

解密AI 绘画与修图

Stable Diffusion+Photoshop

第1章

常用AI图像生成工具简介

AIGC (Artificial Intelligence Generated Content, 人工智能生成内容) 的概念虽然由来已久, 但以ChatGPT为代表的实用性产品落地却是近年来的事情。特别是在AI图片生成领域, 短短两年内涌现了大量模型和工具, 并且以令人瞠目结舌的速度不断迭代演化。本章主要介绍目前主流的图片生成工具, 读者可根据自身需求选择合适的工具。

1.1 Midjourney

2022年8月，在美国科罗拉多州举办的艺术博览会上，一幅名为“太空歌剧院”的作品荣获了数字艺术类别的冠军，如图1-1所示。这一新闻迅速在网络上引起热议，同时也让人们记住了生成这幅作品的工具——Midjourney。



图1-1

Midjourney是目前公认的质量和创意风格最好的AI图片生成工具之一，目前已经更新到V6版本。Midjourney搭载在社交软件Discord的频道上，两者之间的关系类似于微信和小程序。Midjourney的优势主要体现在两个方面：首先，该工具由云端服务器提供算力，用户无须配置高性能计算机，只需打开Discord的网页或客户端即可使用。

其次，AI生成图片需要经历一系列的转化过程，用户首先在脑海中想象一个画面，然后用文字把画面描述出来。接下来，AI需要理解文本描述的内容，最后生成符合描述的图片。在这个过程中，每个步骤都会产生一定的信息损失和偏差。另外，由于AI每次生成的图片都具有很强的随机性，无论使用哪种图片生成工具，用户要想得到特别理想的生成结果，都需要反复生成大量图片，然后从中挑选最符合预期的效果，这就涉及出图效率的问题。相对而言，Midjourney生成图片的下限比较高。用户无须花费很长时间学习，只需使用近乎口语化的提示词和比较少的生成次数，就能得到风格多样且艺术性很强的图片。Midjourney官网画廊如图1-2所示。

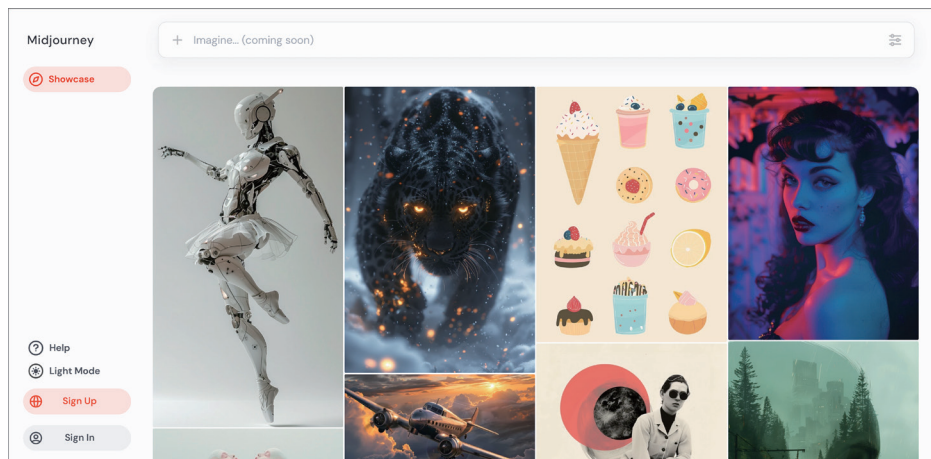


图1-2

在国内，Midjourney的名气很大，但实际使用过的人却很少，这主要是因为国内个人用户无法直接登录Discord，需要购买Midjourney会员才能使用。总体来说，Midjourney更适合生成图像素材和创意参考类图片。对于注重审美和生成效率的用户，可以利用Midjourney快速搭建创意设计流程。

1.2 DALL-E 3

早在2021年，OpenAI公司推出了DALL-E模型。当时的模型还处于研究阶段，对图片的细节处理不够完善，经常出现逻辑和事实错误。如今，最新版的DALL-E 3在语义理解、图片质量、长文本输入等方面都取得了质的飞跃，已经发展成和Midjourney、Stable Diffusion旗鼓相当的图片生成工具。

