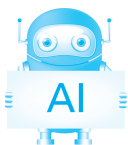


第 5 章

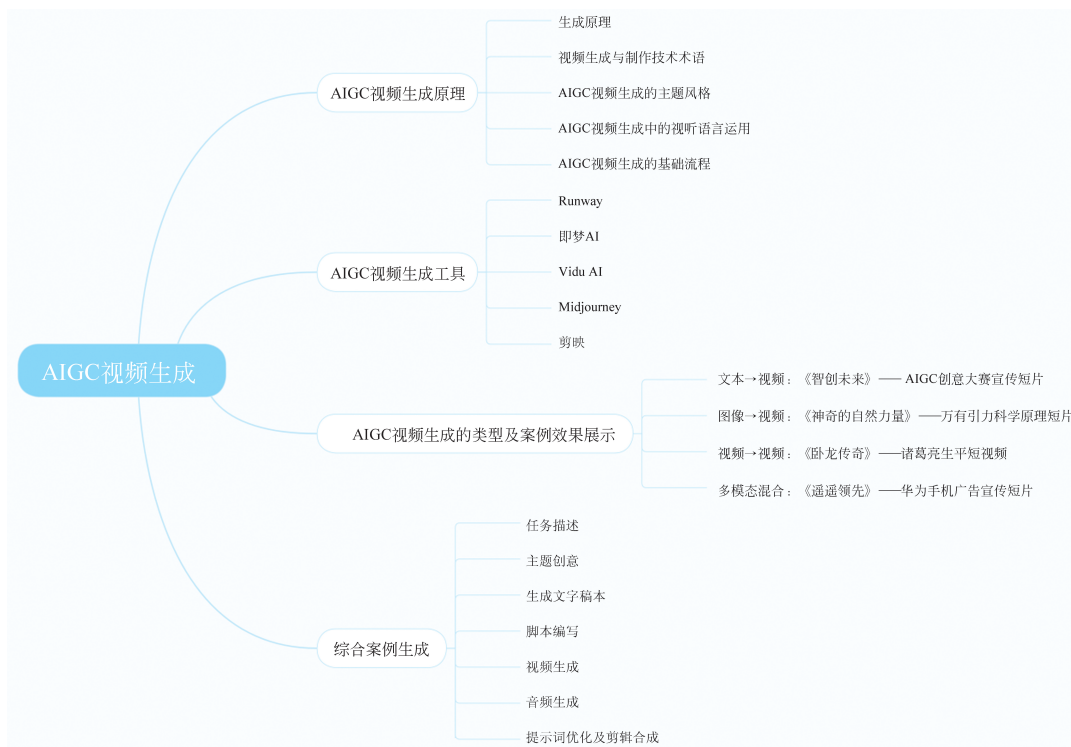


AIGC 视频生成

本章学习目标

- 掌握 AIGC 视频生成的底层技术逻辑、关键术语定义以及核心技术原理,构建对多模态转化、时序建模等核心机制的认识,为实践操作提供理论支撑。
- 熟练掌握 Runway、即梦 AI、Vidu AI、Midjourney、剪映等主流工具,明确不同工具在各场景中的优势,能依据具体创作需求组合工具,实现高效协同创作。
- 掌握 AIGC 视频生成从需求分析、创意策划、工具选型、内容生成到后期优化的标准化流程,具备独立完成从文本、图像、视频等多模态输入到成品输出的全链路创作能力,确保作品质量与创作效率双重提升。

本章思维导图



伴随 AIGC 技术在内容创作领域的深度渗透,视频生成已从早期的技术探索阶段步入规模化商业应用的全新阶段。AIGC 视频生成技术借助 AI 模型对海量数据中的静态视觉规律与动态时序逻辑开展深度学习,搭建起从文本、图像、视频等多模态输入到连贯动态画面输出的高效转化路径,从底层对传统视频制作的生产流程与创作范式进行了重构。

在工具方面,全球 AIGC 视频生成工具已形成“分层竞争、协同共生”的格局:Midjourney 凭借卓越的视觉美学生成能力成为风格化视频内容的创意源泉;Runway 依托全链路视频编辑与动作复刻技术,构建起专业级视频生产的工作流核心;即梦 AI 凭借本土化技术优化与多模态融合能力,实现了中文场景下的高效视频生成;V2Fun 专注于 3D 空间视频重构,突破了传统 2D 视频的维度局限;Vidu AI 以大模型技术创新,拓展了文本生成成长视频的想象范畴;剪映等国民级工具则通过深度整合 AIGC 能力,推动视频创作向全民化方向普及。

