

第1章

从零入门：认识与安装 Stable Diffusion



章前知识导读

Stable Diffusion 是一款热门的 AI 图像生成工具，但对于初学者来说，掌握 Stable Diffusion 却是一项具有挑战性的任务。本章将分享一些入门技巧，帮助大家快速认识并安装 Stable Diffusion。



新手重点索引

-  入门基础：认识与使用 Stable Diffusion
-  快速安装：配置与部署 Stable Diffusion



效果图片欣赏



1.1 入门基础：认识与使用 Stable Diffusion



Stable Diffusion（必要时可缩写为 SD）不仅在代码、数据和模型方面实现了全面开源，而且其参数量适中，使大部分人可以在普通显卡上进行绘画甚至精细调整模型。

毫不夸张地说，Stable Diffusion 的开源对 AIGC（Artificial Intelligence Generated Content，生成式人工智能）的繁荣和发展起到了巨大的推动作用，因为它让更多的人能够轻松上手进行 AI（Artificial Intelligence，人工智能）绘画。本节将深入讲解 Stable Diffusion 的概念以及原理，帮助大家初步认识 Stable Diffusion。

1.1.1 概念：什么是 Stable Diffusion

Stable Diffusion 是一种利用神经网络生成高质量图像的模型（或工具软件），基于扩散过程，能够在保持图像特征的同时增强图像的细节。该模型由 3 个主要部分组成，包括 VAE（Variational Auto-Encoders，变分自编码器）、U-Net 和 CLIP（Contrastive Language-Image Pre-training，对比语言 图像预训练的文本编码器）。详细介绍如下。

（1）VAE。VAE 是一种神经网络结构，主要用于生成模型，通过学习数据的潜在空间表示来生成新的数据。在 Stable Diffusion 中，VAE 被用作基于概率生成模型的编码器和解码器。VAE 通过将输入数据映射到潜在空间中进行编码，然后将编码的向量与潜在变量的高斯分布进行重参数化，这样可以直接从潜在空间中进行采样。

（2）U-Net。U-Net 是一种基于卷积神经网络的图像分割模型，具有特殊的 U 形结构，使输入图像的分辨率逐渐减小，而输出图像的分辨率逐渐增加。在 Stable Diffusion 中，U-Net 能够对图像进行部分特征提取，并在解码过程中对生成的图像进行重构，以获得高质量的生成结果。

（3）CLIP。CLIP 是一种神经网络算法，用于实现“文本 图像”的匹配，可以将输入的文本文本和图像进行语义相关性匹配，从而实现对图像内容的理解。在 Stable Diffusion 中，CLIP 不仅用于评估生成的图像，还可以指导数据的采样方式，以提高生成图像的多样性和相关性。

具体来说，Stable Diffusion 在训练模型时会将原始图像通过不断的随机扩散和反向扩散进行变形处理，将图像的细节信息逐渐压缩到低频区域。这样，Stable Diffusion 不仅能够提取图像的潜在空间表示，而且能够将图像的噪声和细节等信息分离出来。图 1-1 所示为前向扩散过程，能够将图像转换到低维潜在空间。



图 1-1 前向扩散过程

逆概率沿扩散（Inverse Probability Flow Along Diffusion, IPFAD）是用于 Stable Diffusion 模型的逆模型。这个模型是一个自回归模型，可以根据当前帧的噪声和之前帧生成的图像预测下一帧的噪声。通过逆概率沿扩散，Stable Diffusion 可以生成高质量的图像，如图 1-2 所示。

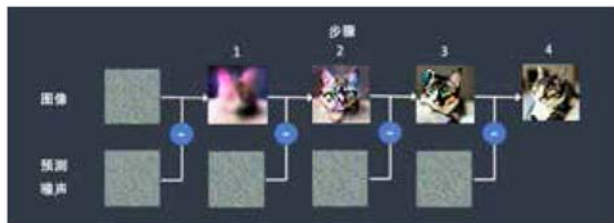


图 1-2 逆概率沿扩散通过逐步减去图像中的预测噪声生成图像

专家指点

在 Stable Diffusion 中，Predicted noise 指的是通过噪声预测器（noise predictor）预测出来的噪声。这个过程发生在去噪步骤之前，首先在潜在空间中生成一张完全随机的图片，然后噪声预测器会估计图片的噪声，并将预测的噪声从图片中减去。

这个过程会重复多次，最后得到一张干净的图片。这个去噪过程也称为采样，因为 Stable Diffusion 在每一步中都会生成一张新的样本图片。采样器决定了如何进行随机采样，不同的采样器会对图像生成结果产生影响。

1.1.2 原理：Stable Diffusion 的 AI 绘画本质

在 Stable Diffusion 中，有两种绘图模式，即通过文本生成图像（即文生图）和通过图像生成图像（即图生图）。Stable Diffusion 中的文生图是指通过输入文本描述（即提示词），利用扩散过程生成与之相关的图像。这种技术基于扩散模型，将文本编码器的输出与噪声相结合，然后通过解码器生成图像。

图 1-3 所示为 Stable Diffusion 文生图的推理流程。首先，使用文本作为输入信息，通过文本编码器（Text encoder）提取文本嵌入，即编码文本（Encoded text）；同时，通过随机数生成器（Random Number Generator, RNG）初始化一个随机噪声，即图中的 64×64 初始噪声碎片（潜在空间上的噪声， 512×512 图像对应的噪声维度为 64×64 ）。然后，将文本嵌入和随机噪声送入扩散模型（Diffusion model）U-Net 中，生成去噪后的潜在空间。最后，将生成的潜在空间送入自编码器的解码器模块（Decoder），得到生成的图像。

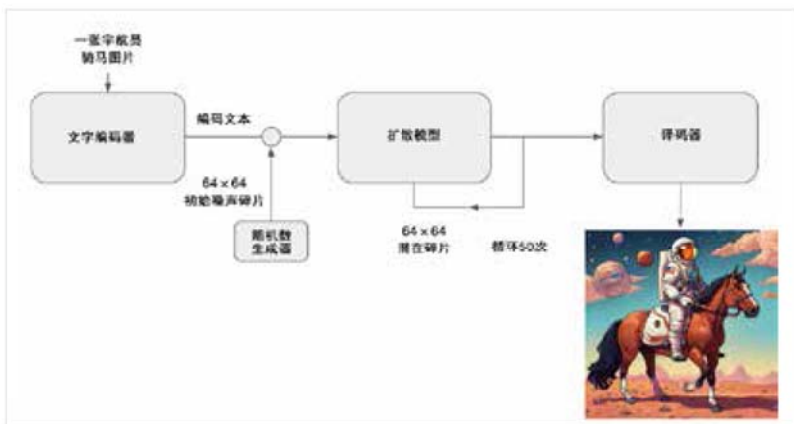


图 1-3 Stable Diffusion 文生图的推理流程



专家指点

64×64 潜在碎片指的是潜在空间中的一个 64×64 像素的区域，它被用作 U-Net 结构的输入。潜在空间指的是在去噪步骤之前，从完全随机的图片中通过噪声预测器预测出来的潜在图片。这个潜在图片可以看作输入文本描述在潜在空间中的一种表示，而 64×64 潜在碎片则是从这个潜在图片中提取出来的一个区域。

循环 50 次指的是在生成图像的过程中，使用 U-Net 结构进行 50 轮的扩散过程。通过多轮的扩散过程，可以使图像更加平滑、细节更加丰富。

Stable Diffusion 的整体操作流程非常简单，共分为 4 个步骤，即选择模型、输入提示词（或上传原图）、设置生成参数和单击“生成”按钮。最终的图像效果是由模型、提示词（或原图）和生成参数三者共同决定的。其中，模型主要决定图像的画风，提示词（或原图）主要决定画面内容，而生成参数则主要用于设置图像的预设属性。通过这个流程，可以轻松地使用 Stable Diffusion 生成符合我们要求的各种图像。

1.1.3 试用：通过 Stable Diffusion 官网绘图

Stable Diffusion 官网是一个非常直观且功能丰富的网站，主要提供了基于 Stable Diffusion 技术的服务和产品，同时还提供了在线 Stable Diffusion 绘图功能，大家可以试用和体验 Stable Diffusion 的绘图效果，如图 1-4 所示。



扫码看视频



图 1-4 绘图效果展示

专家指点

与另一个主流的 AI 绘画工具 Midjourney 相比，Stable Diffusion 的优点如下。

(1) 免费开源。Midjourney 需要登录 Discard 平台上进行使用，并且需要付费；Stable Diffusion 则有大量的免费安装包，用户无需付费即可下载并一键安装，而且将其安装到本地后，生成的图片只有用户自己可以看到，保密性更高。

(2) 拥有强大的开源模型和插件。由于其开源属性，Stable Diffusion 拥有大量免费的高质量外接预训练模型和扩展插件。例如，提取物体轮廓、人体姿势骨架、图像深度信息的 ControlNet 插件，可以让用户在绘画过程中精确控制人物的动作姿势、手势和画面构图等细节。此外，Stable Diffusion 还具备 Inpainting（图像修复）和 Outpainting（输出绘画）功能，可以智能地对图像进行局部修改和扩展，而某些功能是目前的 Midjourney 无法实现的。





下面介绍通过 Stable Diffusion 官网绘图的操作步骤。

STEP 01 进入 Stable Diffusion 官网，在页面下方的 Prompt（提示词）输入框中输入相应提示词，如图 1-5 所示。

STEP 02 单击 Generate（生成）按钮，即可快速生成相应的图像，效果如图 1-6 所示。



图 1-5 输入提示词



图 1-6 生成相应图像效果

1.2

快速安装：配置与部署 Stable Diffusion



Stable Diffusion 是一个开源的深度学习生成模型，能够根据任意文本描述生成高质量、高分辨率、高逼真度的图像效果。为了帮助大家快速入门并充分利用这个功能强大的 AI 绘画工具，本节将详细介绍 Stable Diffusion 的配置要求、安装方法和使用技巧。

1.2.1 前提：Stable Diffusion 的配置要求

Stable Diffusion 是最受欢迎的 AI 绘画工具之一，它快速、直观，并能够生成令人印象深刻的图像效果。如果用户有兴趣尝试使用 Stable Diffusion，则需要检查自己的计算机配置是否符合要求，因为它对计算机配置的要求较高。

不同的 Stable Diffusion 分支和迭代版本可能会有不同的要求，因此需要检查每个版本的具体规格。Stable Diffusion 的基本配置要求如下。

- (1) 操作系统：Windows、MacOS。
- (2) 显卡：不低于 6GB 显存的 N 卡（指 NVIDIA 系列的显卡）。
- (3) 内存：不低于 16GB 的 DDR4 或 DDR5 内存。DDR（Double Data Rate）是指双倍速率同步动态随机存储器。
- (4) 安装空间：不低于 12GB 或更大，最好是 SSD（Solid State Disk 或 Solid State Drive，固态硬盘）。

这是 Stable Diffusion 的最低配置要求，如果用户想要获得更好的出图结果和更高分辨率的

图像，则需要更高配置的硬件，如具有 10GB 显存的 NVIDIA RTX 3080 或者更新的 RTX 4080 和 RTX 4090，它们分别有 16GB 和 24GB 的显存。图 1-7 所示为 2023 年 11 月生成的桌面（台式机）显卡性能天梯图，越往上的显卡性能越好，当然价格也越贵。

