

第 1 章

Chapter

人工智能 绘画概论

AI 绘画
Stable Diffusion
从入门到精通

本章概述：学习人工智能绘画（AI 绘画）的历史、概念、算法、原理，以及常用的 AI 绘画工具，最后简单介绍 Stable Diffusion 软件。通过本章的学习，读者将对人工智能绘画有一个全面的了解。

本章重点：

- 掌握人工智能绘画的概念
- 了解 Stable Diffusion 软件

当时间进入 2023 年，似乎在一夜之间，我们突然进入了 AIGC 时代。ChatGPT、必应、文心一言等各种人工智能工具纷纷进入我们的视野，同时各种 AI 生成的图像也呈现在我们面前。你能分辨出图 1-1 中的哪一幅图是由人工智能工具生成的吗？



图 1-1 人工智能生成的图像

它们都是由人工智能生成的作品，一般人已经很难分辨出哪些是人类的作品，哪些是由AI生成的作品，甚至有时候无法确定是否为相机拍摄的照片。

那么，什么是人工智能绘画（AI 绘画）呢？人工智能绘画是利用计算机技术和人工智能算法生成或转换图像的艺术形式。这些作品包含各种类型，例如照片风格、卡通风格、油画风格、水彩风格、水墨风格等。通常，人工智能绘画使用深度学习技术，让计算机学习艺术家的风格和技巧，并生成新的图像。人工智能绘画可以创造独特的艺术品，也可以为现有的图像添加艺术效果。它已经在艺术、设计和广告等领域得到了广泛应用。图 1-2 为人工智能生成的水墨效果图。



图 1-2 人工智能生成的水墨效果图

人工智能绘画与传统绘画有着本质的区别。如今的人工智能采用深度学习生成式神经网络的原理，模拟了生物的神经网络。其学习方法与人类相似，通过高效学习数亿幅甚至数百亿幅图像，将所学融会贯通，然后生成具有艺术价值的绘画作品。人工智能“师从人类”，受益于人类的启发，然而，它具有超越人类的潜力，所谓青出于蓝而胜于蓝。一个人通常只能从几位老师那里学习，而人工智能却可以通过数以亿计的“老师”进行学习，并且随着硬件计算能力的提升，其学习效率将变得极高。

有些人有这样的疑问：人工智能绘画是否属于拼图？答案当然不是，举个例子，人工智能可以用梵高风格绘制飞机、火车和计算机等现代物品，而这些物品在梵高时代并不存在。因此，人工智能具有智慧和创造力。图 1-3 为人工智能绘制的具有梵高风格的卡车。



图 1-3 梵高风格的卡车

人工智能绘画与传统绘画相比，具有以下特征：

（1）自动化：人工智能绘画可以自动完成一些任务，例如图像生成和颜色填充等，从而减轻了人类艺术家的工作负担。

（2）速度快：人工智能绘画可以在短时间内完成大量的绘画任务，这对于需要高效率完成绘画的场景非常有帮助。

（3）可重复性：由于人工智能绘画是基于算法完成的，因此可以在任何时间重复生成相同的图像，这对于需要大量复制的场景非常有用。

（4）无限可能性：人工智能绘画可以生成各种各样的图像，包括传统绘画风格以外的风格，这使得艺术家可以探索新的艺术形式和风格。

（5）数据质量依赖：人工智能绘画的生成结果很大程度上依赖于训练数据的质量。如果数据集质量不好，生成的图像质量也会受到影响。

（6）需要人类指导：虽然人工智能绘画可以自动生成图像，但在某些情况下，需要人类艺术家对生成的图像进行修饰和调整，以达到更理想的效果。

（7）缺乏情感：人工智能绘画缺乏情感和创造力，无法像人类艺术家一样表达深刻的情感和思想，这在某些情况下可能会限制其艺术价值。

1.1 人工智能绘画的历史

人工智能绘画的起源可以追溯到 20 世纪 50 年代。当时，艺术家和科学家开始使用计算机生成图像和图形，将绘画艺术与技术领域相结合。计算机图像可以被视为人工智能绘画的一部分。下面，我们将按照时间顺序来了解人工智能绘画发展的一些关键时间节点。

