



# 导 论

## 1.1 AIGC 艺术设计基础

如果说 2018 年 AI 绘画作品《艾德蒙·贝拉米》的拍卖成功，还可以归功于人工智能艺术的猎奇标签和高拍卖价格带来的轰动效应，那么到了 2022 年，聊天机器人 ChatGPT 的亮相给公众带来的冲击，则与 1997 年 IBM 的“深蓝”战胜国际象棋冠军卡斯帕罗夫时并无二致。这些事件往往只是成为人们指尖一滑而过的新闻推送，或是社交场合中的闲聊话题。大多数人在短暂关注和讨论后，并不会将自己代入“被机器超越的人类”这一角色之中。但是，像 Midjourney 和 Stable Diffusion 这样商用 AIGC 绘画平台的出现，却真正让所有人，特别是那些以绘画为职业的艺术设计师们，感受到了紧迫的威胁。这不仅仅是人类某项技能被机器取代的问题，而是人工智能触及了人类自豪地认为不可替代的灵感、激情、直觉和创造力等核心精神领域。尽管目前 AIGC 尚无法完全代替人类艺术家和设计师，即使 AI 绘画仍只能在视觉文化内容中占据一小部分，但只要人们开始认可人工智能绘画产品，可以在质量和速度上与人类设计师进行比较，商业美术设计行业、企业、个人作品乃至整个媒介文化都将不可避免地进入一个不可逆转的进程——迈向一个 AIGC 艺术的新时代。

### 1.1.1 AIGC 简介

AIGC (AI generated content, 人工智能生成内容) 是继专业生成内容 (professional generated content, PGC)、用户生成内容 (user generated content, UGC) 之后, 一种更为高效、便捷的新型内容创作方式。它利用人工智能技术, 特别是生成对抗网络 (GAN)、大型预训练模型 (如 Transformer)、变分自编码器 (VAEs), 以及其他深度学习框架来创造文本、图像、音视频、代码、3D 模型等多种媒介形式的内容。这些技术通过模拟人类的创意过程, 不仅能够复现已存在的风格和格式, 还能够创造全新的、未曾存在的创意作品。

AIGC 的技术基础深植于机器学习的两大主要模型: 判别式模型和生成式模型。判别式模型专注于识别和分类数据, 例如识别图像中的对象或对电子邮件进行垃圾邮件分类。相比之下, 生成式模型则致力于生成全新的、看似真实的数据实例。这些模型不仅具有学习数据的特定特征, 还能基于这些学习到的特征产生新的数据点, 这使得它们特别适合内容生成。

生成对抗网络 (GAN) 的引入是 AIGC 领域的一个重要里程碑。GAN 由两部分组成: 一个生成器和一个判别器。生成器的任务是创造逼真的数据实例 (如图像), 而判别器的任务则是区分生成的数据和真实的数据。通过这种对抗性训练, 生成器学习如何产生越来越逼真的数据, 推动了模型在艺术创作、新媒体设计等领域的应用。随着时间的推移, AIGC 已从简单的文本和图像生成, 演进到复杂的音视频合成和 3D 环境模拟。此外, 变分自编码器 (VAEs) 和基于 Transformer 的模型等技术已经被开发和优化, 以处理更大规模的数据集和更复杂的生成任务。2022 年, 扩散模型 (diffusion model) 的兴起为 AIGC 技术带来了重大创新, 通过结合前向扩散与反向生成的过程, 有效地提升了图文生成的效率, 使其成为 AIGC 研究中的一个热点领域。这些技术的进步不仅提升了生成内容的质量, 也极大地扩展了 AIGC 的应用。

AIGC 现已覆盖多个领域, 从艺术和创意产业的自动化创作, 如 AI 绘画和音乐创作, 到专业领域如法律文档自动生成、新闻稿件编写以及个性化教育内容的创建。在商业领域, AIGC 技术支持广告创意生成、社交媒体内容制作和用户体验的个性化。此外, 随着技术的普及, AIGC 也开始在游戏开发、电影制作和虚拟现实等娱乐领域中扮演重要角色。