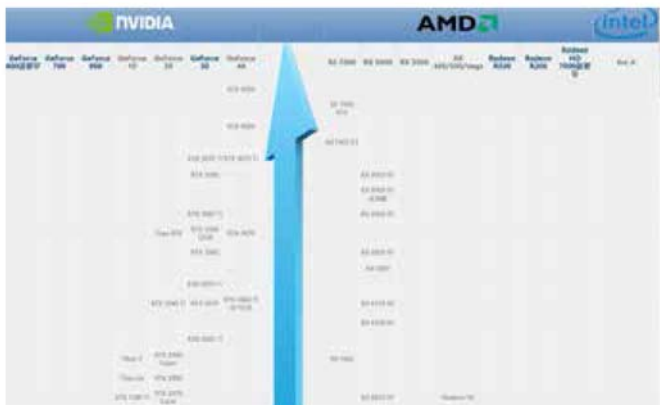


图 1-7 2023 年 11 月生成的桌面（台式机）显卡性能天梯图（部分显卡）

虽然 Stable Diffusion 的官方版本并不支持 AMD (Advanced Micro Devices, 超威半导体公司) 和 Intel (英特尔公司) 的显卡，但是已经有一些支持这些显卡的分支版本，不过安装过程比较复杂。当然，如果用户没有高性能的 GPU (Graphics Processing Unit, 图形处理器)，也可以使用一些网页版的 Stable Diffusion，该版本没有任何硬件要求。

专家指点

要想流畅运行 Stable Diffusion，推荐的计算机配置如下。

- (1) 操作系统：Windows 10 或 Windows 11。
- (2) 处理器：多核心的 64 位处理器，如 13 代以上的 Intel i5 系列或 Intel i7 系列，以及 AMD Ryzen 5 系列或 Ryzen 7 系列。
- (3) 内存：32GB 或以上。
- (4) 显卡：NVIDIA GeForce RTX 4060TI (16GB 显存版本)、RTX 4070、RTX 4070TI、RTX 4080 或 RTX 4090。
- (5) 安装空间：大品牌的 SSD 硬盘，500GB 以上的可用空间。
- (6) 电源：为了保证显卡能够稳定运行，建议选择额定功率为 750W 或以上的大品牌电源。

此外，如果使用笔记本电脑运行 Stable Diffusion，需要注意散热问题，因为在 Stable Diffusion 运行过程中 GPU 可能会满载运行，温度会非常高。

1.2.2 本地：在计算机上安装 Stable Diffusion

随着人工智能技术的不断发展，许多人工智能绘画软件应运而生，使绘画过程更加高效、有趣。Stable Diffusion 是其中备受欢迎的一款，它使用有监督深度学习算法来完成图像生成任务。下面以 Windows 10 操作系统为例，介绍 Stable Diffusion 的安装流程。

1. 下载 Stable Diffusion 程序包

首先需要从 Stable Diffusion 的官方网站或其他可信的来源下载该软件的程序包，文件名通常为 Stable Diffusion 或者 sd-xxx.zip/tar.gz，xxx 表示版本号等信息。下载完成后，将压缩文件解压



到想要安装的目录下，如图 1-8 所示。

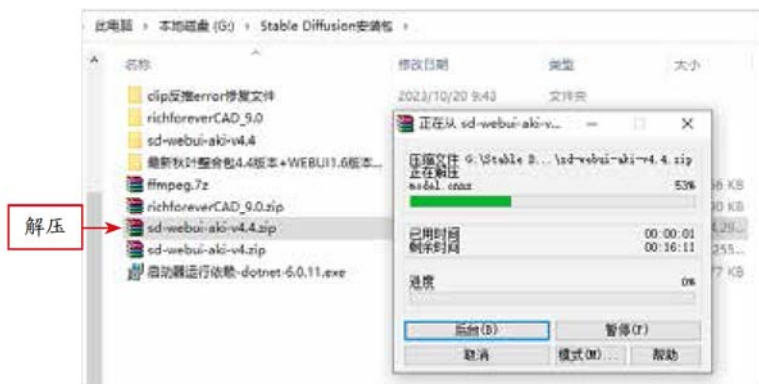


图 1-8 解压 Stable Diffusion 的安装文件

2. 安装 Python 环境

由于 Stable Diffusion 是使用 Python 语言开发的，因此用户需要在本地安装 Python 环境。用户可以从 Python 的官方网站上下载 Python 解释器，如图 1-9 所示，并按照提示进行安装。（注意，Stable Diffusion 要求使用 Python 3.6 以上的版本。）

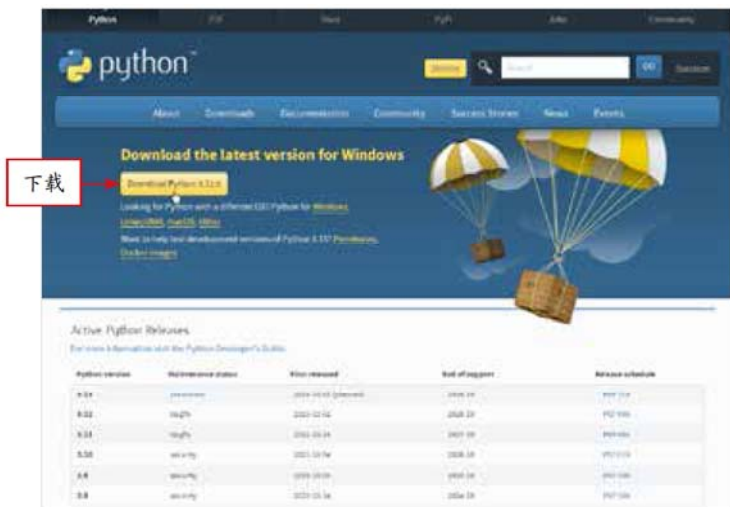


图 1-9 从 Python 的官方网站上下载 Python 解释器

3. 安装依赖项

依赖项指的是为了使 Stable Diffusion 能够正常运行，需要安装和配置的其他相关的软件库或组件，这些依赖项可以是编程语言、框架、库文件或其他软件包。在安装 Stable Diffusion 之前，用户需要确保下列依赖项已经正确安装。

（1）PyTorch。PyTorch 是一个开源的 Python 机器学习库，它提供了易于使用的张量（tensor）和自动微分（automatic differentiation）等技术，使它特别适合于深度学习和大规模的机器学习。

（2）numpy。numpy 是 Python 的一个数值计算扩展，它提供了快速、节省内存的数组（称为 ndarray），以及用于数学和科学编程的常用函数。

(3) pillow。pillow 是 Python 的一个图像处理库，可以用来打开、操作和保存不同格式的图像文件。

(4) scipy。scipy 是一个用于 Python 的数学、科学和工程库，它提供了许多数学、统计、科学和工程方面的工具。

(5) tqdm。tqdm 是一个快速、可扩展的 Python 进度条库，它可以在长循环中添加一个进度提示，让用户知道程序的运行进度。

在安装这些依赖项之前，用户需要确保计算机中已经安装了 Python，并且可以通过命令行执行 Python 命令。用户可以使用 pip（Python 的包管理器）来安装这些依赖项，具体安装命令为：
`pip install torch numpy pillow scipy tqdm`。

当然，用户也可以使用由 B 站大咖秋叶 aaaki 分享的“秋叶整合包”，一键实现 Stable Diffusion 的本地部署，只需运行“启动器运行依赖-dotnet-6.0.11.exe”安装程序，然后单击“安装”按钮即可，如图 1-10 所示。执行操作后，等待出现“控制台”窗口，不必在意“控制台”窗口中的内容，保持其处于打开状态即可。稍待片刻，将会出现一个浏览器窗口，表示 Stable Diffusion 的基本软件已经安装完毕。



图 1-10 单击“安装”按钮

专家指点

在安装 Stable Diffusion 的过程中，用户还要注意以下事项。

(1) 由于 Stable Diffusion 是一个复杂的模型库，因此安装和运行时可能需要较高的系统资源，如内存、显存和存储空间等，用户需要确保计算机硬件配置满足要求。

(2) 关闭其他可能影响 Stable Diffusion 安装的程序或进程。

(3) 理论上来说，4GB 显存的 N 卡甚至仅用 CPU（Central Processing Unit，中央处理器）都可以安装和运行 Stable Diffusion，但出图速度极慢，不推荐。

(4) Stable Diffusion 的安装目录尽可能不要放在 C 盘，同时安装位置所在的磁盘要留出足够的空间，建议至少 100GB。



1.2.3 云端：在云平台上部署 Stable Diffusion

随着云计算技术的发展，将 Stable Diffusion 部署到云端成为了可能，使更多的人能够享受到 AI 绘画工具带来的便利。用户可以在飞桨、阿里云、腾讯云、Colab 云等常用的云平台上部署 Stable Diffusion，这样只需在云端输入自己的文本描述，即可得到 AI 生成的图像。

专家指点

飞桨（PaddlePaddle）是一个集深度学习核心训练、推理框架、基础模型库、端到端开发套件以及大量的工具组件于一体，由百度研发的产业级深度学习平台，具有自主研发、功能丰富、开源开放的特点。在飞桨上租用 Stable Diffusion 的运行环境时，需要计算能力的支持。其中，GPU（Graphics Processing Unit）最好的 V100 四卡需要 8 点算力/小时，建议正常部署和运行 Stable Diffusion 时可以考虑使用 1 点算力/小时或 4 点算力/小时。

V100 显卡是 NVIDIA 推出的一款高端专业显卡，被广泛应用于人工智能、深度学习、虚拟现实等领域。V100 四卡是指在一台计算机上安装了 4 块 V100 显卡，每块显卡拥有 32GB HBM2（High-Bandwidth Memory 2，第二代高速缓存内存）内存和 5120 颗 CUDA（Compute Unified Device Architecture，统一计算设备架构）核心，能够提供强大的 GPU 计算能力，适用于 Stable Diffusion 等大规模的深度学习训练和推理任务。

例如，Colab 是谷歌推出的一款在线工作平台，可以让用户在浏览器中编写和执行 Python 脚本，最重要的是，它提供了免费的 GPU 来加速深度学习模型的训练。用户可以先启动 Colab Notebook 文件，进入 Colab 页面，再执行“代码执行程序”|“更改运行时类型”命令，如图 1-11 所示。



图 1-11 执行“更改运行时类型”命令

执行以上操作后，弹出“更改运行时类型”对话框，确保“硬件加速器”为 GPU，单击“保存”按钮即可，如图 1-12 所示。



图 1-12 单击“保存”按钮

接下来按图 1-13 所示的序号单击“运行”按钮 依次运行相应代码，每项代码执行完成后会显示“Done（结束）”信息，然后继续执行下一项代码。

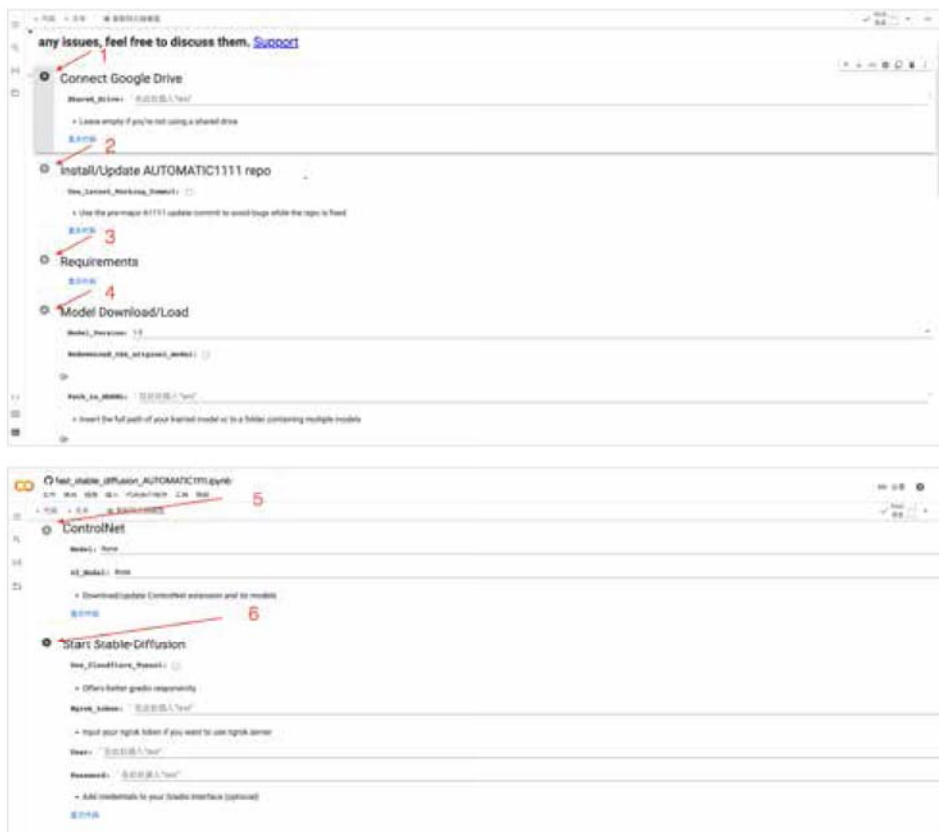


图 1-13 依次运行相应代码

全部代码正常执行完成后，在 Start Stable-Diffusion 下方会出现访问 WebUI 的链接，如图 1-14 所示。单击该链接，即可成功启动 Stable Diffusion。