DALL-E 3作为多模态的一环已集成在ChatGPT中，目前只对ChatGPT Plus用户和企业用户开放使用功能，即只有付费的ChatGPT用户可以使用DALL-E 3。另外，有条件的用户也可以在微软的Copilot（即New Bing）的CHAT中免费体验DALL-E 3。总体来说，DALL-E 3和Midjourney的功能类似，它们就像手机里的APP那样，学习门槛很低，打开即用，同时还能得到效果不错的生成结果。对比DALL-E 3和Midjourney的生成结果，可以发现Midjourney的光效和画面质感更好，DALL-E 3对长文本提示词的理解更准确，同时支持自定义文本生成。DALL-E 3官网示例图如图1-3所示。

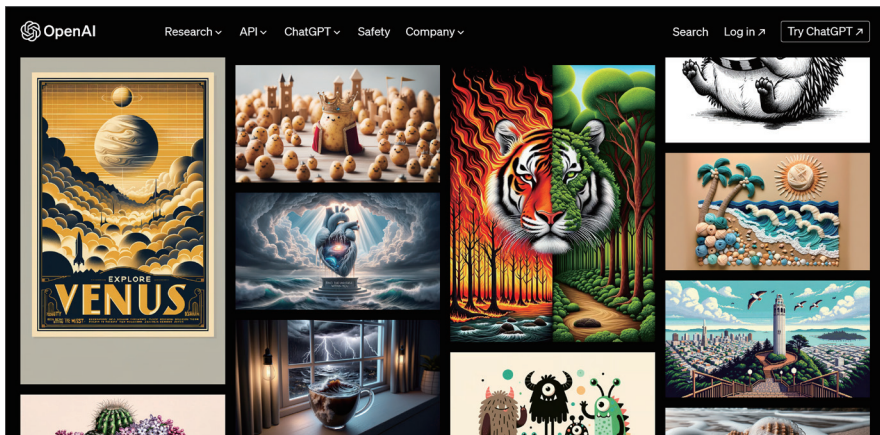


图1-3

对于同时有文案处理和图片生成需求的用户来说，DALL-E 3是理想的工具形态。而对于从事自媒体或设计类行业，每天需要生成大量素材和商业落地图片的用户来说，Stable Diffusion则是更佳的选择。

1.3 Stable Diffusion

Stable Diffusion是一款免费、开源的图片生成工具。虽然它的诞生比Midjourney和DALL-E晚，但由于拥有发展良好的开源社区，Stable Diffusion受到的关注度和用户广度都远超两位“前辈”。

Stable Diffusion最大的特点是可以本地部署，只需一台计算机和一块8GB及以上显存的独立显卡，就可以无限制地生成可商用的图片。此外，开源社区中的大量用户为Stable Diffusion开发了数以百计的插件和数十万个模型，这些插件和模型极大地扩展了Stable Diffusion的图片生成能力。

在生成效果方面，最近发布的SDXL模型具备了直接生成高清图片的能力，图片中人物手部和文字生成方面的问题也得到了很大改善。而已经开启申请测试的Stable Diffusion 3（见图1-4）则更上一层楼，除了持续提高生成质量和对提示词的理解能力外，还支持新的文字生成功能以及通过提示词修改画面内容的功能。



图1-4

与Midjourney、DALL-E 3相比，Stable Diffusion的学习成本比较高。这是因为Stable Diffusion生成的图片风格主要取决于用户下载的模型，而模型的种类和版本繁多且不能混用，所以用户需要对此有一定的理解。此外，生成结果的构图、光照、色彩等方面都可以控制，需要熟悉并准确运用的参数也非常多，即便使用相同的提示词和模型也可能出现天差地别的效果图。这些因素使得人们觉得Stable Diffusion生成的图片质量上限非常高、下限非常低。

需要说明的是，目前使用的Stable Diffusion界面都是由开源社区中的开发者自行制作的，其中WebUI和ComfyUI的用户数量最多。这两个界面都运行在浏览器页面中，WebUI采用了传统软件的设计思路和布局方式，界面由读者熟悉的选项卡、对话框、单选按钮和复选框等控件组成，如图1-5所示。

