



AIGC 分镜概述

1.1 AIGC 概述

1.1.1 AIGC 的定义及历史发展

AIGC (artificial intelligence generated content, 人工智能生成内容) 是一种新兴的内容生产方式, 是指利用生成式人工智能技术生成文本、图片、音频、视频等内容。2022 年, AIGC 技术发展迅猛, 得益于深度学习模型的完善、开源模式的推动和大模型商业化的探索, 其迭代速度呈指数级增长。人工智能在绘画和写作领域的成就, 如 AI 绘画作品获奖和超级聊天机器人 ChatGPT 的出现, 标志着智能创作时代的开启。

在人工智能的发展历程中, 机器创作曾被视为难以实现的目标, “创造力” 被视为人类独有的特质。然而, 随着技术的进步, 机器也被赋予了创造力, 引领我们进入智能创作的新时代。从机器学习到智能创造的演进, 以及从 PGC (专业生成内容)、UGC (用户生成内容) 到 AIGC 的转变, 标志着生产力的深刻变革, 这一变革将深刻影响我们的工作和生活方式。

AIGC 的历史发展大致经历以下三个阶段。

1. 萌芽及实验阶段 (20 世纪 50 年代—90 年代)

1950 年, 艾伦·图灵提出了著名的“图灵测试”, 这可以看作是 AIGC 概念早期的萌芽。由于该阶段的技术所限, AIGC 的应用范围较小, 直到 20 世纪 90 年代, 研究者开始

2 AIGC 动画分镜设计

在新闻、音乐、诗歌等方面进行人工智能的实验尝试，在自然语言处理（NLP）方面，研究者通过应用规则和语法知识构造语句。例如，通过制定规则和算法控制语法结构和词汇选择，生成语法正确的句子或段落。这一时期已经开始尝试使用规则生成技术自动撰写新闻稿件，这些稿件基于人工模板和对事实的语法处理，实现了新闻内容的部分自动化生成。

然而，这一阶段的 AIGC 技术受限于规则和模板，导致生成内容在个性化和创意性上有所欠缺，未能达到高度智能化和自主化。正如斯坦福大学计算机科学系教授莫妮卡·拉马努亚曼（Monica Ramaswamy）所指出的，“AIGC 技术需要进一步优化和智能化，以实现真正的内容生成和创新。”尽管存在局限，但这些早期尝试为 AIGC 技术的未来发展奠定了基础。

2. 大规模应用阶段（20 世纪 90 年代—21 世纪第二个十年）

2006 年深度学习算法取得进步，同时 GPU、CPU 等算力设备日益精进，互联网快速发展，为各类人工智能算法提供了海量数据进行训练，AIGC 技术正逐步实现大规模应用，并渗透到新闻、广告、音乐、电影、游戏等多个领域。AIGC 技术能够快速生成内容，且在很多情况下，生成的作品可与人类创作相媲美。这不仅极大提高了生产效率，还有效降低了成本。

业界领先的公司和机构纷纷投入重金和人才资源，2012 年微软公司展示了全自动同声传译系统，主要基于“深度神经网络”（deep neural network, DNN）自动将英文讲话内容通过语音识别等技术生成中文。计算机科学家和人工智能专家吴恩达（Andrew Ng）表示：“AIGC 可以帮助人类创造更多高质量的内容，并帮助人们更好地理解复杂的数据和信息。”随后，AIGC 技术逐步走向成熟和实用化，受到业界的广泛关注和应用。同时，不同领域的专家和学者对 AIGC 技术发展及其应用影响的讨论，也进一步凸显了这一技术的重要性和深远影响。

3. 快速发展阶段（21 世纪第二个十年至今）

在 AIGC 技术的第三阶段，主要表现在技术革新、算法迭代、产品创新、应用范围的拓展，以及市场规模和商业化的迅猛发展。

自 2014 年起，生成对抗网络（GAN）的提出及其后续的迭代更新，为 AIGC 技术的发展提供了坚实的技术基础。2018 年，OpenAI 推出了 GPT-1，这是首个基于变换器架构的生成式预训练模型，它在多种自然语言处理任务中展现了预训练与微调相结合的强大效能，为后续模型的发展奠定了基石。随后，OpenAI 在 2019 年发布了 GPT-2，2020 年推出了具有里程碑意义的 GPT-3 模型，2022 年发布了基于 GPT 模型的人工智能对话应用服务 ChatGPT，到了 2023 年，GPT-4 的发布进一步推动了 AIGC 技术的发展。

