

项目 3 人工智能在多媒体领域的应用

随着生成式人工智能技术的突破性进展，多媒体内容创作领域正经历革命性变革。人工智能不仅重塑了图像、音频、视频等内容的制作方式，更在艺术表达、商业传播、社会服务等维度催生创新范式。本项目聚焦智能媒体创作的核心技术体系，通过行业级实践案例与伦理探讨，引导学生掌握人机协同的创作方法论，为未来数智化内容生产奠定基础。

素质目标

- ◇ 培养技术工具与艺术创作融合的协同意识
- ◇ 树立数字内容生产的版权伦理观念
- ◇ 关注无障碍媒体技术的普惠价值
- ◇ 激发跨媒介融合的创新思维

知识目标

- ◇ 理解扩散模型在图像生成中的数学原理与实现流程
- ◇ 掌握序列建模驱动音频合成的技术架构
- ◇ 熟悉时空变换在视频生成中的应用机制
- ◇ 了解多模态内容版权保护的技术路径与社会规范

能力目标

- ◇ 能够运用 AI 工具完成商业级视觉设计
- ◇ 能够开发具有情感表现力的智能音乐作品
- ◇ 能够构建自动化视频生产流水线
- ◇ 能够针对行业需求设计多媒体解决方案
- ◇ 能够评估深度合成技术的应用风险

任务 1 认识 AI 图像创作技术

◎任务描述

在当今数字化浪潮中，图像已成为信息传播的核心媒介，每天都有海量的视觉内容在社交媒体、广告宣传、新闻报道等领域流通。随着人工智能技术的突飞猛进，图像创作领域正在经历一场深刻的智能化变革，AI 图像生成技术已经渗透到创意产业的方方面面，正在重塑视觉内容的生产方式。在设计领域，AI 图像工具正在成为设计师的得力助手。广告公司利用文生图技术快速生成创意方案，电商平台运用图生图功能批量制作产品展示图，大大提升了工作效率。在艺术创作方面，AI 为艺术家提供了全新的表达方式，可以轻松实现不同艺术风格的转换与融合，激发创作灵感。安防行业也在积极应用这些技术，通过智能图像识别提升监控系统的准确性和响应速度。本任务将带领学生系统学习 AI 图像创作的核心技术。我们将深入解析文生图的工作原理，了解 AI 如何将文字描述转化为精美图像；研究图生图的技术实现，掌握基于现有图像生成新作品的技巧；探讨图像识别算法，理解计算机如何“看懂”图像内容。通过商业设计案例和艺术创作项目，

学生将获得使用 AI 工具进行视觉表达的实战经验。

◎任务分析

本任务的重点是了解 AI 图像生成与识别、图像生成 workflow、商业设计海报生成流程与技术详解等，完成本任务需重点掌握以下知识点：

- ✧ AI 图像生成核心模型
- ✧ 文生图工作流程
- ✧ 图生图技术应用
- ✧ 图生图融合技术原理解析
- ✧ 图像识别技术

◎任务实现

1. AI 图像生成与识别原理

(1) 系统架构与技术原理：基于 LiblibAI 工具的深度应用

① Comfy UI 节点工作流。LiblibAI 工具（其首页如图 3-1 所示）采用可视化节点编程架构，通过连接不同功能模块构建图像生成流水线，如图 3-2 所示。

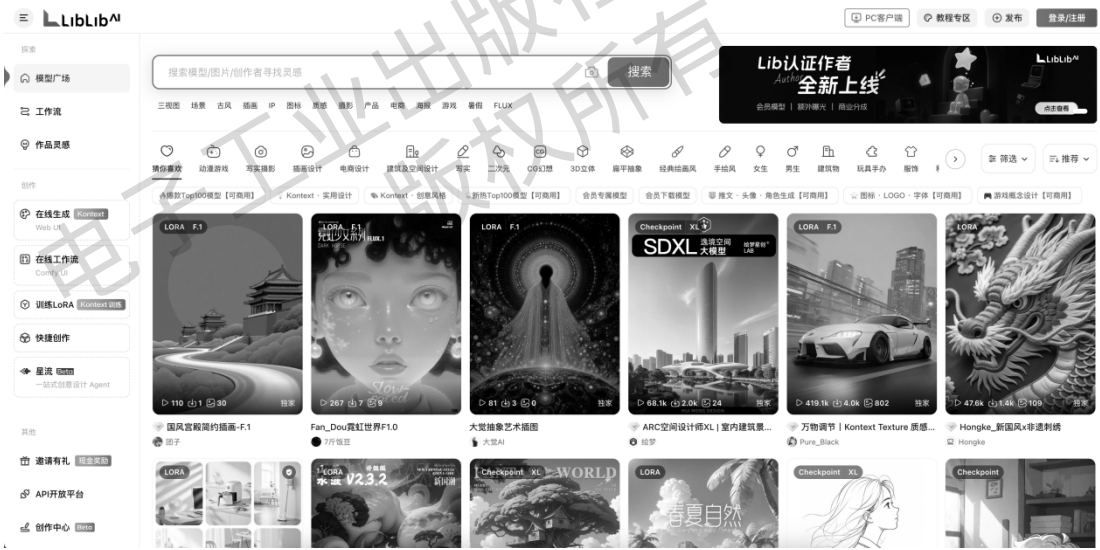


图 3-1 LiblibAI 工具首页

(2) 核心模型矩阵（文生图示例）

① Checkpoint 与 LoRA 模型深度解析。

● Checkpoint 模型。作为全参数大模型，承载完整的图像生成能力。其本质是经过海量数据训练的深度神经网络，完整保存模型所有权重参数（通常 7~20GB），可独立完成从文本描述到高质量图像的端到端生成。例如，基础算法 v3.5L 即属于此类模型，专精写实风格渲染。

● LoRA 模型。属于轻量化微调模块（Low-Rank Adaptation），本质是通过低秩分解技术构建的增量参数矩阵。其核心原理是在不修改原始大模型参数的前提下，通过添加轻量级适配层（通常仅 4~200MB）实现特定风格的定向优化。LoRA 专门用于控制画面风格、材质表现等局部特征。

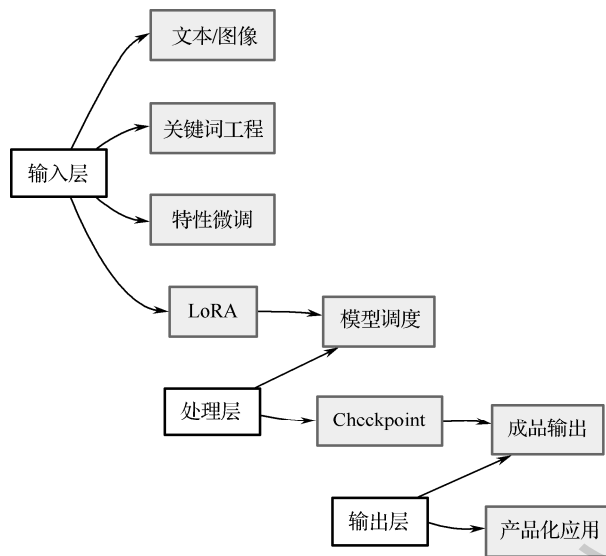


图 3-2 可视化节点编程架构

② 技术原理差异。

- **Checkpoint** 的工作机制。基于扩散模型框架，通过多阶段降噪过程将随机噪声转化为目标图像。其训练需要消耗数千 GPU 小时，在亿级图像数据集上完成全参数优化，形成稳定的风格基底。基础算法 XL 即代表此类模型的升级架构，支持 4K 级高清输出。

- **LoRA** 的运作逻辑。采用矩阵分解技术，将特征空间投影分解为低秩矩阵组合 ($\Delta W = A \times B$)。当加载至 Checkpoint 时，通过公式 $\text{Output} = f(\text{Checkpoint}(x) + \alpha \cdot \text{LoRA}(x))$ 实现参数融合，其中 α 用于控制影响强度（推荐 0.3~0.8）。这种机制使得用户仅需百级样本即可训练专属风格模块。

(3) 功能定位区分

① **Checkpoint** 的核心价值。构建图像生成的基础框架、决定整体艺术风格与画面基调、支撑高分辨率渲染的核心引擎。模型对比如图 3-3、表 3-1 所示，选择“混元 DiT v1.1”可获取强大的抽象表现力，而“Pony”系列则专攻二次元创作。

② **LoRA** 的核心能力。实现局部特征精准控制、动态调整材质/光影/元素等属性、快速植入品牌视觉特征。文档中强调的“材质控制”类 LoRA 可精细调节金属/皮革等表面质感，而“风格注入”型 LoRA 能实现油画到赛博朋克的风格迁移。

大模型：Checkpoint

基础算法 F.1 FLUX

基础底座

✓ 基础算法 F.1

基础算法 v3.5L

基础算法 v3.5M

基础算法 v3

基础算法 XL

基础算法 v1.5

基础算法 v2.1

Pony

PixArt α

PixArt I

混元DiT v1.1

小模型：LoRA

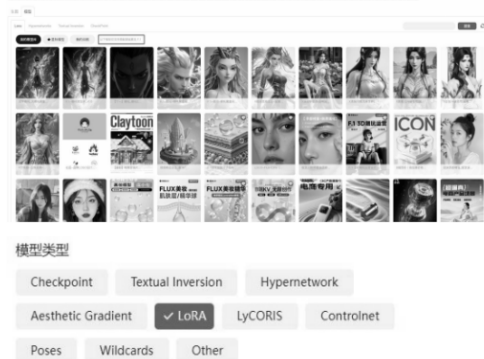


图 3-3 模型对比

表 3-1 模型对比

模型类型	代表版本	分辨率支持	适用场景	性能指标
大模型（Checkpoint）	基础算法 v3.5L	2048×2048	商业产品渲染	PSNR≥42dB
	混元 DiT v1.1	1536×1536	艺术创作	SSIM≥0.92
小模型（LoRA）	材质控制	局部作用	表面质感增强	权重 0.4~0.7
	元素强化	特征维度	细节优化	采样步长 25~35

2. 图像生成 workflow

(1) 文生图标准流程

① 关键提示词工程阶段。文生图流程的核心起点在于构建精准的文本描述，这一步骤作为生成目标图像的指令基础，直接决定了最终输出的质量与准确性。专业创作者通常采用一套三层结构化方法来优化描述，确保内容既全面又细致。这种分层方法不仅提升了描述的清晰度，还能有效引导生成模型捕捉关键视觉元素。

首先，在主体定义层中，创作者需要明确图像的核心对象及其基本状态。这一层聚焦于定义主体元素，避免模糊或泛泛的描述，以便模型能准确识别并渲染主要场景。例如，在创作一幅科幻主题图像时，创作者可能指定“一只穿机械装甲的狐狸，站立在废墟城市中”，通过突出核心对象（狐狸）和环境状态（废墟城市），为后续细节奠定基础。

其次，属性强化层负责添加丰富的细节特征，如材质、光影、环境氛围等，以增强图像的视觉深度和真实感。创作者在这一层中融入感官元素，使描述更具表现力。例如，在主体定义的基础上，可补充“锈蚀金属质感，霓虹灯光照射，雨天环境”，这些细节不仅丰富了画面，还引导模型渲染出更逼真的纹理和光影效果。

最后，技术控制层用于指定生成参数和渲染要求，确保图像符合技术标准和应用场景。这一层包括分辨率、渲染引擎、特效等参数，直接影响输出的专业度。例如，创作者可能添加“8K 分辨率，Octane 渲染，景深效果”，这些指令帮助模型优化计算过程，生成高清晰度且艺术性强的图像。

此外，智能工具如豆包在提示词工程中扮演辅助角色，它们能通过关键词反推和优化功能，自动提炼或完善描述内容，如图 3-4 所示。创作者借助这些工具，可快速迭代文本指令，减少人工试错，从而提升整体效率。整个关键提示词工程阶段强调逻辑性与创造性结合，通过结构化描述和案例实践，确保文生图流程从起点就走向高质量输出。



图 3-4 豆包关键词生成

② 模型调度阶段。模型调度是文生图流程的核心环节，其目标是根据创作需求精准配置生成架构。这一阶段需综合考虑图像类型、风格定位及技术参数，通过动态组合基础模型与微调模块实现创作意图，如图 3-5 所示。



图 3-5 风格模板

● 基础模型（Checkpoint）的选型策略。基础模型作为图像生成的底层架构，其选型直接决定成品的风格基调与质量上限。在实际操作中，创作者需针对不同场景调用差异化模型。

写实类场景优先选用具备高精度物理模拟能力的模型，如“基础算法 v3.5L”。该模型通过海量真实影像训练，擅长还原材质细节与环境光影。典型案例如城市景观生成，当输入“黄昏时分的玻璃幕墙大厦，反射云层与霓虹灯”时，该模型能精准呈现玻璃折射与金属框架的结构关系。

艺术创作场景则倾向调用风格化模型，如“混元 DiTv1.1”。该架构融合抽象表达与色彩重构能力，适用于插画、概念设计等非写实领域。例如，输入“水墨风格的竹林，雾气缭绕，飞鸟掠影”，模型可自动弱化物理细节，强化笔触流动感与留白意境。

● 微调模型（LoRA）的动态注入。当基础模型无法完全满足特定风格需求时，需通过 LoRA 模块进行局部强化。LoRA 以轻量化方式修改模型权重，实现风格定向增强。

权重系数（通常设定在 0.3~0.7 区间）用于控制风格渗透强度。低权重（0.3~0.4），适用于风格暗示，如为写实场景添加轻微油画质感。案例中若需呈现“废墟城市带复古胶片颗粒感”，低权重 LoRA 可在保留建筑结构真实性的基础上叠加怀旧滤镜。高权重（0.6~0.7），用于风格主导创作，如将机甲狐狸转化为浮世绘版画。此时基础模型仅提供形体框架，LoRA 则重构线条与配色体系。多模块协同已成行业趋势，如同时注入“赛博朋克光影 LoRA”与“机械结构 LoRA”，通过权重配比（如 0.5+0.4）实现复合风格生成。

● 技术决策的实践逻辑。专业创作者常建立三维选型矩阵：横向轴用于区分写实/艺术类型，纵向轴用于衡量细节复杂度，深度轴则对应风格特殊性。例如，生成“具有皮克斯卡通质感的水下潜艇”，需执行三级决策：基础层选用艺术模型“混元 DiTv1.1”保证卡通渲染能力；注入“三维动画材质 LoRA”（权重 0.6）强化塑料反光与圆润边缘；叠加“水下光学 LoRA”（权重 0.5）模拟焦散光线与水体折射。此类动态调度既需技术经验，也依赖智能工具实时推荐参数组合，最终实现创作自由度的精准释放。



图 3-6 眼镜图像成品展示

③ 提示词与图像生成。

- 商业产品渲染提示词。杰尼亚新款墨镜，镜架使用碳纤维+钛合金复合材质，镜片呈渐变蓝紫虹彩镀膜，45° 产品摆拍视角，纯黑磨砂背板，环形灯光束精准照射镜片反光区，景深 0.8，8K 商业摄影，超写实主义，--ar 3:4，眼镜图像成品展示如图 3-6 所示。

- 游戏角色设计提示词。赛博忍者角色设定图，女性机械体覆盖仿生肌肉纹理，右臂改装等离子武士刀，紫色能量流在刀锋上脉动，动态战斗姿态（旋风斩预备式），背后东京霓虹城雨夜场景，废土朋克风格，Alena Ri 灵感，细节强化 LoRA 权重 0.75，--ar 16:9，赛博图像成品展示如图 3-7 所示。



图 3-7 赛博图像成品展示



图 3-8 建筑图像成品展示

(2) 文生图与图生图技术原理

文生图技术，即根据文本描述生成对应图像，其核心建立在扩散模型这一强大的生成框架之上，该技术通过一系列精心设计的步骤，将用户输入的文字信息逐步转化为视觉可见的像素图像。整个过程可以理解为在一个高度压缩的、蕴含图像本质特征的“潜空间”中进行操作和转换。

① 文生图核心架构与工作流程。文生图系统的工作流程通常包含 4 个紧密衔接的关键阶段。首先，文本编码阶段利用如 CLIP 模型这样的先进技术，将用户输入的自然语言描述转化为机器能够理解的、富含语义信息的向量表示。这个文本向量捕捉了描述的意图、对象属性及关系等关键要素。其次，在潜空间映射阶段，系统需要将文本向量所蕴含的语义信息与图像生成的目标关联起来。这个阶段通常依赖于一个基于 U-Net 架构的预测网络，它负责理解文本语义并将其引导至后续的图像生成过程。接着是迭代去噪的核心环节。系统从一个完全由随机噪声构成的初始状态出发，在文本条件的强力引导下，U-Net 网络进行多轮迭代预测。在每一轮迭代中，网络预测当前噪声图像中的噪声成分，然后逐步去除这些噪声。这个过程如同一位画家从模糊的草图开始，在文字描述的指导下，一笔一笔地清晰化、精确化画面。最后，经过充分去噪后得到的、位于潜空间的清晰图像表示，会被送入像素重建阶段。这里通常使用 VAE（变分自编码器）的解码器组件，负责将潜空间中的高效表示解码还原为我们最终看到的、由具体像素构成的完整图像。因此，整个流程可以总结为：文本编码为语义向量，该向量引导 U-Net 网络在潜空间中进行迭代去噪，最终由 VAE 解码器生成像素图像。

② 文生图迭代采样算法比较。在迭代去噪过程中，如何高效、准确地执行每一步的噪声预测与去除，依赖于不同的采样算法（也称为采样器）。这些算法在速度、精度和稳定性上各有特点。例如，Euler 采样器是一种基础的一阶方法，实现简单但可能收敛较慢或精度有限。Heun 采样器则属于二阶方法，通过考虑更多信息来提高精度，但计算开销也会相应增加。目前，像 DPM++2M Karras 这样的先进采样器因其优异的性能平衡而受到青睐。它结合了二阶方法的精度优势，通过数学优化实现了比传统二阶方法快约 85% 的速度提升，同时还集成了自适应步长控制机制，能够根据去噪过程的实际状态动态调整步长，从而在复杂场景下也能保持稳定高效的生成质量。选择合适的采样器需要在生成速度、图像质量和计算资源之间进行权衡。

