

第 1 章

Chapter

AI 绘画入门

AI 绘画：
Stable Diffusion ComfyUI
的艺术

本章概述

本章将介绍 AI（人工智能）绘画的基础知识，包括当前 AI 绘画的发展现状、提示词的用法、流行的工具、线上工具的操作流程，以及 AI 在艺术设计中的作用。

本章重点

- 掌握提示词的用法
- 掌握线上 AI 绘画工具的操作流程

2023 年，以 ChatGPT 3.5 的发布为标志，AIGC（AI 生成内容）时代正式来临。在此背景下，AI 绘画作为人工智能的一个重要分支，迅速引起了广泛关注和热议。它的发展对各个领域，特别是文化创意产业，产生了巨大的影响。

AI 绘画是一种利用计算机技术和人工智能算法来生成或转换图像的艺术形式。它与传统的绘画有着本质的区别，当前的人工智能通过深度学习生成式神经网络，模拟生物神经网络的运作原理，其学习方法与人类相似。AI 能够通过高效学习数亿甚至数百亿幅图像，将所学知识融会贯通，生成具有艺术价值的绘画作品。它得益于人类的启发，但其潜力却有可能超越人类。

1.1 AI 绘画的发展

从 2022 年起，具有行业实用价值的 AI 绘画程序进入百家争鸣的时代。流行的程序包括

DALL-E、Midjourney、Stable Diffusion 和 Firefly 等。2022 年 8 月，百度公司正式发布了“文心一格”，这是国内首个具备文字生成图像功能的绘画程序。这些程序各有优势，而 Stable Diffusion 凭借其开源免费的特性、众多的模型和扩展功能，以及对图像的精确控制，赢得了广泛的用户群体。本书以 Stable Diffusion ComfyUI 电脑版为例，详细介绍它的各项功能和技巧。为了行文方便，后续将 Stable Diffusion 简写成 SD。由 Black Forest Labs（黑森林工作室）团队开发的 Flux 模型于 2024 年 8 月初开源，随后立即获得巨大的关注，本书也将大量应用该模型完成教学案例。

2023 年 7 月底，SD 正式发布了开源模型 SDXL。让我们来欣赏 3 幅由 SD 生成的图像，这 3 幅图像分别展示了写实照片、2.5D 风格和二次元风格，如图 1-1 所示。

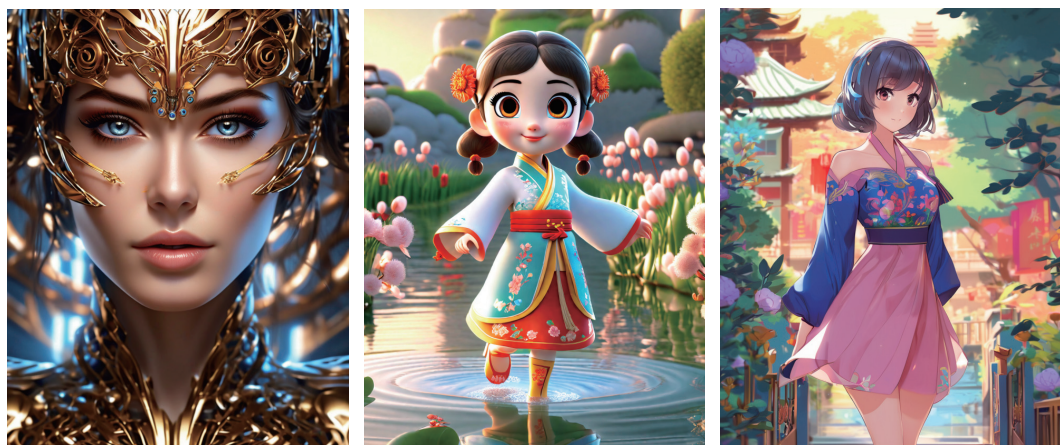


图 1-1 由 SD 生成的图像

在使用 SD1.5 模型时，生成这些不同风格的图像通常要选择 3 个不同风格的模型，而 SDXL 模型则可以通过一个模型生成多种风格的图像，而且画面更清晰，色彩更靓丽，细节更丰富。

