在处理男孩形象时会优先关注面部细节, 而头发、装饰等则可能被视为环境元素继续渲染。

下面是一段正向提示词的示例, 旨在生成一张高质量、细节丰富的图像:

extremely detailed CG unity 8k wallpaper, (((masterpiece))), (((best quality))), ((ultra-detailed)), (best illustration), (best shadow), ((an extremely delicate and beautiful)), dynamic angle, standing, solo, [impasto:1.3, a detailed cute girl with blue eyes and long wavy curly black hair wearing a detailed red dress with a white belt, beautiful detailed eyes:1.2, (cute face:1.2), expressionless, (upper body, legs), (red umbrella:1.3)]:0.5, : (flat color), (dark rainy background), ((medium saturation))), (surrounded by raindrops), ((surrounded by puddles)), surrounded by city lights, (shining), Rain:0.5]

这段提示词旨在让 AI 在前 70% 的渲染步骤中绘制一个穿着红裙子、手持红色雨伞的女孩; 而在后 30% 的步骤中添加雨滴和城市灯光的元素。这样, AI 会首先勾勒出女孩和红裙子的基本轮廓, 然后在画面上增添雨滴和城市的夜景, 最终效果如图 2-30 所示。

正向提示词为: (extremely detailed CG unity 8k wallpaper), (((masterpiece))), (((best quality))), ((ultra-detailed)), (best illustration), (best shadow), ((an extremely delicate and beautiful)), dynamic angle, standing, solo, [impasto:1.3, a detailed cute girl with blue eyes and long wavy curly black hair wearing a detailed blue dress with a white belt, beautiful detailed eyes:1.2, (cute face:1.2), expressionless, (upper body, legs)]:0.5, : (flat color), (bright garden background), ((high saturation))), (surrounded by flowers), ((surrounded by butterflies)), surrounded by trees, (shining), Sunshine:0.5].

在这段提示词的指导下, AI 将在前 50% 的渲染步骤中绘制一个穿着蓝色连衣裙的女孩; 而在接下来的 50% 步骤中增添花园、花朵、蝴蝶和树木等元素。这意味着 AI 会首先勾勒出女孩和连衣裙的基本轮廓, 然后在背景中添加上述自然元素, 最终效果如图 2-31 所示。



图 2-30



图 2-31

2.3 用 Stable Diffusion 生成彩色小鸟

本节在使用 Stable Diffusion 的基础模型的前提下，尝试用最简单的文生图功能生成一只鸟的图像。

(1) 打开 Stable Diffusion 的 Web UI 界面，如图 2-32 所示。

(2) 进入“文生图”页面，在提示词区域输入 1 colorful bird（一只彩色的鸟），设置“采样迭代步数”为 25、“采样方法”为 DPM++ 2S a Karras、“宽度”为 768、“高度”为 1024、“生成批次”为 4，其他保持初始状态，如图 2-33 所示。