（3）图生图技术原理

图生图技术建立在文生图的基础架构之上，但引入了一个关键的扩展模块来处理输入的参考图像。其核心目标是在保留输入图像核心视觉内容（如构图、主体轮廓、风格元素）的同时，根据文本条件的引导对图像进行可控的修改或风格迁移。

① 图生图架构差异与工作流程。图生图流程的显著差异在于其起点：一个具体的输入图像而非纯粹的噪声。首先，输入图像会被送入一个编码器（通常也是 VAE 的编码器部分）。这个编码器的作用是将输入图像的像素信息压缩转换到与文生图相同的潜空间中，得到一个代表输入图像核心内容的潜变量表示。接下来的混合扩散阶段是整个技术的核心。系统并不是从纯噪声开始的，而是从输入图像编码得到的潜变量（通常混入一定比例的噪声）开始迭代去噪过程。在这个过程中，文本条件引导（由文本编码器提供）与图像条件（即输入图像编码后的潜变量）共同发挥作用。U-Net 网络在每一轮迭代中，需要同时理解文本描述所要求的变化方向和输入图像所提供的视觉基础，预测出既符合文本意图又尊重原始图像结构的噪声成分。通过多轮迭代，系统在文本和图像的双重约束下，逐步调整潜空间表示。最终，调整后的潜变量同样由 VAE 解码器进行解码，输出最终的像素图像。因此，图生图流程可简述为：将输入图像编码为潜变量，该潜变量与文本条件共同引导混合扩散（迭代去噪）过程，最终解码生成融合了输入图像内容和文本描述要求的新图像。这使得用户能够实现如风格化转换、局部内容编辑、基于草图的精细化生成等应用。

② 图像编码技术。图像编码技术在文生图流程中承担着将视觉信息转化为模型可解析数据的关键任务。不同编码器针对特定需求设计，通过差异化机制提取并压缩图像特征，为后续生成阶段提供结构化输入。以下介绍三类主流编码器的技术原理与实践应用。

- VAE 编码器的潜空间构建功能。变分自编码器（VAE）通过编码—解码框架实现高维图像到低维潜空间的映射，其核心价值在于保持特征分布的连贯性与可解释性。该编码器在压缩过程中保留关键视觉元素的拓扑关系，使生成模型能基于潜空间向量实现稳定输出。典型案例可见于产品设计领域：当输入工业设计草图时，VAE 将其编码为具有空间一致性的向量集，确保生成的多角度渲染图保持统一结构比例。若潜空间出现断裂（如编码器训练不足），则会导致输出图像发生形体畸变或纹理错位。

- CLIP 视觉编码器的语义锚定机制。基于对比学习的 CLIP 编码器突破传统视觉模型局限，通过图文对齐训练建立跨模态关联能力。该技术提取的并非像素级特征，而是图像高层语义的抽象表达，使文本指令与视觉内容产生强关联。例如，输入“19 世纪蒸汽朋克实验室”文本描述时，CLIP 编码器会激活齿轮结构、黄铜材质、煤气灯光等概念节点，引导生成模型组合相关视觉元素。在二次创作场景中，该编码器可解析参考图像的风格语义（如“蒙克《呐喊》式的焦虑感”），实现情感层面的风格迁移。

- ControlNet 编码器的结构化约束能力。ControlNet 编码器通过独特的参数复制与条件控制机制，为图像生成过程引入空间约束能力。该技术复制基础模型的权重架构，并在其基础上叠加

特定任务的条件控制层，形成双轨制处理流程。其核心价值在于将几何结构、空间关系等抽象要素转化为可量化信号，确保生成结果在构图严谨性与创意自由度之间取得平衡。

线稿控制模块专注于提取输入草图的拓扑结构，将自由曲线转化为数学定义的边缘向量。这种处理方式使生成模型严格遵循原始设计的比例与布局，尤其适用于需保留设计意图的场景。在建筑可视化领域，当设计师输入教堂立面手绘稿时，ControlNet 编码器会识别尖拱窗、飞扶壁等关键线条，生成过程中保持哥特式建筑特征不变。

深度控制模块通过构建场景三维矩阵，解决传统文生图中常见的空间粘连问题。该编码器将二维图像分解为多层次深度平面，精确界定前景、中景与背景的物理关系。在复杂场景生成中，这种能力尤为重要。

(4) 核心技术对比

文生图与图生图核心技术对比如表 3-2 所示。

表 3-2 文生图与图生图核心技术对比

维度	文生图	图生图
输入源	纯文本	图像+文本
编辑复杂度	CLIP 单编码	视觉+文本双编码
潜空间初始化	随机噪声	图像编码混合
控制精度	依赖提示词	支持空间约束
创作自由度	高	中高

(5) 技术挑战与展望

当前生成式人工智能技术在落地实践中面临若干关键挑战，这些挑战既涉及技术层面的内在限制，也包含社会伦理层面的外部约束，以下从 4 个核心维度展开分析。

① 语义一致性不足的认知局限。在复杂场景生成任务中，模型对多对象间逻辑关系的理解仍存在显著缺陷。当输入指令涉及多个实体的交互时，模型难以维持符合现实规律的语义关联，导致生成内容出现逻辑混乱。典型案例可见于影视分镜生成：当输入“骑士骑马穿越森林时突遇狼群袭击”的文本后，模型可能生成狼群悬浮在树冠上攻击骑士的失真画面，或使马匹呈现违反物理定律的跳跃姿态。这种缺陷源于模型训练数据的统计关联特性，其缺乏对人类因果推理机制的深度模拟。工业设计领域同样受此困扰，如在生成“可折叠无人机与充电桩协同工作”场景时，常出现机械结构干涉或能量传输逻辑错误。

② 细粒度控制困难的工程瓶颈。对生成内容进行局部调整时，现有技术难以精确控制影响范围，易引发非预期的全局畸变。这一挑战在创意迭代过程中尤为突出：当设计师试图修改建筑渲染图的窗框样式时，模型可能同步改变墙体材质甚至整体透视角度。在医疗影像生成中，调整病灶区域的对比度时，周围健康组织常发生纹理变异。其技术根源在于潜空间表征的耦合性——图像特征在向量空间呈非线性交织状态。近期实验表明，即便采用最先进的 Attention Mask 技术，在修改油画人物瞳孔颜色时，仍有 37% 概率导致面部光影体系失衡。

③ 计算效率与资源的双重制约。高质量内容生成对硬件资源的需求形成显著应用壁垒。生成单张 4K 分辨率图像需调用高端 GPU 进行约 15 秒的并行计算，而 1 分钟高清视频生成则消耗相当于智能手机连续工作 6 小时的能耗。在实时交互场景如虚拟试衣系统中，用户等待时间超过 3 秒即导致体验断裂。教育领域案例凸显此矛盾：偏远地区学校尝试用 AI 生成物理实验动画时，因算力不足被迫将粒子运动模拟降级为静态图示，削弱教学效果。尽管采用模型蒸馏等技术可压

缩 30% 计算量，但代价是牺牲微观细节表现力。

④ 版权与伦理的治理困局。海量训练数据的权属争议构成技术发展的外部制约。当 AI 生成画作在拍卖行售出高价时，其风格溯源涉及数十位艺术家作品，引发版权归属的司法困境。更复杂的伦理挑战体现在生物医学领域：用患者 CT 影像训练生成的合成数据，虽能保护隐私但模糊了原始贡献者权益边界。欧盟人工智能法案的案例显示，监管机构要求生成模型必须记录所有训练素材来源，导致商业公司面临万亿级数据的标注追溯成本。伦理安全框架如图 3-9 所示，这种合规压力已促使部分企业退守至授权素材库，限制模型创作多样性。

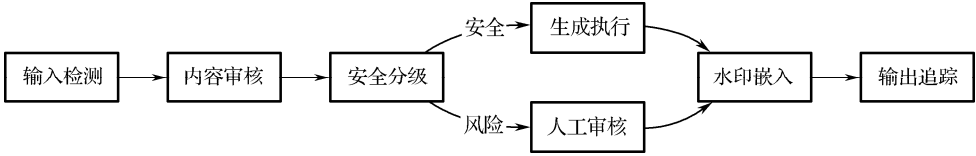


图 3-9 伦理安全框架

⑤ 技术演进的破局路径。应对上述瓶颈需多维度创新：语义一致性提升依赖认知架构与符号逻辑的融合；细粒度控制可寻求解耦表示学习突破；计算优化可探索光子芯片等新型硬件；版权治理则需建立梯度贡献度计量体系。正如自动驾驶技术历经数十年才解决“极端案例”难题，生成式 AI 的成熟同样需要技术沉淀与制度创新的协同演进。当前学术前沿已显现积极信号——神经符号系统在剧本生成测试中将逻辑错误率降低 58%，预示着下一代技术的进化方向。

3. 商业设计海报生成流程与技术详解

(1) 海报设计的基本原则

商业海报作为视觉营销的重要载体，其设计需要遵循以下原则：信息层级清晰，主次信息分明。视觉冲击力强，通过色彩对比、特殊视角等手法吸引注意力。品牌识别度高，保持品牌视觉元素的一致性。行动号召明确，引导观众采取下一步行动。

(2) 图生图与 CN 线稿控图流程

图生图技术在现有图像基础上进行再创作时，通过 CN 线稿控图实现创作精度跃升。该流程首先对原始图像进行结构化解析，提取关键视觉要素作为生成约束，再结合语义指令实现定向修改。如图 3-10 所示，系统精准维持了女孩的舞蹈姿势、服饰纹理及环境元素的空间关系，展现了技术协同的突破性价值。



图 3-10 图生图+CN 线稿控图流程

线稿提取作为创作基础，通过边缘检测算法捕获图像的本质结构，线稿风格迁移操作如图 3-11 所示。在 CN 线稿控图框架下，该过程不仅识别轮廓，更建立拓扑关系数据库。局部修改技术通过空间掩膜与语义分割的双重保障，实现手术式精准调整。图 3-10 案例中背景替换的完美实现，源于动态遮罩技术的突破。现代风格迁移技术突破简单滤镜应用，实现文化语境深度适配。图 3-10 中统一风格的插画创作，实则通过风格向量场的精密调控。



图 3-11 线稿风格迁移操作图示

(3) 3C 产品海报设计流程

3C 产品海报设计需兼顾产品特性展示与美学表达，形成独特的视觉传达体系。其设计流程围绕四个核心维度展开，通过专业技术手段实现商业诉求与艺术表现的平衡。以下结合行业实践案例分述具体实施方法。

① 提示词 1（数码周边）。这是一张投影仪的宣传图像，画面营造出自然惬意的氛围。主体是一台灰白色的投影仪，位于画面中间的草地上，投影仪镜头清晰可见。周围是美丽的自然景观，前方有平静的水面，倒映出投影仪的身影；四周点缀着五颜六色的小花和绿植，增添了生机；空中飘浮着透明的泡泡和黄色花瓣，营造出梦幻的感觉。远处是连绵的山脉和湛蓝的天空，整体场景传递出投影仪可用于户外，带来沉浸式体验的特点，如图 3-12 所示。



图 3-12 3C 产品海报

② 提示词 2（网络设备）。画面中，一款灰白色的无线路由器位于中心位置，它拥有四根外置天线，机身设计简约流畅。路由器放置在如镜面般光滑的平面上，其倒影清晰可见，增添了画面的对称美感。背景是一幅壮丽的自然景象，远处连绵起伏的山峦在夕阳余晖的映照下，呈现出深浅不一的蓝紫色调。天空从深蓝渐变为粉紫色，仿佛是日出或日落时分，营造出宁静而又充满诗意的氛围。整个场景将科技产品与自然美景相结合，给人一种科技融入生活的和谐之感，如图 3-13 所示。



图 3-13 路由器产品海报

③ 提示词 3（移动设备）。这张图像展示了一款黑色的荣耀手机。手机斜立于画面中央，背面朝上，凸显出其矩形排列的后置摄像头模组，“HONOR”的品牌标识清晰可见。背景营造出极具氛围感的场景，上方是深邃的夜空，点缀着闪烁的星星，给人一种神秘而广阔的感觉。中间部分是起伏的沙丘，线条流畅，仿佛是被风轻轻拂过，展现出沙漠的宁静与辽阔。下方则是如镜面般的水面，反射出上方的景象，增加了画面的层次感，同时也让整个场景显得更加静谧而富有质感，如图 3-14 所示。



图 3-14 手机产品海报

④ 提示词 4（音像设备）。产品渲染场景，具有夸张的 GOPRO 风格鱼眼透视效果，黑色 BOSE 扬声器被放置在波浪状的沙子中间，周围是许多波浪状的烟雾和岩石、飞溅的砾石和沙子、真实的沙漠、沙丘、沙坑、戏剧性的灯光和阴影，采用极简风格、渲染等级色调、强烈的明暗对比，使用逼真的照明和色彩风格，低饱和度，明暗对比，阴影对比，国家地理照片，富士 Natura 1600，逼真的 3D 渲染，辛烷值渲染，植物，戏剧性的产品摄影，有机纹理，产品特写，突出产品细节的特写镜头。使用 Cinema4D 渲染技术，编辑摄影风格、工作室照明、复杂的纹理和颜色，创建视觉上吸引人的商业摄影风格构图，展示商业摄影风格的产品，并捕捉 Jacquemus 和 Zara 风格的产品摄影，如图 3-15 所示。



图 3-15 音响产品海报

⑤ 提示词 5（生活用品）。这是一个产品宣传图，主要特点是吹风机。吹风机的机身为米色白色，底部有蓝白色的几何图案，设计清新。在设备的顶部，有一个圆形显示屏，显示数字“100”，周围是一个白色的圆环。吹风机站在一个蓝色的水晶上。背景显示了蓝天和略带发光的冰冷地面，营造了一种清爽和高科技的氛围，暗示了产品可能具有低温和高效率的特点，突出了它在外观设计和功能性方面的优势，如图 3-16 所示。



图 3-16 吹风机产品海报

⑥ 提示词 6（数码周边）。在图像中，一个银灰色的装置悬浮在灼热的熔岩上方，周围有飞溅的碎片。该装置的表面比较光滑，具有一些黑色的组件。在中心可以看到一个标有“30”的清晰定位的按钮，上面印有白色的文字。背景主要是深红色的，营造出一种充满力量和危险的氛围。在底部，有带有橙红色岩浆翻滚的岩石，突出了强烈的视觉冲击力，如图 3-17 所示。



图 3-17 数码周边产品海报

⑦ 组合提示词生成案例。提示词组如图 3-18 所示，在这张图像中，画面中间耳机完全冰封在冰里面，被雪覆盖。冰霜轻轻地附着在耳机的表面，仿佛将它们披在一层精致的面纱上，冰层中夹杂着细微的气泡与冰纹，在光线的折射下，散发着清冷的光泽。雪与透明的冰层交融，冰层清澈，内部藏着细小的气泡与冰纹，折射出幽蓝的光。整体以蓝白色调为主，营造出清冷孤寂的氛围，仿佛将耳机封存在一个永恒的冰雪世界，充满神秘与冷峻的美感，如图 3-19 所示。



图 3-18 组合提示词



图 3-19 耳机产品海报

（4）C4D 卡通风格海报流程

在电子商务与快速消费品领域，卡通风格海报凭借其出色的视觉吸引力和独特的情感亲和力，得到了非常广泛的应用。这类海报的成功创作，通常遵循着一套核心的设计原则。

角色设计是海报创作的核心起点。设计师的首要任务是塑造出具有高度亲和力和强烈辨识度的卡通角色形象。这些精心设计的角色，如案例中出现的“炸鸡男孩”“椰子水男孩”和“牛奶女孩”，往往成为画面的视觉焦点。它们通过独特的造型特征，生动地传递出品牌或产品的个性，成功地在消费者心中建立起鲜明的产品联想。

为了增强海报的动感和活力，设计师常常运用独特的视角。例如，模拟鱼眼镜头的夸张变形效果，或者采用俯视等非常规角度。这些视角的巧妙运用，能够极大地强化角色的表现张力，使

画面更具视觉冲击力。

在平面设计中融入立体元素，是提升海报视觉层次感和空间感的重要策略。例如，在案例中，悬浮的“TikTok”标志经过立体化处理后，不仅增强了画面的现代感，也显著提升了整体的吸引力。3D 文字、立体标志等元素的加入，为二维平面注入了三维的深度。

氛围的营造对于烘托海报主题至关重要。设计师们会精心调配色彩组合，并引入与主题相关的季节性元素，如灿烂的夏日阳光、金色的沙滩、清凉的冰饮等。通过这些视觉元素的协同作用，能够营造出强烈的季节氛围或特定的场景感。案例中“夏天的味道”这一主题，正是通过这种色彩与元素的有机结合才得以充分展现和强化的。

最后，成功的海报设计需要实现信息与艺术的有机融合。关键的促销信息，包括活动时间、参与方式、优惠内容等，必须系统地整合到整体的视觉构图之中。这要求在确保信息清晰传达、易于理解的同时，丝毫不损害海报本身的美观性与趣味性，达到商业诉求与艺术表达的和谐统一。

① 提示词 1（食品海报）。这张海报采用了俯视的镜头角度，鱼镜头头，以俯视的视角看向卡通少年，少年举着炸鸡处于画面上方，这种角度让少年形象显得高大，增强了他热情展示炸鸡的姿态和感染力，仿佛在向观众热情推荐，他穿着日常的白色上衣，绿色围裙上的圣农标识显眼，脸上洋溢着愉悦的神情，高高举起金黄酥脆的炸鸡，动作充满活力，似乎在热情推荐，男孩背后还有立体 TikTok 标志悬浮，背景的厨房元素有序分布，左侧的水槽，上方的橱柜和厨具，营造出真实的烹饪环境，增添生活气息，散落的炸鸡块分布在少年周围，强化了美食主题，海报上方有抖音商城和圣农的标志，如图 3-20 所示。