到了 2024 年，随着时间的推移，SDXL 已经成为一个成熟的模型类型，衍生出了许多优秀的优化模型，并形成了良好的“生态系统”。如果说 SD 以前的版本主要被 AI 绘画用户用于娱乐，那么 SDXL 作为一个综合性模型，使得 AI 绘画的应用领域更加广泛。

2024 年 6 月，Stability AI 发布了新模型 SD3.0，它采用了 Diffusion Transformer (DiT) 架构，这是一种全新的架构设计，能够更好地理解提示词和物理世界。如图 1-2 所示，该模型在对人物和场景的提示词上非常准确地表达出意图。



图 1-2 花园里的女孩

SD3.0 提示词: 1girl, yellow long princess dress, long brown hair, hair accessories, jewelry, in the garden, best quality, masterpiece, movie poster, full body.

对应的含义: 一个女孩, 黄色的长公主裙, 棕色的长发, 发饰, 珠宝, 在花园, 最好的质量, 杰作, 电影海报, 全身照。

第二个案例主要展示 SD3.0 模型对方位和色彩的理解能力, 如图 1-3 所示。



图 1-3 瓶子

SD3.0 提示词: Three transparent glass bottles on a wooden table. The one on the left has red liquid. The one in the middle has blue liquid. The one on the right has green liquid.

对应的含义: 一张木桌上摆放着三个透明的玻璃瓶。从左至右, 每个瓶子内分别装满了鲜艳的红色、蓝色、绿色液体。

第三个案例展示 SD3.0 在文字生成方面的显著进步, 如图 1-4 所示。



图 1-4 文字生成

SD3.0 提示词：Photo of a rectangular orange neon sign with the text “Hello”, the sign is on the wall in a metro station.

对应的含义：一张矩形橙色霓虹灯标志的照片，上面写着“Hello”，标志位于地铁站的墙上。

SD3.0 是一个全新的模型，功能强大，但目前尚未形成完善的“生态系统”。本书主要以 SDXL 模型为基础进行写作。读者仍可以使用 SD3.0 模型进行各类提示词的测试和学习。

2024 年 6 月，快手开源了 Kolours（可图）模型，这是一个基于 SDXL 架构的模型，能够更好地支持中文提示词。这对于中国用户而言，无疑降低了使用门槛。Kolours 兼容支持 SDXL 的 LoRA 和 ControlNet 模型，具有较好的应用生态前景。Kolours 生成的图像如图 1-5 所示。



图 1-5 Kolours 生成的水墨老虎

Kolours 提示词：中国水墨画，山水，老虎。

Kolors 模型对中国用户非常友好，支持中文提示词，线上和线下都能使用，对中国元素的理解也很出色。本书后续的许多案例将采用该模型完成。

2024 年 3 月，SD 的母公司 Stability AI 经历了动荡，一些核心成员离职，并成立了新公司 Black Forest Labs（黑色森林工作室）。2024 年 8 月 1 日，该公司开源了新模型 Flux，该模型具有 120 亿个训练数据，其中 FLUX.1 [schnell] 版本是完全开源的。Flux 在提示词理解、手部处理和文字书写等方面有了显著提升。Flux 模型对手部的处理效果，如图 1-6 所示。



图 1-6 Flux 模型的手部处理效果

Flux 提示词：A herdsman in Tibet, China, is waving goodbye.

对应的含义：一个中国西藏的牧民，正在挥手告别。

Flux 对文字的处理也非常准确，如图 1-7 所示。



图 1-7 Flux 模型的文字处理效果

Flux 提示词：On the table, an ancient map with the words “Flux’s Treasure”, sunshine, morning.