AIGC 的优势在于其生成内容的高效率和低成本, 这正在改变传统内容产业的格局。随着 AIGC 的快速发展, 接踵而来的挑战也日益凸显。内容的原创性、版权归属和道德责任成为亟待解决的问题。此外, 随着生成内容的质量和逼真度日益提高, 区分人工和 AI 生成内容的难度也在增加, 这对传统的内容审核和监管机制提出了新的挑战。未来, AIGC 的发展需要在创新与责任之间找到平衡, 确保其对社会产生积极影响。理想的人工智能应当与人类创作者合作, 而非取代他们, 以确保技术促进而不是破坏人类的文化和艺术价值。这种平衡是技术发展的必由之路, 也体现了科学技术对未来社会的深刻责任感。

## 1.1.2 AIGC 主流工具和软件平台

随着人工智能技术的不断进步，AIGC 已经成为创意产业和多个专业领域中不可或缺的驱动力。全球各地的开发者和企业纷纷推出了一系列创新的 AIGC 工具和平台。在文本生成领域，有 ChatGPT、Claude、Gemini、文心一言等；在图片生成领域，有 Stable Diffusion、Midjourney、DALL-E、文心一格等；在视频生成领域，有 Sora、Runway、Stable Video Diffusion、Pika 等；在音乐生成领域，有 Suno、Stable Audio 等；在数字人领域，有 HeyGen、D-ID、阿里巴巴的 EMO 等。此外，AIGC 也在编程开发、企业运营、办公、教育训练等其他领域展示其创新潜力，这些工具和平台正在塑造内容创作的未来。

本节将介绍几款领先且成熟的 AIGC 工具和软件平台，这些平台已成为设计师、艺术家不可或缺的资源。我们将探讨如 OpenAI 的 Chat GPT、DALL-E 和 Sora，以及 Stable Diffusion、Midjourney 和 Runway 等工具的功能和应用领域。通过了解这些 AIGC 工具的技术细节和实际应用，我们可以更好地评估它们对未来内容创作领域的潜在影响，并探讨这些技术如何继续影响创意产业的发展方向。

### 1. ChatGPT

ChatGPT 是 OpenAI 开发的一款聊天机器人，基于大型语言模型（LLM），允许用户精细控制对话的长度、格式、风格、细节程度和语言。ChatGPT 自从推出以来，其用户迅速增长，到 2023 年 1 月，该平台已成为历史上增长最快的消费软件应用，用户超过 1 亿。ChatGPT 的发布也迫使其他竞争产品加紧推出和更新。

此外，ChatGPT 在对话应用中，结合了监督学习和来自人类反馈的强化学习。尽管核心功能是模拟人类交谈者，但 ChatGPT 的应用非常多样化，包括写作和调试计算机程序、创作音乐、翻译和总结文本、模拟 Linux 系统等。ChatGPT 可像搜索引擎一样提供信息搜索和建议，例如解释笑话之所以有趣的原因，或提出解决特定编程错误的建议。作为一个文本生成器，它能够创建大量清晰、逻辑性强但可能较为普通的文本。此外，它可以帮助创作者克服创作障碍，提供灵感并协助修改和扩充内容。根据 ChatGPT 自身的描述，“它是一个适用于任何任务的工具，既智能又快速；从编写笑话到撰写文章，展现出极大的灵活性”。但 ChatGPT 也有其限制，如有时会产生听起来合理但实际上错误或无意义的回答，这种现象被称为“幻觉”。

ChatGPT 已经集成了对插件的支持，包括 OpenAI 自己制作的插件和第三方开发的插件，如 WebPilot、Wolfram、Code Guru 等。它也支持通过 API 与其他应用程序集成，扩展了应用场景和功能。ChatGPT 操作界面如图 1-1 所示。

### 2. DALL-E

DALL-E 是由 OpenAI 开发的一系列文本到图像模型，包括 DALL-E、DALL-E 2 和 DALL-E 3。这些模型使用深度学习方法从自然语言描述中生成数字图像。DALL-E 的第一个版本在 2021 年 1 月发布，其后继产品 DALL-E 2 于 2022 年发布，能够生成更真实、更高分辨率的图像。最新的 DALL-E 3 于 2023 年发布，集成进了 ChatGPT Plus 和 ChatGPT