(3) 单击“生成”按钮，等待几分钟，就完成了图像的生成，如图 2-34 所示。

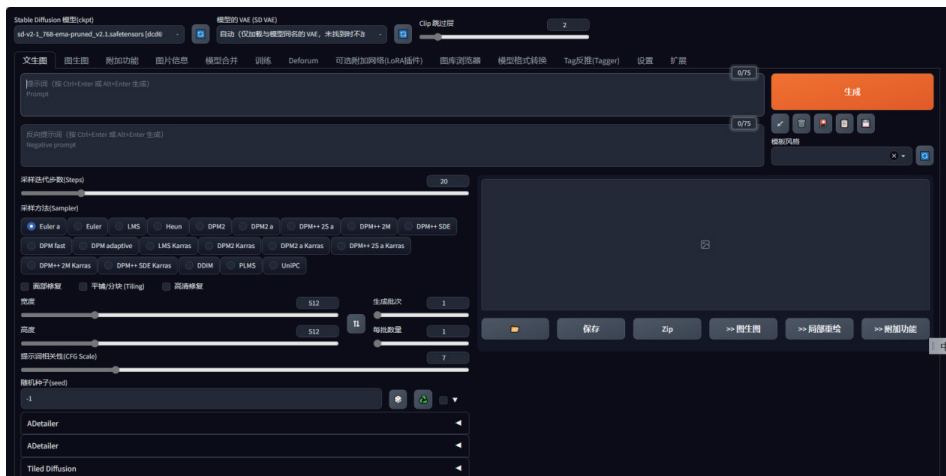


图 2-32

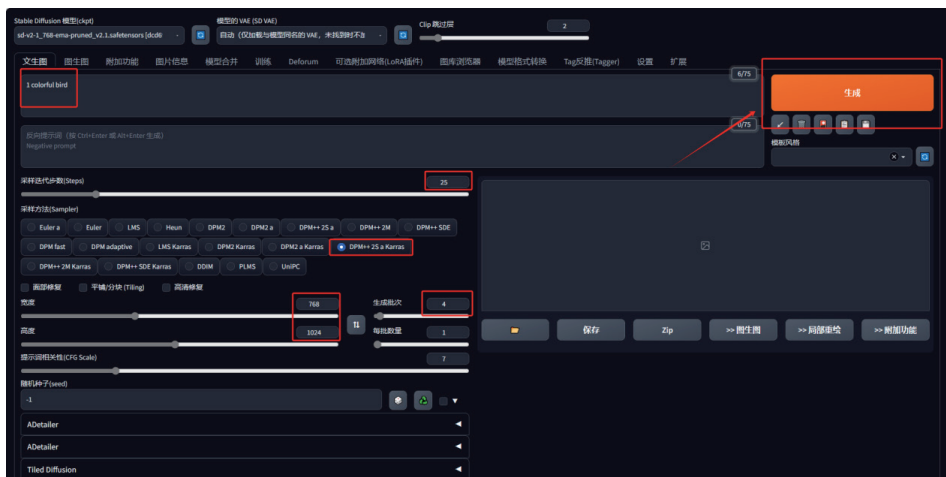


图 2-33

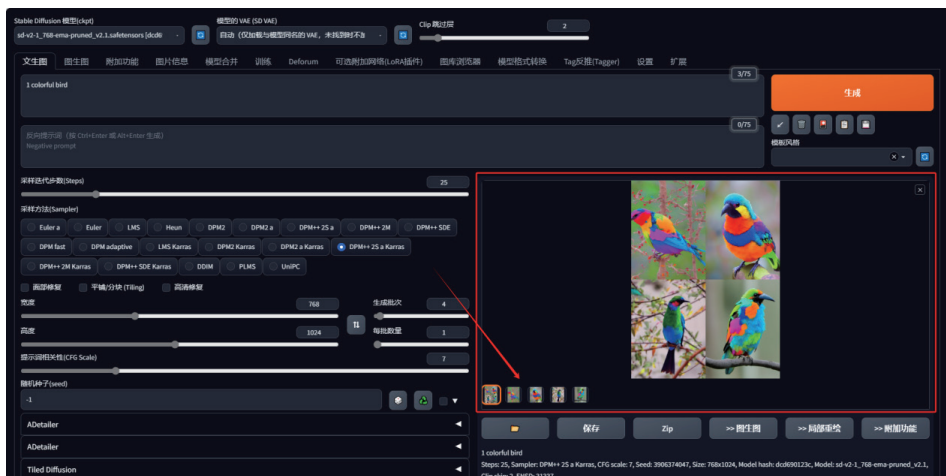


图 2-34

(4) 在生成的结果中，第一张是生成结果的缩略图，如图 2-35 所示。



图 2-35

(5) 通过对比看到图 2 和图 4 比较接近人们对画面的审美需求，如图 2-36 所示。这样就完成了彩色小鸟的创作过程。



图 2-36

2.4 用 Stable Diffusion 生成童话城堡

本节在使用 Stable Diffusion 的基础模型的前提下，尝试用图生图功能生成一个七彩的美丽城堡图像。