对应的含义：桌上放着一张古老的地图，上面写着“Flux 的宝藏”，阳光，早上。

Flux 模型在生成的画质和对文字的理解方面全方位超过了 SD3.0，具有强大的应用前景，后续章节将对此进行详细介绍。

1.2 AI 绘画的提示词技巧

提示词是用户给 AI 下达的绘画指令。随着绘画模型的进步，AI 能够根据简单的提示词生成具有较好艺术效果的图像。SD1.5 版本的提示词通常使用关键字组合，自 SDXL 开源以来，提示词已经能够理解人类的自然语言。特别是 Flux 模型，在语义理解方面有了显著提高。

1 提示词的组成

一般来说，画面提示词的提炼需要包括三个方面的内容：

第一方面：画质和风格

画质描述词包括“大师作品”“杰作”“高清晰度”等；风格描述词包括“摄影”“插画”“水墨”“3D”等。

如图 1-8 所示，本例的画质描述词采用了“大师作品”“高清晰脸部”“高细节皮肤”等，风格描述词则选择了“摄影”。



图 1-8 小男孩

Kolors 提示词：大师摄影作品，高清晰脸部，高细节皮肤，一个中国的小男孩，正在弹吉他。

画质描述词，如“大师作品”“超精细”等，对于确保 AI 作品的品质是非常有效的。

风格描述词可以参考不同的艺术风格，如“印象派风格的花园”“写实主义的人物肖像”“卡通风格的城市街道”等，以使 AI 生成具有特定艺术风格的图像。也可以参考一些著名的艺术家或艺术作品，在提示词中提及这些艺术家的名字或作品的特点，如“模仿梵高的星空风格”“像莫奈的花园一样美丽”等。还可以参考不同的艺术流派，如“巴洛克风格的建筑”“现代主义的绘画”等。

结合文化和历史元素可以为图像增添深度和内涵，如“中国传统风格的山水画”“古希腊神话中的场景”“中世纪的城堡”等，这些元素可以赋予图像独特的文化氛围和历史感。

此外，还可以参考不同国家和地区的文化特色，如“日本动漫风格的角色”“印度传统服饰的女子”等。

第二方面：画面主体对象和所处环境

如图 1-9 所示，兔子行走是主体，树林是环境。



图 1-9 兔子

Kolors 提示词：一只兔子在树林中行走。

描述主体时不要使用过于宽泛的词汇，而应尽量具体地描述想要绘制的对象。例如，不要仅仅说“动物”，而是明确指出“一只毛色金黄的狮子”或者“一只蓝色羽毛的鹦鹉”。这样可以让 AI 更准确地理解绘画需求，生成更符合预期的图像。对于复杂的场景，可以逐个描述其中的主要元素。例如“一座古老的城堡，城堡上有飘扬的旗帜，城堡前有一条清澈的河流，河流上有一座木桥”。

主体语言应包含明确的内容，如希望图像中出现的主要颜色。例如“一幅以蓝色和紫色为主色调的星空图”或者“一个穿着红色连衣裙的女孩”。如果对颜色有特定要求，可以更加详细地描述，例如“淡蓝色的天空中飘着几朵洁白的云”。

对于需要表现特定材质和纹理的对象，可以在提示词中加以描述，如“一个光滑的玻璃花瓶”“一只毛茸茸的小狗”“一块粗糙的石头”等。这样可以让生成的图像更加真实和生动。此外，还可以描述材质的光泽度、透明度等属性，如“闪亮的金属表面”“半透明的水晶球”等。

如果绘制的是人物或动物，可以描述他们的动作和表情，使图像更加富有生命力，如“一个面带微笑的女孩在花丛中奔跑”“一只警惕的猫咪蹲在角落里”。

对于动态的场景，可以描述物体的运动状态，如“飞舞的蝴蝶”“流淌的溪水”等。

第三方面：艺术描述词

艺术描述词包括“光影”“构图”“镜头视角”“色彩”“景别”“艺术大师风格”等，用以提升画面的细节和艺术效果。这部分提示词体现了艺术家的美术素养。

图 1-10 的案例用 Kolors 完成。提示词中，第一段为画质和风格描述词，第二段为主体和环境，第三段是艺术描述词。Kolors 很好地理解了“HDR”“亮调摄影”“模糊背景”等偏向艺术语言的提示词。