#### 4 AIGC 三维艺术设计

Enterprise。该模型名称的灵感来源于超现实主义艺术家萨尔瓦多·达利和皮克斯动画中的机器人 WALL-E，象征其在艺术与科技融合方面的创新。

DALL-E 使用的是 GPT-3 模型的多模态版本，能够理解和生成与给定文本描述相符的图像。这些模型不仅可以生成写实图像，还能创造带有绘画风格的图像或表情符号。DALL-E 能够在没有明确指令的情况下，正确地在图像中放置设计元素，表现出广泛的对视觉和设计趋势的理解。DALL-E 在多个领域中都有应用，包括数字艺术创作、广告设计、媒体内容生成等。由于其在视觉内容创作方面极为出色的能力，它常被用于生成独特的艺术作品和商业图像，帮助设计师和内容创作者扩展他们的创意边界。DALL-E 演示效果如图 1-2 所示。

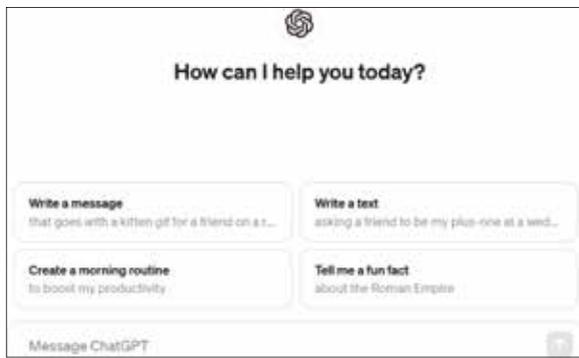


图 1-1  
ChatGPT 操作界面



图 1-2  
DALL-E 演示效果

### 3. Sora

Sora 是由 OpenAI 开发并于 2024 年 2 月在美国正式发布的先进 AI 视频生成模型。该模型名字来源于日语单词“空”，意指其无限的创造潜能。Sora 基于 OpenAI 早期的文本到图像模型 DALL-E 进行开发，能够根据用户的文本描述生成长达 60s 的高真实感视频。Sora 的开发利用基于扩散转换器（diffusion transformer）的技术，这是一种去噪潜在扩散模型，在潜在空间中生成视频后，将其转换到标准空间。这种方法让 Sora 能够生成视觉细节丰富的视频内容，包括复杂的摄像机运动和表情丰富的角色，并能无缝扩展现有短视频，生成前后衔接的新内容。

Sora 继承了 DALL-E 3 的高画质和对指令的精确响应能力，可以准确捕捉用户在文本提示中的详细要求。该模型的推出，为视频制作领域的艺术家、电影制作人员和学生等提供了广阔的创作空间，标志着 AI 在模拟和理解动态物理世界方面的显著进步。Sora 的发布不仅展示了其在复杂场景生成方面的能力，还建立了让 AI 通过学习理解模拟现实世界场景方面重要的里程碑。

尽管 Sora 尚未提供给公众使用，但从 OpenAI 放出的信息来看，用户将能通过简单的文本提示来生成视频，这使得非专业用户也能轻松创建复杂的视频内容。OpenAI 官方已通过演示体现其生成视频的能力，如图 1-3 所示。



图 1-3  
OpenAI 官方演示 Sora 生成视  
频的能力

#### 4. Stable Diffusion

Stable Diffusion 是由 Stability AI 开发的一种深度学习文本到图像生成模型，首次发布于 2022 年。该模型基于扩散技术，特别是潜在扩散模型（latent diffusion model, LDM），能够根据文本描述生成详细的图像。它的开发涉及慕尼黑大学的 CompVis 组和 Runway，以及来自 Stability AI 的捐赠和非营利组织提供的训练数据。

Stable Diffusion 的特点是其代码和模型权重已经公开发布，使其能在大多数具有适中 GPU 性能支持的消费级硬件上运行。这体现了 Stable Diffusion 与以往只能通过云服务访问的专有文本到图像模型（如 DALL-E 和 Midjourney）的显著区别。