在图像和视频生成领域，2018 年 NVIDIA（英伟达）发布的 StyleGAN 模型能够自动生成图片，2019 年 DeepMind 发布的 DVD-GAN 模型能够生成连续视频，而 2021 年 OpenAI 推出的 DALL-E 及其迭代版本 DALL-E-2，则主要用于文本与图像的交互生成内容。Midjourney、Stable Diffusion、ComfyUI 的出现，标志着 AIGC 在图像和视频生成方面的技术日益成熟。

此外，AIGC 技术催生了众多创新产品和应用，如 AI 机器人、AI 建模、AI 动作捕捉等技术，它们正逐步融入内容生产领域，迅速改变着内容创作的格局。这些技术的发展不

仅提高了内容生产的效率，也为创作者提供了更多的创意空间和可能性。随着 AIGC 技术的不断进步，我们可以预见，未来的内容创作将更加多元化和个性化。

1.1.2 AIGC 分镜的模型训练

模型训练是一个反复迭代的过程，它要求我们不断地进行调整和优化，以实现最佳的模型性能。此外，模型生成的分镜图必须经过专业团队的细致审核和校正，以确保它们既符合制作标准，又能够体现艺术的意图和创意。

在 AIGC 分镜模型训练的实践中，虽然面对设备和专业知识的限制，但采用 LoRA (low-rank adaptation) 技术，我们可以高效地进行模型微调。LoRA 是一种专为大规模语言模型设计的微调技术，最初在自然语言处理领域得到应用，尤其适合对像 GPT-4 这样参数量巨大的模型进行微调。由于直接训练这些模型成本极高，LoRA 通过仅训练低秩矩阵来显著降低这一成本。

在 Stable Diffusion (SD) 模型中，LoRA 作为一个插件，允许通过训练少量数据来满足定制化的需求，例如特定的画风、知识产权 (IP) 或人物。与传统模型训练方法相比，LoRA 不需要改变原有的 SD 模型结构，而是通过注入低秩矩阵参数来调整生成风格，这种设计大幅减少了训练所需的资源，特别适合社区用户和个人开发者。

考虑到模型训练对计算机配置的特定要求，本节将提供两种训练 LoRA 的案例：一种是本地部署训练器，另一种是基于 Web-UI 的训练方法。对于本地部署，推荐的计算机配置至少应具备 8GB 及以上的显卡内存。如果配置不满足要求，可以选择基于 Web-UI 的训练方法。

(1) 本地部署训练：适用于具备较高配置的计算机，可以充分利用本地资源进行训练。

(2) Web-UI 训练：适合配置较低或希望简化训练流程的用户，通过图形用户界面进行操作。

接下来将详细介绍两种方法的具体配置要求、操作步骤和注意事项，以帮助读者根据自己的实际情况选择合适的训练方式。

1. 本地部署训练

1) 下载源码或整合包

为了训练使用 LoRA，需要自行下载安装 Stable-Diffusion-WebUI，以及训练 LoRA 的训练器，如图 1-1 所示。

2) 模型选择

准备对应 LoRA 风格的模型，例如训练分镜风格或者动漫人物，就需要找到相应的大模型，当作底模训练，按照如图 1-2 所示进行基础模型的选择。

3) 数据准备

(1) 图片数据。需要准备 10~100 张风格相同的图片，图片风格就是要训练的 LoRA 模型的风格。在训练 LoRA 模型之前，需要将所有的图像调整到同一尺寸。图像尺寸的选择取决于模型和数据，基于 SD-1.5 或 XL，一般选择图像大小为 512×512 像素。如果需要更清晰的画质，则可以修改成更高的分辨率，但相应的训练时长也会增加。将裁剪好的