图 1-10 Kolors 生成的唐朝公主

Kolors 提示词：大师摄影照片，超高清，超精细，详细的皮肤纹理，高清晰的脸，面部细节。一个中国唐朝的公主，在书房。HDR，亮调摄影，柔和，低对比度，低饱和度，冷色调，模糊背景。

艺术语言涉及美术知识，使用色彩搭配的词汇，如“冷暖对比强烈的画面”“色彩鲜艳的組合”等；加入绘画技巧，如“厚涂法”“泼墨法”“点彩”等；加入摄影设备，如“尼康全画幅相机”“F1.4 光圈”“景深”等；加入“光影斑驳”“逆光”“自然光”等光影知识；还可以从一些概念和主题中获取灵感，将其融入提示词中，例如“环保主题的绘画”“未来科技的想象”“爱情的象征”“时尚的装饰”等，这些主题可以引导 AI 生成具有特定意义和价值的图像；甚至可以参考一些哲学、文学或电影中的概念，如“存在主义的思考”“《红楼梦》中的场景”“科幻电影中的宇宙飞船”等。

提示词没有固定格式，每个人都有自己的用法。用户只有在不断的尝试中，才能提升自

己的技术和技巧。

2 用大语言模型辅助生成提示词

目前，大部分 AI 程序已经能较好地理解自然语言，尝试用 ChatGPT、Kimi、豆包等大语言模型生成提示词，不仅提高了效率，还能产生丰富的修饰语言，提升艺术效果。

Kimi 是北京月之暗面科技有限公司于 2023 年 10 月 9 日推出的一款智能助手，打开 Kimi 网页（<https://kimi.moonshot.cn>）就可以使用该大语言模型。文字语言辅助生成是它的功能之一。Kimi 的主页如图 1-11 所示。



图 1-11 Kimi 的主页

笔者的一般用法是：“用一段话描述某风格的画面 + 主体和环境 + 具体细节”。例如，“用一段话描述一个电影的画面：一个中国古代的将军站在城墙上”。输入后，Kimi 生成的自然语言如下：

在夕阳的余晖中，一位身着战甲、头戴铁盔的中国古代将军站在高耸的城墙之上，目光如炬地凝视着远方。城墙下是连绵的山脉和辽阔的平原，而城墙上旌旗猎猎，士兵们严阵以待。将军的面容刚毅，眼中闪烁着坚定与智慧的光芒，仿佛在思考着即将到来的战事或是守护着这片土地的安宁。他的身影在落日的映照下显得格外高大，成为这片古老土地上不屈不挠的守护者的象征。

用 Kolors 生成的图像如图 1-12 所示。



图 1-12 Kolors 生成的将军图像

1.3 线上体验 AI 绘画

国内外的 AI 绘画程序已进入百花齐放的时代，许多大模型开设了网站，使用户能够方便地进行线上绘画，从而摆脱了对专业计算机硬件的限制。

我们主要通过两种方式进行体验。

1 利用大语言模型自带的绘画功能生成画面

部分大语言模型具有绘画功能，如 ChatGPT、文心一言、豆包等。

豆包是字节跳动公司开发的一款人工智能工具。用户可以通过网页或手机上的 App 进行体验。

豆包的网址是 <https://www.doubao.com>，页面如图 1-13 所示，它是一个功能全面的大语言模型，图像生成是其中的一项功能。



图 1-13 豆包页面

选择“图像生成”进入 AI 绘画页面后，页面中会显示大量的网友创作的作品，如图 1-14 所示。将鼠标指针移动到某幅生成的图像上，随即会显示出这幅图像的提示词，这是学习优秀提示词的一种方法。