图 1-14 出现访问 WebUI 的链接

1.2.4 运行：快速一键启动 Stable Diffusion

运行 Stable Diffusion 的方式取决于用户使用的具体软件版本和安装方式。下面以“秋叶整合包”为例，介绍一键启动 Stable Diffusion 的操作方法。

STEP 01 打开 Stable Diffusion 安装文件所在目录，进入 sd-webui-aki-v4 文件夹，找到并双击“A 启动器.exe”图标，如图 1-15 所示。



扫码看视频

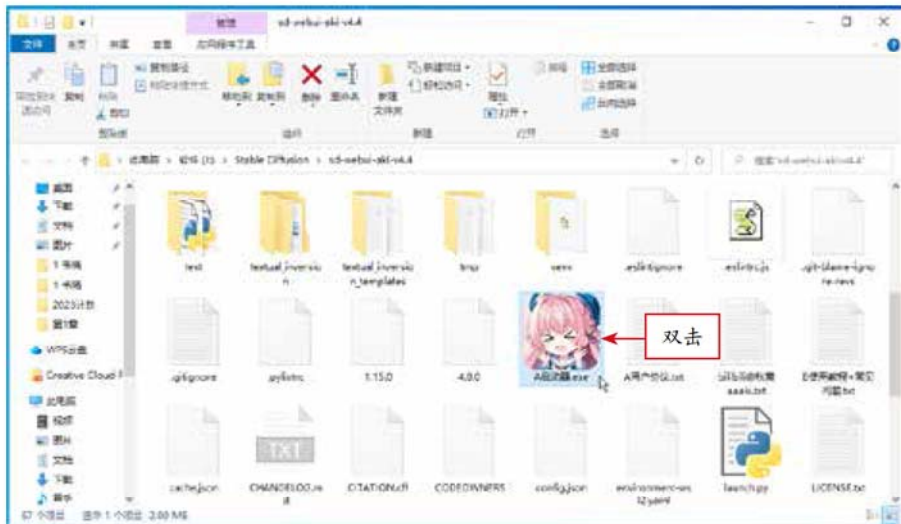


图 1-15 双击“A 启动器.exe”图标

专家指点

如果用户安装的是原版 Stable Diffusion，可以在系统中按 Win + R 组合键，执行 cmd 命令，或者在“开始”菜单的“Windows 系统”列表框中选择“命令提示符”选项，即可打开“命令提示符”窗口。在命令行中进入 Stable Diffusion 程序包的目录，执行以下命令：`python run_diffusion.py --config_file=config.yaml`，也可运行程序。

STEP 02 执行上一步操作后，即可打开“绘世”启动器程序，在主界面中单击“一键启动”按钮，如图 1-16 所示。



图 1-16 单击“一键启动”按钮

STEP 03 执行上一步操作后，即可打开“控制台”窗口，让它运行一会儿，耐心等待命令执行完成，如图 1-17 所示。

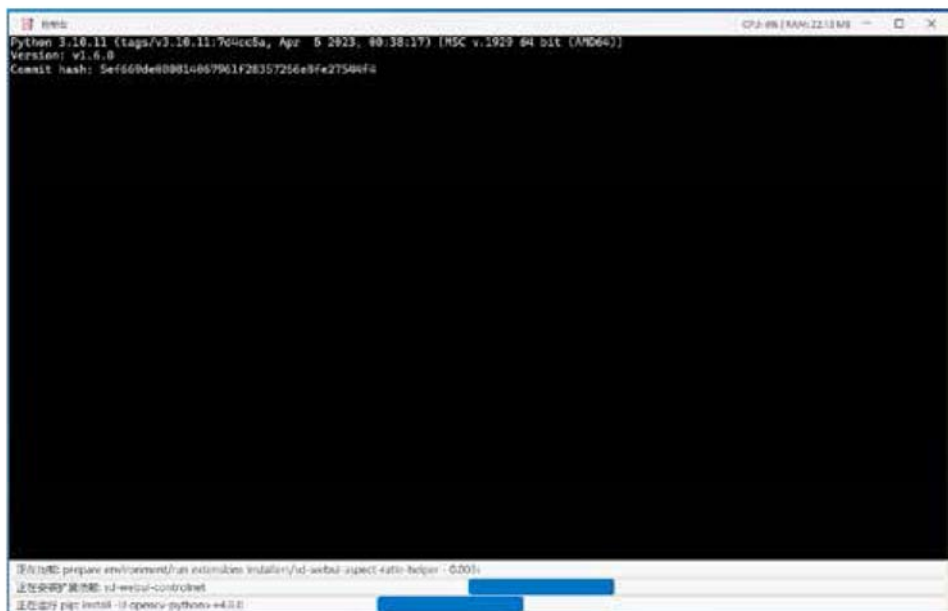


图 1-17 “控制台”窗口

STEP 04 稍等片刻，即可在浏览器中自动打开 Stable Diffusion 的 WebUI 页面，如图 1-18 所示。另外，用户也可以在“控制台”窗口中找到 Stable Diffusion 的运行链接，即 URL（Uniform Resource Locator，统一资源定位系统）后面的 IP（Internet Protocol，互联网协议）地址，将其复制到浏览器窗口中打开即可。

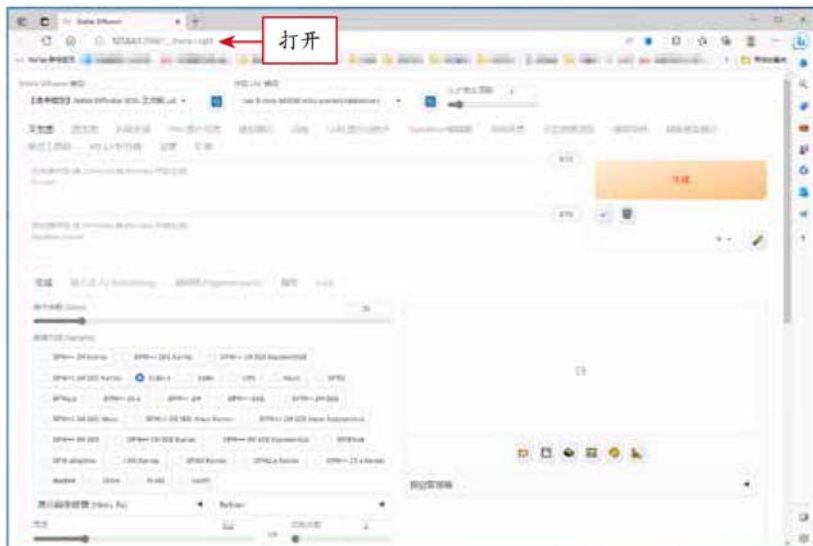


图 1-18 打开 Stable Diffusion 的 WebUI 页面

1.2.5 页面：看懂 Stable Diffusion 的 WebUI

简单来说，Stable Diffusion 的 WebUI 页面就像一间装满了先进绘画工具的工作室，用户可以在这里尽情发挥自己的创作灵感，创造出一件件令人惊艳的艺术作品。图 1-19 所示为 Stable Diffusion 的 WebUI 页面基本布局。

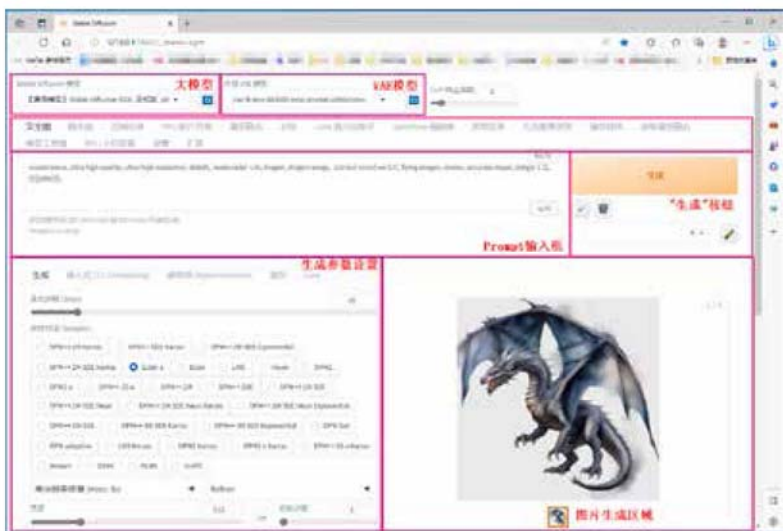


图 1-19 Stable Diffusion 的 WebUI 页面基本布局

其中，“大模型”可以理解为你给 Stable Diffusion 学习的数据包，只有给它学习过的内容，它才能够根据提示词画出来。每个大模型都有其独有的特点和适用场景，用户可以根据自己的需求和实际情况进行选择。“图片生成区域”用于显示生成的图片，在此处可以看到生成过程的每一步迭代图像。其他区域将在后面章节进行具体介绍，这里不再赘述。





1.2.6 出图：用 Stable Diffusion 绘制第一张效果图片

使用 Stable Diffusion 可以非常轻松地进行 AI 绘画，只要输入一个文本描述，它就可以在几秒之内为我们生成一张精美的图片。下面通过一个简单的案例，向大家展示如何使用 Stable Diffusion 快速绘制出一张你喜欢的图片，图片效果如图 1-20 所示。



扫码看视频



图 1-20 图片效果展示

下面介绍使用 Stable Diffusion 绘制第一张效果图片的操作方法。

STEP 01 进入 CIVITAI（简称 C 站）主页，在 Images（图片）页面中找到一张喜欢的图片，单击图片右下角的  按钮，如图 1-21 所示。

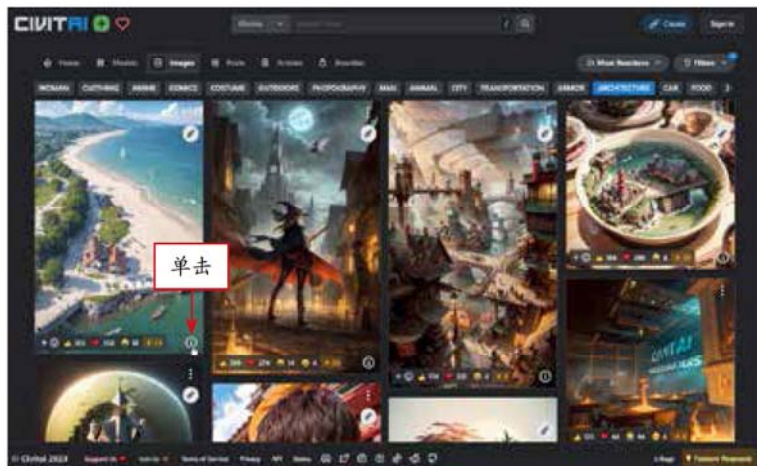


图 1-21 单击图片右下角的  按钮

专家指点

CIVITAI 是一家专注于 AI 生成内容的创业公司，旗下自主研发的 Diffusion 模型可以进行多模态的图像、视频等内容的生成。除了公共模型外，CIVITAI 还支持用户上传数据进行模型微调和优化，以提升图像的生成质量。