4 AIGC 动画分镜设计

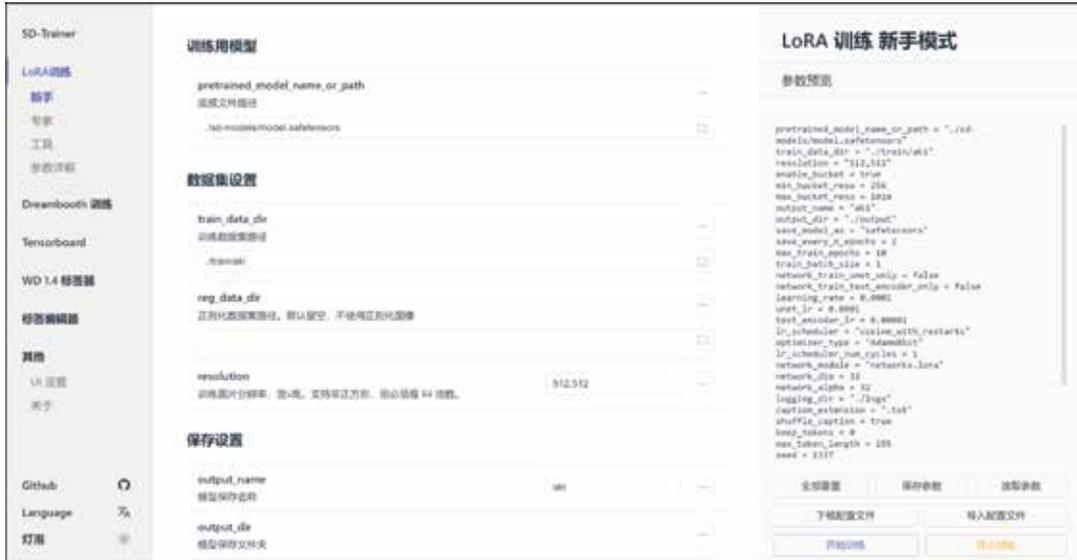


图 1-1
LoRA 训练器

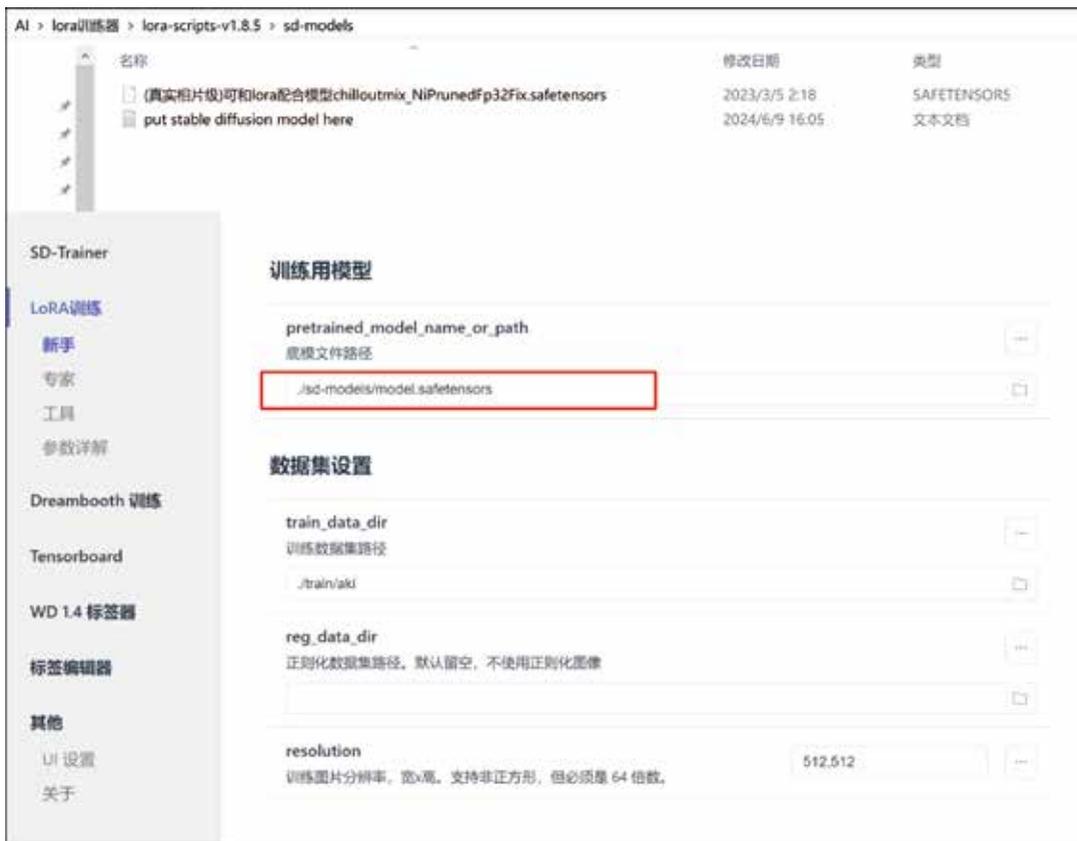


图 1-2
选择基础模型

图片放置在训练根目录文件夹中，如图 1-3 所示。

(2) 添加标签。在数据准备阶段，首先将裁剪好的图片地址复制到标签参数设置路径中，然后进行标签的设置。如果图片质量不够清晰，则可以使用 Stable Diffusion 的高清化模块，将图片高清化处理，图片质量比数量更重要；如果图片数据集不够优质，那么模型训练效果相应地也不会达到理想的效果，既浪费时间也消耗算力。