(1) 打开 Stable Diffusion 的 Web UI 界面，进入“图生图”页面，如图 2-37 所示。

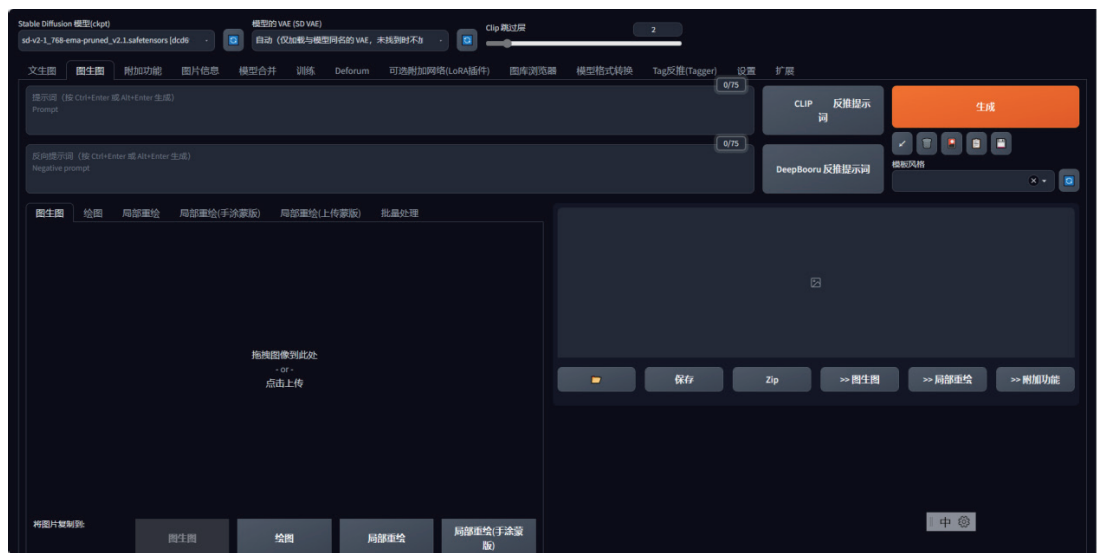


图 2-37

(2) 在上传图像区插入准备好的图像，如图 2-38 所示。



图 2-38

(3) 在提示词区域输入 Panoramic view, a colorful and beautiful castle, a fairy tale world (全景画面, 一个七彩的美丽城堡, 童话世界), 设置“采样迭代步数”为 35、“采样方法”为 DPM++ 2S a Karras、“宽度”为 1024、“高度”为 575、“生成批次”为 4、“提示词相关性”为 8.5、“重绘幅度”为 0.5, 如图 2-39 所示。