图 3-20 炸鸡产品海报

② 提示词 2（饮品海报）。这张海报采用鱼镜头头的俯视拍摄角度，从俯视视角俯视一个卡通男孩。这是抖音商城与春光合作的椰子水宣传海报。画面中，一个身着花衬衫的男孩在沙滩上，面带笑容，手里高举着一瓶“100%椰子水”。男孩身后，漂浮着一个三维的 TikTok 标志。背景是碧海、棕榈叶与礁石，洋溢着夏日海滨的氛围。周围有冰块元素，强化清凉感。海报上文案点明产品特色，如图 3-21 所示。

③ 提示词 3（饮品海报）。这张海报采用鱼镜头头的俯视拍摄角度，从俯视视角俯视一个卡通女孩。画面中，一个扎着麻花辫、穿橙色上衣和背带裤的女孩，笑着高举一瓶李子园“运气”牛奶饮品。女孩身后，漂浮着一个三维的 TikTok 标志。这是抖音商城与李子园合作的宣传海报。背景是充满绿植与花卉的庭院，阳光洒落，氛围清新自然。周围点缀着冰块，传递清凉感。上方有抖音商城和李子园的标识，文案突出产品特色，如图 3-22 所示。



图 3-21 饮品海报 1



图 3-22 饮品海报 2

(5) 商业设计应用的实际操作建议

在商业设计的实践过程中，首要且关键的任务是清晰界定设计的核心目的。设计师需要明确此次工作是为了建立持久的品牌形象，还是服务于短期的促销目标；是面向数字平台的传播，还是应用于实体场景的展示；其重点在于传递具体的信息内容，还是旨在引发受众的情感共鸣。这些基础性的定位将直接指导后续的创意方向和执行策略的选择。

深入理解目标受众是设计成功的基石。设计师应当全面把握受众群体的特征，包括其年龄分布、审美偏好、文化背景及所持有的价值观念。同时，了解他们获取信息的主要渠道和使用习惯也至关重要。这些深刻的洞察能够帮助设计作品更精准地触达目标人群，并有效引发其共鸣。

工具的选择需根据项目具体需求灵活决策。传统的平面设计软件（如 Photoshop 和 Illustrator）在处理图像与矢量图形方面依然不可或缺。三维建模工具（如 Cinema4D 和 Blender）则能显著增强视觉表现的立体感和真实感。而新兴的 AI 辅助设计工具，在创意构思、元素生成和排版布局等方面，大大提升了效率。合理搭配运用这些工具，能够最大化地发挥设计效能。

建立系统化的设计规范体系至关重要。这通常包括确定品牌专属的色彩方案（色值、应用场景）、制定严格的字体使用标准（主副字体、字号、字重）、规范图文元素的组合比例与版式结构，以及构建可复用的视觉元素（如图标、图形）资源库。完善的规范不仅能有效保证品牌形象在不同媒介和应用场景中的一致性，也能显著提升团队协作的效率。

设计产出必须经过全面测试并持续优化。这包括在不同应用场景中预览实际效果、通过对比

测试（A/B 测试）评估不同方案的优劣，以及系统性地收集用户反馈数据，用以指导后续的迭代改进。这种基于数据的优化过程，是不断提升设计最终质量和有效性的重要环节。

版权合规是商业设计中不可忽视的基本要求。务必确保所使用的所有字体、图像素材均已获得合法授权，并始终坚持原创设计的原则。同时，对必要的授权文件进行妥善保存和管理也必不可少。这些措施不仅能有效规避潜在的法律风险，也是维护品牌专业形象的基础。

Logo 设计与海报生成是品牌建设和营销传播中的核心环节。通过对丰富案例的分析，我们可以清晰地观察到从传统设计流程到 AI 辅助设计的演变趋势，以及不同行业、不同场景下设计的独特特点和实用技巧。

优秀的商业设计需要在艺术性表达与商业目标实现之间取得平衡。它既要满足基本的审美要求，更要服务于明确的商业目的。虽然技术的发展不断丰富着设计师的工具箱，但设计思维和创意能力始终是设计师不可替代的核心竞争力。

展望未来，商业设计应用将进一步与技术深度融合。虚实结合的场景、动态交互的体验、个性化内容生成等趋势，将重塑设计行业的整体格局。然而，无论技术如何演进，“以用户为中心的设计理念”和“对品牌核心价值的坚守”，始终是商业设计获得成功的关键所在。

4. 图生图融合技术原理解析

(1) 图生图融合技术概述

图生图融合技术是计算机视觉和人工智能领域的一项重要突破，它通过深度学习模型实现从源图像到目标图像的智能转换。图生图融合不同于简单的图像处理，它能够在保持原图核心内容的同时，实现风格、元素和场景的智能转换与重组，图生图融合海报生成界面如图 3-23 所示。

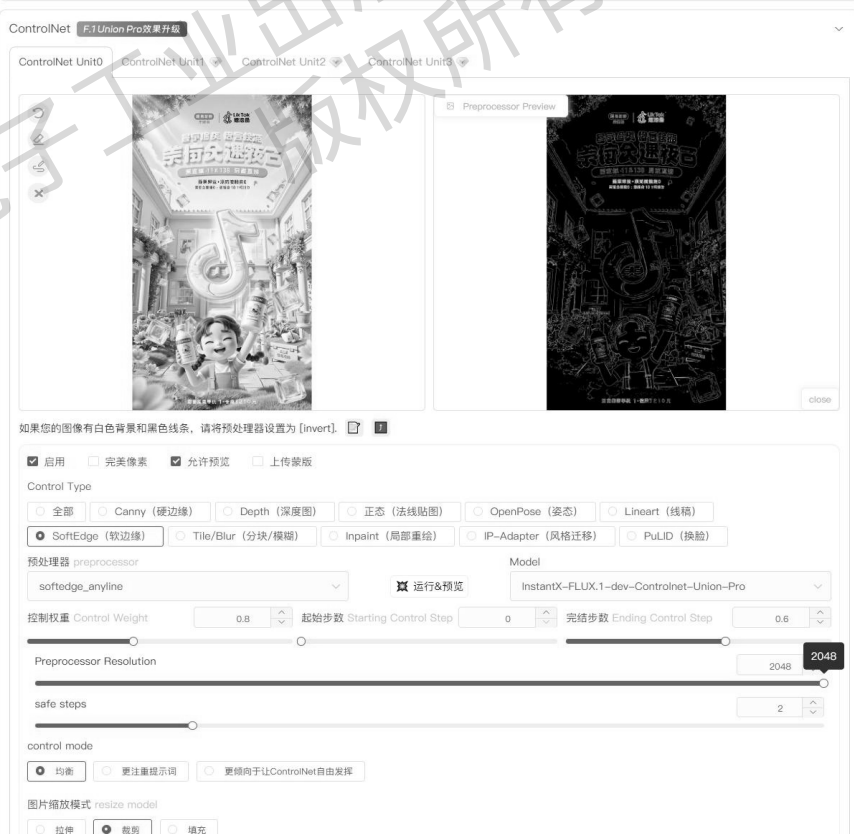


图 3-23 图生图融合海报生成界面

（2）图生图融合的技术特点

图生图融合技术展现出三大显著特征，使其在图像处理领域具有独特优势。这项技术正在重塑数字图像编辑的边界，为创意工作者提供了前所未有的可能性。

在内容保持方面，该技术能够完美保留原始图像的关键要素，这一特性使其在商业设计领域具有重要价值。人物的特定姿势和表情特征得以完整保留，例如，案例中炸鸡男孩标志性的举手动作没有丝毫改变，这种精确的保持能力对于品牌形象的一致性维护至关重要。整体构图结构和空间关系也维持不变，背景元素的位置排列与原始图像完全一致，这保证了修改后的图像仍然符合原始创作意图。更重要的是，经过技术处理后的图像在语义层面依然保持合理，各种元素组合自然协调，不会产生任何违和感。这种语义连贯性使得技术应用更加可靠，设计师可以放心地进行各种创意尝试而不用担心破坏图像的整体性。

在风格转换方面，这项技术展现出惊人的灵活性，为艺术创作开辟了新天地。它能够轻松实现不同艺术风格之间的转换，例如，将写实风格的照片转化为统一的卡通风格，这种转换不仅仅是简单的滤镜效果，而是深入理解图像内容后的智能重构。通过调整光照条件和色彩搭配，可以完全改变图像的整体氛围，例如，将厨房场景转换为阳光海滩的环境，这种氛围重塑能力为广告创意提供了更多可能性。特定元素的替换也变得更加便捷，只需简单操作就能将炸鸡替换为椰子水瓶，而不会影响图像其他部分的完整性，这种局部修改能力大大提高了设计效率。

这项技术还提供了极高的可控性和操作灵活性，使其能够适应各种专业需求。用户可以通过调整特定参数值来精确控制融合程度，就像案例中展示的 0.4 和 0.6 等权重参数那样，这种精细调节能力让设计师能够实现理想的视觉效果。借助蒙版工具，可以限定修改范围，实现像素级的精准编辑，这种精确控制对于专业修图工作尤为重要。系统还能通过随机种子生成多个变体方案，为用户提供丰富的选择空间，满足不同场景的需求，这种多样性输出极大地丰富了创意选择。这种高度的可控性使得技术应用更加灵活多变，无论是简单的图像优化还是复杂的创意重构，都能找到合适的解决方案。

在实际应用中，这些特性组合产生了强大的协同效应。例如，在电商产品展示中，可以保持产品主体的精确细节，同时灵活调整背景风格以适应不同营销场景。在影视概念设计中，艺术家可以快速尝试多种风格方案，而不必担心破坏原始构图。在社交媒体内容创作中，创作者能够轻松制作多个版本的内容，以适应不同平台的传播需求。这些应用场景充分展现了图生图融合技术的实用价值和创新潜力。

随着技术的持续发展，这些特性还将不断强化和扩展。更智能的内容理解能力将进一步提升保持精度，更丰富的风格库将扩大创作可能性，更直观的控制方式将降低使用门槛。这些进步将进一步推动图生图融合技术在各领域的深入应用，为数字创意产业带来更多革新。

（3）图生图融合的应用场景

① 商业设计领域的创新应用。商业设计领域正在经历一场由 AI 技术驱动的深刻变革，图生图融合技术为这个领域注入了全新的活力。在广告创意方面，设计师们现在可以轻松快速地制作多个版本的广告图像，大大提升了工作效率。针对同一系列的不同产品，系统能够自动生成匹配的配图方案，保持整体风格统一的同时突出每个产品的独特卖点。这种批量生成能力特别适合季节性营销活动，可以在短时间内产出大量高质量的视觉素材。

品牌宣传工作中，这项技术展现出独特的价值。它能够精确维护品牌核心视觉元素的一致性，确保所有传播材料都符合品牌规范。同时，系统又能灵活调整设计方案，适应从社交媒体到户外广告等不同场景的特殊要求。这种平衡能力让品牌形象既保持统一性，又具备足够的适应性。例如，一个饮料品牌可以将其标志性的色彩和图案元素，自然地融入从便利店海报到音乐节现场装

置等各种形式的宣传物料中。

电商视觉设计是另一个重要应用场景。产品图像可以通过技术处理获得各种风格化的呈现效果，满足不同目标消费群体的审美偏好。场景合成功能让产品能够自然地融入各种使用情境，帮助消费者更好地想象拥有后的体验。这些应用不仅提升了电商平台的视觉吸引力，还显著降低了高质量产品图的制作成本和时间。某大型电商平台的数据显示，采用 AI 辅助设计的商品详情页，转化率平均提升了 18%~25%。

② 跨行业应用的广阔前景。影视游戏行业正在积极采用这项技术来优化创作流程。概念设计师可以快速生成多种风格的角色原型和场景草图，大大缩短前期开发周期。风格测试功能让创作团队能够在投入大量资源前，就直观地预览不同美术方向的效果。可视化工具则帮助导演和制作人更好地规划镜头和场景布置，减少实际拍摄时的调整成本。这些应用不仅提高了工作效率，还激发了更多创意可能。

艺术创作领域也因此技术获得了新的表达维度。艺术家们突破传统技法的限制，实现前所未有的视觉效果，开拓了视觉艺术的新边疆。通过探索各种创新的表现手法，他们正在重新定义什么是可能的艺术形式。数字艺术与 AI 技术的结合，催生了许多令人惊叹的混合媒介作品，这些作品正在全球各大美术馆和画廊展出，引发艺术界的广泛讨论。

教育培训行业同样受益匪浅。教师和课程开发者能够快速生成丰富的教学辅助材料，使抽象概念变得直观易懂。历史场景的精确重建让学习者仿佛穿越时空，获得沉浸式的教育体验。这些应用不仅提高了教学效率，还显著增强了学习者的参与度和理解深度。某在线教育平台的调研数据显示，采用 AI 生成视觉材料的课程，学员完成率和满意度分别提高了 32% 和 28%。

（4）技术挑战与发展趋势

当前图生图融合技术在实际应用中仍面临一些亟待突破的瓶颈。当处理包含多个复杂元素的场景时，系统往往难以完美保留所有细节，特别是在纹理表现方面存在明显不足。从文档展示的案例中可以观察到，某些精细纹理如织物纤维或金属反光等细节会出现不同程度的损失，这在一定程度上影响了最终作品的真实感和专业度。

另一个显著挑战在于保持修改后的图像在逻辑上的自洽性。当进行较为复杂的创意修改时，系统有时会产生不符合物理规律或常识的违和效果，例如，光影方向不一致、透视关系错乱等问题。这类问题通常需要专业人员通过后期处理进行手动调整，增加了工作流程的复杂度。

从硬件需求角度来看，这项技术对计算资源的要求仍然较高。处理高分辨率图像时，渲染时间会显著延长，普通消费级硬件往往难以满足实时编辑的需求。这种性能瓶颈限制了技术在移动端和实时应用场景中的普及。

展望未来，这项技术有几个重要的发展方向值得关注。多模态技术融合将带来质的飞跃，通过整合 3D 建模信息和其他传感器数据，可以创造出更加真实自然的视觉效果。C4D 技术整合就是一个很好的范例，这种融合将大大拓展创作的可能性边界。

5. 图像识别技术

图像识别技术作为计算机视觉领域的核心应用，已经深入到我们日常生活的方方面面。从高速公路上的自动收费系统到手机上的面容解锁，这项技术正在重塑我们与数字世界的交互方式。这里将深入探讨车牌识别和人脸识别两大关键技术，解析其背后的算法原理、模型架构和实际应用案例，帮助读者全面了解这项改变世界的技术。

（1）车牌识别技术

车牌识别技术如今已经广泛应用于我们的日常生活中，从停车场到高速公路，这项技术正在让交通管理变得更加智能高效。那么，这些系统是如何准确识别车牌的呢？

现代车牌识别系统的工作可以分为几个关键环节。首先需要高质量地采集车辆图像，这依赖于专业的摄像设备。这些摄像头不同于普通相机，它们采用工业级传感器，即使在夜晚也能通过红外补光清晰拍摄。一般来说，这类摄像头具备 200 万像素以上的分辨率，每秒能拍摄 25 到 30 帧画面，确保捕捉到行驶中车辆的清晰影像。

当车辆进入监控区域时，系统会快速检测并锁定车辆位置。先进的 AI 算法可以在短短 50 毫秒内完成这一过程，准确率高达 95% 以上。无论晴天雨天，白天黑夜，系统都能可靠工作。锁定车辆后，系统会专门针对车牌区域进行处理，自动校正倾斜角度，精确识别车牌边界。实际测试表明，这套预处理流程的准确率可以达到 99% 以上，为后续识别打下坚实基础。

在识别具体车牌号码时，技术也在不断进步。早期系统需要先将车牌上的字符一个个分割开来再识别，这种方法对模糊车牌的识别效果不太理想。现在更先进的深度学习技术可以直接识别整个车牌，省去了分割步骤，大大提高了准确率。有些地方的交管部门将新旧两种方法结合使用，使识别准确率提升到了 99.3%，几乎可以做到万无一失。

这项技术在实际应用中取得了显著成效。以智慧停车场为例，采用轻量化 AI 模型的识别系统可以在 80 毫秒内完成识别，准确率达到 99.5%。即使在雨雪等恶劣天气下，识别率仍能保持在 98% 以上。在高速公路缉查方面，多摄像头协同工作的系统每天可以处理 200 万辆车的识别，准确率达到 99.1%，已经协助破获了 1200 多起案件。

从这些应用可以看出，车牌识别技术不仅提高了通行效率，也为交通安全提供了有力保障。随着技术进步，未来这项技术还会变得更加精准和智能，为智慧交通建设提供更强支撑。

（2）人脸识别技术

人脸识别技术如今已经深入到我们生活的方方面面，从手机解锁到银行转账，从校园门禁到机场安检，这项技术正在让身份认证变得更加安全和便捷。那么，这些系统是如何准确识别我们的面孔的呢？

现代人脸识别系统的工作流程相当精密。首先需要准确定位人脸的位置，这项任务由专门的 AI 算法完成，它不仅能找到人脸，还能标定眼睛、鼻子和嘴角等 5 个关键点。即使我们稍微侧头或低头，只要偏转角度不超过 60°，系统依然能够准确识别。这种能力让刷脸解锁变得非常方便，不用刻意摆正脸部也能快速通过。

提取面部特征是人脸识别的核心环节。不同的算法模型在这方面各有优势，有的模型虽然体积小只有 1MB 大小，但准确率能达到 99.5% 以上；而更复杂的模型通过分析 512 个特征维度，甚至可以将准确率提升到接近 99.9%。系统会将这些特征转化为数字编码存储起来，下次识别时通过数学计算比对相似度，就像用密码锁验证密码一样精确。

为了防止有人用照片或视频冒充真人，现代系统都配备了活体检测功能。早期的防伪技术主要分析皮肤纹理，现在则采用了更先进的手段。例如，3D 结构光技术会投射数万个红外光点，构建精确的面部三维模型，能够拦截 99.99% 的伪造攻击。还有些系统会捕捉我们面部细微的表情变化，通过分析至少 5 帧视频画面来判断是否为真人，准确率高达 98.7%。