Stable Diffusion 使用的 LDM 架构通过在训练图像上连续移除高斯噪声（视为一系列去噪自编码器的过程）来训练。模型包含三部分：变分自编码器（VAE）、U-Net 和一个可选的文本编码器。VAE 将图像从像素空间压缩到较小的潜在空间，捕获图像更本质的语义。随后，高斯噪声被迭代地应用到压缩的潜在表示上，最后，U-Net 通过反向扩散过程去除噪声，再由 VAE 解码器将潜在表示转换回像素空间，生成最终图像。

此模型支持通过文本提示生成新图像，也允许对现有图像进行重绘，来包含由文本提示描述的新元素。此外，还可以通过修图和扩图等方式，部分修改现有图像。目前 Stable Diffusion 有两种用户界面，分别是 Web UI 和 Comfy UI。Web UI 是一个基于网络的界面，允许用户通过简单的文本输入来生成图像。这个界面通常托管在云服务器上，用户可以通过任何标准的网络浏览器访问它。Comfy UI 是 Stable Diffusion 的另一种用户界面，它提供了一些额外的功能和定制选项，使得用户体验更加“舒适”和个性化。这两种界面都是为了让用户能更容易地接触和使用 Stable Diffusion 模型，无论是通过直接在网页上操作，还是通过更为复杂和功能丰富的节点化界面进行高级图像生成和编辑。用户可以根据自己的需求和偏好在二者之间进行选择。Stable Diffusion 的 Web UI 操作界面如图 1-4 所示。

#### 5. Midjourney

Midjourney 是由位于旧金山的独立研究实验室 Midjourney, Inc. 创建和托管的生成式人工智能程序和相关服务。该工具由 Leap Motion 的联合创始人，David Holz 领导的团队



图 1-4  
Stable Diffusion 的 Web UI 操作界面

开发。Midjourney 通过 Discord 机器人命令生成图片，用户可以使用文本描述（提示词）来生成图像，类似于 OpenAI 的 DALL-E 和 Stability AI 的 Stable Diffusion。它能够快速实现概念艺术原型设计，并在开始实际工作之前向客户展示。此外，Midjourney 经常更新其算法版本，不断改进其图像生成能力。Midjourney 的使用展示如图 1-5 所示。

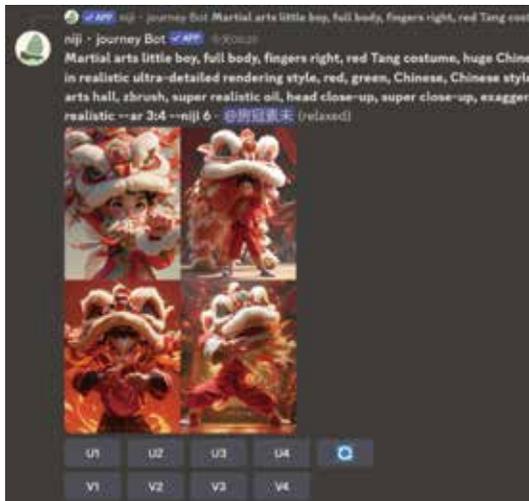


图 1-5  
Midjourney 的使用展示

Midjourney 被广泛应用于艺术和设计领域，帮助艺术家、设计师和创意专业人员快速生成创意视觉内容。它也被广告行业用于创建原创内容和快速头脑风暴，为广告制作提供新的可能性。Midjourney 主要通过 Discord 平台访问，用户通过发送特定的命令来激活图像生成过程。这种接入方式使得 Midjourney 可以轻松集成到现有的社交和工作流程中，为用户提供便捷的图像创作工具。同时，其持续更新的算法保证了图像质量和创新性，推动了创意产业的发展。

## 6. Runway

Runway 是一个由人工智能推动的创意平台，致力于通过提供一系列广泛的 AI 工具和功能来革新艺术、娱乐和设计领域。该平台为用户从概念到成品提供多种创意工具，帮助将创意转化为现实。Runway 特别适合电影制作人、视频编辑、数字艺术家、设计师，以及教育工作者。