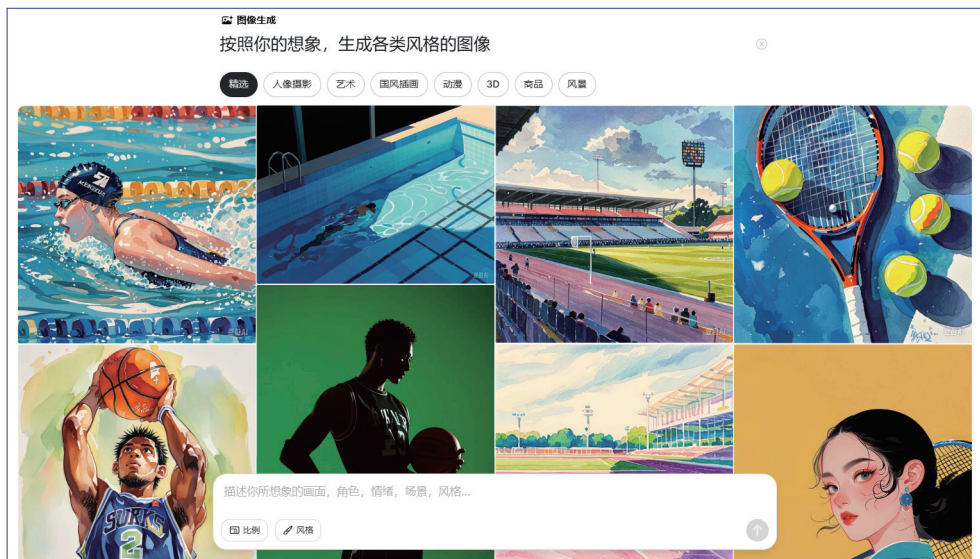


图 1-14 豆包 AI 绘画页面

中间的文字框用于输入提示词。例如，输入提示词“中国水墨，工笔画，一个唐朝的仕女，豪华的装饰”。输入框底部的“比例”按钮可用于更改画面的长宽比，如 1:1、2:3、4:3 等，以满足不同媒介的需求；“风格”按钮可用于设置画面的艺术类型，如油画、水彩、动漫、插画等。

豆包生成了四幅画面供我们选择，如图 1-15 所示。



图 1-15 豆包生成的工笔仕女

2 用 AI 图像模型在线生成图像

使用 AI 图像模型在线生成图像是一个有趣且富有创造性的过程。以下是国内部分图像生成工具和平台，用户可以使用它们来在线生成图像。

1) 文心一格

文心一格是百度推出的 AI 艺术和创意辅助平台，它基于百度的文心大模型技术，能够根据用户输入的文本描述智能生成创意画作。文心一格特别适合在中文环境下使用，因为它在中文语义理解方面具有优势。

文心一格的网址是 <https://yige.baidu.com>。

进入网站后，单击“AI创作”进入绘画面板，如图 1-16 所示。在绘画面板中，输入提示词“中国水墨画，凤凰，丰富的色彩”。设置“画面类型”为“中国风”，“比例”为“方图”，“数量”为“1”，然后单击“立即生成”按钮生成图像。

生成的图像如图 1-17 所示。文心一格能够生成国外模型很难生成的中国风图像，如龙凤、神兽、生肖、工笔、剪纸等。



图 1-16 文心一格绘画面板



图 1-17 文心一格生成的凤凰

2) 即梦 (Dreamina)

即梦是由字节跳动旗下的剪映推出的生成式人工智能创作平台，支持用户通过自然语言描述或图片输入来生成高质量的图像和视频内容。

即梦的网址为 <https://jimeng.jianying.com>。

进入网站后，单击“AI 作图”中的“图片生成”按钮，进入 AI 绘画面板，如图 1-18 所示。在绘画面板中，输入提示词“静物摄影，装置概念设计，蓝色毛茸茸蓬松汽车，毛茸茸，公园，草地，自然光线，超现实主义空间，幻想，抽象，明亮的颜色，黄橙色。”“模型”选择“即梦通用 V2.0”，其他“精细度”“比例”参数保持不变。然后单击“立即生成”按钮生成图像。