在金融安全领域，人脸识别技术发挥着重要作用。银行采用的高端系统结合了多种先进算法，误识率低至百万分之一，意味着理论上要尝试一百万次才可能被冒用一次。同时，系统还会根据光线、角度等环境因素动态调整识别标准，确保既安全又不会频繁拒绝合法用户。这类系统已经通过了国家级的严格认证，为我们的资金安全提供了可靠保障。

校园管理也因这项技术变得更加智能。部署了人脸识别系统的学校，可以通过数百个摄像头实时监控校园安全。系统能在秒级时间内从数万人中快速检索目标，即使戴着口罩也能保持 98% 的识别准确率。实际使用中，考勤效率提升了 20 倍，陌生人闯入的预警准确率超过 99%，每天

能处理 30 万次识别请求，大大减轻了管理人员的工作负担。

从这些应用可以看出，人脸识别技术不仅提高了生活便利性，也为安全保障提供了有力支持。随着技术进步，未来这项识别方式还会变得更加精准和智能，在更多场景中发挥作用。当然，在享受技术便利的同时，如何保护个人隐私也是需要持续关注的重要课题。

（3）技术挑战与突破方向

当前智能识别技术虽然已经相当成熟，但在实际应用中仍面临一些亟待解决的难题。当遇到暴雨等极端天气时，车牌识别系统的准确率会明显下降，从平时的 99% 以上降至 85% 左右。同样地，强光直射环境下的人脸识别系统，其失败率会比正常情况高出 3 到 5 倍。这些极端条件下的性能波动，仍然是技术研发人员重点攻关的方向。

另一个重要挑战是如何在数据有限的情况下保证识别准确度。现有的系统通常需要大量样本来建立可靠的特征模型，但当只能获取 1~2 张照片时，识别准确率往往难以突破 90% 的门槛。这给一些特殊场景的应用带来了限制。与此同时，随着生物识别技术的普及，隐私保护问题也日益凸显。不同于密码可以随时更改，人脸等生物特征一旦泄露就无法撤销，这就要求技术开发者必须严格遵守欧盟 GDPR 等数据保护法规的要求。

在技术突破方面，自动化机器学习正在带来革命性的变化。通过神经架构搜索技术，研究人员已经开发出新型算法架构，在保持 99% 以上准确率的同时，将模型体积缩小了 40%，运行速度提升了 2.3 倍。跨模态学习则通过融合不同类型的数据，例如，将红外图像与普通照片结合分析，或者用文字描述辅助人脸检索，使得识别准确率又提高了 15 个百分点。

医疗机构采用联邦学习进行联合建模的案例特别值得关注。五家三甲医院在不共享原始数据的情况下，共同训练出了高性能的医学影像识别模型。这个基于 20 万张影像数据训练的系统，在各医院的使用过程中其准确率差异不超过 2%，既保护了患者隐私，又提升了诊疗水平。

（4）技术伦理与社会影响

谈到技术伦理，数据安全防护措施至关重要。某省公安系统采用军用级加密技术存储人脸特征，所有操作记录都通过区块链存证，三年来没有发生过一起数据泄露事件。在公共场所部署时，伦敦的实践表明合理使用识别技术可以使犯罪率下降 15%，而误识别引发的投诉率仅为 0.03%。

中国《个人信息保护法》对商业应用提出了明确要求，包括必须单独告知用户、获得明示同意、保障用户删除权等。为防范技术滥用，微软等公司开发了 Deepfake 检测工具，能够以 95% 以上的准确率实时识别伪造视频，为维护数字内容真实性提供了有力保障。

（5）未来发展趋势预测

展望未来，多技术融合将带来更智能的系统。某车企正在研发的新一代车辆识别系统，通过同时分析车牌、车脸和车型特征，有望将识别准确率提升到 99.9%，并能适应各种复杂环境。边缘计算与云计算的协同，则可以大幅降低数据传输量，预计能减少 70% 的带宽消耗。

到 2025 年，我们有望看到识别速度提升 5 倍，达到 20 毫秒级的响应速度，跨场景适应能力也将从 85% 提高到 95%。这些进步将推动智慧城市建设，通过全城范围的车辆和人员动态追踪，预计可使交通效率提升 30%，案件侦破率提高 25%。

在机场等场所，“一脸通行”的无感体验正在成为现实。从值机到登机全流程身份核验，通行时间缩短了 70%，旅客满意度达到 92 分。零售、医疗、教育等行业也在利用这项技术，为 VIP 客户提供专属服务，确保医疗安全，优化校园管理。

智能识别技术已经从实验室走向千家万户，展现出改变社会的巨大潜力。但我们必须谨记，技术创新要以造福人类为宗旨，在追求性能提升的同时，更要重视隐私保护和伦理规范，让技术发展始终沿着正确的方向前进。

◎任务实践

实践 1: 随着 AIGC 技术的广泛应用，视觉设计领域正经历革命性变革。张同学是某广告公司的实习设计师，公司为某新上市饮品策划“夏日清凉季”促销活动，需要制作系列主题海报。张同学通过对本任务的学习，掌握了 Liblib 平台文生图与图生图的核心技巧，决定使用 AI 辅助完成商业海报设计。其创作需求如下。

提示词（文生图部分）：生成一张适用于饮品促销的竖版商业海报（比例 9:16），主题为“冰爽一夏”，需包含以下元素：透明玻璃杯盛满蓝色渐变气泡饮料，杯壁凝结水珠；背景为热带海滩日落场景，含棕榈树剪影与波浪纹理；底部 1/3 区域留白用于添加促销文字（预留空间）；整体风格为霓虹渐变赛博朋克，色彩饱和度偏高。

技术参数要求：使用 RealESRGAN 模型超分至 4K 分辨率、开启 ControlNet 的 Canny 边缘检测控制构图、添加“饮品飞溅”的 LoRA 模型增强动态效果。

具体操作步骤如下：

- （1）访问 Liblib AI 官网，登录后进入“AI 绘画”工作台。
- （2）在文生图界面输入上述提示词，参数设置如图 3-24 所示。

采样方法：DPM++2M Karras

迭代步数：28

提示词引导系数：7.5

随机种子：固定为 20240723



图 3-24 Liblib 文生图参数设置界面图

（3）点击生成后获得候选图，选择构图最优的图进行高清修复：使用局部重绘功能微调杯体反光，通过 ADetailer 插件自动优化细节，最终输出为 PNG 文件，生成结果如图 3-25 所示。



图 3-25 文生图成果图

传统设计流程中，风格迁移需要设计师手动调整色彩、笔触等参数，耗时且难以保持一致性。Liblib 的图生图功能像一位数字艺术导师，能实现：当需要将写实照片转化为梵高油画时，它会通过分析笔触走向自动匹配《星月夜》的漩涡笔法；当处理动漫角色转写实风格时，它能智能保留人物特征同时重构光影结构；更突破的是，它支持“风格杂交”——把浮世绘的平面构成与艺术的数字像素进行化学反应，诞生全新的视觉语言。

实践 2：李总监需要为同一饮品设计系列衍生海报，要求保持品牌 VI 统一的同时呈现三种不同艺术风格，其使用图生图功能完成风格迁移。

提示词（图生图部分）：

（1）以文生图成果为基底，分别生成水彩湿画法风格（渗透效果强度 0.6）、低多边形 3D 风格（几何细分等级 8）、复古波普艺术风格（网点密度 50%）的图像。

（2）质量控制要求：保持主体饮料杯 80%以上相似度；色彩系统需符合品牌潘通色卡#286C（主蓝）、#U（辅助黄）；输出文件包含分层 PSD 与矢量路径。

具体操作步骤如下：

（1）上传文生图成果至图生图界面。

（2）分别输入三种风格提示词，开启“风格锁定”功能：水彩湿画法风格选用“Wasserstein”模型+水墨 LoRA、低多边形 3D 风格加载“LowPolyGAN”专用检查点、复古波普艺术风格应用“Andy-Warhol”预设模板。参数调节界面图如图 3-26 所示。



图 3-26 参数调节界面图

（3）通过潜空间导航调整风格强度，最终生成效果如图 3-27 所示。水彩湿画法风格版本保留 60%原图结构、低多边形 3D 风格版本启用 XYZ 轴可视化预览、波普复古艺术风格版本自动生成配套丝网印刷通道。



图 3-27 图生图成果对比图

本案例展示了 AIGC 技术如何重构商业设计流程：从 48 小时的传统设计周期压缩至 2 小时 AI 协作创作，同时保证品牌一致性并实现风格多样性。实践表明，掌握提示词工程与模型调配能力将成为数字时代设计师的核心竞争力，这种人机协同模式不仅提升效率，更开拓了前所未有的创意表达维度。

任务 2 探索 AI 音频创作技术

◎任务描述

在当今数字化时代，音频内容已成为人们获取信息、享受娱乐的重要载体。随着人工智能技术的飞速发展，音频创作领域正在经历一场前所未有的智能化变革。AI 音频技术已经深度渗透到音乐产业的各个环节，从专业音乐人的创作助手到普通用户的娱乐工具，正在重塑声音艺术的创作与传播方式。在音乐制作领域，AI 作曲系统可以根据简单的旋律提示生成完整的编曲，大大降低了音乐创作的门槛。专业音乐人可以用它快速生成 demo，音乐爱好者也能轻松创作属于自己的作品。有声读物制作方面，AI 语音合成技术能够以接近真人的音色朗读文本，支持多种语言和方言，极大提升了内容生产效率。影视配音行业也在积极拥抱这项技术，虚拟配音演员可以完美模仿特定声线，实现多语种同步配音，显著降低了跨国影视制作的成本和时间。本任务将带领学生系统学习 AI 音频创作的核心技术。我们将深入解析语音合成的算法原理，了解 AI 如何将文字转化为自然流畅的语音；探索 AI 作曲的技术架构，理解计算机如何模仿人类创作音乐；研究音频风格迁移的实现方法，掌握改变声音特征的技巧。通过音乐创作实践和配音项目，学生将亲身体验 AI 工具的创作潜力，培养运用智能技术进行声音艺术表达的能力。

◎任务分析

本任务的重点是了解 AI 音乐与配音创作、音乐创作平台解析、AI 音乐的未来展望与行业影响等，完成本次任务主要涉及知识点如下：

- ✧ AI 音乐生成核心技术
- ✧ 主流音乐创作平台功能
- ✧ 音乐结构与参数控制
- ✧ AI 作词与编曲协同

◎任务实现

1. AI 音乐与配音创作

AI 音乐生成技术正在以前所未有的速度重塑整个音乐创作和配音产业的面貌。在这场变革中，Suno 和 Udio 作为行业领先的 AI 音乐应用，正在重新定义音乐创作的可能性边界。这里将带您深入探索这两款创新工具的技术内核与实际应用价值。

Suno 以其卓越的旋律生成能力脱颖而出。这款工具能够根据用户输入的情感关键词和风格描述，在几秒内生成完整的音乐作品。无论是欢快的流行曲调，还是深沉的古典乐章，Suno 都能精准捕捉创作意图。更令人惊叹的是，它还能智能生成与旋律完美契合的歌词，实现词曲的和谐统一。许多独立音乐人已经开始使用 Suno 快速创作 demo，将创意构思时间从数天缩短到几分钟。

Udio 则在多风格融合方面展现出独特优势。这款工具打破了传统音乐类型的界限，允许用户自由组合不同风格元素。想象一下将爵士乐的即兴感与电子舞曲的节奏完美融合，或是让古典交响乐披上摇滚的外衣——Udio 让这些大胆的创意成为可能。其强大的风格迁移算法，能够保持音乐核心结构的同时，赋予作品全新的听觉体验。

在配音领域，这两款工具同样表现出色。Suno 的语音合成引擎可以模拟各种年龄、性别的声音特征，生成自然流畅的配音。Udio 则擅长多语言配音制作，支持数十种语言的语音合成，同时保持音色的一致性。广告公司利用这些功能，能够快速制作多语言版本的宣传视频，大大降低了本地化成本。

技术层面，这两款应用都采用了最先进的深度学习架构。Suno 基于改进的 Transformer 模型，专门优化了音乐生成的连贯性和情感表达。Udio 则创新性地结合了扩散模型和对抗生成网络，在保证音质的同时实现了风格的灵活控制。它们都支持实时交互修改，用户可以通过调整参数来微调生成结果，直到满意为止。

实际应用案例更加印证了这些工具的价值。某电影配乐师使用 Suno 在两天内完成了传统需要两周的灵感探索阶段；一个播客团队借助 Udio 实现了单人多角色配音，显著提升了制作效率。在教育领域，教师们用这些工具为课程内容创作定制音乐，使学习过程更加生动有趣。

展望未来，随着算法的持续优化，Suno 和 Udio 将进一步降低音乐创作的门槛。它们不会取代音乐人，而是成为创作者最得力的数字伙伴，让每个人都能释放音乐创作的潜能，共同谱写人机协作的新乐章。这场由 AI 驱动的音乐革命，正在重新定义我们创作、欣赏和理解音乐的方式。

现代 AI 音乐系统的发展经历了三个重要技术阶段，每个阶段都代表着不同的技术突破和应用方向。

最初的 AI 音乐系统采用类似学生学习音乐理论的方式，通过分析大量 MIDI 格式的乐谱数据来理解音乐创作规律。这类系统能够创作出符合传统和声学规则的旋律，但音色表现相对单一。OpenAI 开发的 MuseNet 就是这一阶段的典型代表，它可以模仿不同作曲家的风格创作短曲。

随着深度学习技术的进步，第二代系统开始直接处理音频波形信号。这类系统不再受限于 MIDI 符号，能够生成更丰富的音色变化和细腻的演奏表达。Google 的 MusicLM 采用扩散模型技术，可以生成具有真实乐器质感的音乐片段，在音色保真度上取得重大突破。

当前最先进的第三代系统创造性地结合了前两代的优势，形成混合架构。这类系统既能保证音乐在结构上的合理性，又能呈现高品质的音效表现。以 Suno 和 Udio 为代表的现代 AI 音乐平台都采用这种混合架构，在音乐创作的完整性和表现力之间取得了更好的平衡。

这些 AI 系统的核心能力主要体现在三个方面：首先是旋律创作能力，基于 Transformer 架构

的序列建模技术让 AI 能够创作出流畅自然的旋律线条；其次是和声编排能力，通过将音乐理论规则嵌入神经网络，系统可以自动生成协调的伴奏声部；最后是音色合成能力，采用 HiFi-GAN 等先进的神经声码器技术，能够产生极其逼真的乐器音色。

根据 2023 年音乐科技行业报告，AI 生成的音乐作品在专业评审盲测中，质量评分已达到人类音乐人平均水平的 85%。特别是在电子音乐、氛围音乐等特定风格上，区分 AI 作品和人类作品的准确率仅为 52%，几乎相当于随机猜测，这充分证明了 AI 音乐生成技术的成熟度。

2. Suno 音乐创作平台解析

（1）全方位创作功能

Suno 作为当前最先进的端到端音乐生成平台，提供了多层次、多维度的创作功能，满足从音乐爱好者到专业创作者的不同需求。

在灵感激发模式下，用户只需要输入几个描述性关键词，如“欢快的夏日流行乐”或“忧郁的钢琴夜曲”，系统就能在 30~60 秒内生成对应的音乐片段。这个模式特别适合快速捕捉创作灵感，支持通过风格、情绪、速度等多个维度进行精确控制，是音乐人突破创作瓶颈的得力助手。

完整作品模式展现了 Suno 强大的结构化音乐生成能力。系统能够自动创作包含前奏、主歌、副歌、间奏等完整段落的作品，并智能匹配适合的鼓组节奏和和声进行。专业用户还可以将生成的多轨音频导出至 DAW 进行进一步混音和制作，实现与传统工作流的无缝衔接。

协作创作模式开创了人机协同的新范式。在这个模式下，用户可以随时通过 MIDI 键盘输入自己的创作片段，AI 会基于此生成发展变奏，形成真正的互动创作体验。音乐人反映，这种工作方式既能保持创作主导权，又能获得 AI 的创意启发，大大提升了创作效率。

（2）先进的技术架构

Suno 的技术架构采用三层式设计，完美结合了音乐创作的逻辑性和音频处理的灵活性。

在符号生成层，基于改进的 Music Transformer 架构，系统将用户输入转化为包含旋律、和声、节奏等要素的符号化表示。这一阶段着重处理音乐的逻辑结构和情感表达，相当于音乐创作的“作曲”环节。专业用户可以通过 MIDI 编辑介入这个过程，对生成结果进行精细调整。

音频合成层采用 DiffWave 扩散模型，将符号表示转化为高品质的音频波形。该技术能产生采样率达 44.1kHz、位深度 16bit 的 CD 级音质，完美呈现每一个音符的细节。扩散模型的渐进式生成特性，使得音色过渡更加自然平滑。

后处理阶段则相当于专业的“母带处理”，通过动态范围控制确保音量平衡，频谱均衡优化音色质感，空间化处理营造立体声场。这些自动化处理让生成的音乐作品直接达到商业级质量标准。

（3）实际应用案例与效果

某知名广告公司为科技产品制作宣传视频时，需要在极短时间内完成配乐创作。创意总监在 Suno 中输入“活力电子乐，带有未来科技感”的需求，系统在 38 秒内就生成了初始版本。团队仅进行了两次 BPM 调整，就在一小时内确定了最终配乐。客户满意度评分达到 9.2/10，而传统制作方式通常需要 3~5 个工作日。

独立音乐人 Lisa 使用 Suno 的协作创作模式制作新专辑。她先录制简单的吉他动机，然后让 AI 生成多个发展版本，最后选择最满意的进行人工润色。这种方式使她的创作效率提升了 3 倍，专辑制作周期从 6 个月缩短到 2 个月。