Runway 的研究工作涉及 Gen-1 和 Gen-2 的先进 AI 生成技术，以及其他关于图像合成、视频合成、计算机视觉和音频生成的领域。Gen-1 技术可以使用文本和图像从现有视频中生成新视频，而 Gen-2 则允许用户根据文本提示生成任何风格的视频。此外，Runway 提供了包括文本到图像、音频处理、视频抠图等在内的全面 AI 工具，并基于这些工具开发了一个以协作和速度为卖点的 Web 端视频编辑软件，旨在提升视频后期编辑和特效制作的效率。Runway 的产品理念是始终保持 AI 技术发展的最前沿，探索图像和视频编辑的全新方法。他们的目标是利用云计算和 AI 的力量定义新的工作流程，而不仅仅是提供一个更优的 Photoshop 或 Premiere 替代品。Runway 旨在成为未来内容创作的核心工具，尤其是在视频创作领域。Runway 的官网界面如图 1-6 所示。

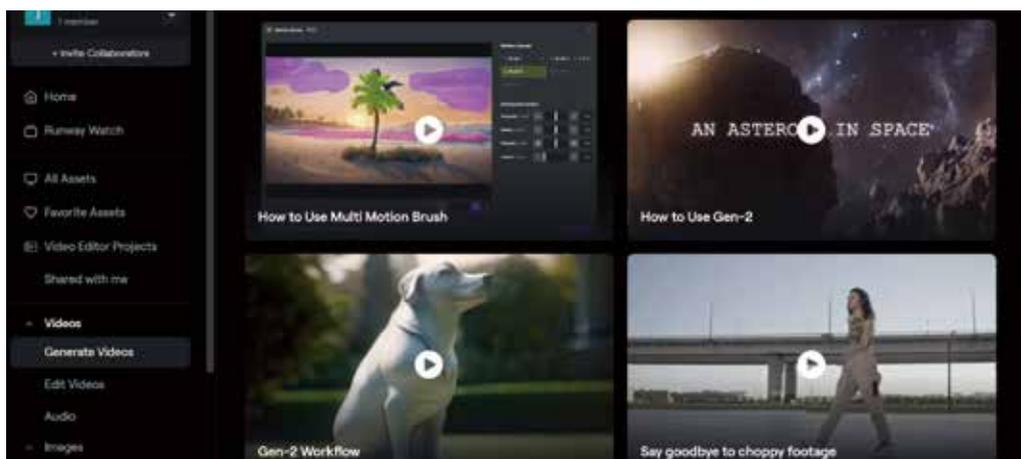


图 1-6  
Runway 的官网界面

AIGC 作为一种新型内容创作方式，正在重塑内容产业的格局。AIGC 衍生出许多平台，在多个领域提供了强大的创意工具。通过了解和应用这些先进的 AIGC 工具和平台，我们能够更好地把握这一技术潮流，为创意产业和其他专业领域带来更多创新和机遇。同时，关注技术伦理和社会责任，将确保 AIGC 的健康发展，促进社会的整体进步。

## 1.2 艺术与技术的融合：AIGC 与三维艺术的崭新时代

中国的游戏和影视产业正在迅速发展，涵盖虚拟仿真、文化创意产业等多个领域，对高质量三维艺术作品的需求持续增长。在游戏、影视剧和动画制作中，大量优秀的三维模

型、动画和特效得到了广泛应用。AIGC 在相关艺术制作领域已经突破静态图片生成，开始向动态视频和智能三维生成技术拓展。一方面，融入原有制作环节，如 AI 建模、AI 动作控制等；另一方面，也开始从渲染风格上突破原有三维艺术常规的美术表现，支撑甚至跨越部分需要漫长制作周期和大量人力的技法流程。

与直面 AIGC 挑战的绘画艺术相比，三维艺术设计和影视动画行业早已习惯与计算机图形学领域的前沿技术同步发展。相比于二维静态艺术“机器战胜人力”的传统工业神话，三维艺术家们目前还处在成熟的专业技术范式和工业化流程构筑起来的堡垒庇护之下。即便如此，AIGC 仍是从业者在原有的三维图形软硬件迭代之外，必须随时关注和积极融入的视觉内容生产技术大趋势。