本章聚焦于 AIGC 视频生成实践,通过解析工具特性、拆解制作流程以及剖析典型案例,全方位展现 AIGC 在视频生成领域的技术实力与应用价值。首先,系统梳理 AIGC 视频生成的核心技术逻辑与标准化制作流程,构建视频生成知识体系;其次,围绕 5 大核心场景案例进行深度解析:文本驱动的《智创未来》创意大赛宣传短片、图像衍生的《神奇的自然力量》科学原理可视化作品、视频重构的《卧龙传奇》人物传记短视频、多模态融合的《遥遥领先》商业广告,以及 AI 赋能非遗传承的《泉韵·AI 织非遗》。这 5 大案例宛如 5 块关键拼图,共同描绘出 AIGC 视频生成从基于规则的模板化生产,到数据驱动的智能合成,再到跨模态的创意融合的完整技术演进图谱,为不同场景下的创作者提供兼具理论深度与实践价值的可实施指南。

5.1 AIGC 视频生成原理

5.1.1 生成原理



5-1

AIGC 视频生成技术的本质,是将文本、图像等静态输入数据转化为具备动态视觉效果和时序连贯性的视频内容。这一技术的实现极大程度上依赖深度学习模型的突破,其中扩散模型(Diffusion Model)与生成对抗网络(GAN)是当前应用最为广泛、成效最为显著的两类架构。其核心生成原理可分为以下 4 个关键环节。

1. 基于海量数据的模式学习

深度学习模型的“智能”来源于对海量数据的学习与归纳。在 AIGC 视频生成任务中,模型会通过剖析数百万甚至数十亿小时的视频数据,从中提炼出物体运动规律、场景切换逻辑、光影变化模式以及时间序列特征等基础视觉语言。

以扩散模型为例,其训练过程采用“正向破坏+逆向重建”的独特机制:首先,模型会向清晰的原始视频帧中逐步添加随机噪声,直至帧画面彻底变为无意义的噪声;随后,模型通过学习逆向过程,掌握如何从充满噪声的帧中逐步去除干扰,重建出清晰、连贯的视频帧序列。这种训练方式使模型能够精准把握视频的连续性和动态变化规律,为后续生成高质量视频奠定基础。

2. GAN 的博弈优化

GAN 的核心是通过生成器与判别器的“对抗博弈”,实现生成效果的迭代提升。在视

频生成任务中,生成器的目标是依据输入的文本、图像等信息,尽可能逼真地生成视频帧序列;判别器则负责将生成的视频帧与真实视频帧进行对比,判断其“真假”并反馈给生成器。二者在持续的对抗训练中不断优化自身性能:生成器会根据判别器的反馈调整生成策略,增强视频的真实感;判别器则会在与生成器的博弈中,不断增强对细节的辨别能力。经过数万甚至数百万轮的迭代后,生成器最终能够生成足以“以假乱真”的视频内容,尤其在人物表情、物体纹理、光影细节等方面表现优异。图 5-1 是利用 AIGC 工具生成视频帧的效果。



图 5-1 利用 AIGC 工具生成视频帧的效果

3. 多模态信息的融合与对齐

AIGC 视频生成的显著优势之一在于能够处理文本、图像等多模态输入,并将其转化为视听一体化的视频内容。此过程的关键在于跨模态对齐技术,即模型能够把不同模态的信息映射至同一语义空间,以实现精准的内容转化。

例如,当输入“日落海滩”的文本描述时,模型会先将文本语义拆解为视觉元素(金色沙滩、橙红色落日、翻涌的海浪)和听觉元素(舒缓的海浪声、轻柔的背景音乐);接着,通过跨模态对齐技术,将这些元素与训练数据中对应的视觉、听觉特征进行匹配,并按照时序逻辑组合成完整的视频帧序列,最终达成“所见即所听”的视听同步效果。图 5-2 是“日落海滩”作品的效果展示。



图 5-2 “日落海滩”作品的效果展示

4. 时序一致性的保障机制

与静态图像生成不同,视频生成的核心挑战之一在于维持帧与帧之间的时序连贯性,防止出现物体运动不自然、画面闪烁、跳帧等状况。为应对这一问题,AIGC 视频生成模型通常会引入光流估计或 3D 场景重建等技术,对视频帧的时序一致性加以约束。

以人物行走视频生成为例,模型会先借助光流估计技术,预测人物肢体在每一帧中的运动轨迹与姿态变化;接着,依据这些预测结果,生成符合物理规律的连续帧画面,保证人物的行走动作自然流畅,避免出现肢体瞬移、姿态突变等不协调现象。对于复杂场景的视频生成,模型还会运用 3D 场景重建技术,构建虚拟的三维空间,确保物体在场景中的运动符合透视原理和空间逻辑,人物行走视频生成的原理如图 5-3 所示。