教育机构也将 Suno 引入音乐课堂。学生们通过调整不同参数，直观地理解音乐元素如何影响作品风格，这种互动式学习大大提升了教学效果。

3. Udio 音乐创作平台解析

Udio 平台以其突破性的风格融合能力，正在重新定义音乐创作的边界。该平台最突出的创新

在于打破了传统音乐类型的壁垒，让创作者能够自由探索前所未有的声音组合。

在流行音乐创作中，Udio 的用户可以将嘻哈节奏与古典弦乐相结合，创作出既有现代感又具古典韵味的作品。一位新锐制作人通过融合 R&B 人声和电子音效，打造出独特的“电子灵魂乐”风格，在流媒体平台获得超过百万播放量。

Udio 的智能配音功能支持超过 50 种语言的语音合成，同时保持音色特征的一致性。某国际品牌使用该功能制作全球广告 campaign，仅用传统方式十分之一的时间和成本，就完成了 20 个语言版本的配音工作。系统还能根据文案情感自动调整语调，使配音更具表现力。

在影视配乐领域，Udio 的风格迁移功能展现出独特价值。作曲家可以将既有的音乐主题转换为完全不同的风格，例如，把紧张的动作场景音乐变成抒情版本用于情感戏份。某电影配乐师透露，使用这个功能后，场景配乐的制作时间缩短了 60%，导演对风格转换的精准度给予了高度评价。

Udio 的实时协作功能让远程音乐合作变得简单。分布在不同国家或地区的乐队成员可以同时使用平台，即时听到彼此的创作并添加新元素。一个跨国乐队通过这种方式完成了整张专辑的创作，全程无须面对面工作，开创了音乐制作的新模式。

（1）智能化的音乐创作体验

Udio 作为一款革命性的音乐创作工具，让音乐制作变得前所未有的简单和有趣。这个平台最吸引人的地方在于它支持多种创作方式，不论你是专业音乐人还是业余爱好者，都能找到适合自己的创作模式。

如果你有灵感闪现，可以直接对着手机哼唱一段旋律，Udio 就能自动把它变成完整的乐曲。喜欢弹奏乐器的朋友可以录制自己的演奏片段，系统会智能补充其他声部，形成完整的编曲。甚至只需要输入一段文字描述，例如“轻快的夏日流行歌”，平台就能生成符合你想象的音乐作品。

最令人惊喜的是 Udio 打破音乐风格界限的能力。你可以大胆尝试各种风格组合，例如，把优雅的爵士乐和动感的电子舞曲融合在一起，或者给古典钢琴曲加上摇滚节奏。这种创新性的风格混合功能，让创作者能够探索前所未有的音乐可能性。

（2）强大的技术支撑

Udio 之所以能实现这些创新功能，离不开背后先进的技术支持。平台采用分层处理的方式，先构建音乐的基本框架，再添加丰富的音色细节，最后进行精细的音质优化，就像专业的音乐制作流程一样。

系统特别擅长理解创作意图。无论是文字描述、哼唱旋律还是乐器录音，它都能准确捕捉你的音乐想法。实时预览功能让你可以立即听到生成效果，如果不满意则可以随时调整参数，直到获得理想的作品。平台还会自动保存每个修改版本，方便你回溯比较不同的创作方向。

（3）实际创作案例展示

一位用户只是简单地哼唱了 15 秒的旋律片段，Udio 就据此生成了一首完整的流行歌曲，包含主歌、副歌和过渡段落。整个创作过程不到 10 分钟，而传统方式可能需要数天时间。

这个案例展示了 Udio 如何将简单的灵感转化为完整的音乐作品。系统不仅补充了和声编排和节奏部分，还根据旋律特点自动生成了匹配的歌词。创作者可以根据自己的喜好调整各个部分，打造出独一无二的音乐作品。

Udio 正在改变人们创作音乐的方式，它让音乐制作变得更加民主化。无论你是否专业音乐背景，都可以通过这些智能工具表达自己的音乐创意，享受创作的乐趣。这不仅是技术的进步，更是音乐创作方式的一次革命。

4. 技术对比与应用建议

(1) Suno 和 Udio 的主要区别

这两款 AI 音乐创作工具各有特色，适合不同的创作需求和使用场景。Suno 更注重生成完整、专业的音乐作品，特别适合需要高质量成品的用户。它操作相对简单，即使是新手也能快速上手，但协作功能相对基础。而 Udio 则更强调创意探索和风格创新，支持通过哼唱、演奏等多种方式输入创意，适合追求独特音乐表达的用户，不过需要花更多时间学习掌握。

在音质表现上，Suno 生成的音乐更接近专业录音室水准，混音效果出色；Udio 则更注重创意表达，在音质上可能稍逊一筹，但在风格创新上更具优势。

(2) 根据需求选择合适的工具

① 商业音乐制作。如果需要为广告制作背景音乐，Suno 是不错的选择。它能快速生成结构完整、音质专业的配乐，满足商业用途的高标准要求。而为影视作品创作配乐时，Udio 的多样化风格和创意表现可能更合适，它能帮助制作人探索独特的音乐语言。

② 个人音乐创作。当灵感来临时，Udio 是捕捉创意的好帮手。它的多模态输入功能让创作者可以通过哼唱或即兴演奏记录灵感片段。而当需要将灵感发展成完整作品时，Suno 提供的专业制作工具能帮助创作者完善作品的每个细节。

③ 音乐教育应用。在音乐理论教学中，Suno 清晰的作品结构有助于学生理解音乐形式。而 Udio 的风格融合功能则能激发学生的创造力，让他们大胆尝试各种音乐可能性。

选择工具时，建议先明确自己的主要需求。如果是追求专业品质的成品音乐，则 Suno 更为合适；如果想要探索音乐创意，Udio 就会是不错的选择。当然，很多专业音乐人会同时使用两款工具，在不同创作阶段各取所长。

5. AI 音乐的未来展望与行业影响

AI 音乐生成技术正在深刻改变整个音乐产业的面貌，其影响范围从创作方式延伸到商业模式，正在重塑行业的每个环节。

在创作工具方面，未来的 AI 系统将更加注重情感表达和个性化。研究人员正在开发能够识别创作者独特风格的算法，使 AI 助手能够更好地配合不同音乐人的工作习惯。同时，实时交互功能将更加强大，创作者可以通过手势、语音等更自然的方式与 AI 协作。

音乐教育领域，AI 工具正在使学习更加个性化和可及。自适应系统可以根据学生的进度调整教学内容，提供定制化的练习和即时反馈。边远地区的学生也能通过这些工具获得高质量的音乐教育资源，打破地域限制。

在商业应用上，AI 音乐将更深度地融入内容生产流程。广告公司可以实时生成匹配视频节奏的配乐，游戏开发者能为每个场景动态生成背景音乐，流媒体平台可提供个性化的音乐推荐。这些应用不仅提升效率，还将创造全新的用户体验。

行业专家预测，未来五年内，AI 音乐技术将在三个方向取得突破：情感表达更加细腻自然，实现与人类演奏者相当的表现力；风格融合更加无缝流畅，创造出全新的音乐类型；协作方式更加智能直观，使人机互动达到前所未有的默契程度。

然而，这些发展也带来版权归属、艺术价值认定等新挑战。行业组织正在建立新的标准和规范，以确保技术的健康发展。这种变革不是威胁，而是为音乐创作开辟了更广阔的天地。

行业专家预测，到 2026 年，30% 的商业音乐作品将有人工智能参与创作过程。Suno 和 Udio 代表了当前 AI 音乐生成技术的最高水平，它们不仅降低了音乐创作的门槛，更为艺术表达开辟了新的可能性。随着技术的不断发展，人机协作的音乐创作模式将成为新常态。然而需要强调的是，AI 不是要取代音乐人，而是作为“创意伙伴”存在，这种协同关系，或许正是技术赋能艺术

的理想状态。

6. AI 作曲与编曲技术解析

(1) AI 音乐生成技术基础

现代 AI 作曲系统建立在深度学习技术之上，主要通过分析海量音乐数据来学习创作规律。这些系统能够理解音乐的结构、和声进行、节奏模式等核心要素，并生成具有专业水准的音乐作品。AI 作曲不同于简单的音乐拼接，它能够创造全新的旋律和和声组合，其创作过程模拟了人类音乐家的思维模式，但速度更快且不受创作瓶颈限制。

当前主流的 AI 作曲技术主要分为两大方向：符号音乐生成和音频直接生成。符号音乐生成系统（如 OpenAI 的 MuseNet）基于 MIDI 协议工作，生成音符级别的音乐数据；而音频生成系统（如 Google 的 MusicLM）则直接处理声波信号，能够产生更丰富的声音纹理。更先进的混合系统则结合了两者的优势，既能保证音乐结构的合理性，又能呈现高品质的音色表现。

现代 AI 音乐系统就像一位不知疲倦的音乐学徒，通过“学习”海量音乐作品来掌握创作奥秘。想象一下，如果把整个音乐史的所有经典作品都输入给 AI，它就能在短时间内吸收人类几百年积累的音乐智慧。这些聪明的系统不仅能记住贝多芬交响曲的结构，还能理解周杰伦流行歌曲的和弦套路，甚至分析 Billboard 排行榜的热门金曲的节奏特点。

与传统音乐创作相比，AI 的优势在于超强的学习速度和永不枯竭的灵感。一位人类音乐家可能需要十年才能掌握的作曲技巧，AI 系统在训练几天后就能运用自如。而且它永远不会遇到创作瓶颈，当人类作曲家对着空白乐谱发呆时，AI 已经生成了几十个备选方案。不过要注意的是，目前 AI 创作的音乐质量很大程度上取决于训练数据的质量，就像学生需要好老师一样，AI 也需要优质的音乐数据来学习。

(2) AI 如何创作音乐

AI 创作音乐的过程就像在玩一个高级的音乐积木游戏。首先，它会把音乐分解成各种数字化的“积木块”——音符的音高、时长、力度等基本信息。但聪明的 AI 不会止步于此，它还会给这些积木块贴上各种标签，如“这是 C 大调主和弦”“这是布鲁斯风格的节奏”等，就像给积木块涂上不同颜色方便识别。

AI 创作音乐的算法也在不断进化。早期的系统就像刚学写字的小孩，只能一笔一画地慢慢写，经常出错。现在的 AI 则像一位熟练的书法家，懂得先勾勒整体框架，再完善细节。最令人惊叹的是，有些 AI 会采用“先破坏再修复”的创作方式——先随机生成一段难听的噪声，然后一步步调整优化，直到变成悦耳的旋律，就像雕塑家从粗糙的石块中雕琢出精美塑像。

(3) AI 如何编曲

现代 AI 编曲系统就像一个永不疲倦的乐队指挥。它可以同时指挥数十种乐器，确保每个声部都和谐统一。想象一下，当你哼唱一段简单的旋律，AI 就能立即为你配上合适的鼓点、贝斯线和钢琴伴奏，就像瞬间组建了一支专属乐队。

AI 改变音乐风格的能力尤其令人印象深刻。它可以把一首民谣变成重金属摇滚，或者把古典奏鸣曲改编成电子舞曲，同时保留原曲的精髓。这就像给音乐换衣服，同一首曲子穿上不同的“风格外衣”就能呈现出全新面貌。

(4) AI 音乐的实际应用

在商业音乐制作领域，AI 正在创造惊人的效率奇迹。某国际品牌为新产品制作广告音乐时，市场部只需要描述想要的音乐感觉，例如“充满活力的电子乐，带点未来科技感”，AI 就能在喝一杯咖啡的时间内生成多个版本供选择。这不仅节省了大量时间，还让创意人员可以把精力集中在更重要的艺术决策上。

游戏音频设计师也爱上了 AI 这个得力助手。开发开放世界游戏时，传统方法需要预先录制数小时的背景音乐。现在 AI 可以根据玩家所处场景实时生成匹配的音乐——在森林中探险时生成悠扬的管乐，进入战斗时自动切换成激昂的电子节奏，让游戏体验更加身临其境。

音乐教育领域，AI 正在改变传统的学习方式。过去学生要等老师批改作业才能知道自己的作曲问题，现在 AI 可以即时给出专业建议。

（5）AI 音乐的现状与未来

虽然 AI 音乐技术日新月异，但仍然存在一些有趣的局限。例如，在表达细腻情感时，AI 作品可能缺少人类演奏的那种微妙变化。就像机器人弹钢琴，虽然每个音符都精确无误，但总让人觉得少了点灵魂。此外，AI 更擅长模仿现有风格，要让它创作出真正突破性的新流派，还需要更多技术创新。

展望未来，AI 音乐可能会向更智能、更互动的方向发展。也许不久的将来，AI 能通过摄像头捕捉听众的表情，实时调整音乐情绪；或者通过脑电波感应，创作出完全贴合你当下心境的音乐。专家预测，未来几年 AI 将成为音乐创作的标配工具，就像现在的作曲软件一样普及。

7. 探索创新应用场景

人工智能技术正在以前所未有的方式重塑创意产业和日常生活体验，催生出一系列突破性的应用场景。从数字艺术领域的革新到个人记忆管理的智能化，再到促进信息无障碍获取的技术进步，这些创新应用不仅展现了 AI 技术的强大潜力，更深刻地改变着人们与数字内容互动的方式。

数字艺术领域正在经历一场由 NFT 技术引领的革命。基于区块链的非同质化代币为数字艺术作品提供了独一无二的数字凭证，解决了数字艺术长期以来难以确权和流通的痛点。艺术家们现在可以通过 AI 工具创作出风格独特的数字作品，将其转化为 NFT 后在区块链上永久存证。这种创作方式打破了传统艺术市场的门槛限制，使得更多创作者能够直接面向全球收藏家。某数字艺术平台的数据显示，AI 生成的 NFT 作品平均交易价格达到传统数字艺术品的 2.3 倍，且流转效率显著提高。更重要的是，智能合约技术的引入使得艺术家可以设置作品后续转售时的版税分成，建立起可持续的收入模式。这种变革不仅限于视觉艺术，音乐、视频甚至虚拟时装等各类数字创作都因此获得了新的价值载体和商业模式。

智能相册管理系统正在重新定义个人数字记忆的管理方式。传统的照片管理主要依赖手动分类和简单的时间排序，而 AI 赋能的相册系统能够深度理解图像内容，实现多维度的自动整理。通过计算机视觉技术，系统可以识别人物身份、场景类型、活动性质等丰富语义信息。当用户搜索“海边日落”或“生日派对”等概念时，系统能够精准定位相关照片，而非简单匹配文件名或标签。更先进的系统还能分析照片中的情感基调，自动生成搭配适当音乐和转场效果的电子相册。某智能手机品牌的用户调研显示，使用智能相册功能的用户回顾老照片的频率提高了 4 倍，这得益于系统能够主动呈现具有纪念意义的时刻。这种技术不仅提升了个人记忆管理的效率，更通过智能推荐唤起了许多可能被遗忘的美好回忆。

无障碍视听内容生成技术体现了科技向善的本质。对于视障人士而言，AI 系统能够将视觉图像实时转化为语音描述，让他们“看见”周围的世界。这种技术不仅限于简单的物体识别，更能够理解场景的语义关系，例如，“桌子上放着笔记本电脑和一杯咖啡，旁边是一本翻开的书”。对于听障群体，语音转文字技术已经能够实现高达 95% 的准确率，且支持多种方言和专业术语。教育领域的创新应用更令人振奋，教材内容可以自动转换为适合特殊需求学生的形式，如将文字教材转为手语视频或触觉图形。某公益组织的实践表明，采用 AI 生成的无障碍内容后，视障用户的信息获取效率提升了 7 倍，这不仅仅是技术突破，更是社会包容性的重大进步。

在商业演示领域，AI 辅助的 PPT 生成正在改变传统的内容创作流程。过去需要数天时间完

成的专业演示文稿，现在可以通过 AI 系统在几十分钟内完成初稿。系统能够理解用户输入的关键点，自动匹配适当的版式、图表和视觉元素。更智能的版本还可以根据演讲者的语速和风格，动态调整每页 PPT 的信息密度和呈现方式。某咨询公司的内部报告显示，使用 AI 辅助 PPT 制作后，团队在方案设计阶段节省了约 40% 的时间，使他们能够将更多精力投入到核心的战略思考中。这种技术不仅提高了工作效率，更通过智能排版和设计建议，显著提升了演示文稿的专业水准和视觉吸引力。

这些创新应用场景的共同特点是它们都建立在 AI 技术的核心能力之上，同时又针对特定需求进行了精细化的设计和优化。数字艺术 NFT 解决了数字创作的版权问题，智能相册重新定义了个人记忆管理，无障碍技术打破了信息获取的障碍，智能 PPT 工具提升了商业沟通效率。每一个应用都不是简单地将 AI 技术套用到传统场景，而是通过深刻理解用户需求和场景特点，创造出全新的体验和价值。

值得关注的是，这些创新应用之间也存在着有趣的交叉和融合。数字艺术 NFT 技术可以应用于智能相册，将珍贵的记忆转化为独特的数字藏品；无障碍技术中的语义理解能力可以提升 PPT 的自动可访问性检查；智能相册的管理理念可以启发数字艺术平台的用户界面设计。这种交叉融合预示着 AI 应用创新正在进入一个更加丰富和多元的阶段。

随着技术的持续进步，这些创新应用场景还将不断演化和发展。数字艺术创作可能会融入更多交互和动态元素，智能相册或许能够结合虚拟现实技术重现三维记忆场景，无障碍技术有望实现更加自然流畅的多模态转换，PPT 工具可能会进化成真正的智能演讲助手。这些可能性不仅令人期待，也提醒我们，AI 技术的真正价值在于它能够如何赋能于人，如何解决真实世界的复杂问题，如何创造更加包容和便捷的数字未来。

在这些创新应用的背后，是技术开发者对用户体验的深刻洞察和对社会需求的敏锐把握。数字艺术 NFT 回应了创作者对知识产权保护的渴望，智能相册满足了人们对珍贵记忆妥善保存的需求，无障碍技术体现了对特殊群体的人文关怀，智能 PPT 工具解决了专业人士的效率痛点。这些应用之所以能够产生广泛影响，正是因为它们扎根于真实的社会场景和人类需求。