STEP 02 执行上一步操作后，弹出一个包含图片信息的面板，单击 Prompt 右侧的 Copy prompt（复制提示词）按钮，如图 1-22 所示，即可复制正向提示词。



图 1-22 单击 Copy prompt 按钮

STEP 03 将复制的提示词输入 Stable Diffusion 的“正向提示词”输入框中，使用同样的操作方法，复制 Negative prompt（否定提示）并输入 Stable Diffusion 的“反向提示词”输入框中，同时根据图片信息面板中的生成数据对“文生图”的相应参数进行设置，如图 1-23 所示。

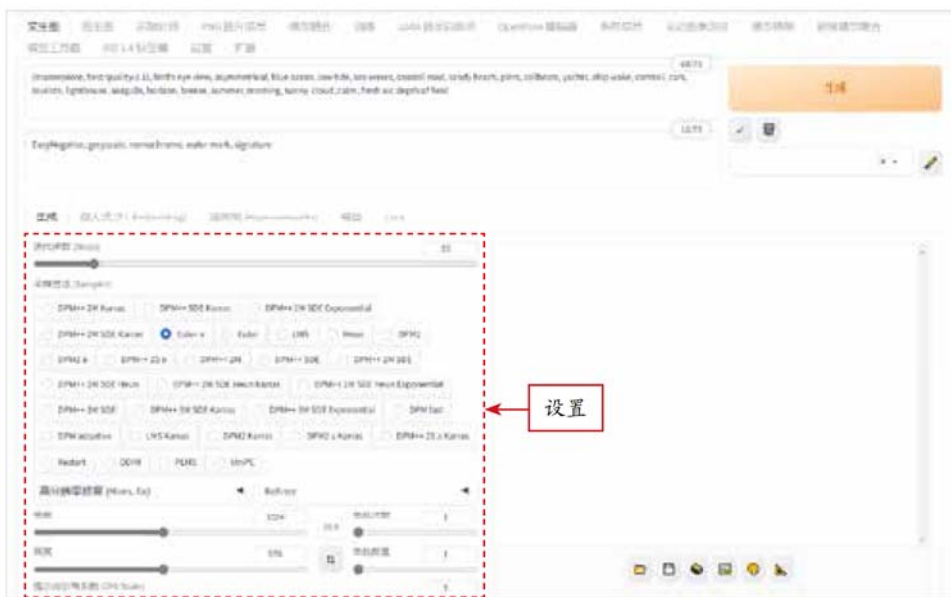


图 1-23 输入提示词并设置“文生图”参数

STEP 04 单击“生成”按钮，即可快速生成相应的图像，图片效果如图 1-20 所示。由于使用了不同的主模型，因此出图效果会有些差异。



第2章

生成参数：提升 AI 绘画的效果精美度



章前知识导读

Stable Diffusion 具有强大的图像生成能力，让更多的人对这个领域充满了无限遐想。本章主要介绍文生图和图生图时的基本生成参数设置技巧，帮助大家提升 Stable Diffusion 的绘画效果精美度。



新手重点索引

-  以文生图：通过文本描述生成图像
-  以图生图：通过参考图片生成图像



效果图片欣赏



2.1 以文生图：通过文本描述生成图像



Stable Diffusion 作为一款强大的 AI 绘画工具，可以通过文本描述生成各种图像，但是其参数设置比较复杂，对新手来说不容易上手。如何快速了解和掌握 Stable Diffusion 的基本参数，使生成图片更符合预期呢？本节将带你快速了解 Stable Diffusion 文生图中各项关键参数的作用，并掌握相关的设置方法。

2.1.1 参数 1：迭代步数

迭代步数（Steps）是指输出画面需要的步数，其作用可以理解为“控制生成图像的精细程度”，Steps 越高生成的图像细节越丰富、精细。不过，增加 Steps 的同时也会增加每张图片的生成时间，减少 Steps 则可以加快图片生成速度。



扫码看视频

Stable Diffusion 的采样迭代步数采用的是分步渲染的方法。分步渲染是指在生成同一张图片时，分为多个阶段使用不同的文字提示进行渲染。在整张图片基本成型后，再通过添加文本描述的方式来渲染和优化细节。这种分步渲染的方法，需要对照明、场景等采用一定的美术处理技巧，才能生成逼真的图像效果。

Stable Diffusion 的每一次迭代都是在上一次生成的基础上进行渲染。一般来说，Steps 保持在 18 ~ 30 之间，即可生成较好的图像效果。如果 Steps 设置得过低，可能会导致图像生成不完整，关键细节无法呈现；而 Steps 设置得过高则会大幅度增加生成时间，但对图像效果提升的边际效果较小，仅对细节进行轻微优化，因此可能会得不偿失。图 2-1 所示为不同迭代步数生成的图像效果对比。

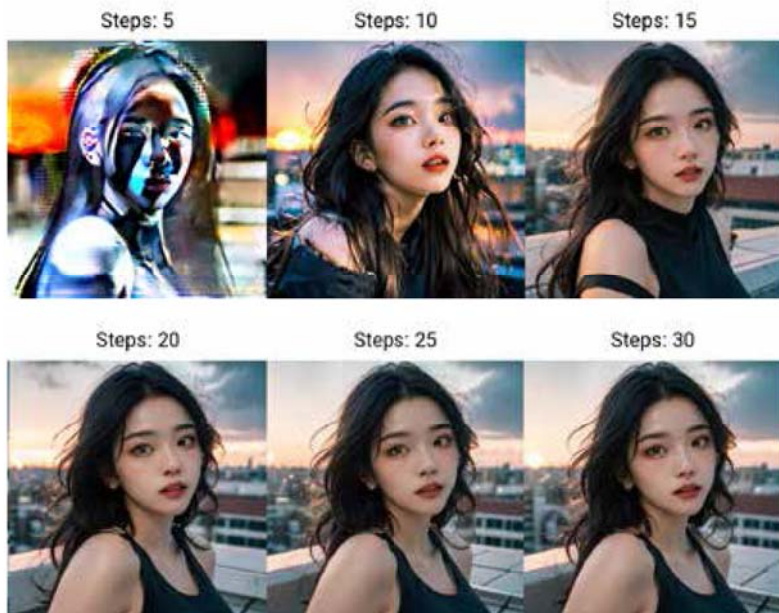


图 2-1 不同迭代步数生成的图像效果对比

下面介绍设置迭代步数的操作方法。

STEP 01 在 Stable Diffusion 的“文生图”页面中输入相应的提示词，将“迭代步数”设置为 5，单击“生成”按钮，可以看到生成的人物图像效果是非常模糊的，且面部不够完整，如图 2-2 所示。

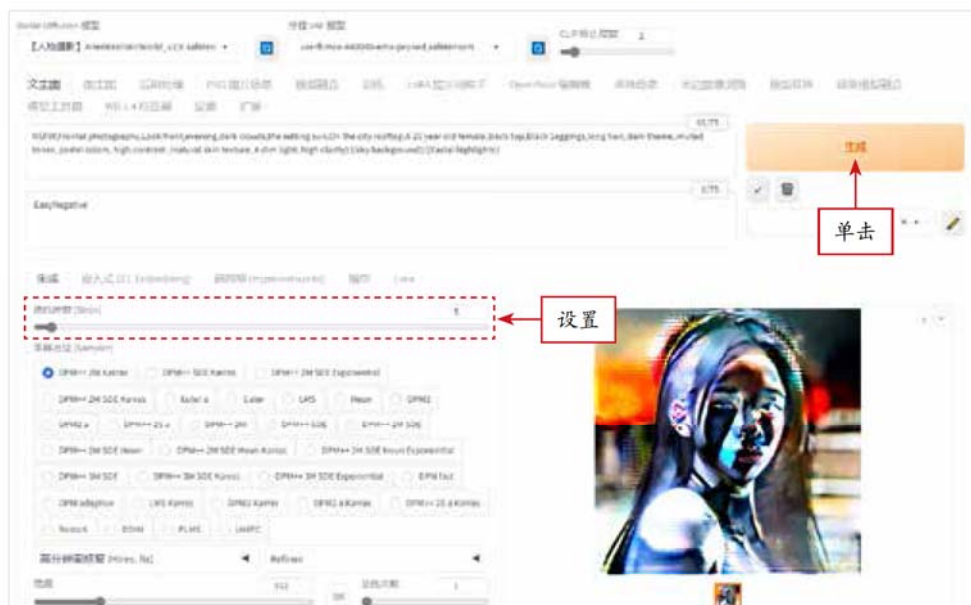


图 2-2 “迭代步数”设置为 5 生成的图像效果

STEP 02 锁定图 2-2 中的“随机数种子”数值，将“迭代步数”设置为 30，其他参数保持不变，单击“生成”按钮，可以看到生成的图像效果是非常清晰的，而且画面效果是完整的，如图 2-3 所示。

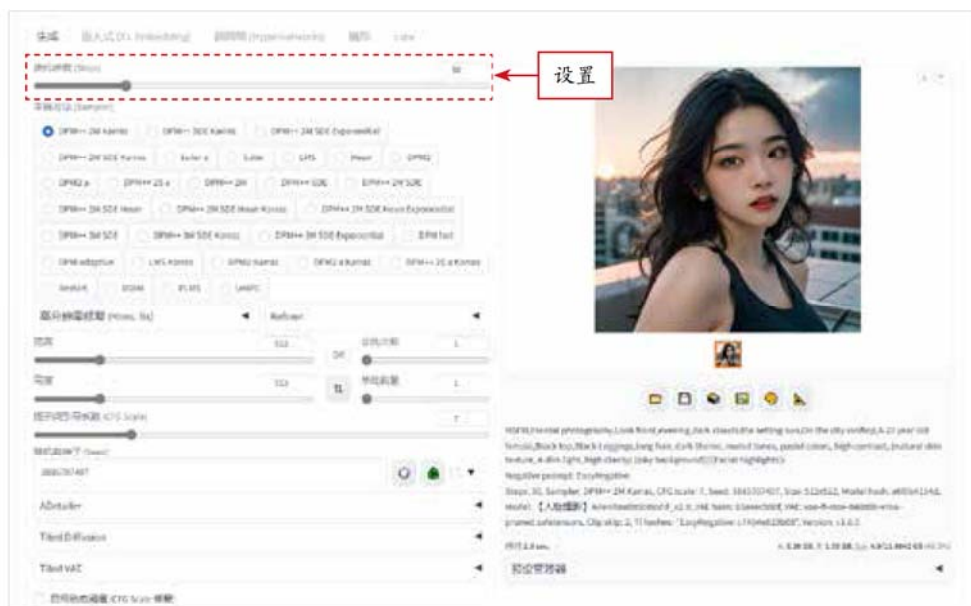


图 2-3 “迭代步数”设置为 30 生成的图像效果

扫码看视频

- 速度快: Euler 系列、LMS 系列、DPM++ 2M、DPM fast、DPM++ 2M Karras、DDIM 系列。
- 质量高: Heun、PLMS、DPM++ 系列。
- 标签 (Tag) 利用率高: DPM2 系列、Euler 系列。
- 动画风: LMS 系列、UniPC。
- 写实风: DPM2 系列、Euler 系列、DPM++ 系列。

下面介绍设置“采样方法”的操作方法。

[illegible]