图 1-18 即梦的绘画面板

生成的图像效果如图 1-19 所示。



图 1-19 即梦生成的小汽车

当鼠标指针移动到其中一幅图像上时，会出现如图 1-20 所示的功能按钮。

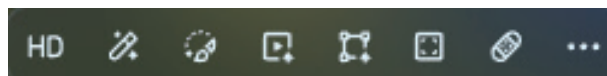


图 1-20 功能按钮

其功能分别是：

- **HD** HD 画质提升：用于提高画面的分辨率，将它变成超高清图。
- **修复** 细节修复：用于重新生成一幅图像，让细节更加丰富。如图 1-21 所示，左边是修复前的图像，右边是经过了细节修复的图像。



图 1-21 即梦的细节修复

- **局部重绘**：单击该按钮后会弹出“局部重绘”面板，通过画笔把轮胎部分涂上蒙版，并输入提示词“绿色的轮胎”。重绘后，被遮罩涂抹的部分就变成了绿色，如图 1-22 所示。

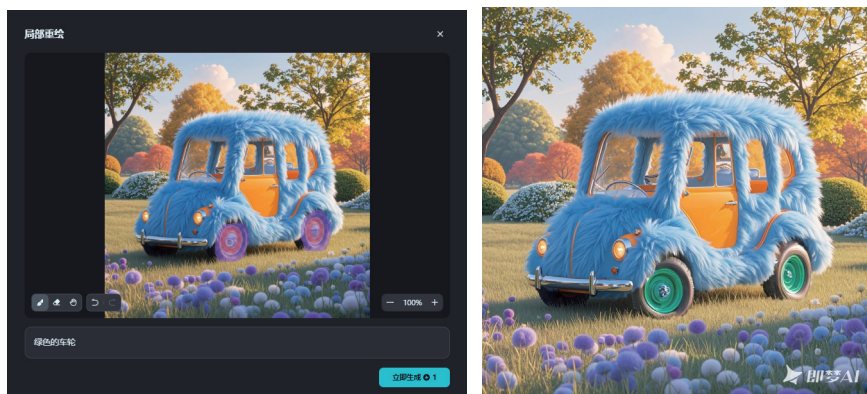


图 1-22 即梦的局部重绘

-  生成视频：单击该按钮进入“视频生成”面板，将这幅图像作为底图来引导视频的生成，如图 1-23 所示。输入提示词“毛茸茸的汽车在公园行驶”，参数设置如下：“运镜控制”用于设置摄像机的运动，这里选择“随机运镜”；“运动控制”用于更改运动的幅度，这里选择“适中”；“时长”决定动画的时间长度，这里选择“3s”；其他参数保持默认设置，生成后即可得到汽车行驶的动态画面。