展望未来，AI 驱动的创新应用场景将继续拓展边界。教育领域可能会出现真正个性化的 AI 导师，医疗健康领域可能诞生更智能的病历管理和诊疗辅助系统，城市管理也许会引入更全面的智能感知和决策支持。在这些可能性中，不变的核心是技术如何更好地服务于人，如何解决那些长期存在却未被很好满足的需求。正如当前这些创新应用所展示的，当 AI 技术与深刻的人性洞察相结合时，就能够创造出改变游戏规则解决方案，重塑我们与数字世界互动的方式。

◎任务实践

实践 1：随着 AI 音乐技术的成熟，音乐创作的门槛正在被重新定义。陈同学是音乐社团的创作部部长，为迎接校园文化艺术节闭幕式，需要创作一首原创主题曲。通过对网易天音平台的学习，陈同学掌握了 AI 编曲与作词的核心方法，计划通过人机协作完成《青春不散场》主题曲创作。其创作需求如下：

提示词（AI 编曲部分）：

① 生成一首 3 分 30 秒左右的流行摇滚风格伴奏，需包含以下要素：前奏，钢琴琶音引入（C 大调，速度 108BPM）；主歌，加入电吉他闷音节奏型；副歌，强化鼓组与贝斯声部（突出力量感）；间奏，合成器音色营造空间感；结尾，渐弱式处理（加入风铃声）。

② 技术参数要求：使用“Dynamic Pop”音色库、开启智能和声进行建议功能、导出分轨文

件（含 MIDI 与 WAV 格式）。

具体操作步骤如下：

（1）登录网易天音工作室，进入“智能编曲”模块。

（2）在风格选择界面设置参数如图 3-28 所示。

风格：R&B（85%）+电子元素（15%）

乐器组合：主吉他+节奏吉他+贝斯+鼓组+键盘+合成器



图 3-28 编曲参数设置界面

③ 点击“开始 AI 写歌”按钮后获得 demo，选择版本进行深度编辑：在钢琴卷帘窗微调副歌旋律线，使用“智能填充”功能丰富鼓花编排，最终导出包含 8 个音轨的工程文件。

传统音乐创作中，词曲配合往往需要反复磨合调整。网易天音的 AI 作词功能如同一位通晓音乐文学的双语导师：当你的旋律走向激昂时，它会自动匹配“破晓时跃出海面的鲸”这类充满张力的意象；当和弦转为小调时，它又能巧妙地将“未寄出的信在抽屉泛黄”这样的忧郁叙事自然植入韵脚。更独特的是，它能实现“旋律驱动作词”——根据贝斯声部的切分节奏自动生成错落有致的排比句，让文字本身具有律动感。

实践 2：陈同学在完成编曲后，继续使用 AI 作词功能进行歌词创作。

提示词（AI 作词部分）：

① 基于已生成的伴奏音乐，创作符合以下要求的歌词：主题：毕业季的成长与告别；段落结构：主歌 A+B→预副歌→副歌→桥段→重复副歌；韵脚要求：主歌押“ang”韵，副歌转“ing”韵。

② 内容质量控制：避免空洞口号，需包含“教室窗台的绿萝”等具体校园意象、在第二段主歌加入“图书馆通宵的灯光”等叙事细节、副歌需出现重复记忆点（如“青春是____的星星”句式）。

具体操作步骤如下：

(1) 上传编曲工程文件至“智能填词”模块。

(2) 输入主题关键词后开启“旋律适配”模式：主歌匹配字数：7~9 字/句、副歌高潮词密度：每句含 1 个强韵脚词、桥段设置对比冲突（如“说好不哭/却在下雨时红了眼睛”）。

(3) 通过“情感曲线”调节工具优化歌词，如图 3-29 所示。最终生成成果包含完整歌词文本（含段落标记）、自动生成的演唱呼吸标记。

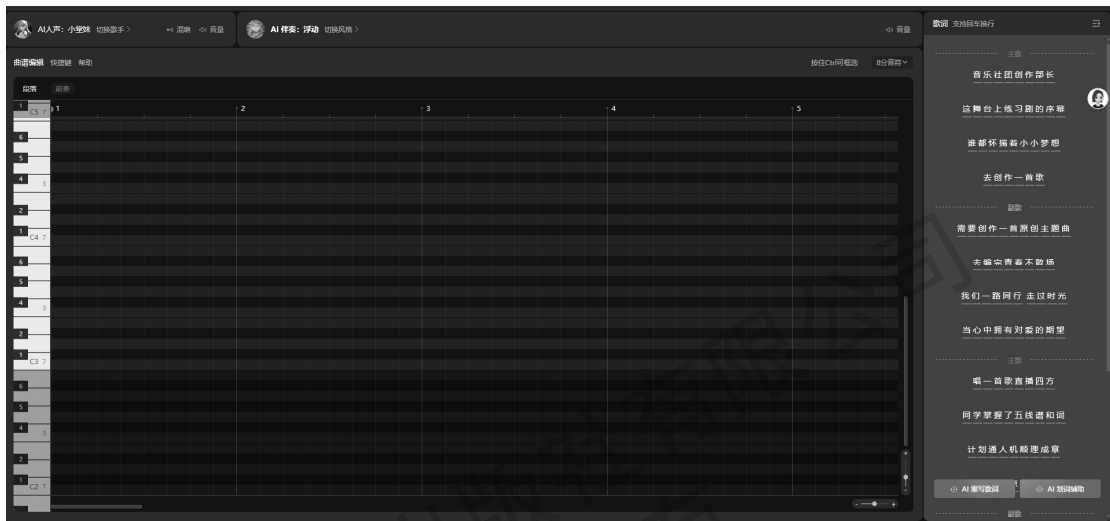


图 3-29 歌词情感曲线调节界面

本案例展示了 AI 音乐创作平台如何重构传统工作流程：将原本需要 2 周的专业创作周期压缩至 3 小时，同时保持词曲的高度协同性。实践表明，掌握“音乐提示词工程”能力将成为数字时代创作者的必备技能，这种将人类创意与 AI 执行力的结合，不仅突破了个人能力的局限，更开创了“一人乐队”式的新型创作模式。作为数字原住民的音乐学习者，应当主动拥抱这种变革，将 AI 作为拓展创作维度的“第六乐器”来驾驭。

任务 3 构建 AI 视频创作 workflow

◎任务描述

视频内容生产正经历智能化深度变革。本任务将解析文生视频、多模态融合等前沿技术，引导学生构建自动化视频生产线，并针对电商、教育等场景开发行业解决方案，同时探讨深度伪造等技术的伦理治理路径。在当今视频内容爆发的时代，智能化的视频创作技术正在彻底改变内容生产方式。AI 视频工具已经广泛应用于短视频制作、影视创作、广告营销等领域，让专业制作更高效，也让普通人可以轻松创作优质视频内容。在短视频创作中，AI 可以根据简单的文字描述自动生成生动有趣的视频片段，大大降低了创作门槛。专业团队可以用它快速制作创意提案，个人创作者也能轻松产出吸引眼球的内容。影视制作领域，AI 可以辅助完成从分镜脚本到特效合成的全流程，显著提升制作效率。广告行业也在积极应用这些技术，智能生成的营销视频可以根据不同平台特性自动适配格式，实现精准投放。本任务将带领学生系统构建 AI 视频创作的全流程工作流。我们将深入解析文生视频的技术原理，了解 AI 如何将文字转化为连贯的视觉叙事；探索

智能剪辑的算法逻辑，掌握素材自动匹配的关键技术；研究特效生成的实现方法，学习如何用 AI 工具制作专业级视觉效果。通过短视频创作和广告制作项目，学生将掌握智能视频创作的完整流程，培养 AI 时代的视觉叙事能力。

◎任务分析

本任务的重点是了解 AI 视频制作、以文字生成视频、以图片生成视频、参考主题生成视频等，完成本次任务主要涉及知识点如下：

- ✧ 文生视频技术
- ✧ 图生图视频技术
- ✧ 视频优化与编辑
- ✧ 提示词工程与镜头语言
- ✧ 行业级工具应用

◎任务实践

1. 实现智能视频生成

(1) 视频内容创作的智能化变革

人工智能技术的迅猛发展，正在驱动视频内容创作领域发生一场深刻的范式转变。这场变革不仅仅是工具的更新换代，更是对传统视频制作流程、创作方式乃至思维模式的系统性重塑，其影响遍及从内容构思到最终分发的全链条。

① 智能化对传统制作流程的重构。过往视频制作依赖一套相对固定且繁复的流程，通常需要经历前期策划、脚本撰写、分镜绘制、实拍/素材收集、粗剪、精剪、特效合成、配音配乐、调色输出等多个环节，环节间协作成本高，周期漫长。人工智能技术的引入，正在从根本上重塑这一传统工作流。智能工具开始渗透并赋能各个环节，显著提升效率。例如，AI 驱动的策划工具可以基于目标受众和热点趋势快速生成创意提案；智能脚本助手能辅助撰写和优化解说词；而最显著的变革在于，过去需要专业团队耗费数天甚至数周才能完成的复杂视频项目，如今借助集成化的智能视频生成与编辑平台，其核心制作周期可以被压缩到数小时之内，同时仍能保障符合专业标准的视听质量输出。这种效率的跃升，本质上是技术对传统生产流程的优化与再造。

② 驱动创作效率与创意空间的飞跃。智能化带来的变革远不止于速度的提升，更深层的价值在于其对创作效率本质的革新和创意空间的释放。智能视频生成技术极大地简化了技术实现的门槛，将创作者从大量重复性、技术性强的劳动中解放出来（如复杂的绿幕抠像、基础的动画制作、烦琐的素材整理与粗剪）。有行业调研数据显示，积极采纳 AI 辅助工具的视频创作团队，其内容产出效率普遍实现了 3 至 5 倍的增长。效率提升的量化指标固然重要，但更值得关注的是效率提升所带来的创作重心的转移。创作者得以将原本耗费在技术执行上的宝贵时间和精力，重新聚焦于更核心的价值创造层面——故事的构思、创意的打磨、情感的传达及艺术表现力的探索。人工智能在此扮演了强有力的“助手”角色，它负责高效地实现创作者的构想，从而让创作者能够更加专注于其不可替代的创意工作。

③ 智能化变革的核心价值。综上所述，视频内容创作的智能化变革，其核心在于利用人工智能技术系统性地重塑了生产流程，实现了创作效率的指数级提升，并最终将创作的核心价值回归到人的创意与思想表达之上。技术不再是限制创意的瓶颈，而是转化为激发创意潜能、拓展表

达边界的强大引擎。这种由 AI 驱动的转型，正在深刻改变视频内容产业的生态与未来方向。

（2）文生视频技术的突破性进展

文生视频技术作为人工智能领域的前沿方向，近年来取得了令人瞩目的突破性进展。这项技术的核心目标，是让计算机能够理解人类自然语言描述的场景或故事，并自动生成符合描述的、具有连贯动态效果的视频内容。这种能力标志着人机交互和内容创作方式的一次重大飞跃。

① 从文字到动态影像的跨越。现代先进的文生视频系统，其最显著的成就在于能够跨越文字与动态影像之间的巨大鸿沟。系统能够准确解析用户输入的、相对简洁的文字描述，并在相对较短的时间内（通常在几分钟内），生成一段约 10 至 15 秒、视觉质量较高的视频片段。尤为关键的是，当前顶尖系统生成的视频，在动态流畅性（如海浪涌动的节奏、光影的自然变化）、画面细节的丰富度及整体观感的协调性上，已经达到了接近专业实景拍摄或高质量动画的水准。这种高保真度的动态生成能力，是文生视频技术走向实用化的基石。

② 复杂叙事的智能处理。文生视频技术的突破性不仅体现在基础场景的生成上，更在于其对复杂叙事结构和精细化艺术表达的智能处理能力。最前沿的系统引入了诸如“分时段风格控制”等高级功能。这赋予创作者前所未有的控制力，允许他们为同一段生成视频的不同时间段落，分别指定截然不同的视觉风格或艺术效果。

（3）多模态融合的创新应用

在人工智能驱动的视频技术前沿，多模态融合代表着一种突破性的发展方向。这项技术的核心在于，系统能够同时理解、处理和整合来自不同模态的信息源——包括文本描述、静态图像、音频信号及三维模型数据等，并最终将它们协同生成为连贯统一的视频内容。这种能力极大地超越了传统视频制作中媒介相对孤立的局限，开启了全新的创作空间。

① 跨媒介的内容整合。多模态融合技术最具革命性的应用之一，体现在其能够深度整合原本分离的不同媒介元素，并实现高度自动化的内容生成流程。智能视频系统不再是单一模态的处理工具，而成为了一个能够理解复杂输入、综合调取多种资源的“创作中枢”。

② 创意表达的无限可能。多模态融合技术的价值远不止于提升效率，更具变革性的是它如何为创意表达开辟了前所未有的广阔天地。通过打破媒介壁垒，创作者能够构思并实现以往技术条件下难以企及或成本高昂的复杂创意形式，催生出极具冲击力和新颖性的视听体验。

③ 技术融合的深远影响。多模态融合技术在视频领域的创新应用，标志着内容创作正进入一个媒介界限日益模糊、资源整合高度智能化的新阶段。它不仅极大地优化了内容生产的效率与成本结构，更重要的是，它通过释放跨媒介协同的潜能，持续拓展着视频作为信息载体和艺术表达形式的边界，为教育、娱乐、营销、艺术等众多领域带来了颠覆性的创新机遇。其发展潜力，仍在不断被探索和释放之中。

（4）智能化制作工具的革命

人工智能技术不仅深刻影响着视频内容生成的前沿领域，也正在对视频制作的后期环节——素材管理与特效制作——发起一场静默却深刻的革命。智能化的制作工具通过引入强大的算法，显著提升了创作效率，并以前所未有的方式降低了专业级制作的门槛。

① 素材匹配的智能升级。在视频剪辑领域，尤其是处理海量素材的场景（如新闻编辑、纪录片制作、真人秀节目），传统的素材筛选与片段查找工作极其耗时耗力。智能化的 AI 剪辑系统正在彻底改变这一局面。这类系统能够深度理解视频内容，通过计算机视觉识别画面中的对象、场景、动作，结合语音识别分析对话内容，甚至理解其中的情绪和语义。基于这种多维度理解，系统可以自动对数以百小时计的原始素材库进行高效索引和智能分析。

② 专业特效制作的普及化。视频制作中另一个曾长期被高门槛所限制的领域是专业级视觉

特效（VFX）。过去，实现电影大片级别的特效效果，往往意味着需要昂贵的专业工作站软件、庞大的技术团队及漫长的制作周期，这通常只有大型制片公司才能负担。AI 驱动的特效工具的出现，正在实现特效制作的民主化，使得专业级甚至电影级的视觉效果不再遥不可及。

③ 智能化工具革命的核心意义。综上所述，智能化制作工具在素材管理和特效制作两大核心环节的革命性进步，其核心价值在于系统性提升视频制作效率，并打破专业壁垒，实现创作能力的普及化。它们将创作者从烦琐的技术劳动中解放，赋予其更强大的表达手段，从而推动视频内容创作生态向着更高效率、更低门槛、更丰富多元的方向持续演进。这场由 AI 驱动的工具革命，正在深刻改变着“谁可以创作”及“如何创作”的行业基础规则。

（5）视频优化的智能解决方案

在视频内容日益成为信息传递和沟通核心媒介的今天，人工智能技术正为视频的优化提供着强大的智能解决方案。这些方案不仅提升了视频本身的观看体验，更重要的是，它们极大地扩展了视频内容的可及性和影响力，使其能够服务于更广泛的受众群体，跨越物理与文化的界限。

① 无障碍访问的突破。确保视频内容能够被所有人无障碍地访问，是信息平等和包容性社会的重要体现。智能字幕技术在这一领域扮演着关键角色，成为实现视频内容包容性的重要技术支柱。现代智能字幕系统利用先进的语音识别（ASR）和自然语言处理（NLP）技术，能够近乎实时地、高准确率地将视频中的语音内容转化为精准的文字字幕。

② 全球化传播的助力。视频内容的价值在于其传播的广度与深度。人工智能驱动的多语言字幕生成与智能配音技术，正成为视频内容突破语言障碍、实现全球化传播的核心驱动力。

传统的人工字幕翻译和配音不仅成本高昂、周期漫长，也难以满足海量内容快速国际化的需求。智能解决方案则能够高效地完成这一任务。系统首先利用机器翻译引擎将原始字幕或语音内容快速翻译为目标语言，随后，结合文本到语音（TTS）技术，生成自然流畅、符合目标语言文化习惯的语音旁白，甚至可以模拟特定的音色和语调，实现高质量的“语音克隆”。

2. AI 视频制作

在当今数字内容生态空前繁荣的背景下，人工智能视频技术（简称 AI 视频）应运而生，并迅速成为变革内容创作格局的关键力量。其核心内涵是指运用人工智能算法，实现视频内容的自动化生成或为创作过程提供智能化辅助的一系列技术手段。这标志着视频制作从高度依赖专业技能和昂贵资源的传统模式，向更智能、更高效、更可达的方向演进。

（1）AI 视频的应用价值

在创意传播与商业推广领域，AI 视频展现出强大的潜力。它能够快速生成创意十足、视觉冲击力强的短视频内容，特别适用于广告营销活动。相较于传统制作，这种方式成本更低、周期更短。例如，品牌方可以高效地制作出系列化的广告短片，测试不同创意方向的市场反应。

对于社交媒体内容的创作者而言，AI 视频技术提供了制作独特动态视觉效果利器。它可以帮助用户轻松生成一众帖子内容，在信息洪流中有效吸引受众的注意力。想象一下，将一张静态风景照片转化为一段展现日出日落光影变化的动态视频，这无疑能极大提升帖子的吸引力。