(1) 在 20 世纪 50 年代，电子艺术先驱 Ben Laposky 使用计算机示波器生成了一幅艺术作品，如图 1-4 所示。

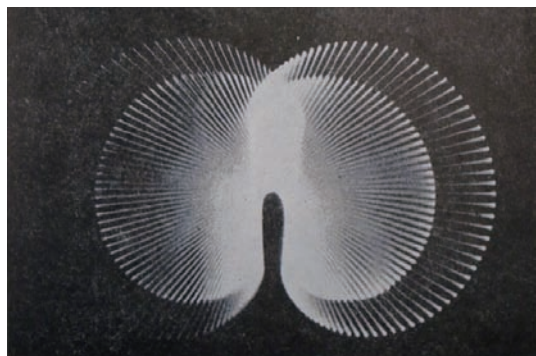


图 1-4 Oscillon (1953) Ben Laposky 作品

(2) 在 20 世纪 60 年代，艺术家和科学家开始使用计算机的计算能力进行创作。图 1-5 是计算机艺术之父 Charles Csuri 用计算机生成的作品，名为“蜂鸟”。



图 1-5 蜂鸟 (1968) Charles Csuri 作品

(3) 在 20 世纪 70—90 年代，随着计算机硬件和软件的不断进步，计算机图形学技术取得了重大突破。在这期间，Harold Cohen 开发了 AARON 绘画程序，并在随后的几十年中不断改进和完善，成为人工智能绘画领域的先驱之作。图 1-6 为 AARON 的一件作品。



图 1-6 AARON 的作品 (1974)

(4) 在 21 世纪初, 随着深度学习技术的发展, 计算机视觉和图像生成算法取得了显著进展, 为人工智能绘画提供了新的可能性。在 2009 年, 基于深度学习的艺术风格迁移算法 DeepArt 发布, 使得将某种艺术风格应用于图像成为可能。图 1-7 为基于 DeepArt 算法绘制的作品。



图 1-7 DeepArt 作品

真正意义上的人工智能绘画指的是基于深度学习模型进行自动作图的计算机程序, 这种绘画方式的发展在时间上是较晚的。

(5) 在 2012 年, Google 公司的吴恩达 (Andrew Ng) 和 Jef Dean 进行了一项实验, 使用 1.6 万个 CPU 训练一个当时世界上最大的深度学习网络, 用于指导计算机绘制猫脸图像。他们使用来自 YouTube 的 1000 万幅猫脸图像进行训练, 历时 3 天, 最终用得到的模型生成了一幅非常模糊的猫脸图像, 如图 1-8 所示。



图 1-8 猫脸 (2012)

这个模型的训练效率和输出结果对于当时的 AI 研究领域来说是一次具有突破意义的尝试。它正式开启了支持深度学习模型的人工智能绘画这个全新的研究方向。人工智能科学家们纷纷投入到这个新的具有挑战性的领域中, 探索如何利用深度学习技术来生成具有艺术性的图像和绘画作品。这项实验的成功为后续的研究和发展奠定了坚实的基础, 并推动了人工智能绘画的进一步发展。

(6) 在 2014 年, AI 学术界提出了一个非常重要的深度学习模型, 那就是著名的对抗生成网络 (Generative Adversarial Network, GAN)。这个深度学习模型核心理念是通过让两个内部程序, 即生成器 (generator) 和判别器 (discriminator), 相互对抗平衡来获得结果。

生成器的目标是生成逼真的样本, 如图像、音频等, 而判别器的目标是尽可能准确地区分生成器生成的样本和真实样本。通过不断地进行对抗训练, 生成器和判别器相互竞争、学习和提升, 最终达到生成高质量样本的目的。

GAN 模型一问世就风靡 AI 学术界, 在多个领域得到了广泛的应用。它也随即成为许多 AI 绘画模型的基础框架, 其中生成器用来生成图像, 而判别器用来评估图像质量。GAN 模型的引入极大地推动了图像生成、风格迁移等领域的发展, 从而推动了 AI 绘画的发展。图 1-9 为基于 GAN 模型的 AI 绘画作品。

但是, 使用基础的 GAN 模型进行 AI 绘画也存在明显的缺陷。一方面, 对于输出结果的控制力较弱, 往往容易产生随机图像, 而 AI 艺术家的输出应该是稳定可控的。另一方面, 生成图像的分辨率较低。



图 1-9 GAN 模型作品 (2018), 图像来源于 Brock

(7) 在 2015 年, 人工智能绘画领域取得了新的突破。Gatys 等人提出了著名的神经风格迁移论文, 通过将卷积神经网络 (CNN) 应用于艺术风格迁移, 使得人工智能绘画的技术更加成熟。这项研究将艺术风格与内容分离, 并利用 CNN 的特征表示来实现图像的风格迁移。这一方法在艺术创作和图像处理领域引起了广泛的关注和应用, 为人工智能绘画的发展带来了重要的进步。

(8) 在 2015 年, Google 发布了一个名为深梦 (Deep Dream) 的图像工具, 该工具引起了广泛的关注。深梦通过对图像进行迭代处理, 强调和增强图像中的纹理和模式, 创造出