在标签参数选项设置中可以根据训练画面内容添加提示词，如图 1-4 所示。



图 1-3
准备训练的图像

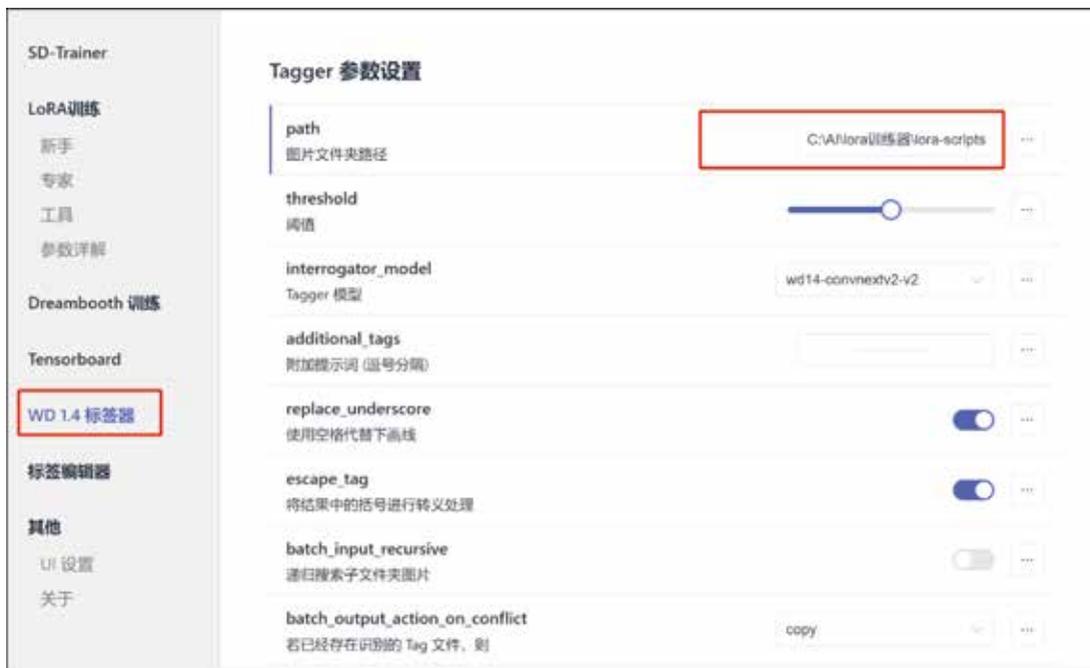


图 1-4
添加标签

6 AIGC 动画分镜设计

(3) 保存设置，开始训练。在使用训练器训练 LoRA 时，需要为新生成的 LoRa 模型设置保存名称和保存文件夹的位置，并设置分轮数保存模型、模型训练轮次，如 save_every_n_epochs 填写 2、训练 10 次（图 1-5），那么就会得到 4 个中间模型和一个最终模型。如果没有特殊的参数调整，则可以直接开始训练模型。