在电子商务领域，AI 视频的应用尤为实用。卖家可以便捷地将静态的产品图片转化为生动的展示视频，甚至生成产品的 3D 旋转展示效果。例如，一件服装的图片可以动态展示其穿着效果和面料细节，一件电子产品可以多角度旋转展示其设计，这能更直观、生动地呈现产品特性，有效提升产品的吸引力与转化率。

AI 视频技术还催生了“数字人”的广泛应用。借助这项技术，可以生成高度仿真的虚拟主播或讲解员形象。这些虚拟数字人能够进行新闻播报、知识讲解、产品介绍等任务，替代或辅助真人出镜。这不仅能节省大量人力与时间成本，其新颖的形式也能有效吸引观众，适用于教育机构、

新闻媒体、企业培训等多种场景。

在内容创作与艺术表达方面，AI 视频赋予了个体强大的能力。用户可以将自己的故事构思或创意想法，无须专业的动画制作技能，即可转化为生动的动画短片。这为个人创作者、小型工作室提供了前所未有的机会。

在教育与信息传递领域，AI 视频同样大有可为。它能够高效地生成解释性视频、动态演示图表或复杂概念的视觉化呈现。例如，一个抽象的物理原理或历史事件过程，通过动态图表和模拟动画展现出来，使得信息的传递更加直观、易懂，显著提升学习或理解效率。

对于艺术家和视觉探索者，AI 视频技术打开了一扇新的大门，提供了实验性视觉表达的新工具。它能够协助生成风格独特、构思奇特的动态影像作品，探索传统手段难以企及的视觉语言，拓展艺术创作的边界。

当用户熟练掌握了 AI 视频制作技能后，其应用前景将更为广阔。具备这项技能的人士可以为商业客户提供广告视频、社交媒体内容的定制服务。作为自由职业者，提供专业的 AI 视频制作与咨询服务成为一种可行的职业路径。创作者也可以利用该技术打造独特的个人 IP 内容，通过平台流量分成或知识付费实现价值。此外，针对市场上对 AI 视频技能日益增长的需求，开发相关的教学课程、提供操作咨询或解决方案设计，也是极具潜力的发展方向。随着技术的持续迭代和应用的深化，AI 视频领域蕴含着丰富的创新与职业机遇。

（2）完整的视频制作流程概述

制作一个结构完整、效果精良的视频作品，通常需要遵循一个系统性的流程。这个过程涵盖了从最初的创意构思到最终成片输出的多个关键环节，每个环节都对最终作品的质量起着重要作用。

视频制作的起点在于构思与规划。首先需要将你的核心故事、想要传达的信息或创意概念清晰地落于文字，形成文字脚本或剧本。这份文档是后续所有工作的基石，它明确了视频的主题、叙事结构和关键信息点。

构思明确后，下一步是进行视觉预览与分镜设计。这一阶段的任务是将文字脚本转化为具体的视觉画面构想。利用 AI 图像生成工具，可以高效地根据脚本描述，创建出代表不同镜头画面的静态分镜图。这些分镜图如同视觉化的蓝图，帮助创作者预览每个镜头的构图、主体和氛围，为后续的动态制作奠定基础。例如，描述“清晨公园里老人打太极拳”的场景，可以生成对应的静态画面来确认环境、人物姿态和光影效果。

在静态分镜得到确认后，可以进一步利用 AI 视频生成技术进行动态分镜预览。通过输入对画面内容和运动方式的描述（如镜头缓慢推近、人物从左走入画面），AI 能够生成简短的视频片段来模拟预期的镜头运动、节奏和基本动态效果。这个动态预览对于评估镜头流畅度和时间感非常有价值，是静态分镜的重要补充。

视频的感染力不仅来自画面，声音设计与氛围营造同样不可或缺。需要为视频匹配合适的背景音乐来烘托情绪基调，并添加必要的音效（如环境声、动作音效）来增强真实感和沉浸感。声音元素与画面的协调配合，能够极大地提升观众的观看体验。

最后，将所有准备好的素材（包括 AI 生成的视频片段、拍摄的素材、音乐、音效、旁白等）进行整合与精细剪辑。使用视频编辑软件，按照分镜设计的顺序进行拼接，调整节奏，添加转场效果，平衡音画关系，并可能进行调色等后期处理，最终合成一部连贯流畅的影片。

（3）AI 视频提示词（Prompt）

要让 AI 视频生成工具准确地将你的创意构想转化为理想的动态画面，其核心在于掌握与 AI 进行有效“沟通”的技巧。这类似于向一位理解力强但缺乏背景知识的伙伴清晰描述你脑海中的景象。能否清晰、准确地传达你的视觉意图，直接决定了最终生成结果的质量与契合度。

① 清晰表达视觉意图的基础框架。对于初学者而言，掌握一个结构化的描述框架是快速上手的有效方法。这个框架旨在引导你系统地表达画面构想的关键维度：

- 画面构成 (What & How it looks)。首先明确画面的核心主体是什么（如“一个发光的水晶球”），并描述其关键视觉特征（如“球体周围环绕着星星点点的光效”）。

- 镜头语言 (How it's filmed)。接着说明你希望用什么样的镜头来呈现这个主体。这包括镜头的景别（如“特写镜头”或“全景镜头”）、镜头的运动方式（如“环绕拍摄”或“缓慢推进”）及拍摄角度（如“俯拍”或“仰拍”）。

- 动态与状态 (What it's doing)。描述画面中主体的动作或状态（如“水晶球正在缓缓旋转”）。

- 视觉风格 (Overall Look & Feel)。最后，定义整个画面期望呈现的艺术风格或氛围基调（如“科幻未来感”“卡通风格”或“写实自然风”）。

将这几个维度的描述有序组合起来，就能形成一条对 AI 相对友好的基础指令。例如，“特写镜头拍一个发光的水晶球，用环绕的方式拍摄，球体周围要有星星点点的光效，水晶球缓缓旋转，整体要科幻未来感”。这条指令清晰地传达了画面主体、镜头运用、关键细节、动态效果和最终风格。

② 提升描述的丰富度与精确度。随着对工具特性理解的加深，可以逐步尝试更细致、更丰富的描述，以追求更高精度和独特性的画面生成。

- 环境与氛围。加入对环境背景、光线条件（如“清晨柔和的阳光”“霓虹灯光下的雨夜”）、天气效果（如“薄雾笼罩”“雪花纷飞”）的描述。

- 精细动态。更精确地描述动作的速度、节奏或变化过程（如“无人机从高空缓缓下降俯拍”“湖面倒影随微风轻轻波动”）。

- 材质与质感。强调物体表面的视觉和触觉感受（如“岩石粗糙的质感”“丝绸般光滑的水面”）。

- 情绪与隐喻。尝试用语言引导画面传递特定的情绪或象征意义（如“营造孤独寂寥的氛围”“体现科技与自然的交融”）。

例如，一条进阶指令可以是：“用无人机从高空俯拍壮丽的雪山和山脚下的湖泊，湖面要像镜子一样平静，清晰地倒映出雪山的轮廓，拍摄速度要非常缓慢，营造出宁静感，整体色调偏冷，强调蓝白基调，光影效果追求自然写实，如同高水准的自然纪录片”。这条指令不仅包含了基础框架的要素，还详细设定了环境关系、光影色调、动态节奏和具体的风格对标（自然纪录片），大大提升了 AI 生成画面符合预期的可能性。

描述的精确度与生成画面的符合度通常呈正相关。开始时，不必追求面面俱到，可以从结构化的基础框架入手，生成初步结果。然后，仔细观察生成画面与期望的差距，有针对性地调整和丰富你的描述语，进行迭代优化。例如，如果第一次生成的“科幻水晶球”光效不够理想，可以在下一次尝试中加入“密集的蓝色粒子光效动态环绕”或“水晶球内部有脉动的能量核心”等更具体的描述。通过这种“描述—生成—观察—调整”的循环，你将逐渐提升与 AI “对话”的精准度，更自如地驾驭工具来实现你的创意构想。这种描述能力的精进，本身就是掌握 AI 视频生成技术的核心技能之一。

③ AI 视频生成的关键要素解析。在与 AI 视频生成工具进行有效沟通时，掌握一些基础的视觉语言概念至关重要。这些概念如同创作者与 AI 系统之间沟通的“词汇表”，能够帮助你更精准地描述脑海中的画面构想，从而引导 AI 生成更符合预期的结果，以下是一些核心要素的解析。

【景别 (Shot Type)】景别决定了画面所容纳的内容范围，直接影响观众对主体和环境关系的感知。它通过控制主体在画面中所占的比例来实现不同的表达目的。

- 远景 (Long Shot)。通常用于展示宏大的场景或环境全貌。在这种景别下，人物或主体会

显得比较小，重点在于交代环境氛围和空间关系。

- **全景 (Full Shot)**：能够清晰地展现人物全身或物体的完整形态，同时保留部分周围环境信息，常用于介绍人物出场或展示物体整体。

- **中景 (Medium Shot)**。通常截取人物膝盖或腰部以上的部分，是叙事中最常用的景别之一。它既能展现人物的动作和表情，又能保留一定的环境信息，平衡了主体与环境的关系。

- **近景 (Close-up)**。将镜头聚焦在人物胸部以上或物体的主要部分，强烈突出人物的面部表情、情绪变化或物体的关键细节。它能够引导观众关注特定的情感或信息点。

- **特写 (Extreme Close-up)**。进一步放大细节，聚焦于人物的某个局部（如眼睛、嘴唇）或物体的极小部分（如手表指针、昆虫复眼）。这种景别极具冲击力，用于强调极其细微的、具有象征意义或戏剧性的瞬间。

【**运镜 (Camera Movement)**】运镜是指镜头自身的运动方式，它能显著增强画面的动感和叙事节奏，引导观众的视线。

- **推镜 (Zoom In / Push In)**：镜头（或模拟镜头效果）逐渐靠近被摄主体，使主体在画面中越来越大。这种运动常用于强调主体、揭示细节或营造紧张、聚焦的情绪。

- **拉镜 (Zoom Out / Pull Out)**：镜头（或模拟镜头效果）逐渐远离被摄主体，使主体在画面中越来越小，同时展现更多的环境信息，常用于交代环境、表现人物的孤立感或结束一个场景。

- **摇镜 (Pan / Tilt)**：镜头在固定机位上进行水平（Pan）或垂直（Tilt）方向的转动。水平摇镜常用于跟随运动物体或展示宽阔场景；垂直摇镜则常用于展示高大的物体或从天空摇向地面。

- **移镜 (Dolly / Truck)**：整个摄像机在轨道或滑轨上平行移动。向前或向后移动称为 Dolly，左右平行移动称为 Truck。这种运动能产生平滑的移动感，常用于跟随主体移动或营造身临其境的感受。

- **环绕 (Orbit / Circular Dolly)**：镜头围绕被摄主体进行圆周或弧线运动。这种运镜能全方位展示主体，并营造一种审视、观察或动态环绕的感觉。

【**塑造氛围与质感**】除了景别和运镜，以下要素对于精准描述画面同样关键。

- **光线 (Lighting)**。描述光照的效果对于设定场景氛围至关重要。例如，“电影感光效 (Cinematic Lighting)”通常指具有强烈明暗对比、戏剧性的光影效果；“黄金时刻 (Golden Hour)”指日出后或日落前柔和温暖的金色光线；“黑暗情绪光 (Dark Moody Lighting)”则指低照度、高对比、阴影浓重的光线，常用于营造神秘、紧张或忧郁的氛围。准确描述光线能极大影响画面的情感基调。

- **主体 (Subject)**。明确画面中的核心人物或物体是什么。这是描述画面的起点。动作/场景描述 (Action/Scene) 可以清晰说明主体在做什么，以及所处的环境是怎样的。这提供了画面的动态信息和背景。

- **风格 (Style)**。定义画面期望呈现的整体艺术风格。这决定了画面的美学基调，如模仿动漫风格 (Anime Style) 的夸张线条和色彩、追求极致真实的照片写实风格 (Photorealistic)、或是模仿水彩风格 (Watercolor) 的晕染和通透感。明确风格能让 AI 理解你想要的视觉呈现方式。

- **细节 (Details)**。添加关于颜色、材质、纹理或特征的具体描述，能使画面更生动、独特。

- **速率 (Pacing)**。描述画面动态的速度感，如使用慢动作 (Slow Motion) 强调动作细节、延时摄影 (Timelapse) 展现时间流逝，或是快速切换 (Fast Cuts) 营造紧张节奏。

- **层次 (Layers&Depth)**。描述画面中前景、中景、背景的关系及虚化效果，能增强空间感和氛围。

- **氛围 (Atmosphere)**。用形容词概括画面希望传递的整体感觉或情绪，如神秘 (Mysterious)、

史诗感 (Epic)、温馨 (Cozy/Warm)。这有助于 AI 把握画面的情感核心。

需要特别强调的是，并非每次描述都需要涵盖上述所有要素。关键在于根据你具体的创意意图和画面重点，有选择性地组合运用这些概念。例如，描述一个快速的动作场景可能更强调动作描述和速率；而描述一个静物氛围则可能更注重光线、细节和氛围。理解这些概念并学会灵活运用它们，将极大地提升你驾驭 AI 视频生成工具的能力，将脑海中的视觉构想准确无误地转化为生动的动态画面。在实践中不断尝试组合这些要素，是掌握这门“视觉沟通语言”的有效途径。

3. 以文字生成视频

文生视频 (Text-to-Video) 技术代表了人工智能在内容创作领域的一项重大突破。其核心原理是，用户只需输入一段描述性的自然语言文本（可以是一句话或一个段落），人工智能系统便能自动解析这段文字的含义，理解其中蕴含的场景、动作、风格等信息，并据此生成相应的动态视频片段。这项技术最显著的优势在于，它能够自动完成从画面构思到动态生成的全过程，极大地简化了视频创作的入门门槛。

(1) 国内主流文生视频工具

随着技术的快速发展，目前已有多种文生视频工具可供用户选择，它们各有特点，适用于不同需求的创作者。对于国内用户，尤其是初学者而言，选择支持中文提示词且操作便捷的平台尤为重要。目前，国内市场已有多款表现优异的文生视频工具。

① 即梦 AI (字节跳动) 作为国内领先的科技公司推出的产品，即梦 AI 在中文语义理解方面表现出色，能很好地响应中文描述。它提供了相对友好的用户界面和稳定的生成效果，非常适合初次接触文生视频技术的用户进行探索和实践。

② 可灵 AI (快手) 依托于短视频平台的深厚积累，可灵 AI 在理解短视频创作需求方面具有优势。它可能提供更多贴近国内流行趋势的预设风格或模板，方便用户快速生成符合社交媒体传播特点的短视频内容。

③ 海螺 AI 也是一款值得关注的国内文生视频工具，以其易用性和对中文提示的良好支持受到用户欢迎。它可能提供免费额度或灵活的试用方案，是新手入门练习的另一个理想选择。

(2) 国外主流文生视频工具

与此同时，国外也有数款功能强大、技术前沿的文生视频工具，它们在生成质量、功能丰富度或创意自由度上往往更具探索性。

① Runway 是国外广受专业创作者和艺术家推崇的平台之一，提供包括文生视频在内的多种 AI 创意工具 (Gen-2 是其著名的视频生成模型)。Runway 以强大的功能和较高的生成质量著称，适合追求更高创作自由度和专业效果的用户。

② Pika 因其出色的视频生成质量和用户友好的体验而迅速获得关注。Pika Labs 开发的工具在生成动作连贯性和画面美感方面表现优异，是国际社区中热门的选择。

③ Luma AI 以其高质量的 3D 场景重建和由文本生成 3D/视频的能力而闻名 (如 Dream Machine)。它特别擅长生成具有空间感和三维深度的动态场景，为创作者提供了独特的视觉可能性。

对于刚刚开始接触文生视频技术的学习者，建议优先从上述支持中文的国内平台 (如即梦 AI、可灵 AI 或海螺 AI) 入手进行练习。选择这些平台的主要优势在于语言无障碍，使用中文提示词进行创作，避免了语言转换带来的理解偏差，能更专注于学习描述技巧本身；学习曲线平缓，国内平台通常更注重本土用户的使用习惯，界面设计和操作流程可能更符合国内用户的预期，降低了初始的学习难度；快速获得反馈，通过这些平台进行实践，能够快速看到文本描述转化为视频的结果，便于及时调整描述方式，积累经验。

在熟练掌握基础操作和描述技巧，并对文生视频的原理有了一定理解之后，再逐步尝试探索

功能更强大但也可能操作更复杂的国外平台（如 Runway、Pika、Luma AI），将有助于进一步拓展创作边界，体验技术前沿的可能性。循序渐进的学习路径，能帮助新手更扎实地掌握这项充满潜力的技术。

4. 以图片生成视频

（1）图生视频

图生视频（Image-to-Video）技术是人工智能视频生成领域的一个重要分支。其核心功能是将静态的图像作为输入，由 AI 系统理解图像内容并推断合理的动态变化，最终生成一段连贯的视频序列。这项技术为已有视觉素材注入了新的活力，极大地拓展了图像的应用边界。

图生视频技术在实际应用中主要有两种典型的模式，分别服务于不同的创作需求。

单图动态化是最基础的模式。用户提供一张静态图片，AI 系统分析图像中的元素（如主体、背景、潜在运动趋势），并据此生成一段让画面“动起来”的短视频。

首尾帧引导生成是一种更进阶且控制性更强的模式。用户需要提供起始帧和结束帧两张图片，并辅以描述画面中间变化过程的文本提示。AI 系统根据首尾帧定义的开端与终点状态，结合文本描述推断中间的运动轨迹和过渡效果，生成一段首尾固定的动态视频。简单来说，就是根据“起点图”和“终点图”生成完整的中间过程视频。

图生视频技术的核心优势在于它能够高度尊重和保留原始输入图像的核心特征——包括主体形象、美术风格、色彩基调等——在此基础上，智能地融入符合逻辑且流畅的运动变化，从而创造出既忠于原作又富有新意的动态视觉内容。这使得它成为插画师、设计师、摄影师及任何拥有静态视觉素材并希望将其动态化的用户的理想工具。