图 1-23 即梦的视频面板

-  画布编辑：此功能可进一步编辑画面，并支持实时绘画。如图 1-24 所示，在左图的车顶上画上两笔后，右图对应的位置就变成了汽车的盖子。



图 1-24 画布编辑

-  扩图：用于对图像进行扩展，展示更多的环境，如图 1-25 所示。



图 1-25 扩图

-  消除笔：用于清除画面中的特定内容。如图 1-26 所示，在左图中用遮罩涂抹保险杠后，

右图中的保险杠就消失了。

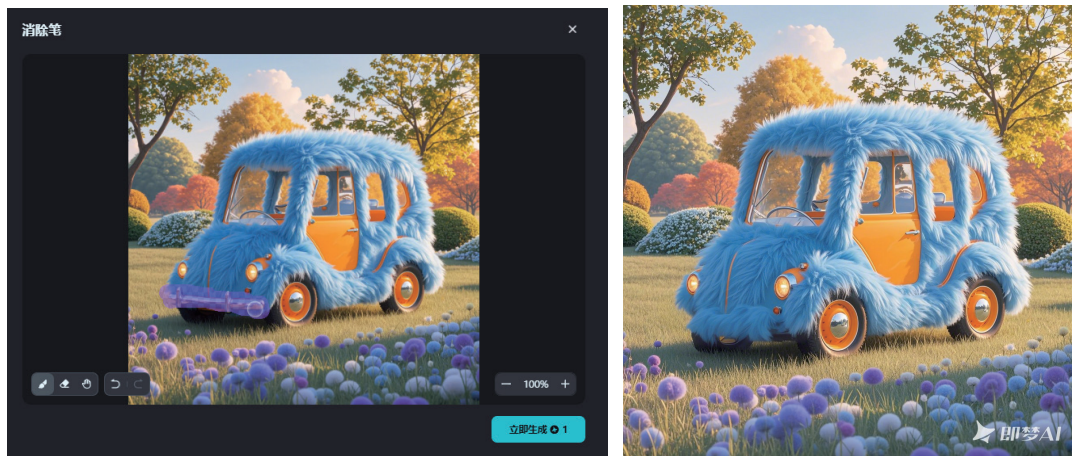


图 1-26 消除笔

-  更多：提供删除和举报画面内容的选项。

通过以上演示可以看出，即梦能够对画面进行全面的处理。虽然它是一个在线程序，但其功能非常完善，基本具备了 ComfyUI 对图像处理的常用功能。即梦背后的公司是字节跳动，该公司拥有强大的技术能力和丰富的训练数据，相信在不久的将来，该平台的功能会有更大的提升。

3) Kolors

Kolors 是快手公司开发的大规模文本到图像生成模型，经过数十亿个文本 - 图像对（Text-Image Pairs）的训练，在视觉质量、复杂语义准确性以及中英文文本渲染方面展现出了显著优势。

Kolors 的网址是 <https://klingai.kuaishou.com>。

进入网站后，选择“AI 图片”，进入绘画面板，如图 1-27 所示。在绘画面板中，输入提示词“在宁静的夜晚，一座古桥在月光的照耀下显得格外神秘而庄严，桥身的石块在柔和的光线中泛着淡淡的银光，仿佛被时间的魔法轻轻抚摸。桥下的河水在夜色中泛着波光，反射出桥影和星光，增添了几分梦幻般的色彩。桥上的灯笼随风轻轻摇曳，发出温暖的光亮，为夜行的人们指引着方向。周围的古建筑和树木在夜色中若隐若现，与古桥共同构成了一幅宁静而美丽的夜景画卷。整个画面充满了古典美和宁静的氛围，让人在凝视中感受到历史的沉淀和自然的和谐。”其他参数保持默认设置。生成的图像效果如图 1-28 所示。



图 1-27 Kolors 的绘画面板



图 1-28 Kolors 生成的夜景

Kolors 还支持通过垫图生成图像。垫图即 AI 绘画中的“图生图”。进入网站后，选择“参考图 / 垫图”，导入图 1-29 中的左图，输入提示词“中国女孩，汉服”，将“参数设置”更改为更适合参考图的“2:3”竖图，生成后得到图 1-29 中的右图。左边的参考图在画面风格、姿势、服装上的影响是显著的。垫图在 AI 绘画中应用非常广泛。



图 1-29 Kolors 的图生图

Kolors 是一个非常优秀的模型，后面会详细介绍它在 ComfyUI 中的使用方法。Kolors 生成的图像能够直接转成视频，我们将在“第 8 章 ComfyUI 视频”中详细介绍。

4) 智谱清言

智谱 AI（北京智谱华章科技有限公司）是一家专注于打造新一代认知智能大模型的公司，致力于中国创新的大模型技术。智谱清言是该公司推出的一款功能全面的生成式 AI 助手。

智谱清言的网址是 <https://chatglm.cn>。进入网站后，选择“AI 绘画”进入绘画界面，如图 1-30 所示。它的特点是能够通过聊天的方式进行绘画。