The image shows a web-based configuration interface for training a LoRA model. It is divided into two main sections: "保存设置" (Save Settings) and "训练相关参数" (Training Parameters).
In the "保存设置" section:
- "output_name" (模型保存名称) is set to "Test".
- "output_dir" (模型保存文件夹) is set to "./output".
- "save_every_n_epochs" (每 N epoch (轮) 自动保存一次模型) is set to 2.
In the "训练相关参数" section:
- "max_train_epochs" (最大训练 epoch (轮数)) is set to 10.
- "train_batch_size" (批量大小) is set to 1.

图 1-5
模型训练前的设置

2. Web-UI 训练

第二种训练方式是在 AIGC 平台使用 Web-UI 训练 LoRA，无须特别设置参数，适合计算机配置普通的使用者，只需把收集好的图片上传即可。以 LibLib 为例，登录该平台，在首页中选择训练我的 LoRA，进入如图 1-6 所示的界面，在这一界面中上传图片数据。

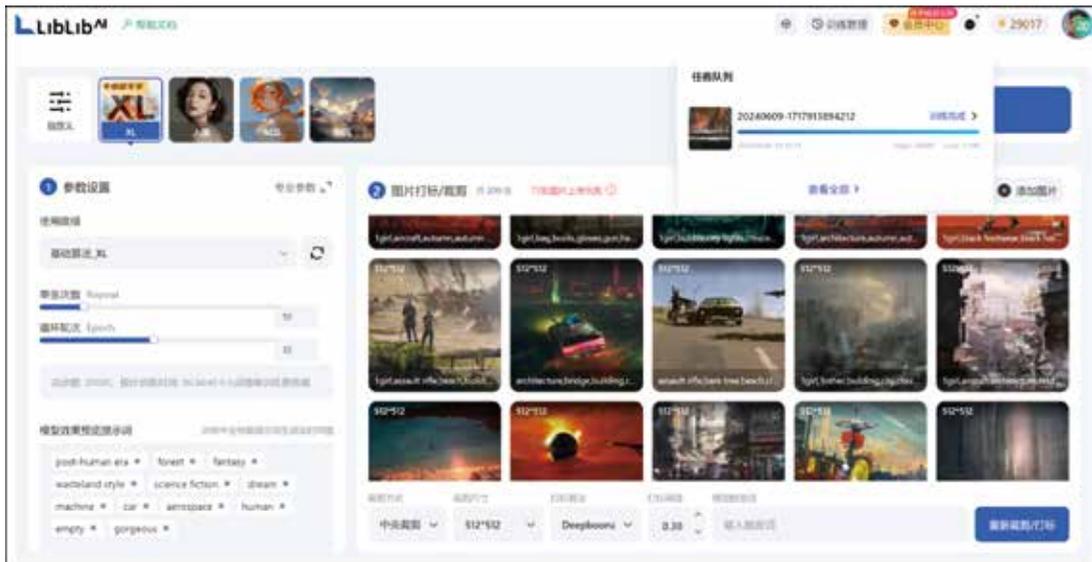


图 1-6
上传图片数据界面

等待训练，根据图片数据的质量或数量的不同，训练所需时间也会不同，耐心等待即可。训练好 LoRA 后，就可以进行效果测试，如图 1-7 所示。

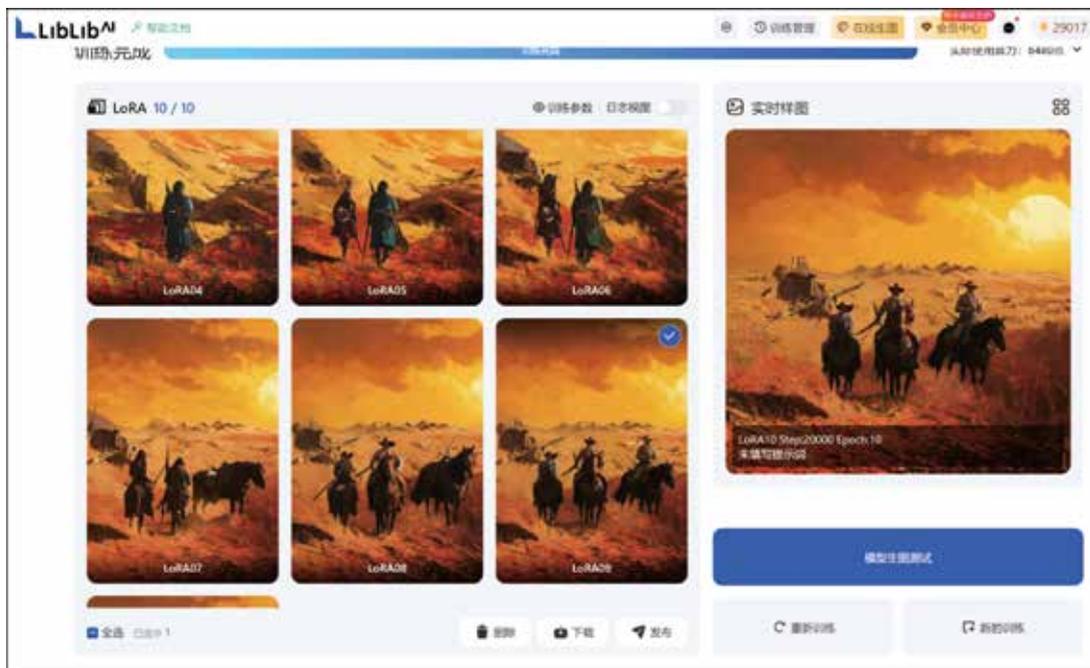