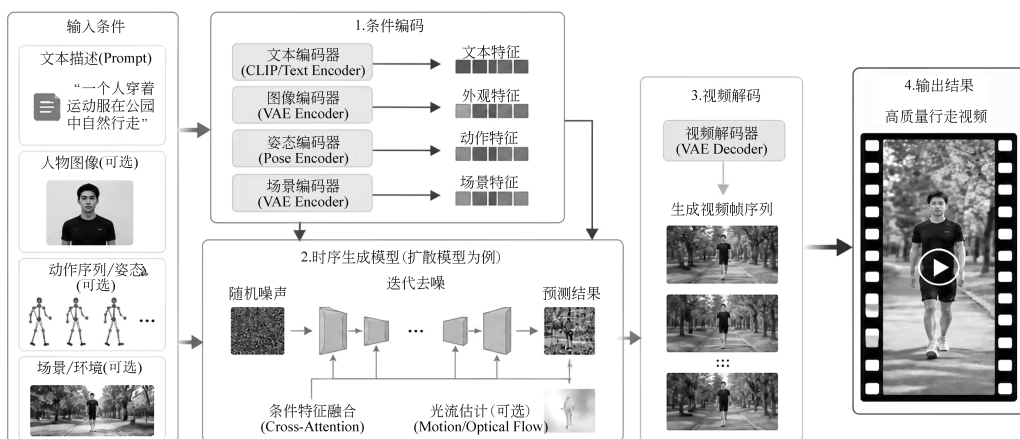


图 5-3 人物行走视频生成的原理

5.1.2 视频生成与制作技术术语

(1) 分镜头。

分镜头是视频拍摄与制作的核心脚本,它详细规定了每个镜头的景别、拍摄手法、画面内容、音效配置以及时长参数。通过精心设计分镜头,能够保障整个视频项目的逻辑连贯性,达成精准的视觉叙事。

(2) 动态捕捉。

动态捕捉技术借助传感器系统记录表演者的动作轨迹、力度变化和节奏特征,将真实的人体运动转化为数字化信息。这些数据可以直接驱动由 AI 生成的虚拟角色,实现虚拟场景与真实表演的精准同步,使虚拟角色的动作自然且流畅。

(3) 光流估计。

光流估计技术通过计算相邻帧间像素的运动矢量,预测物体的运动轨迹与速度。在视频生成中,这项技术能够确保动作的物理一致性,避免出现肢体瞬移、姿态突变等不协调现象,使虚拟角色的运动符合真实世界的物理规律。

(4) 3D 场景重建。

3D 场景重建技术能够从二维图像或文本描述中构建出三维空间模型。例如,仅通过一张房屋照片,AI 就能生成该房屋的完整三维模型,支持从任意角度进行观察,为多视角动态

视频的创作提供丰富的素材,如图 5-4 所示。



图 5-4 3D 场景重建

(5) 剪辑点。

剪辑点是视频剪辑中切换镜头的精确时间节点,它直接影响画面衔接的流畅程度。其核心原则是与人物动作、情绪节奏或场景逻辑相匹配,例如在人物抬手的瞬间切换镜头,能够实现动作的无缝衔接,让观众感觉画面过渡自然流畅。

(6) 转场。

转场是连接两个镜头的视觉过渡效果,可分为基础转场与特效转场两类。基础转场包含硬切、叠化、淡入淡出等常规手法,特效转场则包括旋转、划像等创意效果。转场的作用在于引导观众情绪,或者明确区分时空关系,使镜头转换更为自然。