独特而幻觉般的视觉效果。深梦生成的画作吸引了很多人的注意，谷歌甚至为这些作品策划了一场画展，进一步展示了深梦在艺术领域中的潜力和影响。图 1-10 为深梦的作品之一《月球时代的白日梦》。



图 1-10 深梦作品《月球时代的白日梦》（Moonage Daydream）

2018 年，Obvious 艺术团队利用 GAN 创作的《肖像：埃德蒙·贝拉米》在佳士得拍卖中以 43.25 万美元的价格成交。这一事件意味着人工智能绘画正式被认可为一种艺术形式，并得到了市场的承认。这次拍卖成交的高价反映了人工智能绘画作品的独特性和艺术价值，同时也引发了对于人工智能在艺术创作中的探索和潜力的讨论。图 1-11 为 GAN 创造的这幅作品。



图 1-11 《肖像：埃德蒙·贝拉米》(2018)

(9) 2021 年年初, OpenAI 发布了备受关注的 DALL-E 系统, 这标志着人工智能开始具备一个重要的能力, 那就是可以根据文字进行创作。DALL-E 系统利用深度学习模型生成图像, 并能够根据文字描述来创造出与之对应的图像。这一技术的推出引起了广泛的关注和讨论, 为人工智能在创作领域的发展带来了新的可能性。通过输入文字, 人工智能可以生成与之相关的图像, 这为创意产业和设计领域带来了新的创作工具和思路。图 1-12 为 DALL-E 系统创作的作品《戴珍珠耳环的少女》。



图 1-12 DALL-E 作品: 《戴珍珠耳环的少女》

(10) 在 2021 年 1 月, OpenAI 团队开源了他们的深度学习模型 CLIP (Contrastive Language-Image Pre-Training, 对比文本—图像预训练模型), 这是当时最先进的图像分类人工智能模型之一。

CLIP 模型的训练过程可以简单概括为: 使用已标注好的“文字—图像”训练数据, 分别对文字和图像进行模型训练。通过不断调整两个模型的内部参数, 使得模型输出的文字特征值和图像特征值能够准确匹配对应的“文字—图像”关系。CLIP 模型与以往的“文字—图像”匹配模型不同, 它利用了 40 亿个“文本—图像”训练数据。这么多的数据和昂贵的训练时间使得 CLIP 模型终于修成正果。互联网上的图像通常都带有各种文本描述, 例如标题、注释、用户标签等, 这些文本成为可用的训练样本。通过这种巧妙的方式, CLIP 的训练过程完全避免了最昂贵费时的人工标注, 或者说, 全世界的互联网用户已经提前完成了标注工作。这一创新为图像分类和语义理解领域带来了重要的突破, 使得 AI 能够更好地理解和处理图像与文本之间的关系。

(11) 2022 年 3 月, 一个全球范围的非营利机器学习研究机构 LAION 开放了当前最大规模的开源跨模态数据库 LAION-5B。该数据库包含接近 60 亿 (58.5 亿) 个图像 - 文本对, 可用于训练从文本到图像的生成模型以及用于给文本和图像的匹配程度打分的 CLIP 模型。这两种模型都是现代 AI 图像生成的核心。

LAION 不仅提供了大量的训练素材库, 还训练 AI 根据艺术感和视觉美感对 LAION-5B 中的图像进行评分, 并将得分较高的图像归入名为 LAION-Aesthetics 的子集。实际上, 最新的 AI 绘画模型, 包括随后提到的 AI 绘画 Stable Diffusion, 都是基于 LAION-Aesthetics 这个高质量数据集进行训练的。这一数据集的质量和规模为 AI 绘画领域的研究和发展提供了重要的资源和支持。

(12) 扩散模型的引入为 AI 绘画领域带来了新的思路, 并弥补了 GAN 模型的一些不足之处。GAN 模型是生成对抗网络, 它在附加条件方面表现较差。例如, 在生成人脸后, 很难进一步指定发型、细节等特定要求。为了解决这个问题, 扩散模型被提出作为另一种思路。

扩散模型通过将图像加入高斯噪点形成噪点图, 然后通过算法逆过程进行减噪, 生成最终的图像。这种模型可以在噪点图的基础上进行操作, 通过调整和控制噪点的分布, 实现更加精细的图像生成。扩散模型已经成为主流的 AI 绘画软件的基础, 它可以更好地满足用户对于图像的特定要求和细节控制, 提供更灵活和个性化的绘画体验。