22

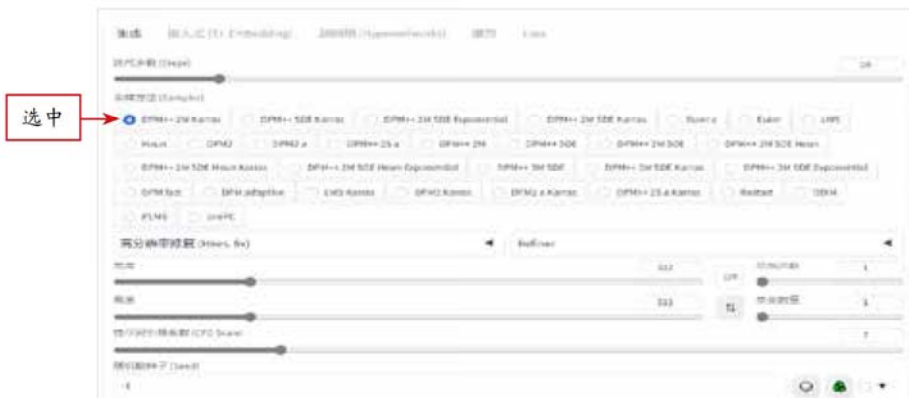


图 2-6 选中 DPM++ 2M Karras 单选按钮

STEP 03 单击两次“生成”按钮，即可通过 DPM++ 2M Karras 的采样方法生成两张图片，图片效果见图 2-4。

专家指点

“采样方法”技术为 Stable Diffusion 等生成模型提供了更加真实、可靠的随机采样能力，从而可以生成更加逼真的图像效果。“采样方法”又称为采样器，除 DPM++ 2M Karras 外，常用的“采样方法”还有 3 种，分别为 Euler a、DPM++ 2S a Karras 和 DDIM。

- Euler a 的采样生成速度最快，但在生成高细节图像并增加迭代步数时，会产生不可控的变化，如人物脸部扭曲、细节扭曲等。Euler a 采样器适合生成图标（icon）、二次元图像或小场景的画面。
- DPM++ 2S a Karras 采样方法可以生成高质量图像，适合生成写实人像或刻画复杂场景，而且 Steps（即迭代步数）越高，细节刻画效果越好。
- DDIM 比其他采样方法具有更高的效率，而且随着 Steps 的增加，可以叠加生成更多的细节。

2.1.3 参数 3：高分辨率修复

高分辨率修复（Hires.fix）功能首先以较小的分辨率生成初步图像，接着放大图像，然后在不更改构图的情况下改进其中的细节。Stable Diffusion 会依据用户设置的“宽度”和“高度”尺寸，并按照“放大倍率”进行等比例放大。



扫码看视频

专家指点

在“高分辨率修复”选项组中，下面几个选项的设置非常关键。

（1）放大倍数。放大倍数是指图像被放大的比例。需要注意的是，当图像被放大到一定程度后，可能会出现质量问题。

（2）高分迭代步数。高分迭代步数是指在提高图像分辨率时，算法需要迭代的次数。如果将其设置为 0，则将使用与 Steps 相同的值。通常情况下，建议将高分迭代步数设置为 0 或小于 Steps 的值。

（3）重绘幅度。重绘幅度是指在进行图像生成时，需要添加的噪声程度。该值为 0 表示完全不加噪声，即不进行任何重绘操作；值为 1 则表示整个图像将被随机噪声完全覆盖，生成与原图完全不相关的图像。通常在重绘幅度设置为 0.5 时，会对图像的颜色和光影产生显著影响；而在重绘幅度设置为 0.75 时，甚至会改变图像的结构和人物姿态。



对于显存较小的显卡来说，可以通过使用高分辨率修复功能，把“宽度”和“高度”尺寸设置得小一些，如默认分辨率为 512×512 的图片，然后将其“放大倍数”参数设置为 2，Stable Diffusion 就会生成分辨率为 1024×1024 的图片，且不会占用过多的显存，图片效果如图 2-7 所示。



图 2-7 图片效果展示

下面介绍设置“高分辨率修复”的操作方法。

STEP 01 在 Stable Diffusion 的“文生图”页面中输入相应的提示词，单击“生成”按钮，生成一张分辨率为 512×512 的图片，效果如图 2-8 所示。

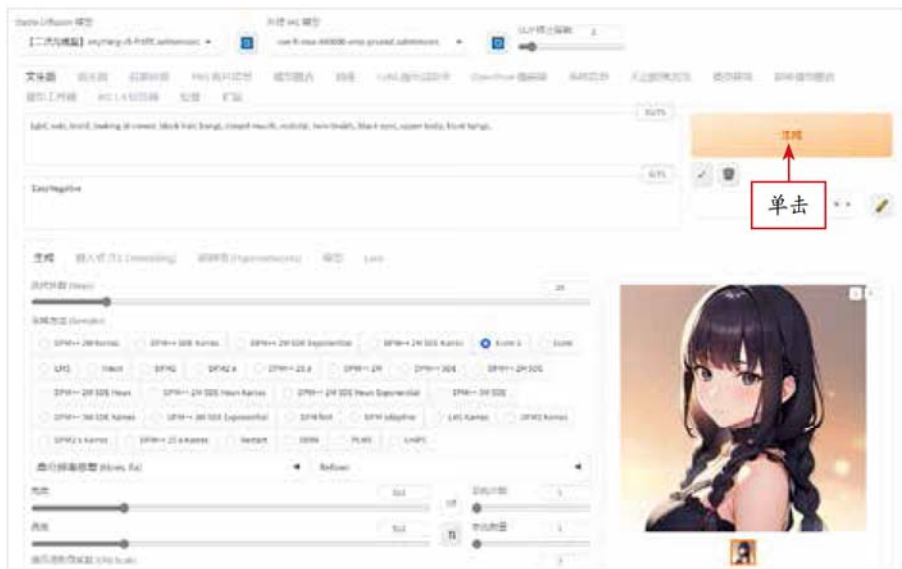


图 2-8 生成一张分辨率为 512×512 的图片效果

STEP 02 展开“高分辨率修复”选项组，设置“放大算法”参数为 R-ESRGAN 4x+ Anime6B，如图 2-9 所示，这是一种基于超分辨率技术的图像增强算法，主要用于提高动漫图像的质量和清晰度。

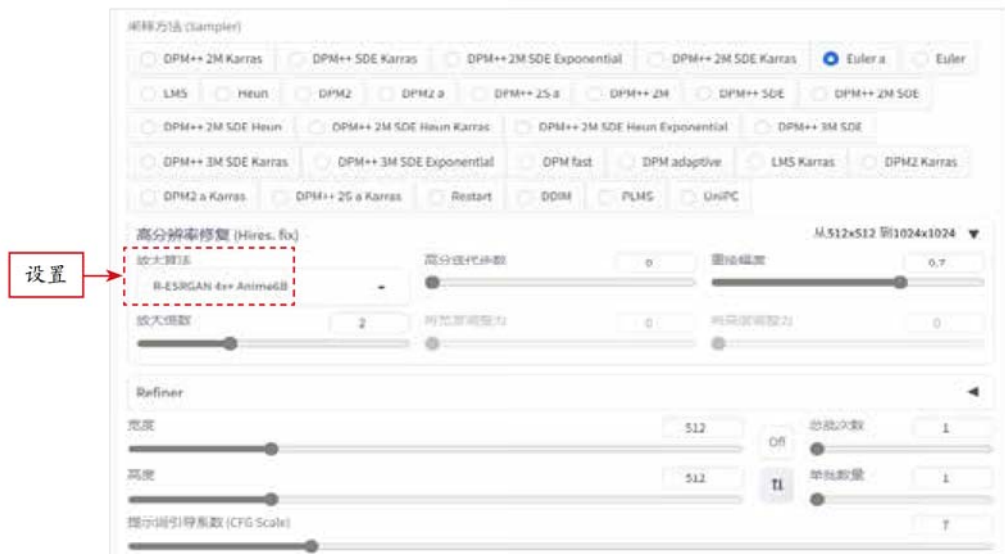


图 2-9 设置“放大算法”参数

STEP 03 其他参数保持默认设置，单击“生成”按钮，即可生成一张分辨率为 1024×1024 的图片，效果如图 2-7 右图所示。

2.1.4 参数 4：图片尺寸

图片尺寸即分辨率，指的是图片宽度和高度的像素数量，它决定了数字图像的细节再现能力和质量。例如，分辨率为 768×512 的图像在细节表现方面具有较高的质量，可以提供更好的视觉效果，图片如图 2-10 所示。



扫码看视频



图 2-10 图片效果展示



下面介绍设置图片尺寸的操作方法。

STEP 01 进入“文生图”页面，选择一个写实类的大模型，输入相应的提示词，指定生成图像的画面内容，如图 2-11 所示。



图 2-11 输入相应的提示词

STEP 02 在“高分辨率修复”选项组中设置“宽度”为 768、“高度”为 512，表示生成分辨率为 768×512 的图像，其他设置如图 2-12 所示。

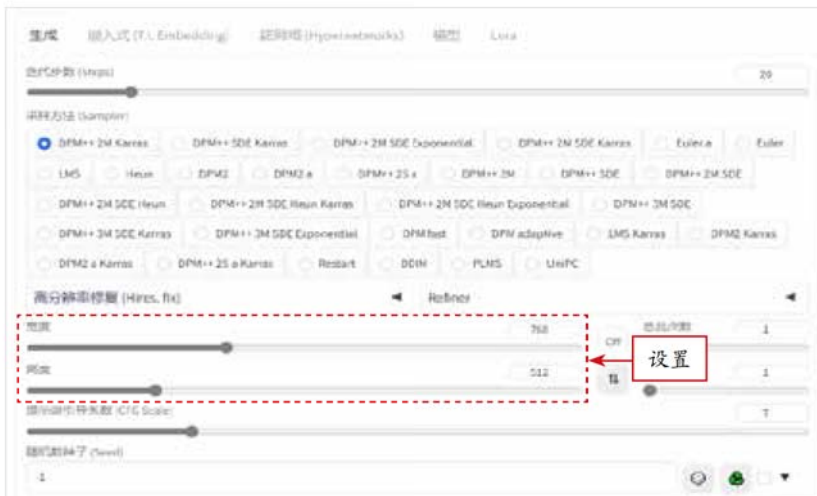


图 2-12 设置相应参数

专家指点

通常情况下，8GB 显存的显卡，图片尺寸应尽量设置为 512×512 的分辨率；否则太大的画面无法描绘清晰，太大的画面则容易“爆显存”。8GB 显存以上的显卡则可以适当调高分辨率。“爆显存”是指计算机的画面数据量超过了显存的容量，导致画面出现错误或者计算机的帧数骤降，甚至出现系统崩溃等情况。

图片尺寸需要和提示词所生成的画面效果相匹配，如分辨率设置为 512×512 时，人物大概率会出现大头照。用户也可以固定一个图片尺寸的值，并将另一个值调高，但固定值要保持在 512～768 之间。

STEP 03 单击“生成”按钮，即可生成相应尺寸的横图，图片效果见图 2-10。

2.1.5 参数 5：总批次数与单批数量

“总批次数”就是在绘制多张图片时，显卡是按照一张接一张的顺序往下画；“单批数量”就是在显卡同时绘制多张图片时，绘画效果通常比较差。例如，在Stable Diffusion中使用相同的提示词和生成参数，可以一次生成6张不同的图片，效果如图2-13所示。