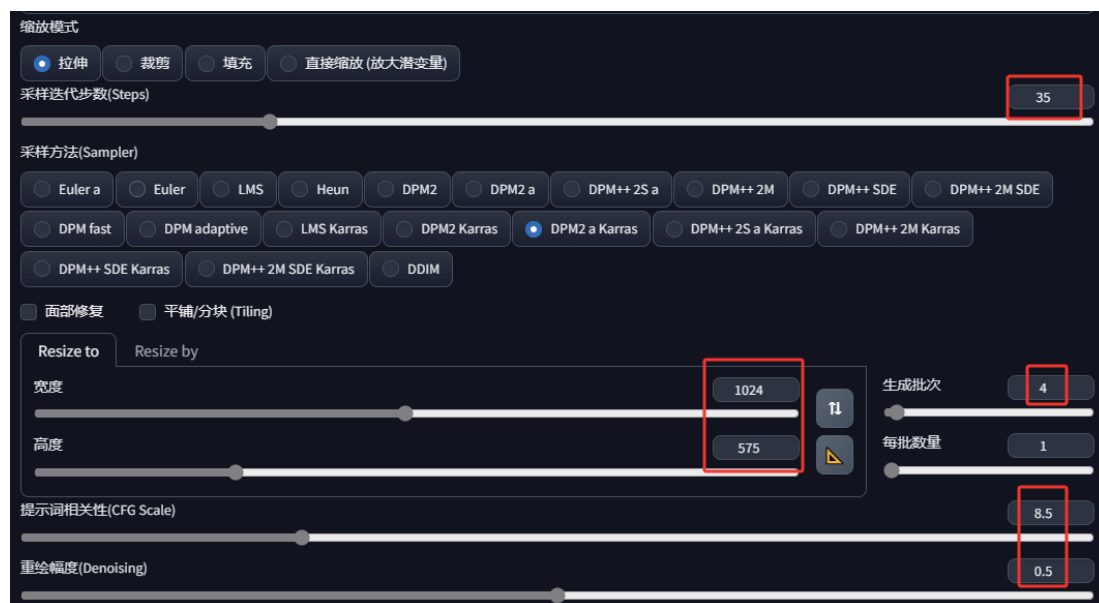


图 2-39

(4) 单击“生成”按钮, 等待几分钟, 就完成了图像的生成, 如图 2-40 所示。

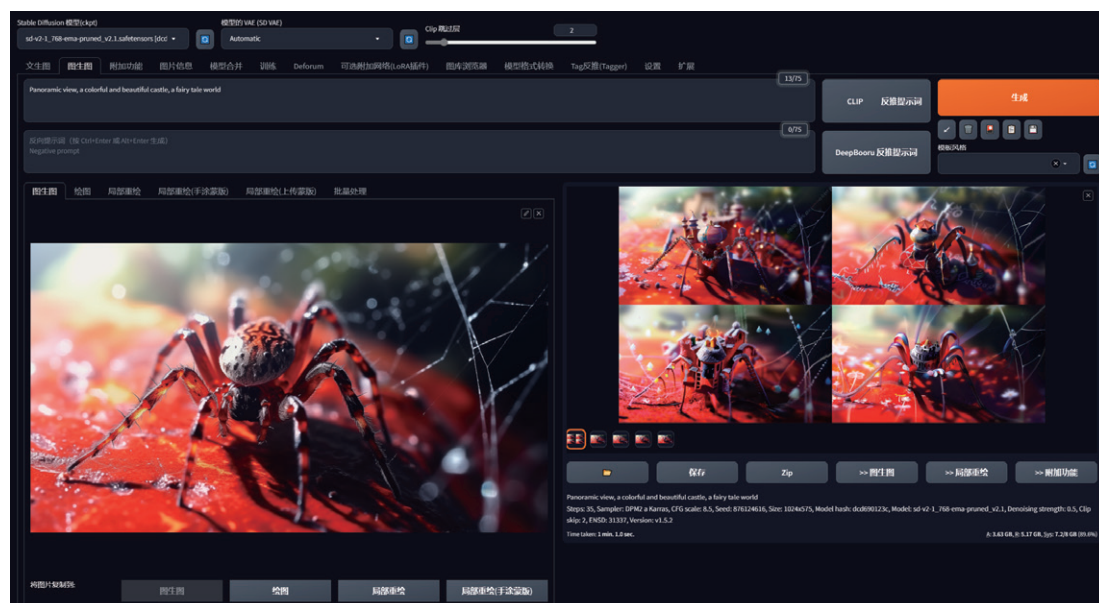


图 2-40

(5) 生成的结果中, 第一张是生成结果的缩略图, 如图 2-41 所示。



图 2-41

(6) 通过对比看到右下角的图比较接近人们对画面的审美需求, 如图 2-42 所示。这样就完成了城堡的创作过程。



图 2-42

(7) 如果大家觉得画面细节有所欠缺, 则需要增加“采样迭代步数”的值, 设置为 65 后再次生成, 如图 2-43 所示。生成的新的缩略图如图 2-44 所示。