图 1-7
完成模型的训练

1.1.3 AIGC 角色动作的自动生成

要实现角色动作的自动生成，可以采用深度学习技术结合 AIGC 辅助工具和动画制作软件。以下是角色动作自动生成的工作流程。

(1) 输入与解析：首先提供角色动作的详细文字描述，AIGC 工具将解析这些描述，理解所需执行的动作细节。

(2) 动作数据库构建：AIGC 辅助工具将访问一个动作数据库，该数据库汇集了多样化的角色动作，可能来源于动作捕捉、手绘动画或 3D 动画模型。

(3) 动作匹配与选择：AIGC 工具在数据库中搜索与描述最匹配的动作数据。对于特定的动作描述，系统可能需组合多个动作片段以满足需求。

(4) 新动作生成：若数据库中缺乏精确匹配，AIGC 工具将利用深度学习模型进行新动作的生成，通过已有数据进行推断和合成。

(5) 生成分镜：AIGC 工具将生成的动作整合到分镜模板中，形成连贯的动作序列，并确保动作在物理和视觉上的合理性。之后，根据角色特性和情景要求进行微调，以提升自然度和表现力。

(6) 预览与修正：生成的分镜和动作需经过细致的检查和必要的手动调整，以确保动作符合创作意图和叙事需求。

整个流程关键在于整合计算机视觉、机器学习、自然语言处理和创意算法，以自动解读动作指令，并创作出既艺术又精确的动画分镜。尽管 AIGC 技术在自动动作生成方面取得了显著进展，但最终效果的质量仍需依赖于人类的创意和专业判断，以辅助生成高质量的动画作品。

以生成平台 DELL · E3 为例，首先提供角色动作的详细文字，如：生成一组正在跑步的男生，画面风格为中国画。为防止 AI 生成风格不一致的人物形象，应在输入提示中强调形象的一致性。示例提示词：生成一组正在跑步的男生，每个男生保持统一的人物形象，以电影故事板格式展示，采用中国画风格（图 1-8）。