扫码看视频



图 2-13 一次生成6张不同图片效果展示

下面介绍设置“总批次数”与“单批数量”的操作方法。

STEP 01 进入“文生图”页面，选择一个写实类的大模型，输入相应的提示词，指定生成图像的画面内容，如图2-14所示。

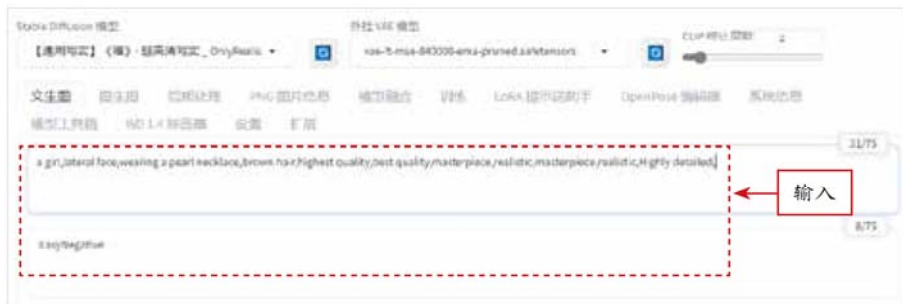


图 2-14 输入相应的提示词

STEP 02 在“高分辨率修复”选项组中设置“总批次数”参数为6，可以理解为一次循环生成6张图片，其他设置如图2-15所示。

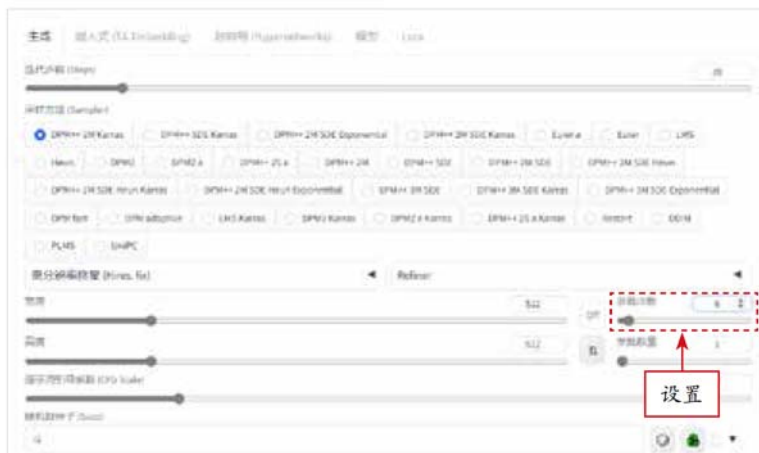


图 2-15 设置“总批次数”参数

STEP 03 单击“生成”按钮，即可同时生成 6 张图片，且每张图片的差异性都比较大，图片效果如图 2-16 所示。



图 2-16 生成 6 张图片效果

STEP 04 在保持提示词和其他参数不变时，设置“总批次数”参数为 2、“单批数量”参数为 3，可以理解为一个批次里一次生成 3 张图片，共生成 2 个批次，如图 2-17 所示。

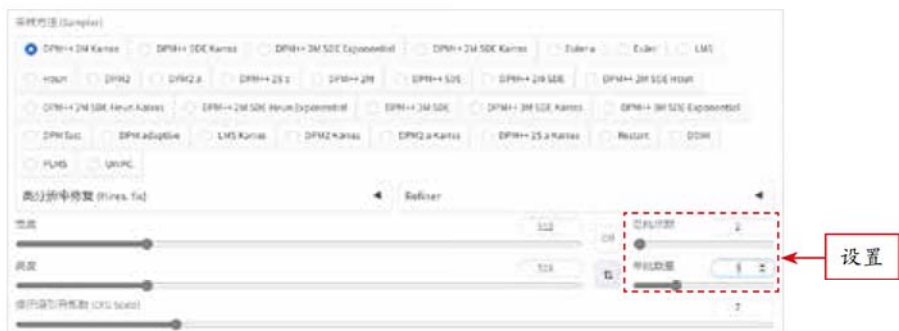


图 2-17 设置“总批次数”和“单批数量”参数

STEP 05 单击“生成”按钮，即可生成 6 张图片，且同批次中的图片差异较小，同时出图效果也比较差，效果如图 2-13 所示。

专家指点

需要注意的是，Stable Diffusion 默认的出图效果是随机的，又称为“抽卡”，也就是说我们需要不断地生成新图，从中挑选出一张效果最好的图片。

如果用户的计算机显卡配置比较高，可以使用“单批数量”的方式出图，速度会更快，同时也能保证一定的画面效果；否则，就加大“总批次数”，每一批只生成一张图片，这样在硬件资源有限的情况下，可以让 AI 尽量画好每张图。

2.1.6 参数 6：提示词引导系数

“提示词引导系数（CFG Scale）”主要用来调节提示词对 AI 绘画效果的引导程度，参数范围为 0 ~ 30，数值较大时绘制的图片会更加符合提示词的要求，效果对比如图 2-18 所示。



扫码看视频



图 2-18 数值大小不同时图片效果对比



下面介绍设置“提示词引导系数”的操作方法。

STEP 01 进入“文生图”页面，选择一个写实类的大模型，输入相应的提示词，指定生成图像的画面内容，如图 2-19 所示。



图 2-19 输入相应的提示词

STEP 02 在“高分辨率修复”选项组设置“提示词引导系数”参数为 2，表示提示词与绘画效果的关联性较低，其他设置如图 2-20 所示。

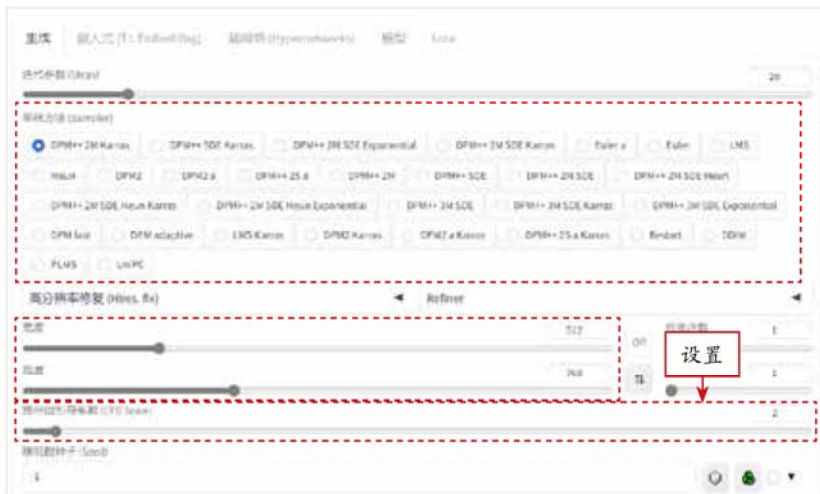


图 2-20 设置“提示词引导系数”参数

专家指点

“提示词引导系数”的参数值建议设置为 7 ~ 12，过小的参数值会导致图像的色彩饱和度和度降低；而过大的参数值则会产生粗糙的线条或过度锐化的图像细节，甚至可能导致图像严重失真。

STEP 03 单击“生成”按钮，即可生成相应的图像，且图像内容与提示词的关联性不大，效果如图 2-18 左图所示。

STEP 04 保持提示词和其他设置不变，设置“提示词引导系数”参数为 10，此时提示词与绘画效果的关联性较高，如图 2-21 所示。

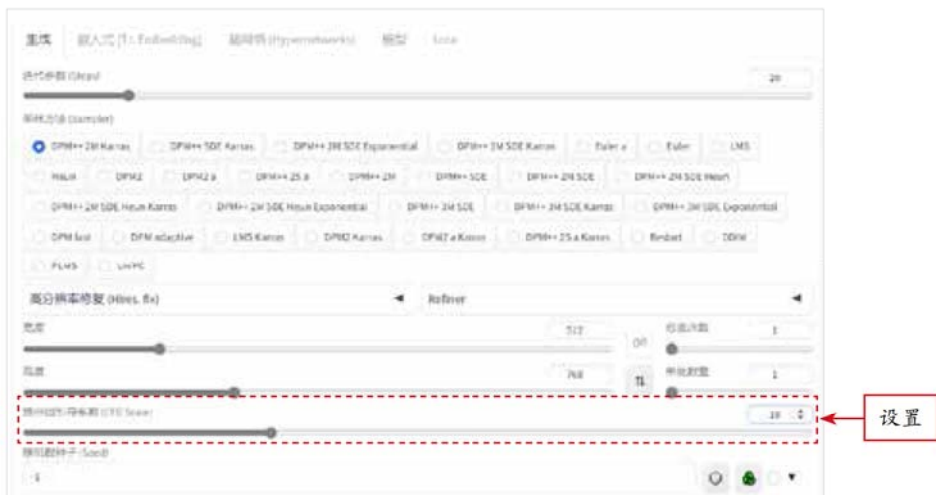


图 2-21 设置较高的“提示词引导系数”参数

STEP 05 单击“生成”按钮，即可生成相应的图像，且图像内容与提示词的关联性较大，画面的光影效果更突出、质量更高，效果如图 2-18 右图所示。

2.1.7 参数 7：随机数种子

在 Stable Diffusion 中，随机数种子（Seed，也称为随机种子或种子）可以理解为每个图像的唯一编码，能够帮助我们复制和调整生成的图像效果。当用户在绘图过程中发现有满意的图像时，就可以复制并锁定图像的“随机数种子”，让后面生成的图像更加符合自己的需求，图片效果如图 2-22 所示。



扫码看视频



图 2-22 图片效果展示



下面介绍设置“随机数种子”的操作方法。

STEP 01 进入“文生图”页面，选择一个二次元风格的大模型，输入相应的提示词，指定生成图像的画面内容，如图 2-23 所示。

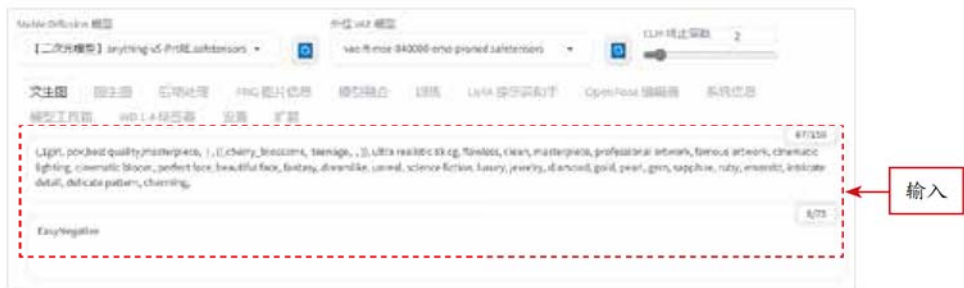


图 2-23 输入相应的提示词

STEP 02 在“生成”选项卡中，“随机数种子”的参数值默认为 -1，表示随机生成图像效果，其他设置如图 2-24 所示。

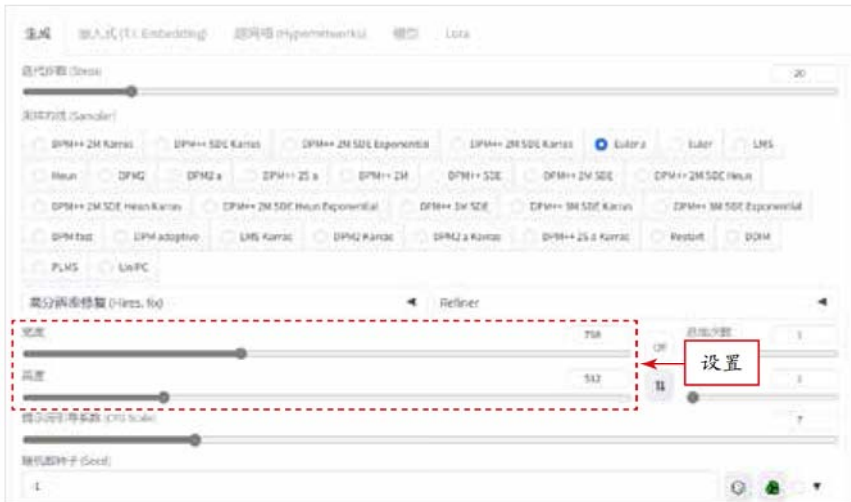


图 2-24 设置其他相应参数

专家指点

在 Stable Diffusion 中，“随机数种子”是通过一个 64 位的整数（指用 64 位二进制数表示的整数）来表示的。如果将这个整数作为输入值，AI 模型会生成一个对应的图像。如果多次使用相同的随机数种子，则 AI 模型会生成相同的图像。