(7) 镜头语言。

镜头语言是通过不同镜头类型的组合运用来传递信息、表达情感的视觉叙事体系。

全景镜头：用于交代场景环境,展示空间关系。

中景镜头：聚焦人物上半身动作,兼顾环境与细节。

特写镜头：突出人物面部表情或物体关键细节,强化情绪表达。

通过合理运用镜头语言,能够使视频作品更具表现力和感染力。全景/中景/特写对比如图 5-5 所示。

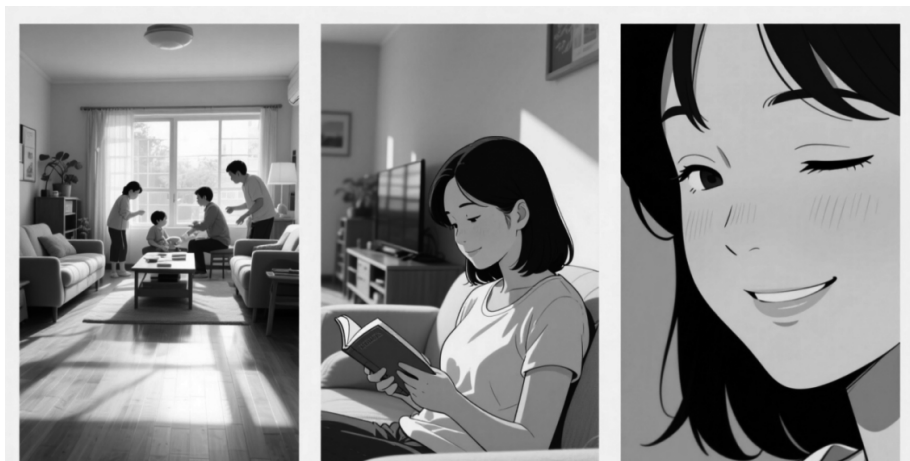


图 5-5 全景/中景/特写对比

(8) 节奏控制。

节奏控制是借助调节镜头时长与剪辑频率达成的视频节奏管理策略。快节奏视频一般采用短镜头与高频剪辑,适用于动作片等需营造紧张氛围的场景;慢节奏视频则运用长镜头与低频剪辑,适用于文艺片等需进行细腻情感表达的场景。节奏控制需与视频主题及情感表达相契合。

(9) 色阶/调色。

色阶/调色是通过调整画面色彩参数与明暗关系,使视频整体色调风格达成统一的后期处理技术。示例如下。

冷色调:营造悬疑、冷峻的氛围。

暖色调:传递温馨、治愈的情感。

高对比度:增强视觉冲击力,凸显主体。

调色是提升视频质感、强化艺术风格的关键环节,可使视频作品更具专业水准。冷色调/暖色调/高对比度画面对比如图 5-6 所示。



图 5-6 冷色调/暖色调/高对比度画面对比

(10) 视频合成。

视频合成是一个后期制作过程,旨在将真实表演素材、AIGC 生成的虚拟背景、视觉特效以及音频元素进行整合。通过专业剪辑工具的处理,把分散的素材加工成叙事完整、视觉风格统一的最终视频作品,恰如将零散的拼图碎片组合成一幅完整的画面。

5.1.3 AIGC 视频生成的主题风格

在 AIGC 视频生成体系里,主题风格是衔接创意理念与受众情感的核心纽带。借助 AIGC 技术对视觉元素、色彩体系以及运镜节奏进行精准调控,创作者能够迅速打造出适配不同传播场景的视频风格,可大大拓展内容创作的边界与可能性。本节将系统剖析主流主题风格的技术特征与应用逻辑,为 AIGC 视频生成提供专业的风格设计框架。

1. 主流主题风格

1) 时尚潮流风格

核心视觉表现:以高饱和度的撞色(例如霓虹粉、电光蓝、荧光绿)作为色彩基础,搭配



5-2

动态光影特效、快节奏的分镜切换以及非线性剪辑,使画面呈现出强烈的动感与视觉冲击力。在构图方面,多运用不规则几何分割、中心聚焦等方式,以强化视觉张力。

AIGC 技术支撑路径如下。

(1) 色彩增强算法:通过 AI 智能识别画面主体,有针对性地强化色彩饱和度与对比度,从而突出视觉焦点。

(2) 动态特效生成:基于关键帧驱动的粒子系统,自动生成光影流动、爆炸、变形等动态视觉元素。

(3) 智能节奏匹配:分析背景音乐的 BPM(每分钟节拍数),自动调整镜头时长与切换频率,实现音画节奏的高度同步。

适用场景:美妆、服饰等快消品牌广告,社交媒体短视频推广,潮流文化活动宣传,电竞与街头文化内容创作。

典型实践案例:某国际美妆品牌 2025 年新品口红推广视频,借助 AIGC 生成的渐变霓虹背景以及 1 秒/帧的快切镜头,在 30 秒内完整展示了 12 款口红的上唇效果与色彩层次。图 5-7 为口红产品展示效果,图 5-8 为上唇效果与色彩层次展示效果。