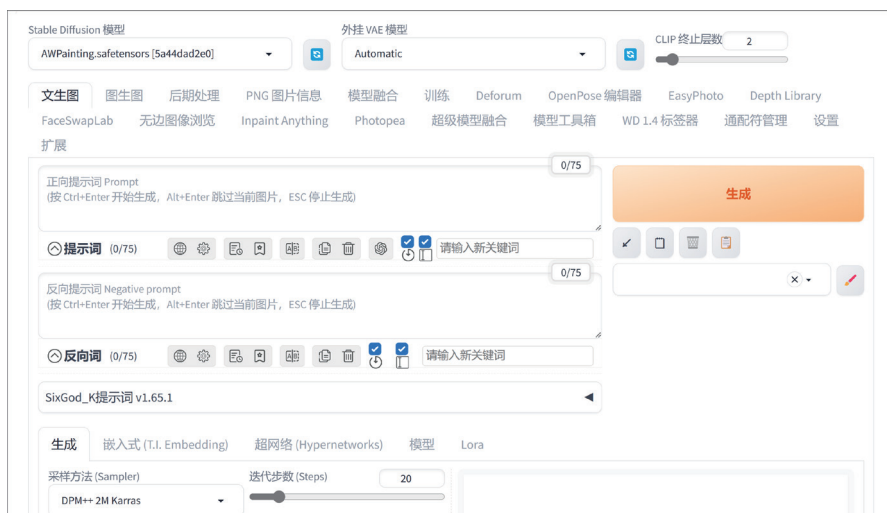


图1-5

ComfyUI采用了节点式的用户界面，通过不同节点的搭配组合可以生成适用于各种需求的工作流程，如图1-6所示。这两种界面并不是互斥的关系，许多用户通常会从更易于理解的WebUI开始学习Stable Diffusion的基本参数，在需要进行复杂工作流程时再使用ComfyUI。

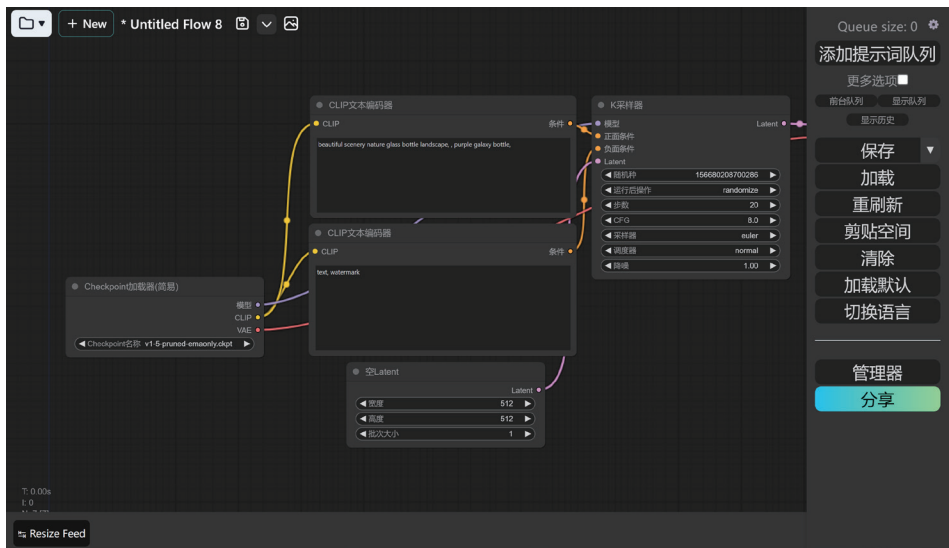


图1-6

1.4 文心一格

随着AI概念的异军突起，国内也掀起了火热的百模大战。在百花齐放的各种AIGC工具中，百度的“文心一格”是一款发布时间较早、完善程度也较高的产品。文心一格能够生成多种风格的图片，同时提供了商品换背景、艺术字设计和画面扩展填充等功能。生成的图片质量中规中矩，其最大优点是可以使用中文提示词，如图1-7所示。

文心一格采用的是会员和签到领取“电量”的模式，用户可以在网页端和微信小程序中使用它，适合图片生成频率不高的轻度用户。另外，目前在手机上使用文心一言可以免费生成图片，如图1-8所示。百度自己解释道：文心一言的画图功能更偏向实用性和便捷性，而文心一格更专注于艺术性和创意性。