在“随机数种子”文本框的右侧，单击  按钮，可以将参数值重置为 -1，则每次生成图像时都会使用一个新的随机数种子。如果复制图像的随机数种子值，并将其输入“随机数种子”文本框内，后续生成的图像将基本保持不变。

STEP 03 单击“生成”按钮，每次生成图像时都会随机生成一个新的种子，从而得到不同的结果，图片效果如图 2-25 所示。



图 2-25 “随机数种子 (Seed)” 参数设置为 -1 时生成的图像效果

STEP 04 在下方的图片信息中找到并复制 Seed 值，将其输入“随机数种子”文本框中，如图 2-26 所示。

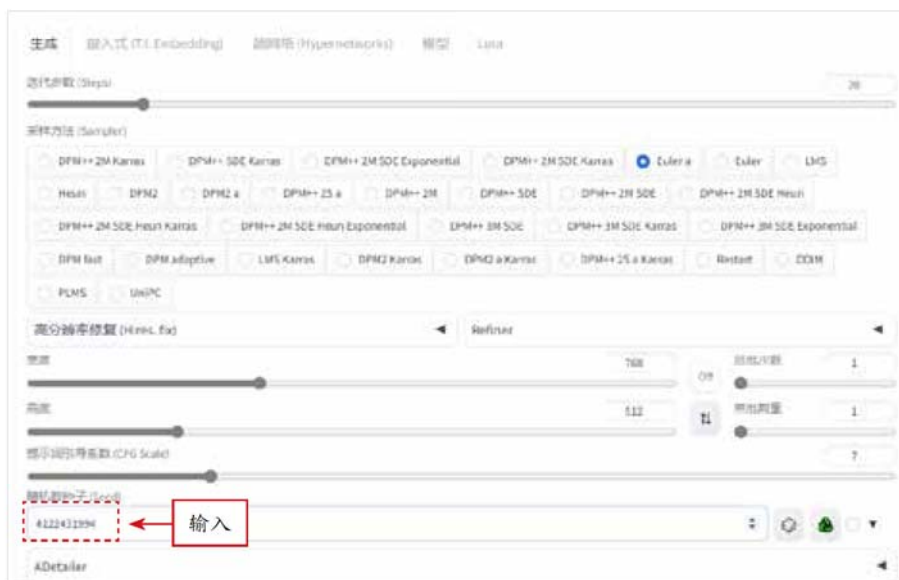


图 2-26 输入 Seed 值

STEP 05 单击“生成”按钮，则后续生成的图像将保持不变，每次得到的结果都会相同，效果如图 2-22 所示。



2.1.8 参数 8：变异随机种子

除了“随机数种子”值外，在 Stable Diffusion 中用户还可以使用“变异随机种子”（different random seed，简称 diff seed）值来控制出图效果。“变异随机种子”是指在生成图像的过程中，每次的扩散过程中都会使用不同的随机数种子，从而产生与原图不同的图像，可以将其理解为在原来的图片上进行叠加变化，图片效果如图 2-27 所示。



扫码看视频



图 2-27 图片效果展示

专家指点

当“变异随机种子”参数设置为 0 时，表示完全按照“随机数种子”的值生成新图像，也就是完全复制输入的原图像，即新图与原图完全相同。在这种情况下，无论输入什么样的图像，只要“随机数种子”相同，生成的图像结果就相同。

当“变异随机种子”参数设置为 1 时，表示完全按照“变异随机种子”的值生成新图像，也就是与输入的原图像有很大的差异，即新图与原图完全不相同。在这种情况下，每次输入相同的图像，都会得到不同的结果，因为每次都会生成新的“变异随机种子”。

下面介绍设置“变异随机种子”的操作方法。

STEP 01 在上一例效果的基础上，选中“随机数种子（Seed）”右侧的复选框，展开该选项组，可以看到“变异随机种子”的参数值默认为 -1，保持该参数值不变，将“变异强度”参数设置为 0.21，如图 2-28 所示。



图 2-28 设置“变异强度”参数

STEP 02 单击“生成”按钮，则后续生成的新图与原图比较接近，只有细微的差别，图片效果如图 2-29 所示。



图 2-29 生成的新图与原图比较接近

STEP 03 将“变异强度”设置为 0.5，其他参数保持不变，如图 2-30 所示。



图 2-30 设置“变异强度”参数

专家指点

“随机数种子”的尺寸通常很少用到，它的概念是尝试生成图像，与同一“随机数种子”在指定分辨率下生成的图像相似。例如，首先使用分辨率为 512×512 生成一张人物图片（将其称为图 1），人物的脸部可能会变形，俗称“脸崩”，这是因为在该分辨率下图片无法承受太多的人物细节。此时，可以再生成一张分辨率为 512×1024 的人物图片（将其称为图 2），并在图像信息中复制其 Seed 值。

接下来锁定图 1 的 Seed 值，并将图 2 的 Seed 值输入“变异随机种子”文本框内，设置“变异强度”为 0.5、“从宽度中调整种子”为 512、“从高度中调整种子”为 1024，单击“生成”按钮，生成相应的人物图片，此时人物的脸部和手部的变形程度稍微降低了。

对于使用低显存显卡的用户来说，这是一个比较实用的功能，可以用 512×512 的分辨率，高效率地生成高度为 1024 的人物全身图片。

STEP 04 单击“生成”按钮，则后续生成的新图与原图差异更大，效果如图 2-27 所示。



2.1.9 参数 9：变异强度

变异强度（diff intensity）表示原图与新图的差异程度。变异强度越大，则变异随机种子对图像的影响就越大，通常可以根据需要灵活调整生成的新图像与原图像之间的相似程度。利用随机数种子和变异随机种子的特点，可以通过控制变异强度，将不同的图片效果进行融合，原图与效果的对比如图 2-31 所示。



扫码看视频



图 2-31 原图与新图的效果对比

在该案例中，固定了图 2 的 Seed 值，而图 1 则作为影响图 2 的一个变量。从图 2-31 中的对比可以直观地看到，当“变异强度”设置为 0.2 时，图 2 带有一点点图 1 的风格；当“变异强度”设置为 0.8 时，图 2 几乎变成了图 1 的风格。

下面介绍设置变异强度的操作方法。

STEP 01 进入 Stable Diffusion 的“PNG 图片信息”页面，在“来源”选项组中单击“点击上传”链接，如图 2-32 所示。



图 2-32 单击“点击上传”链接

STEP 02 弹出“打开”对话框，选择相应的素材图像，单击“打开”按钮上传图像，将其命名为图 1，同时在右侧即可看到图像的提示词等生成参数，单击“发送到文生图”按钮，如图 2-33 所示。



图 2-33 单击“发送到文生图”按钮

专家指点

在“查看 PNG 图片信息”页面中，只有上传用 Stable Diffusion 生成的图片，页面右侧才会显示对应的生成参数信息。如果上传的图片不是由 Stable Diffusion 生成的，或者图片被重新编辑和保存过，可能会无法读取到对应的生成参数信息。

STEP 03 执行操作后，进入“文生图”页面，在“随机数种子”右侧单击按钮，重置“随机数种子”，如图 2-34 所示。



图 2-34 重置“随机数种子”



STEP 04 单击“生成”按钮，即可生成新的图像，将该图像命名为图 2，并复制 Seed 值，将其输入“随机数种子”文本框中，如图 2-35 所示。

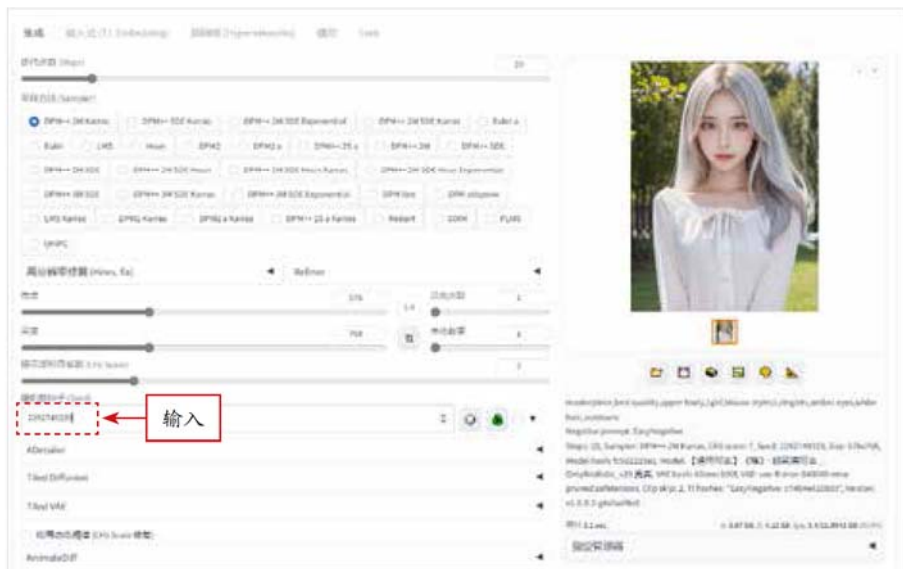


图 2-35 在“随机数种子”文本框内输入图 2 的 Seed 值

STEP 05 返回“PNG 图片信息”页面，复制图 1 的 Seed 值，将其输入“变异随机种子”文本框中，并将“变异强度”设置为 0.2，单击“生成”按钮，生成相应的图像效果，如图 2-36 所示。