(13) Diffusion 模型是一种对于像素空间具有巨大算力需求的模型进行优化的方法。传统的扩散模型在像素空间中操作, 因此需要大量的计算资源和内存。为了解决这个问题, 提出了基于潜在空间的 Diffusion 模型, 通过降低维度来减少计算和内存需求。

基于潜在空间的 Diffusion 模型与像素空间模型相比, 能够显著降低内存和计算要求。例如, Stable Diffusion 模型使用的潜在空间编码缩减因子为 8, 即将图像的长和宽都缩减 8 倍, 一个 512×512 像素的图像在潜在空间中直接变为 64×64 像素, 节省了 $8 \times 8 = 64$ 倍的内存。

这种基于潜在空间的优化能够在保持图像质量的同时, 大幅度减少计算和内存需求, 使得 Diffusion 模型在实际应用中更加高效和可行。这为 AI 绘画领域的发展带来了重要的技术突破, 使得更多人能够在有限的硬件资源下享受到高质量的 AI 绘画体验。

(14) 在 2022 年的 AI 领域, 基于文本生成图像的 AI 绘画模型成为备受关注的主角。其中, Disco Diffusion 是一个在 2 月初开始爆红的 AI 图像生成程序, 它能够根据描述场景的关键词渲染出相应的图像。这个程序的开发者是艺术家兼程序开发员 Somnai_dreams。

Disco Diffusion 的独特之处在于它能够通过文字输入描述来生成具有艺术感的图像, 并且能够根据关键词准确地渲染出所需的场景。这种技术为艺术创作和设计领域提供了新的可能性, 使得艺术家和创作者能够以更直观的方式表达他们的创意和想象。

Somnai_dreams 作为该程序的开发者，通过结合艺术和技术的力量，为 AI 绘画领域带来了新的创新和突破。图 1-13 为 Disco Diffusion 程序的界面。

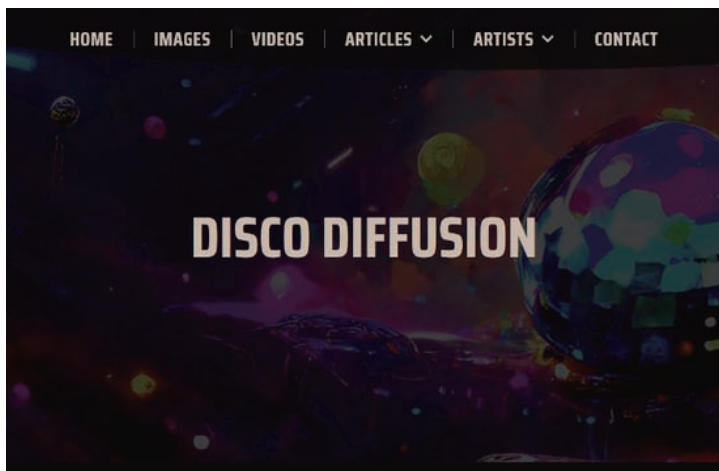


图 1-13 Disco Diffusion

(15) 2022 年 4 月，著名人工智能团队 OpenAI 发布了新一代的模型，名为 DALL-E 2.0。该名称来源于著名画家达利 Dalí 和电影《机器人总动员》（*Wall-E*）。同样类似于前一代的 DALL-E 模型，DALL-E 2.0 也具备从文本描述生成效果良好的图像的能力。DALL-E 2.0 在继承了前一代模型的基础上进行了改进和优化，以提供更高质量、更多样化的图像生成结果。

(16) 2022 年 4 月，人工智能 Midjourney 邀请内测。由 Midjourney 创作的《太空歌剧院》作品一度引起了轰动，并在美国科罗拉多州举办的新兴数字艺术家竞赛中荣获“数字艺术 / 数字修饰照片”类别的一等奖。《太空歌剧院》的获奖彰显了人工智能在数字艺术领域的潜力和创造力。Midjourney 的创作展示了人工智能在图像处理和艺术创作方面的能力。图 1-14 为 Midjourney 创作的《太空歌剧院》。



图 1-14 《太空歌剧院》（2022）

(17) 在 2022 年的 5 月和 6 月，Google 发布了两个重要的模型，分别是 Imagen 和 Parti，并开放了相关的论文。Imagen 模型和 Parti 模型都代表了人工智能图像处理领域的前