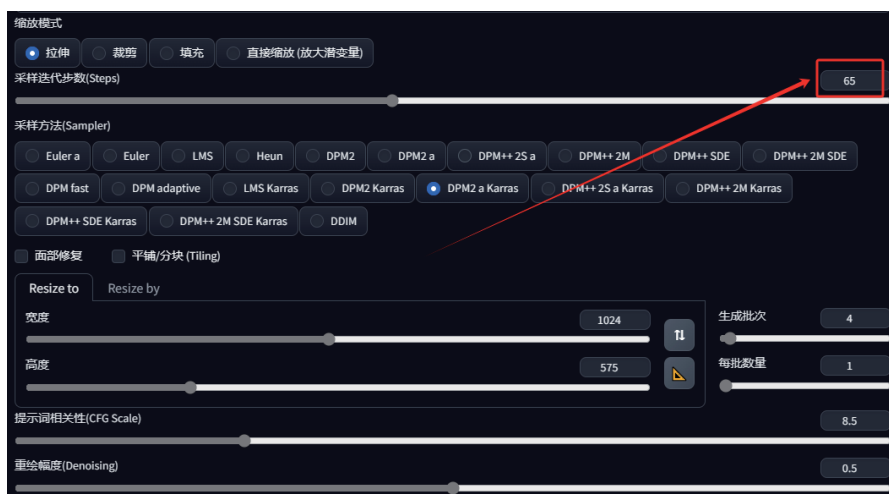


图 2-43



图 2-44

2.5 用 Stable Diffusion 替换人物服装

本节给人物图像进行服装替换操作。Stable Diffusion 的局部重绘功能是其图像生成应用中的一个核心特性，它允许用户对图像的特定部分进行再创作，从而实现更加细致和可控的图像编辑。

局部重绘功能主要依赖于蒙版的使用。用户可以通过画笔工具在图像上涂抹需要修改的部分，这部分区域即被蒙版覆盖。之后，Stable Diffusion 会根据用户的指示和提供的正向提示词对蒙版区域进行重绘，而图像的其他部分保持不变。