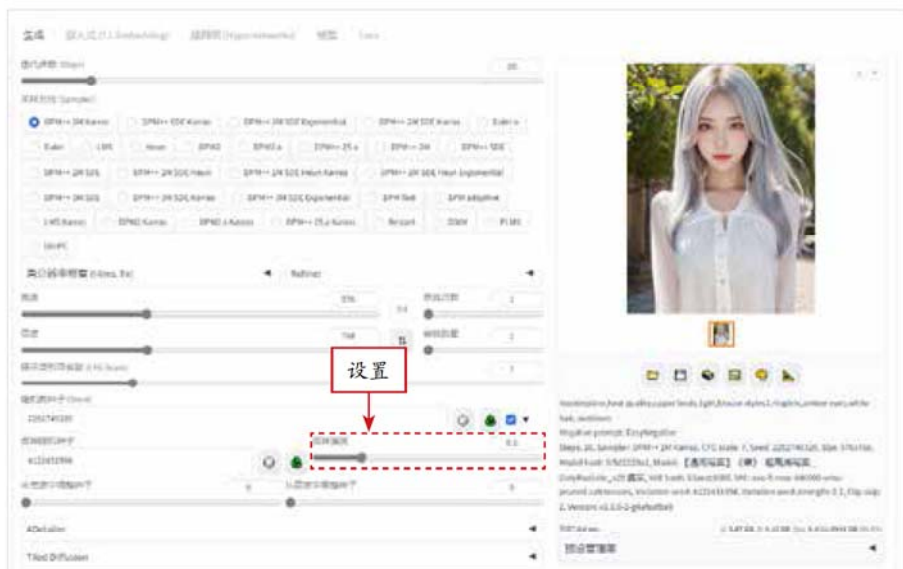


图 2-36 将“变异强度”设置为 0.2 时生成的图像效果

STEP 06 将“变异强度”设置为 0.8，再次单击“生成”按钮，生成相应的图像效果，如图 2-37 所示。

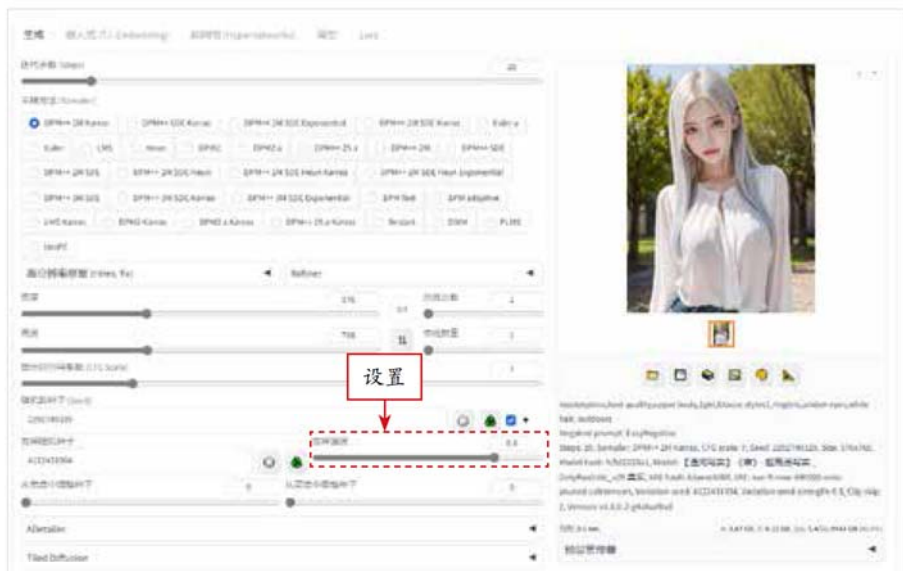


图 2-37 将“变异强度”设置为 0.8 时生成的图像效果

2.1.10 参数 10: X/Y/Z 图表

Stable Diffusion 中的 X/Y/Z 图表是一种用于可视化三维数据的图表，它由 3 个坐标轴组成，分别代表 3 个变量，这个工具的作用就是可以同时查看最多 3 个变量对于出图结果的影响。具体而言，X、Y 和 Z 这 3 个坐标轴分别代表图像的不同生成参数。其中，X 轴和 Y 轴用于确定图像的行数和列数，而 Z 轴则用于确定批处理尺寸。通过在这 3 个坐标轴上设定不同的生成参数，可以将不同的生成参数组合起来生成多个图像网格。



扫码看视频

专家指点

通过 X/Y/Z 图表的对比，可以快速生成一张图片并观察不同生成参数组合下的效果，避免了频繁生成图像后再作对比。同时，所有生成的图像都将在同一界面上展示，以便用户更方便地比较和分析 AI 出图效果。

例如，利用 Stable Diffusion 的 X/Y/Z 图表工具，可以非常方便地对比不同采样方法和大模型的出图效果，如图 2-38 所示。

专家指点

需要注意的是，X/Y/Z 图表中只能显示英文参数，因此模型名中的中文字符会变成乱码。如果用户对此有讲究，可以将模型名改为纯英文，但修改时要谨慎，避免模型名变得混乱和难以理解。



图 2-38 出图效果展示

下面介绍设置 X/Y/Z 图表参数的操作方法。

STEP 01 在“文生图”页面中选择一个二次元风格的大模型，输入相应的提示词，单击“生成”按钮，生成一张卡通图片，效果如图 2-39 所示。

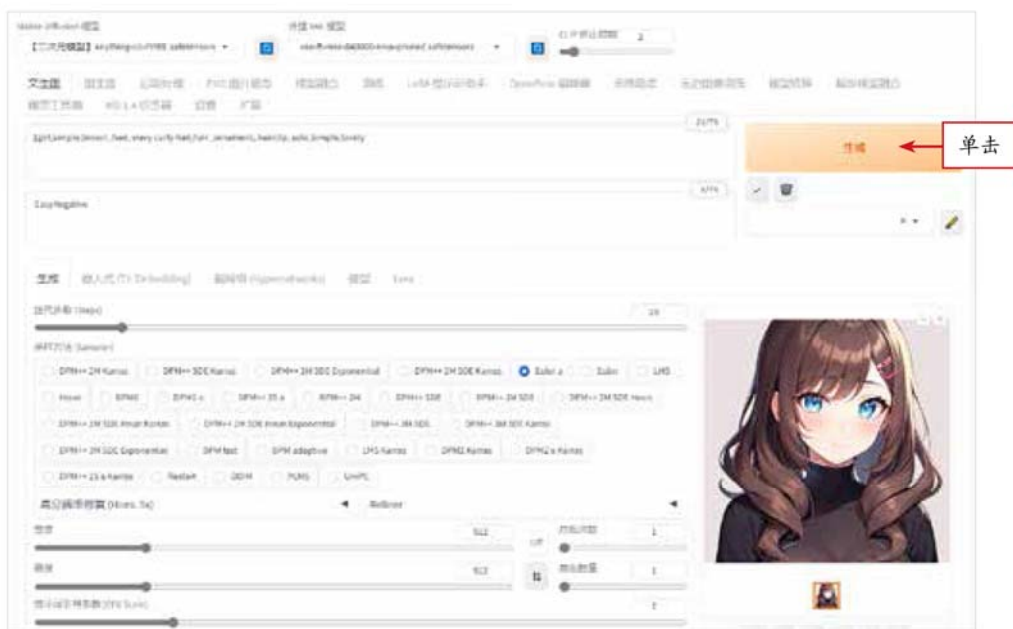


图 2-39 生成一张卡通图片

STEP 02 锁定该图片的 Seed 值，在页面下方的“脚本”列表框中选择“X/Y/Z 图表”选项，如图 2-40 所示。

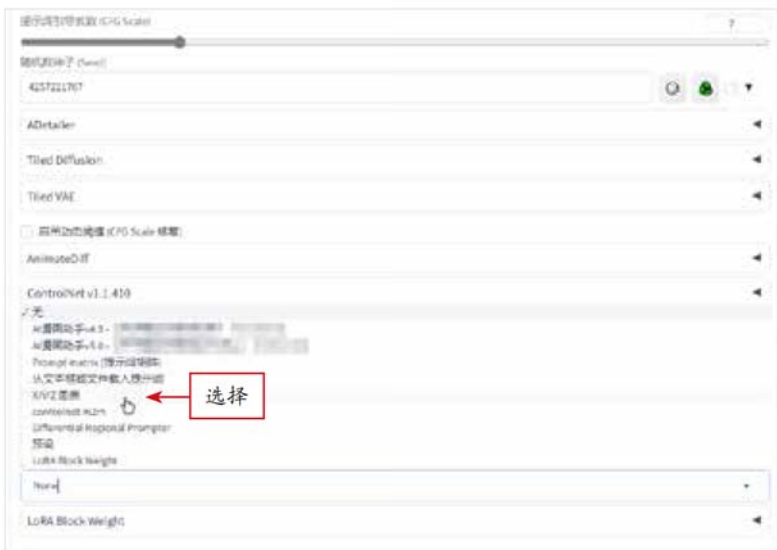


图 2-40 选择“X/Y/Z 图表”选项

STEP 03 执行操作后，即可展开 X/Y/Z plot（图表）选项组，单击“X 轴类型”下方的下拉按钮 ▾，如图 2-41 所示。



图 2-41 单击“X 轴类型”下方的下拉按钮

STEP 04 在弹出的下拉列表中选择“采样方法”选项，即可将“X 轴类型”设置为 Sampler，单击右侧的“X 轴值”按钮，在弹出的列表框中选择想要对比的采样方法，此处选择多种采样方法，如图 2-42 所示。



图 2-42 选择多种采样方法



STEP 05 单击“生成”按钮，即可非常清晰地对比同一个提示词下，6种不同采样方法分别生成的图像，效果如图 2-43 所示。



图 2-43 对比 6 种不同采样方法分别生成的图像效果

STEP 06 在 X/Y/Z plot 选项组中，单击“Y 轴类型”下方的下拉按钮 ▾，在弹出的下拉列表中选择“模型名”选项，如图 2-44 所示。

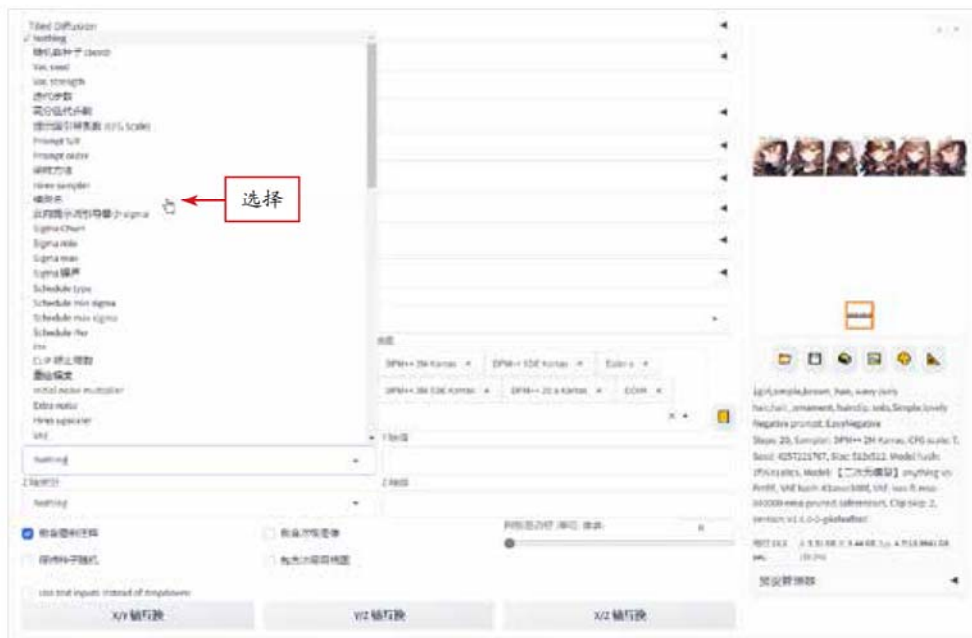


图 2-44 选择“模型名”选项

专家指点

在 X/Y/Z plot 选项组中，通过不同的轴互换操作，可以更加灵活地呈现数据，帮助用户更好地理解不同变量之间的关系，相关技巧如下。

- 单击“X/Y 轴互换”按钮，X 轴和 Y 轴将会互换，即原来在 X 轴上的变量会移动到 Y 轴上，原来在 Y 轴上的变量会移动到 X 轴上。这样可以将两个变量的关系以相反的方向呈现在图表上，方便进行对比和分析。
- 单击“Y/Z 轴互换”按钮，Y 轴和 Z 轴将会互换，即原来在 Y 轴上的变量会移动到 Z 轴上，原来在 Z 轴上的变量会移动到 Y 轴上。这样可以将第 3 个变量从另一个维度中展示出来，方便观察和分析 3 个变量之间的关系。
- 单击“X/Z 轴互换”按钮，X 轴和 Z 轴将会互换，即原来在 X 轴上的变量会移动到 Z 轴上，原来在 Z 轴上的变量会移动到 X 轴上。同样地，这样可以将第 3 个变量从另一个维度中展示出来，方便观察和分析另外两个变量之间的关系。

STEP 07 执行操作后，单击“Y 轴值”下方的下拉按钮 ▾，在弹出的下拉列表中选择需要对比的模型名，选择相应的模型名，同时将“X 轴值”列表框中的一些采样方法适当删除，如图 2-45 所示。



图 2-45 选择相应的模型名并适当删除一些采样方法

STEP 08 单击“生成”按钮，即可非常清晰地对比同一个提示词下，3 种不同采样方法和 2 种不同大模型分别生成的图像，效果见图 2-38。

2.2 以图生图：通过参考图片生成图像



图生图（Image to Image）是一种基于深度学习技术的图像生成方法，它可以将一张参考图片通过转换得到另一张与之相关的新图片，这种技术广泛应用于计算机图形学、视觉艺术等领域。

图生图功能突破了 AI 完全随机生成的局限性，为图像创作提供了更多的可能性，进一步增强了 Stable Diffusion 在数字艺术创作等领域的应用价值。Stable Diffusion 图生图功能的主要特点如下。