图 5-7 口红产品展示效果



图 5-8 上唇效果与色彩层次展示效果

2) 复古怀旧风格

核心视觉表现:运用低对比度的暖黄色调,叠加胶片颗粒质感与暗角效果,模拟老式摄

像机的轻微抖动及镜头畸变,以还原特定年代的影像氛围。画面多呈现柔焦效果,色彩偏向褪色感,从而强化时光沉淀的视觉记忆。

AIGC 技术支撑路径如下。

(1) 胶片滤镜建模:借助机器学习复刻柯达 GOLD 200、富士 Provia 100F 等经典胶片的色彩曲线与颗粒特征。

(2) 抖动模拟算法:基于物理引擎模拟 8mm/16mm 老式摄像机的拍摄抖动与画面畸变效果。

(3) 年代色彩还原:通过 AI 分析特定年代的流行色与影像特征,自动匹配契合时代背景的色彩体系。

适用场景:品牌周年庆宣传片、历史题材纪录片、怀旧主题短视频、老电影修复与重制、传统手工艺传承记录。

典型实践案例:某国民汽车品牌 2024 年复古车型推广视频,AIGC 将现代城市街景转化为 20 世纪 90 年代老城区风貌,叠加老式广播音效与胶片颗粒效果,唤起受众的时代记忆。街角展示效果如图 5-9 所示。

3) 科幻未来风格

核心视觉呈现:以具备冷色调金属质感的场景为基础,融入全息投影、能量光束、悬浮建筑等未来元素,运用抽象几何构图以及对称式布局方式,营造出充满科技感的超现实氛围。在光影运用方面,着重突出强烈的明暗对比以及反射效果,以增强空间的纵深感。

AIGC 技术支撑路径如下。

(1) 3D 场景快速构建:依托 NeRF (Neural Radiance Field,神经辐射场)技术,依据文本描述自动搭建未来城市、太空站等复杂的 3D 场景。

(2) 实时光影追踪:采用路径追踪算法对金属反光、全息投影的光影折射效果进行模拟,达成电影级的渲染质量。

(3) 动态粒子合成:借助 AIGC 生成能量流、星云、数据瀑布等动态粒子特效,增强科幻场景的沉浸感。

适用场景:科幻游戏先导预告、科技产品发布会视频、概念设计展示片、元宇宙场景构建、未来城市规划可视化。

典型实践案例:某 3A 科幻游戏 2026 年先导预告视频中,AIGC 生成的废弃太空站场景与光影流动的全息地图相配合。科幻短片场景展示效果如图 5-10 所示。

4) 自然纪实风格

核心视觉表现:采用柔和且低饱和度的自然色彩,模拟自然光效(例如黄金时刻的光线、阴天的漫射光),运用缓慢跟拍与固定镜头,保留场景的原生质感与细节。画面强调真实感与沉浸感,避免过度修饰与使用特效。

AIGC 技术支撑路径如下。



图 5-9 街角展示效果



图 5-10 科幻短片场景展示效果

(1) 智能降噪算法：基于深度学习的音频降噪技术，在保留自然音效（如鸟鸣声、水流声）的同时消除环境杂音。

(2) 平滑运镜生成：借助 AI 模拟手持稳定器的跟拍效果，实现镜头的平稳移动与场景的自然过渡。

(3) 自然色彩匹配：分析真实自然场景的色彩特征，自动调整画面色彩参数，还原森林、海洋等场景的真实色调。

适用场景：自然生态纪录片、科普教育视频、治愈系短视频、环保公益宣传、户外探险内容创作。

典型实践案例：某国际环保组织 2026 年的公益视频，AIGC 生成的亚马逊雨林生态场景配合真实的鸟鸣声、水流音效，完整展示了雨林的生物多样性。环保公益宣传短片展示效果如图 5-11 所示。

5) 国风古韵风格

核心视觉表现：以水墨晕染效果作为核心视觉语言，融合传统纹样（云纹、回纹、饕餮纹）、古典建筑元素以及低饱和度的古典色调（黛蓝、赭石、石青），营造出东方美学意境。构图遵循中国传统绘画的留白原则与对称美学理念。

AIGC 技术支撑路径如下。

(1) 水墨风格迁移：基于 CycleGAN 等风格迁移算法，将实景或者 3D 场景转化为水墨动画效果。

(2) 传统元素生成：借助 AI 学习中国传统纹样与建筑特征，批量生成契合国风的装饰元素与场景组件。

(3) 古风场景建模：依据文本描述自动生成故宫、苏州园林等古典建筑场景，还原东方



图 5-11 环保公益宣传短片展示效果

美学空间。

适用场景：传统文化宣传视频、国风品牌广告、非遗传承纪录片、古装影视概念设计、国风游戏 CG 动画。

典型实践案例：某非遗剪纸艺术推广视频中，AIGC 把静态剪纸作品转化为动态水墨动画，并搭配古筝作为背景音乐。国风古韵风格展示效果如图 5-12 所示。



图 5-12 国风古韵风格展示效果