(1) 启动 Stable Diffusion 的 Web UI 界面，这是进行图像生成和编辑的主要平台。

(2) 选择图生图功能，在“图生图”页面中用户可以找到局部重绘的选项，如图 2-45 所示。

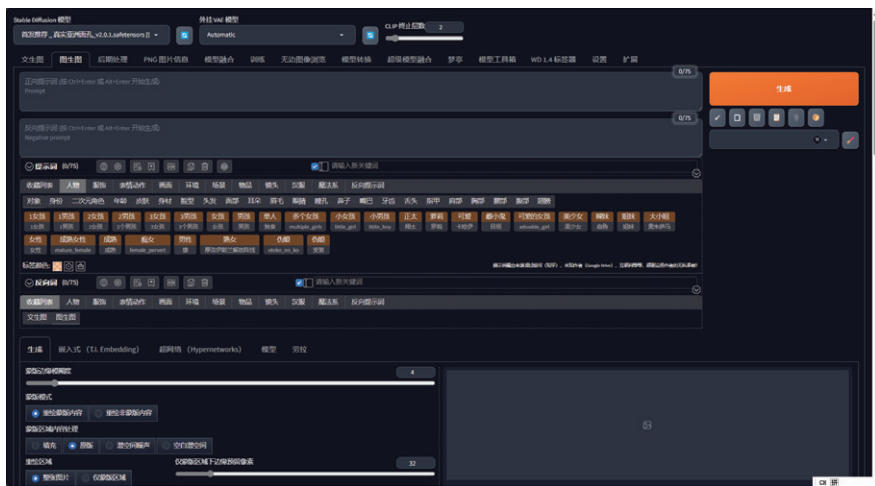


图 2-45

(3) 通过上传自己的图像，或者使用从图生图功能中发送过来的图像进行局部重绘，如图 2-46 所示。

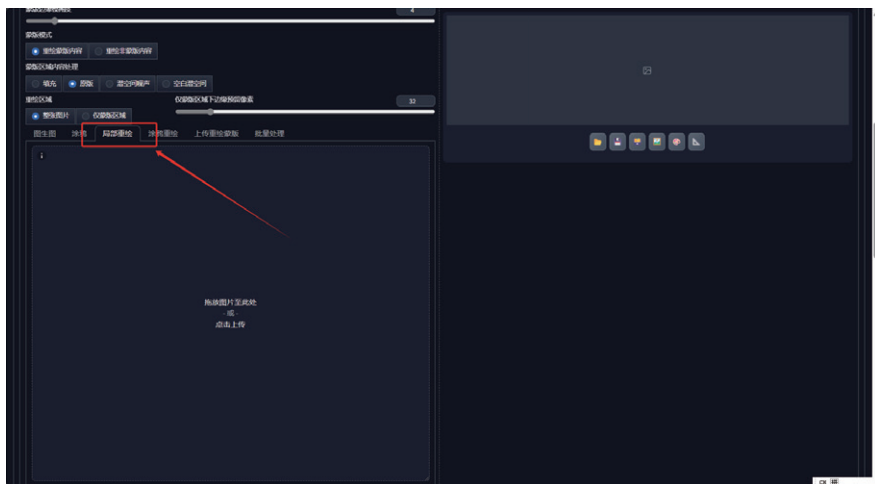


图 2-46

(4) 使用局部重绘功能来修改图像中衣服的颜色及款式。在局部重绘界面中，用户可以使用画笔工具在图像上涂抹需要修改的部分，涂抹后的区域即被蒙版覆盖。用户还可以根据需要进行调整蒙版的模糊程度，以及选择重绘蒙版内容还是重绘非蒙版内容，如图 2-47 所示。

(5) 更改笔刷的大小，如图 2-48 所示。

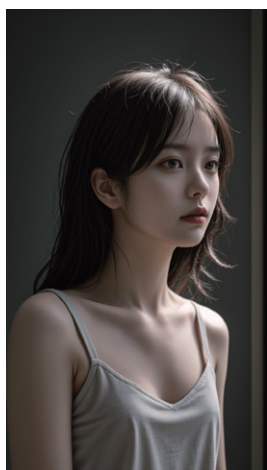


图 2-47

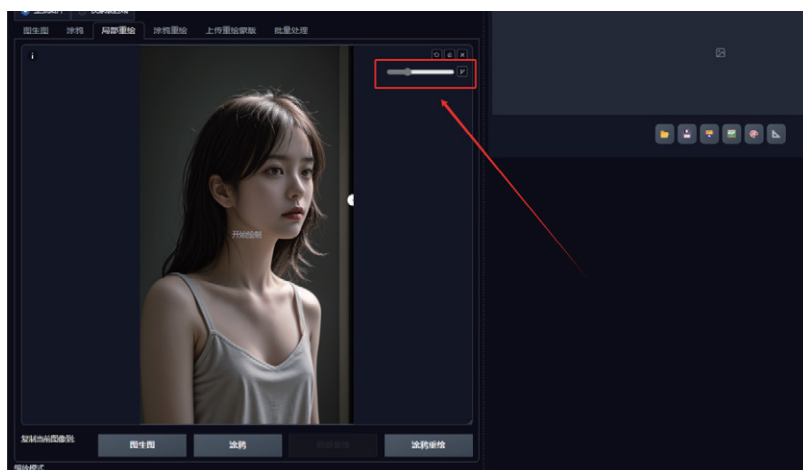


图 2-48

(6) 将要更改的地方使用笔刷进行涂抹，如图 2-49 所示。

(7) 设置“蒙版模式”为“重绘蒙版内容”，设置“蒙版区域内容处理”为“原图”，设置“重绘区域”为“整张图像”，如图 2-50 所示。

(8) 设置“迭代步数”为 30，将“重绘尺寸”设置为上传图像的尺寸，或者设置为图像尺寸的不同倍数，否则图像会出现拉伸情况。设置“重绘幅度”为 0.5，重绘幅度越大，与原图越不相符，反之越符合原图，如图 2-51 所示。