### 1.2.1 AIGC 时代行业对三维艺术设计的需求现状

无论是动画还是 CG 电影，都对三维美术资源有大量的需求。尽管 AIGC 技术已经能够在一定程度上实现静态图像和视频画面的风格化生成，但其质量相比产业一线需求仍然存在较大差距。同时由于游戏交互、沉浸体验、模拟教学、文化资源记录和浏览等领域，对于视觉元素设计和技术实现的复杂性，三维形式和技术流程仍不可或缺。在现有的文化复兴、数字赋能文旅的大背景和趋势下，对三维美术资源的需求将长期存在。AIGC 技术在辅助三维视觉对象的实体化和艺术化方面，有望成为行业内越来越普遍的工作方式。这不仅能够提高生产效率，还能够推动创新，使艺术创作更加多元化和个性化。

#### 1. 岗位需求

和大部分数字艺术类似，三维艺术设计师的职业化，往往必须从属于动画 CG 或数字游戏这样的典型工业化流程。三维动画中期工作流，即建模、材质贴图、骨骼绑定、角色动画、灯光渲染等，基本涵盖了所有与三维内容生成有关的专业性需求。随着技能的专精和工作岗位的细化，每一个设计师往往只能胜任三维艺术制作的某一特定环节。依据商业化目的和专业性评价，最终的三维艺术作品几乎都是分工合作的产物。如今，尽管 AIGC 已经可以生成模拟三维美术风格的静态甚至动态内容。但鉴于三维艺术应用领域的多元化和专业性评价，AIGC 用于三维艺术的价值和可行性，目前还需要在既有的工作流中讨论。毋庸置疑的是，AIGC 融入现有工业化流程的专业岗位，可以代替机械的重复劳动，降低从业人员的技能门槛，使得更多的三维艺术家可以跨领域发展，甚至独立完成整部作品的创作。

#### 2. 应用前景

尽管目前标榜具有三维资源生成能力的 AIGC 平台，其生成质量大幅度落后于 AIGC 用于图片生成的质量，应用前景不尽如人意。但毕竟已经让 AI 技术发展出了机器学习和处理、生成三维内容的能力，形成了技术迭代的基础。以下介绍 AIGC 可用于现有三维工作流的方式和前景。

##### 1) AI 建模

(1) 目前人工智能可以用来自动生成三维模型，包括从二维图像或文本描述中生成模

型。现有 AI 建模平台生成作品的质量尽管相比行业应用需求仍存在一定的差距，但正在快速迭代优化发展中。

(2) 部分参数化建模平台，如虚幻引擎的 Metahuman 可通过真人照片或面部三维扫描数据，智能识别、匹配特征，快速搭建完整的数字虚拟人。

(3) 平面图形和短视频生成是目前 AIGC 发展比较成熟的领域，AI 建模可利用来自其他 AIGC 平台的二维内容来继续智能生成三维内容。

## 2) AI 绑定

(1) 人工智能可以通过对现实生物的运动学规律，以及对已有的三维角色控制效果的学习，来自动完成角色的绑定过程。目前 Adobe 的 Mixamo 平台即可使用较少的关键节点定义快速进行角色绑定。

(2) 目前部分 AIGC 建模平台也引入了自动绑定功能，使得三维制作流程的自动化程度进一步提高。基于文字或单张设计图，AIGC 平台就能生成可动三维角色，并提供可导入三维软件或游戏引擎的标准格式。

## 3) AI 动画

(1) 人工智能技术可以用于自动生成动画，或帮助艺术家进行运动捕捉，从而使动画角色的动作更加自然和流畅。目前很多 AI 动作和表情捕捉平台，允许用户使用单镜头的视频来解算出流畅且合理的动作表演数据。一些传统的人体动力学动画软件（如 Cascadeur）通过对人体动作规律的智能学习，支持用户用较少的关节操作即可获得符合运动习惯的全身姿态。同时也可以为简单的关键帧补间动画制作更自然的细节，如缓冲和次级跟随动作等。