沿技术，它们在图像生成、图像分割、图像处理等方面具有重要的应用价值。Google 的开放论文也为学术界和研究人员提供了宝贵的参考和研究资源。

(18) 在 2022 年的 8 月，Stable Diffusion 模型开源。Stable Diffusion 是一个重要的 AI 绘画模型，通过扩散化和潜在空间的技术，实现了高质量图像的生成。该模型的开源使更多的研究人员和开发者能够了解和应用这一先进的 AI 绘画技术，促进了 AI 绘画领域的进一步发展和创新。这一开源的举措为艺术家和创作者提供了更多的工具和资源，推动了 AI 在艺术创作中的应用和探索。

(19) 2022 年 8 月 26 日，基于家用 GPU 的训练模型 Dreambooth 正式宣布问世。12 天后，该模型的开源端口也被公布出来。随后的 25 天，Dreambooth 的训练所需的内存空间降低了整整 79%。到了 10 月 8 日，Dreambooth 已经能够在仅有 8GB 的 GPU 上进行训练。这些进展意味着 Dreambooth 模型在训练过程中对硬件资源的需求大大降低，使更多的个人用户和研究者能够在家用 GPU 上使用和训练该模型。Dreambooth 的出现为 AI 绘画领域带来了更加便捷和高效的训练方案，推动了 AI 艺术创作的普及和发展。

(20) 2023 年 2 月，Stable Diffusion 基于图像精确控制的 ControlNet 发布。

(21) 2023 年 3 月，Midjourney v5 正式发布。

(22) 2023 年 5 月，著名的图像软件公司 Adobe 发布了 Firefly。

人工智能绘画（AI 绘画）作为一个充满探索和交流氛围的领域，将会在技术的不断发展和应用中不断取得进步。随着人工智能技术的不断成熟和进步，我们可以期待人工智能在艺术领域发挥更加重要的作用。

人工智能绘画不仅为艺术家和创作者提供了新的工具和资源，还激发了更多的创新和创造力。通过人工智能的算法和模型，我们能够以更加智能和高效的方式进行艺术创作，探索出更多新颖、独特的艺术表达形式。

未来，人工智能绘画有望在艺术领域带来更多的创新和发展。它将成为艺术家们的合作伙伴和创作工具，为艺术作品注入新的灵感和想象力。我们可以期待在人工智能的帮助下，艺术领域将迎来更多的突破和进步。

1.2 人工智能绘画的算法和原理

无论是 DALL-E 2、Midjourney 还是 Stable Diffusion，它们的主要算法和原理都基于扩散模型，并且它们之间也存在千丝万缕的联系。

人工智能是一种模拟人脑神经网络的技术。通过训练，它可以学习各种任务，比如绘画。当我们让 AI 学习绘画时，它会结合文字进行训练。通过大量填鸭式的训练，在某个时刻，它会突然领悟，能够根据文字要求进行绘画，并且它的绘画具有类似人类的逻辑性，它的能力得到了快速提升。

目前，扩散模型是最常用的 AI 生成图像的方法之一。扩散模型基于非平衡热力学，这是热力学的一个分支，专门研究不处于热力学平衡中的物理系统。一个典型的例子是一滴墨水在水中扩散。在墨水开始扩散之前，它会在水中某个地方形成一个大的斑点。如果要模拟墨水开始扩散前的初始状态概率分布，将会非常困难，因为这个分布非常复杂，很难进行采样。然而，随着墨水扩散到水中，水逐渐变成淡蓝色，墨水分子会更加简单和均匀地分布。此时，我们可以使用数学公式来描述其中的概率分布。非平衡热力学可以描述墨水扩散过程中每一步的概率分布。由于扩散过程的每一步都是可逆的，因此只要步长足够小，我们就可以从简单的分布中推断出最初的复杂分布。

通过墨水的例子，我们可以得到一个启示：如果我们使用像素图对模型进行存储，那么将需要大量的硬件资源；相反，如果我们使用高斯噪点图进行存储，那么采样分布将更加容易。我们可以轻松地使用数学公式描述高斯分布的概率，并从简单的高斯分布中进行采样。因此，模型库可以采用高斯噪点图进行存储，这样反转过程也相对容易。这种方法在节省存储空间的同时，还能够保留模型的重要特征。

为了便于理解，我们再举一个例子，扩散模型类似于乐高玩具，打散后就是一个个的小方块，相当于变成噪点，小方块组装成各种各样的造型，类似于去噪的过程。

扩散模型可以将材质、色彩、光影、位置关系、透视关系、风格、笔触等视觉元素转换为某个标记，并将这些标记储存在高斯噪点图中的空间位置。

应用于 AI 绘画的过程中，可以通过逐步添加高斯噪声（模拟墨滴在水中扩散）来处理一幅图像。最终，这幅图像会变成高斯分布（模拟墨滴最终均匀扩散到水中）。高斯分布是一种非常容易建模和采样的概率分布，因此它在 AI 绘画的训练过程中起到了重要作用。推理过程则是将这个过程逆向进行，从一个均匀分布的高斯分布中进行采样，并逐步去除噪声，最终得到一幅完整的图像，这也就是将墨水扩散的过程进行逆转的过程。