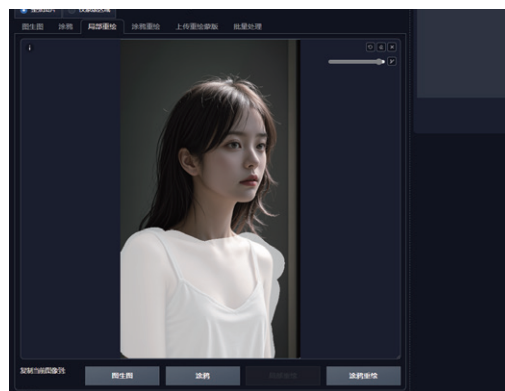


图 2-49

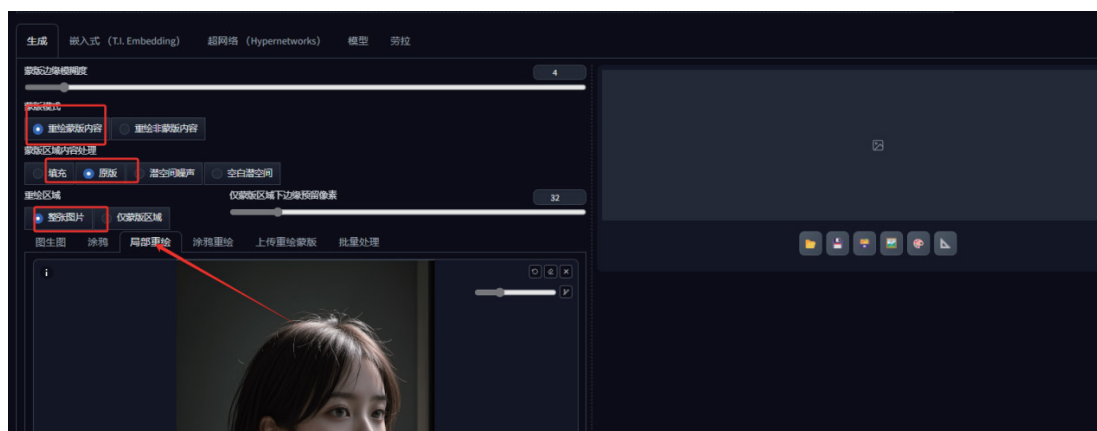


图 2-50

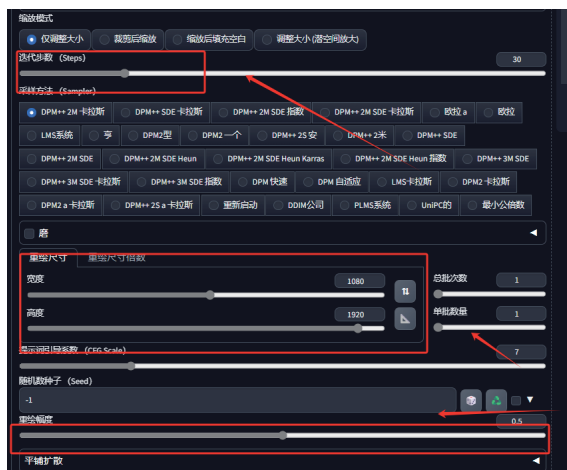


图 2-51

(9) 在正向提示词框中输入 Orange clothes, shirt（橙色衣服，衬衫），如图 2-52 所示。



图 2-52

(10) 单击“生成”按钮，原图将被局部修改（更改了衣服的颜色及款式），如图 2-53 所示。



图 2-53