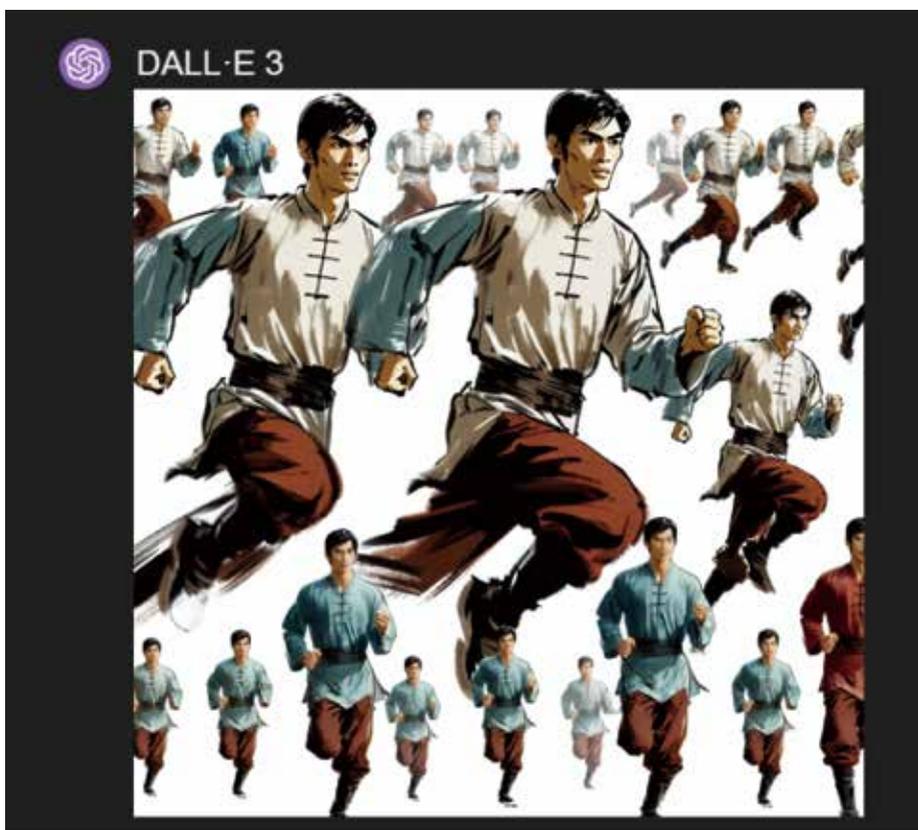


图 1-8
DELL · E3 生
成的角色形象

进一步生成包含两个角色的互动动作场景，确保动作和人物描绘的清晰度。示例提示词：生成一个打太极拳的男孩和一个女生做拉伸动作的电影镜头，保持中国画风格（图 1-9）。为了构建连贯的动作序列，需细化动作步骤并进行反复提示和优化，以确保动作适合动画分镜的需求。示例提示词：生成一个男孩打太极拳的五个分解动作，保持人物形象统一，画面完整，采用中国画风格（图 1-10）。

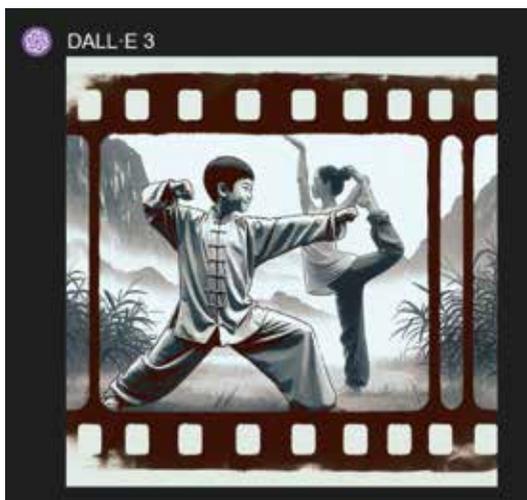


图 1-9
DELL · E3 生成的双人动作



图 1-10
DELL · E3 生成的动作步骤

1.1.4 AIGC 场景的智能化构建

为实现 AIGC 动画分镜中的场景智能化构建，可以遵循以下工作流程，该流程融合了深度学习、计算机图形学和自然语言处理技术。

(1) 数据收集：搜集丰富的动画场景相关数据，涵盖场景描述、构建、布局等关键方面。

(2) 场景描述分析：利用自然语言处理技术对场景描述进行深入分析，提取场景元素、结构、情节发展等关键信息。AI 将分析输入的文本或语音，理解场景的时间、地点、环境、氛围及物体。