图 1-15 展示了扩散模型的加噪与去噪的过程。扩散模型的原理包括两个步骤：首先是正向扩散，逐渐给图像添加高斯噪声，直到获得纯噪声的图像；然后，通过训练一个神经网络进行图像去噪，从纯噪声的图像开始，直到获得最终的图像。

那么，AI 绘画具体是怎么工作的呢？以 Stable Diffusion 为例，首先输入提示词，如“戴眼镜的少女”，然后 Stable Diffusion 开始工作，主要分为三个部分，如图 1-16 所示。

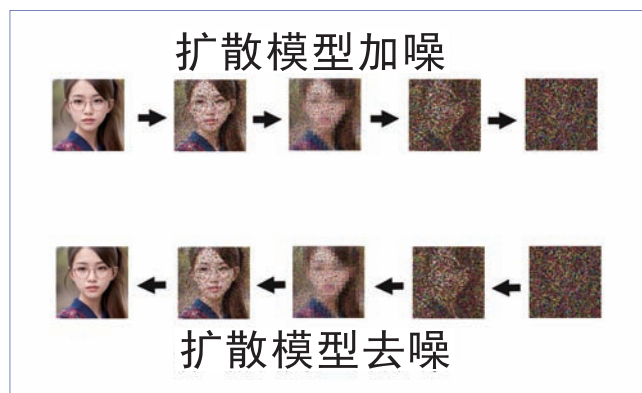


图 1-15 扩散模型的加噪与去噪

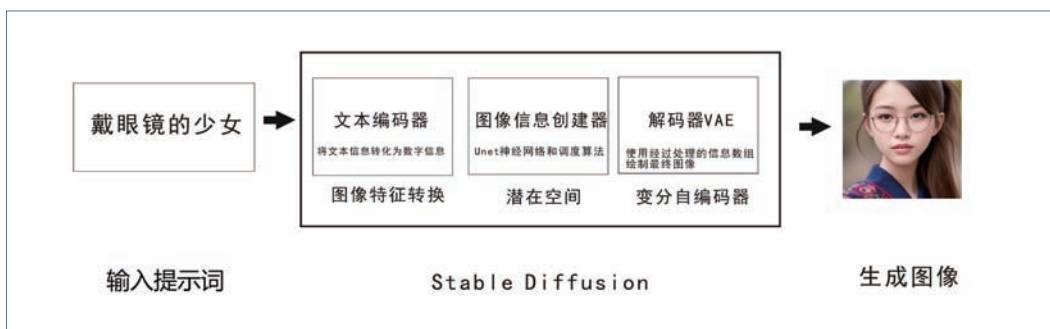


图 1-16 Stable Diffusion 原理

- 第一部分：文本编码。通过图像转换特征，把文本转换为数字信息，并提取出关键标记，如眼镜、少女、金属、年轻等。
- 第二部分：潜在空间生成。使用图像信息生成器，主要使用 U-net 调度算法生成图像。
- 第三部分：变分自编码器（VAE）编码。通过 VAE 图像解码器，把潜在空间 64×64 像素的图像解码成 512×512 像素的图像，从而绘制出训练图像。

注意 给 AI 绘图提供文本的提示词也被称为指令，被用户戏称为“咒语”。

通过 Stable Diffusion 的计算，可以在 WebUI 界面上生成直观的图像。

AI 绘画的过程实际上比描述的更加复杂。作为艺术工作者，我们很难像工程师一样深入了解其具体编码过程。然而，熟悉其基本原理对我们以后无论是生成图像、训练模型还是使用 ControlNet 工具编辑图像都会有很大帮助。这种了解有助于我们更好地创作和提高操作 AI 绘画工具的能力。

1.3 人工智能绘画的工具

目前在 AI 绘画领域，有一些比较知名的工具和模型，如 Stable Diffusion、Midjourney、DALL-E 2、Adobe Firefly 等。同时，国内的文心一格也已经开发并开始测试。这些工具和模型都在 AI 绘画领域有一定的影响力，提供了创作和编辑图像的功能，为艺术家和创作者带来了更多的可能性和创作灵感。随着技术的不断发展，预计会有更多创新的工具和模型涌现，推动 AI 绘画领域的进一步发展。