(2) 一些 AI 动画平台具备提示词生成动画功能，可以根据用户输入的文字描述，利用存储的角色动作数据库，融合、拼接成符合描述连续动作表演。

## 4) AI 贴图

(1) 人工智能可以用来自动生成材质和纹理。多数 AI 建模平台都可以根据输入的设计图生成角色正面的贴图，同时补充设计图中原本不包含的，角色背面的贴图。部分平台也可以为用户上传的无材质模型，根据提示词生成贴图。

(2) 目前一些 AI 贴图独立软件或插件可以和 AIGC 图片生成功能联动，如将当前三维视图中的模型以深度信息采样，提供给 Stable Diffusion 这样的本地部署 AIGC 平台，用作控制信息输入。深度信息结合提示词生成的图片会被投射到三维模型上，形成与该角度匹配的局部贴图。在多个角度重复执行上述操作即可获得完整的 AI 贴图。

## 5) AI 渲染

(1) 视频的风格化渲染是伴随 AIGC 发展的典型应用领域。在可以生成模仿三维渲染风格的静态图片基础上，AIGC 平台可结合图生图或骨骼数据提取等手段，不经过建模、绑定、动画、材质等工序，直接生成动态三维风格视频。该领域的发展在未来有望颠覆动画 CG 领域的三维制作流程，建立新的行业标准和范式。

(2) 目前三维软件的风格化渲染效果，大多仍存在较明显的三维制作痕迹，如三维卡通渲染效果，尚无法完美模拟传统二维动画手绘造型和原动画表演规律。AIGC 基于图片

风格化的视频处理功能，可以为三维渲染结果加强手绘感，以及抽帧插值的动画效果。

(3) 以机器学习能力为核心的 AIGC 算法，既可以为简单的三维粗模快速生成带有完整细节的渲染预览效果；也可以分析已有美术作品的风格特征，并为三维渲染视频添加该风格的滤镜，从而实现任意视觉风格动态影像的三维制作。

### 1.2.2 AIGC 技术和三维艺术风格

2022 年，计算机艺术家格伦·马歇尔 (Glenn Marshall) 凭借其 AI 电影《乌鸦》(The Crow) 赢得了戛纳电影节短片竞赛单元评审团奖。这部作品通过使用计算机算法，将真人舞蹈视频转化为具有乌鸦外观特征的动画角色，模仿了舞蹈者的动作，创造出一段独特的视觉体验。这种兼具视频原型特征与抖动闪烁的视觉效果，代表了此后大多数 AI 动画的共同风格，如图 1-7 所示。

在 2023 年，特效团队 Corridor 采用了 Stable Diffusion 这一开源绘图 AI 技术，以逐帧处理的方式，将真人演员的实拍镜头转化为动画。他们参考了日本动画《吸血鬼猎人 D》的画面风格，创作出了动画短片《剪刀、石头、布》(ANIME ROCK, PAPER, SCISSORS)，如图 1-8 所示。这部短片以其稳定可控的视觉效果和出色的画面表现，迅速成为 AI 动画领域的一个热点，引起了广泛的讨论和关注。



图 1-7  
AIGC 电影短片《乌鸦》



图 1-8  
AIGC 动画《剪刀、石头、布》

迪士尼资深动画师亚伦·布莱斯 (Aaron Blaise) 在观看了这部短片及其制作花絮后，对 AIGC 技术介入动画制作流程的潜力和已经取得的成果表示了肯定和赞赏。他回忆起 20 世纪 90 年代三维 CG 技术刚刚进入动画制作领域时，许多动画师所感受到的不安和威胁。然而，作为一个经验丰富的动画师，亚伦·布莱斯面对动画技术的新一轮变革持有积极和开放的态度，并对新兴技术表示了祝贺和兴奋。

AIGC 技术在生成图像风格方面已经取得了显著进展，特别是在模拟真实材料和体积感的三维艺术创作上。一个突出的例子是图片处理软件 Remini 推出的“照片黏土滤镜”，该滤镜利用 AIGC 技术对图像进行风格化处理，将人物和景物转换成具有手工捏制黏土质感的美