(3) 场景生成模型训练：利用收集的数据训练深度学习模型，采用 GAN、VAE 等生成模型，学习场景生成的规律和特征。考虑到设备限制，可选择使用现有的大型模型或训练 LoRA 模型，同时融入环境纹理、物体模型和光照效果进行训练。

(4) 场景智能化构建：应用训练好的模型，根据场景描述自动生成动画场景，包括布局、角色位置、道具摆放等。AI 将挑选合适的环境元素，并规划场景布局，考虑视觉平衡和叙事需求，生成详细的场景草图或三维模型。

(5) 场景评估与优化：对生成的场景进行评估，根据真实感、合理性等指标不断优化模型和算法，提高场景生成质量。

(6) 与角色动作生成结合：将智能化生成的场景与角色动作相结合，创建完整的动画分镜。根据需求调整相关模型参数，以符合动画分镜的具体要求。AI 还能根据场景的时间、天气和氛围智能调整光照和阴影，增强场景的真实感和情感表达。

通过这一流程，能够高效地构建出既符合艺术标准又满足技术要求的动画分镜场景，同时为动画制作提供强有力的 AIGC 技术支持。

以生成平台 Midjourney 为例，选择赛博朋克风格化电影场景，可以输入相关提示词，如飞船、宇宙、科幻、星际空间等。这里提示词为：cyberpunk movie, on the top of the building, looking out at the entire city, ultra wide shot, Unreal Engine, neon cold lighting, deserted city buildings, hyper quality, high resolution。在生成的图片中可以选择其中的一张继续优化，直到生成适合动画作品的场景，进行放大导出使用，如图 1-11 所示。

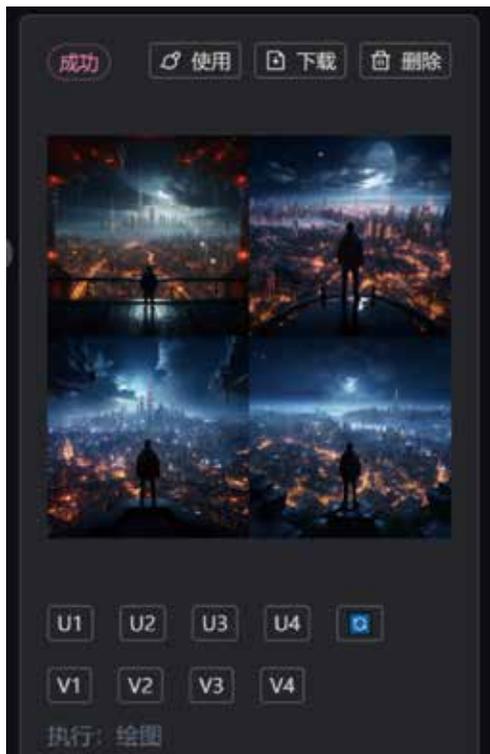


图 1-11
Midjourney 生成场景

1.1.5 基于 AIGC 的情感表达与镜头选择

为了高效地实现基于 AIGC 的情感表达与镜头选择，可以遵循以下优化后的工作流程，确保动画镜头和情节表达能够精准匹配情感和故事需求。

(1) 情感分析与情绪识别：应用自然语言处理技术，对剧本或情节描述进行深入的情感分析，识别故事中的情感状态和情绪变化。这一步骤将帮助我们明确每个场景或镜头所需的情感表达，如喜悦、悲伤或紧张等。

(2) 镜头与情感匹配：根据情感分析的结果，挑选与特定情感状态相匹配的动画镜头和场景。通过训练深度学习模型，学习在不同情感背景下的镜头构建和情节表达规律，实现镜头序列的自动生成。

(3) 情感化镜头构建：依据情感分析结果和镜头选择规则，自动生成与情感状态相匹配的动画镜头和场景。考虑情感的强度和变化，选择合适的镜头构建策略，包括镜头切换频率、角度和运动等。