1 Midjourney

Midjourney 是目前用户较多的 AI 绘画工具。它在 2022 年 4 月推出了第 2 版，7 月 25 日发布了第 3 版，11 月 5 日又发布了第 4 版的 alpha 迭代版供用户使用。而在 2023 年 3 月 15 日，第 5 版的 alpha 迭代版也已经发布，它的写实作品“中国情侣”（见图 1-17）在国内风靡一时。

Midjourney 以其效果良好、简单易上手的特点受到用户的喜爱。它经过模型优化，即使在提示词较为简单的情况下，仍能产生出较好的艺术效果。

此外，Midjourney 还提供了一系列功能，如图像放大 (Upscale)、图像变体 (Variation)、定向修改 (Remix)、图像提示 (Image prompt)、与机器人私聊生成图像 (DM to Bot) 以及个人画廊手机版 (Gallery) 等，并且还在不断进步和创新。

2023 年 5 月，Midjourney 官方宣布将推出中文版，并在 QQ 频道上进行内测，邀请中国创作者参与测试。这次内测将不再需要登录国外服务器，支持中文提示词，为中国用户带来了便利。

2 DALL-E 2

DALL-E 2 是 OpenAI 开发的 DALL-E 的升级版。其名称是由西班牙著名艺术家 Salvador Dalí 和广受欢迎的皮克斯动画机器人 Wall-E 组合而来。图 1-18 为 DALL-E 官网的部分作品。



图 1-17 中国情侣

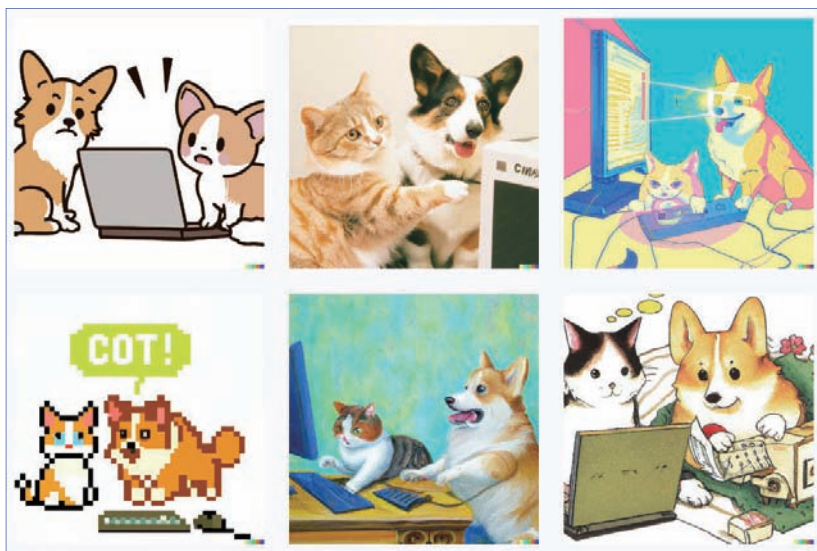


图 1-18 DALL-E 官网作品

2022 年 7 月，DALL-E 2 进入了测试阶段，用户登录 DALL-E 2 的官网并创建账户就可以进行体验。在这个平台上，我们可以使用不超过 400 个字符的描述性文本，让 AI 艺术生成器处理并生成相应的图像。这种体验并非免费的，DALL-E 的用户每月只能免费生成 15 幅图像，额外的生成需要付费。

值得注意的是，DALL-E 的输出经过了过滤，以避免生成包含暴力、裸露或逼真人脸的图像。此外，DALL-E 的生成范围也被限定，不会创建公众人物或名人的图像。

3 Adobe Firefly

Adobe Firefly 是 Adobe 公司旗下的生成式 AI 工具，已经集成到 Photoshop 中，正版用户可以使用 AI 进行创作。Adobe 公司在全球范围内拥有大量的用户群体，而且他们表示 AI 引擎将来还将整合到他们开发的剪辑和合成软件中。图 1-19 为 Firefly 网页。

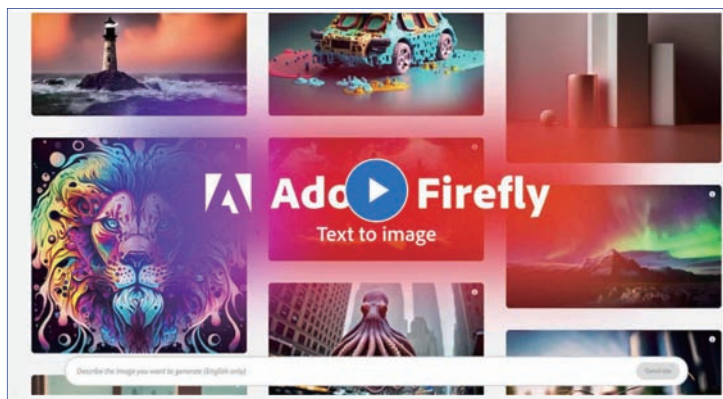


图 1-19 Firefly 网页

4 文心一格

文心一格是百度公司开发的产品，它是基于文心大模型的文生图系统实现的人工智能绘图工具。2022年8月19日，在中国图象图形大会（CCIG 2022）上正式发布了文心一格，这是百度利用飞桨（PaddlePaddle）和文心大模型技术创新推出的首款“AI作画”产品。图 1-20 为文心一格的部分测试作品。