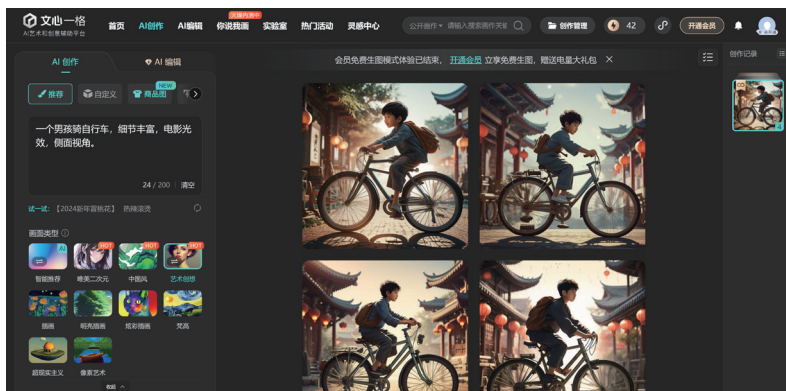


图1-7

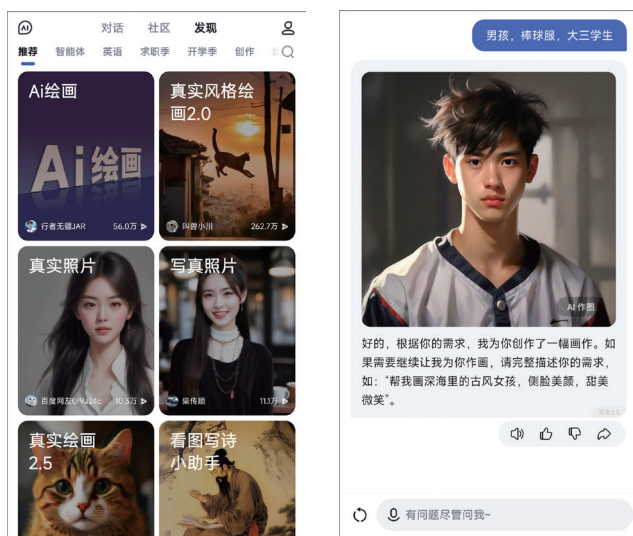


图1-8

1.5 在线版图片生成网站

目前国内有许多调用Midjourney的API接口和部署了Stable Diffusion云端的在线图片生成网站。在这些网站中，调用Midjourney API接口的网站大多采用会员制收费，无论是界面、图片生成指令还是生成结果，它们与Midjourney本身相差不大，如图1-9所示。

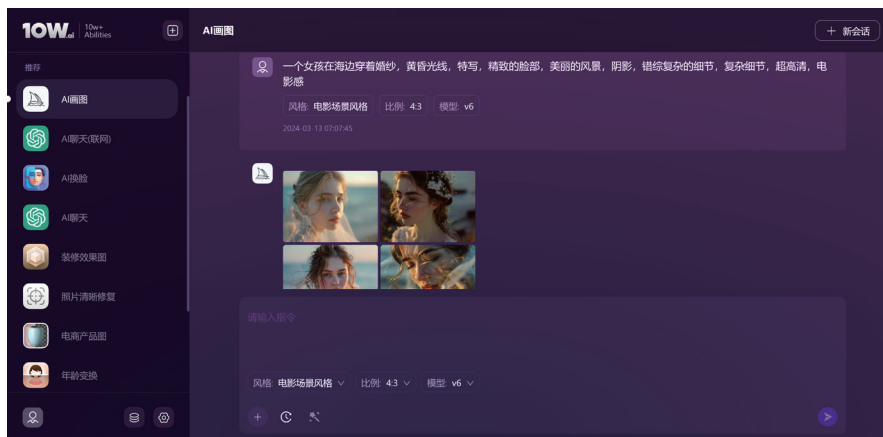


图1-9

当然，这些网站需要把用户的图片生成请求上传到Midjourney，Midjourney生成图片后再传送回来，因此生成图片的速度普遍比较慢，同时模型和指令的更新也会滞后于Midjourney官方网站。

Stable Diffusion则是开源且免费的，而且其云端部署非常简单。对于显卡性能不足的用户，可以通过登录提供在线生图服务的网站来体验新版模型带来的画质提升，如图1-10所示。



图1-10