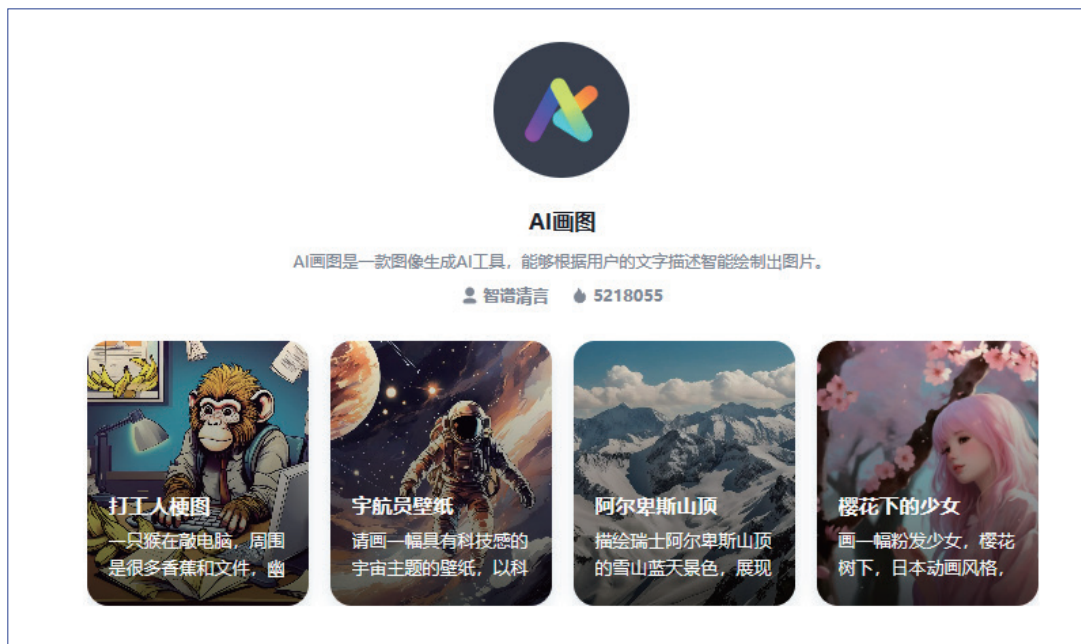


图 1-30 智谱清言的绘画界面

输入提示词：在新海诚动画电影般的细腻笔触下，一幅江南水乡的画卷缓缓展开，16K 的超细节分辨率让每一砖、每一瓦、每一叶扁舟都清晰可见。丰富的色彩交织出清新明亮的视觉效果，天空是清澈的蔚蓝，水面泛着温柔的碧波，两岸的柳树随风轻摆，翠绿的枝叶在阳光下闪烁着晶莹的光点。温馨的光影在古建筑的白墙黑瓦间跳跃，营造出一种宁静而温馨的氛围，让人仿佛能嗅到空气中弥漫的清新水汽和泥土的芬芳，感受到时间在这片古老而又充满活力的土地上静静流淌。

生成的图像如图 1-31 所示。



图 1-31 智谱 AI 生成的江南水乡

用户可以在该图的基础上进一步生成图像。输入提示词“变成冬天的场景”，生成的图像便变为冬季的场景，如图 1-32 所示。然而，这幅图像并不是在前面的垫图基础上生成的，因此场景变化较大。



图 1-32 智谱 AI 生成的冬天的江南水乡

用户可以进一步输入提示词，在该图的基础上改变画面的效果。智谱 AI 的使用非常人性化，部分模型同时发布了适用于 Windows、安卓和 iOS 版本的 App，用法和网站类似。

5) fluxpro.art

随着 Flux 模型的流行，也诞生了体验 Flux 模型的网站——fluxpro.art，网址是 <https://fluxpro.art>，用户可以每天用积分生成图像，观看大量网友的作品，学习图像附带的提示词，感受 Flux 的魅力。

随着各种 AI 模型的不断发展，未来将有更多的网站提供 AI 绘画的体验。易用和方便是 AI 绘画工具未来发展的重要趋势之一。

1.4 人工智能绘画的作用

AI 绘画的发展迭代极为惊人。初期阶段，AI 绘画还很难生成优质的图像，很多艺术工