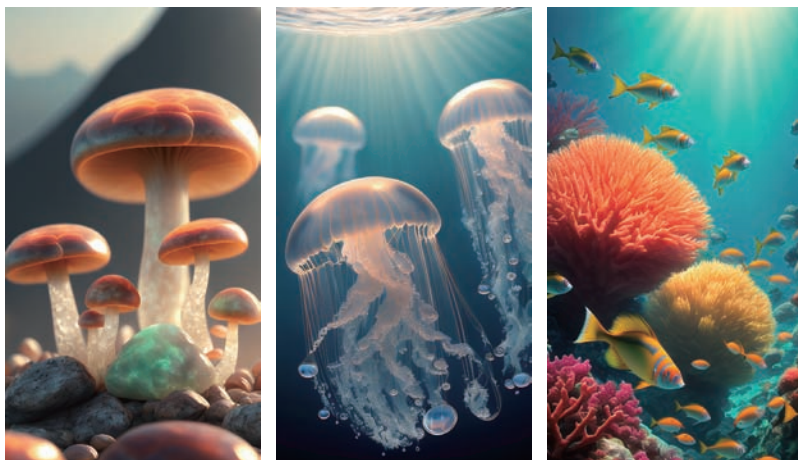


图 1-20 文心一格的测试作品

在文心一格平台上，用户可以通过每天签到和发布作品来获得积分。这些积分可以用于创作，并与其他创作者进行交流和分享。

1.4 Stable Diffusion 简介

Stable Diffusion 在 2022 年 8 月开源，是由慕尼黑大学的 CompVis 研究团队开发的生成式人工神经网络。该项目由初创公司 StabilityAI、CompVis 和 Runway 合作开发，并得到了 EleutherAI 和 LAION 的支持。截至 2022 年 10 月，StabilityAI 已筹集了 1.01 亿美元的资金。Stable-Diffusion-WebUI 是一个能够在浏览器上运行的网页版，它是一个具有跨时代意义的产品，让普通用户能够真正体验到 AI 绘画的无限魅力。图 1-21 为 Stable Diffusion 生成的作品。

Stable Diffusion 具有以下特点：

（1）发展潜力巨大：Stable Diffusion 不仅为专业人士提供高效、精确和富有创意的工具和资源，还让没有美术基础的普通人通过提示词创作出令人惊叹的作品。传统绘画需要长时间的练习和技能积累，而 AI 绘画技术利用机器学习和深度学习算法，能更轻松地表达创意和想法。



图 1-21 Stable Diffusion 生成的作品

（2）丰富的功能：Stable Diffusion 不仅可以通过文字生成图像，还可以通过素材图生成新的图像，用户还可以在软件中训练自己的模型。此外，Stable Diffusion 的功能在逐步提升，不断地加入新的插件和脚本。

（3）多样的模型：作为开源软件，Stable Diffusion 拥有众多爱好者训练的模型，这些模型可以在平台上发布供用户下载和测试。

（4）强大的可控性：2023 年 2 月，精确控制插件 ControlNet 的推出拓展了 Stable Diffusion 的功能，使它能更好地解决 AI 绘画过程中的随机性问题。

（5）免费单机使用：Stable Diffusion 的免费单机使用是其最大的优势，用户可以通过不断探索和使用该工具，从大量的作品中找到自己喜欢的作品。

（6）开放性带来质疑：尽管开放性使得 Stable Diffusion 在改进速度上超过其他竞争对手，但也带来了一些质疑，例如内容过于敏感，有暴力问题和版权问题。它受到 AI 艺术界的赞扬，但也受到传统艺术家的批评。有人喜欢它，有人对它持有批评态度。无论如何，AI 已经打开了大门，让我们紧跟时代的步伐，不断向前发展。

1.5 思考与练习

思考题：什么是人工智能绘画（AI 绘画），它与普通绘画有什么区别？

练习题：请访问文心一格网站，尝试画一幅 AI 绘图作品。