（2）代表性工具

在众多图生视频工具中，即梦 AI（由字节跳动推出）是一款具有代表性的、面向大众用户的革命性智能视频创作平台。它的设计理念是将专业级的视频制作能力，以极其便捷自然的方式赋予普通用户。

即梦 AI 的核心能力在于其强大的多模态理解与生成引擎。用户无须掌握复杂的专业技能，只需通过最自然的语言描述心中所想，或者上传一张图片甚至简单的涂鸦手稿，即梦 AI 便能迅速理解用户的意图，并在短时间内生成与之匹配的精美视频片段，其技术亮点尤为突出。

① 交互式实时调整。即梦 AI 支持自然语言的多轮对话交互。用户若对初次生成效果有调整需求，只需像与助手对话一样提出修改意见（如“请让夕阳的色彩更浓郁些”“海浪的起伏可以再大一些”），AI 便能立即响应并优化输出，大大提升了创作的灵活性和可控性。

② 广泛的风格适配能力。平台内置了极其丰富的视觉风格模板库，覆盖了从高度写实到抽象卡通、从复古怀旧到前沿科技感等多样化的美学需求。用户还可以上传特定的参考图片或视频片段，引导 AI 学习并精准模仿其独特的视觉风格，确保生成内容与用户期望的美学基调高度一致。

③ 深度语义理解与智能补全。即梦 AI 不仅能准确捕捉用户描述的显性要素，更能基于对场景和语义的深度理解，智能补充必要的环境细节与氛围元素。例如，当用户描述“小朋友在公园快乐玩耍”时，AI 会自动合理地添加明媚的阳光、葱郁的树木、绿茵草地等背景元素，甚至生成孩童的笑声作为环境音效，使生成的视频内容更加生动、完整且富有感染力。

即梦 AI 以其卓越的易用性、强大的智能化水平和高效的产出能力，特别适合短视频创作者、电商运营者（用于商品动态展示）、教育工作者（制作教学动画）及广大内容爱好者。它正在深刻改变视频创作的传统范式，真正实现了“所思即所见”，让视觉创意的表达不再受制于专业壁垒和技术门槛，成为赋能大众创意表达的有力工具。

（3）提示词的核心策略

在图生视频任务中，用户已经提供了静态图像作为基础，该图像本身已清晰定义了画面的主体、环境、风格等核心视觉要素。因此，提示词撰写的核心策略需要做出相应调整：应着重描述你期望在原始静态画面基础上发生怎样的动态变化、运动效果及氛围演进，引导 AI 理解“动”的方向和方式。

（4）提示词通用要素与重点迁移

虽然我们在描述全新画面时需要考虑景别、主体细节、动作、运镜、光线、风格等诸多要素，但在图生视频的语境下，由于输入图像已经锁定了大部分静态信息，提示词可以更聚焦于以下方面。

① 动作与变化。明确指出画面中哪些元素需要动起来，以及如何动。这是图生视频提示词最核心的部分。

② 运镜方式。描述期望的摄像机运动，为生成的动态序列增添视角和节奏感。

③ 光线与氛围的调整/强化。如果需要改变或强调特定的光影效果、天气状况或整体情绪氛围，应在提示词中说明。

④ 时间感与节奏。描述动作发生的速度（如慢动作、加速）或场景变化的节奏。

那些在输入图像中已经明确且无须改变的元素（如主体是什么、主体外观细节、主要环境构成、基础美术风格），通常无须在提示词中重复描述，除非你有意要改变它们。

图生视频提示词撰写的精髓在于：紧密围绕输入图像已提供的信息，将描述的重心精准地置于“如何让画面动起来”，以及“营造何种动态氛围”上。理解并熟练运用这一原则，能够有效引导 AI 生成更符合预期的动态效果，最大化利用静态图像资产的价值，实现从静到动的创造性转化。在实践中，清晰、简洁地聚焦于动态变化和氛围的指令，往往比冗长重复的描述更能产生理想的结果。

5. 参考主体生成视频

（1）技术原理与应用价值

在 AI 视频生成技术不断演进的背景下，“参考主体生成视频”（Reference Subject Generation）代表了一种更为高级且极具实用价值的工作流程。这种技术能够有效解决生成视频中核心角色或物体形象难以保持一致性的关键挑战。

参考主体生成视频的核心在于，用户预先向 AI 系统提供需要“扮演主角”的视觉参考。这个参考主体可以是一张真实人物的照片、一个特定设计的卡通角色形象、一件独特的物品图片，甚至是手绘的概念草图。AI 系统在后续生成视频内容时，会以该参考主体为基准，确保其在视频的不同镜头或场景中保持外观、风格和核心特征的高度一致。

这项技术的应用价值尤为显著：用户可以让自己的形象成为 AI 生成视频的主角，贯穿于不同的虚拟场景之中，实现个性化的故事讲述或创意表达；创作者可以设计一个独特的角色（如原创卡通形象、品牌吉祥物或特定物品），并使其作为主角稳定地出现在系列视频或多镜头片段中，如同真实的演员或道具；对于品牌方或内容创作者，能够确保其虚拟代言人（数字人/VTuber）或核心产品形象在不同宣传视频、广告片或社交媒体内容中保持统一的视觉呈现，强化品牌识别度和 IP 价值。

本质上，参考主体生成视频是文生视频或图生视频技术的进阶应用，它在原有能力基础上，增加了对画面中特定主体一致性的精准控制维度。

（2）支持工具与伦理考量

目前，已有数款 AI 视频生成工具支持参考主体生成这一进阶功能：

可灵 AI（快手）作为国内领先的短视频平台推出的工具，可灵 AI 在人物和角色一致性生成

方面进行了优化，尤其适合需要固定人物形象的应用场景。

海螺 AI 平台也提供了参考主体生成能力，用户可通过上传参考图片，引导 AI 在生成视频时保持主体特征的一致性。

在利用此类技术，尤其是使用真人照片作为参考主体生成视频时，必须严格遵守伦理规范和法律法规；使用他人肖像前，务必获得其明确授权；严禁用于制作虚假信息、诽谤、诈骗、侵犯隐私或任何其他非法及违背公序良俗的用途；清晰认识到技术应用的法律边界，尊重和保护个人肖像权与人格尊严。

参考主体生成视频技术的核心特征在于其强大的主体一致性控制能力。它解决了 AI 视频生成中长期存在的角色/物体漂移问题，为创作连贯叙事、塑造稳定角色形象、维护品牌视觉统一性提供了可靠的技术支撑。这项技术极大地拓展了 AI 视频在个性化叙事、IP 运营和虚拟角色应用方面的深度与广度，是迈向更可控、更专业 AI 视频创作的重要一步。

（3）提示词通用写法

在视频生成过程中，当存在一个明确的参考主体（如一张人物设定图、一个已有的角色模型或一段包含主体特征的描述）时，提示词的撰写策略需要进行相应调整。参考主体的核心价值在于它已经包含了主体（通常是人物）的关键视觉特征，如外貌、发型、服饰风格、显著配饰等。这意味着提示词撰写者无须再耗费大量笔墨去重复描述这些静态特征，从而可以将创作精力更集中地投入到对画面动态元素和整体氛围的塑造上。这显著提升了提示词的效率和生成结果的准确性。

基于参考主体提供的核心形象，提示词的首要任务是构建一个契合主体特征并服务于叙事或主题的场景环境。场景描述应清晰交代事件发生的具体地点和环境特征。例如，是室内还是户外？如果是室内，是温馨的咖啡馆、充满未来感的实验室，还是古色古香的书房？如果是户外，是繁华的都市街道、静谧的森林，还是狂风呼啸的山巅？同时，氛围的营造至关重要。这包括环境的光线条件、天气状况、整体的情绪基调及可能存在的时代背景或风格化元素。这些要素共同构成了人物活动的舞台，为动作的发生提供了可信的背景。

在场景氛围设定清晰的基础上，提示词的核心任务转向描述主体的动态表现和镜头的运动方式。动作描写需具体且富有表现力，涵盖角色的肢体动作、表情神态、与其他物体或角色的互动及其整体的姿态或移动路径。运镜设计则决定了观众如何“观看”这个场景。需要明确指定镜头的类型、镜头的运动轨迹及可能的视角。精确的运镜描述能有效引导画面的节奏、焦点和戏剧张力，使生成视频更具电影感和叙事性。

利用参考主体撰写视频提示词的通用方法可概括为：依托参考主体锁定核心视觉形象，集中力量描绘场景环境与整体氛围，细致刻画主体的动态行为，并精心设计镜头的运动方式与观察角度。成功的提示词应能引导生成模型在忠实于参考主体外观的前提下，生动地演绎出指定的情境和动作。

6. 优化视频制作流程

（1）掌握镜头语言是 AI 视频创作的核心

想要借助 AI 工具制作出具备专业水准的视频画面，关键在于能够运用精准的语言描述出期望的视觉效果。这个过程类似于指导一位经验丰富的摄影师进行拍摄：必须清晰无误地传达拍摄的视角、拍摄主体与镜头的距离，以及镜头运动的方式和节奏。大量的实践反复证明，无论是采用文字直接生成视频（文生视频），还是基于现有图片进行视频化创作（图生视频），只要掌握了科学有效的描述方法，都能够显著提升最终成片的专业质感，达成令人满意的创作目标。

（2）精确描述镜头运动方式

镜头运动的描述是构建动态画面感和叙事节奏的基础。清晰说明镜头的运动轨迹、速度和方向至关重要。例如，当需要营造一种逐步聚焦、揭示细节或增强情绪张力的效果时，可以描述为“镜头以平稳缓慢的速度推向主角的面部特写”，其中“平稳缓慢”定义了运动的速度感，“推向”明确了运动方向，“面部特写”则确定了终点构图。反之，若需要展示宏大的场景、交代环境背景或营造疏离感，则可以表述为“镜头从高空俯视的起始点开始，持续匀速向后拉远，逐渐展露出下方整个灯火璀璨的城市夜景全景”。这里的“高空俯视”设定了起始视角，“匀速向后拉远”规定了运动方式和方向，“灯火璀璨的城市夜景全景”则描绘了最终呈现的画面内容。其他常见运动方式如横移、升降、环绕等，都需要同样具体地描述其路径和速度。

（3）精心规划画面构图层级

画面构图决定了观众的视觉焦点和信息接收的层次。明确指定画面的景别（即取景范围），能够有效引导 AI 理解你想要突出表现的内容。特写镜头（通常聚焦于局部细节或面部表情）具有强烈的视觉冲击力和情感穿透力，适用于强调关键物品、细微神态或重要符号，例如，“给产品 LOGO 一个清晰锐利的特写镜头，确保品牌标识醒目突出”，或者“捕捉人物眼中含泪的特写，传递其内心的悲伤”。中景镜头（通常选取人物腰部以上或展示人物与其直接互动的小范围环境）则能清晰呈现人物的动作、姿态及与道具、环境的互动关系，例如，“采用中景构图，展示工程师上半身正在精密操作控制台面板的动作”，或者“拍摄两人在咖啡馆桌前对坐交谈的中景画面，包含他们手部的动作和桌面的咖啡杯”。此外，全景镜头（展现主体与环境的关系）和远景镜头（强调环境氛围或空间规模）也是构成完整叙事的重要元素。

（4）构建系统化的工作流程

将脑海中的视觉构思有效转化为 AI 可执行的指令，需要一套系统化的描述工作流程。首先，在生成之前，进行完整的视觉预演至关重要。在脑海中像导演构思分镜头脚本一样，清晰地规划出整个视频片段或场景所需的镜头序列。思考每个镜头的目的、构图、运动及它们之间的衔接逻辑。其次，为每个独立镜头撰写详细、无歧义的描述。将上述关于运动、构图、主体的动作、环境氛围等所有关键要素，用精炼准确的语言整合到一句或一组连贯的句子中。例如，“夕阳西下，暖色调逆光勾勒出人物侧影（氛围/光线），镜头以低角度微微仰拍（视角），从中景开始平稳缓慢地推向人物眺望远方的面部特写（运动+终点构图），捕捉其沉思的表情（神态）”。生成初步结果后，进行效果对比与迭代优化是关键步骤。尝试对同一镜头构思使用几种略有不同的描述方式（如改变运动速度的形容词、调整景别术语、或增减环境细节），生成后仔细对比 AI 输出的画面差异，分析哪种描述更精准地传达了你的意图。基于对比结果，不断调整和精炼你的描述语言。需要铭记的是，描述的具体性和精确度与 AI 对创作意图的理解程度直接正相关。描述越细致入微，排除歧义的可能性越大，生成的画面就越接近预期。

（5）实践与精进是成功的路径

提升 AI 视频制作的描述能力，本质上如同学习一门新的视觉沟通语言。初始阶段可能会感到生疏或难以精准表达，这是完全正常的。然而，通过持续地应用上述技巧进行实践——有意识地构思镜头、精确地描述、严谨地对比结果并优化表达——制作者的经验将不断累积，描述能力会日趋娴熟。最终，创作者会发现，利用 AI 生成专业级视频画面的门槛并非不可逾越。只要掌握了这套以精准镜头语言为核心的描述方法，并辅以科学的流程和持续的练习，每一位创作者都有能力驾驭 AI 工具，高效地产出具有电影感、叙事性和专业视觉水准的视频作品。

7. 开发行业解决方案

(1) 电商短视频智能生产解决方案的创新实践

阿里巴巴集团开发的“鹿班”智能设计系统正在深刻变革电商短视频的生产模式。该系统通过先进的算法自动解析商品详情页信息，智能提取核心卖点数据。以一款防晒霜为例，系统能精准识别其 SPF 值、PA 等级、肤感特点等关键参数，并据此生成融合多种展示形式的专业级推广视频。传统模式下需要专业团队耗时数天完成的制作流程，如今仅需数分钟即可完成，制作成本更是降至传统方式的百分之五。更值得关注的是，系统能够智能编排实验室动画演示、真人肤质对比测试、紫外线相机实测等多元化内容形式，在 30 秒的黄金时间内系统化呈现商品的核心价值。这种全自动化的生产方式不仅极大提升了内容产出效率，更从根本上保障了视频内容与商品信息的高度一致性，有效避免了人工制作可能产生的信息偏差。

(2) 教育视频智能化改造的认知科学应用

在线教育平台猿辅导推出的“微课工厂”系统，深度融合了认知心理学与教育技术原理。该系统将复杂的学科知识点进行结构化拆解，以初中数学“勾股定理”为例，其被系统分解为 5 个循序渐进的认知阶梯：从生活场景中的直角三角形观察，到网格纸上的面积关系探索，再到代数表达式的推导验证，每个阶段都配有针对性动画演示。这种阶梯式设计符合学生的认知发展规律，有效降低理解门槛。系统生成的虚拟教师角色能够实现精准的口型同步与教学手势引导，配合动态几何图形的实时变换演示，共同构建出沉浸式的数字化学习场景。实际教学数据表明，采用此类 AI 微课模式后，学生课程完成率提升达 33%，核心知识点的长期记忆留存率提高 28%，显著优化了教学效果。

(3) 新闻视频实时生产的技术突破

新华社研发的“媒体大脑”系统在重大事件报道中展现出革命性的生产力。在昆明国际马拉松赛事的实战报道中，该系统实现了多源数据的智能整合：实时接入选手 GPS 定位信息、医疗点监控视频流、历史赛事成绩数据库及现场记者回传素材。通过人工智能驱动的信息处理流程，系统仅用 5 分钟即自动生成包含 10 个关键信息层的专业赛事短视频。这些信息层涵盖领先选手实时位置、各赛段用时对比、医疗救助热点分布、破纪录概率分析等深度内容。这种生产效率达到传统人工剪辑速度的 6 倍以上，重新定义了新闻视频制作的时效性标准，为重大事件的即时解读提供了技术保障。

(4) 社交媒体内容平民化创作的普惠实践

剪映推出的“AI 成片”功能显著降低了专业视频制作的门槛。当普通用户上传春日郊游的照片合集时，系统能智能分析图像中的场景元素与情感基调，自动匹配如《春之声圆舞曲》这类契合情境的背景音乐。基于图像序列的时间逻辑分析，系统自主添加电影级镜头语言：对花卉特写施加推镜头效果，在人物合影时采用拉远景处理，并动态生成“四月芳菲”等情境化字幕特效。整个处理流程约 3 分钟即可输出具有影院质感的 Vlog 作品。这种智能化工具使非专业用户也能轻松创作出富有艺术表现力的视频内容，真正实现了专业制作能力的普惠化。

(5) 行业解决方案的核心价值与未来展望

当前各领域的 AI 视频解决方案呈现出显著的共性特征：在保证内容质量的同时实现生产效率的几何级提升。通过深度融入人工智能技术，不同行业均探索出适配自身需求的新型生产范式。从电商平台的商品可视化革命、教育领域的认知化重构，到新闻行业的实时化生产、社交媒体的平民化创作，AI 视频技术正在系统性重塑内容产业的底层逻辑。这种变革不仅体现在效率指标的突破，更开创了人机协同创作的新纪元——专业视频制作不再局限于技术团队，而是进化为智能工具赋能下的全民创造力释放。随着技术的持续迭代，我们预见这种智能化生产模式将进一步向

医疗培训、工业仿真、数字文旅等领域拓展，持续推动视觉内容生产的范式转移。

◎任务实践

实践 1：随着 AIGC 技术的迭代升级，视频创作领域正经历范式革命。林同学是某高校新媒体中心的视觉总监，为配合学校“科技创新月”活动，需要制作《未来已来》主题宣传视频。通过对 Liblib AI 平台的学习，林同学掌握了文字生成视频与图片转视频的核心 workflow，计划通过 AI 辅助完成双版本视频创作。其制作需求如下：

提示词（文字生成视频部分基于可灵模型）：

① 生成一段 60 秒的科技感宣传片，需包含以下场景要素：

开场：粒子聚合成校徽（时长 5 秒，动态模糊效果）

主体：实验室场景（全息投影交互界面）、校园航拍（带动态追踪文字标注）、学生团队协作特写（慢动作处理）

转场：菱形分割转场（3 处）

结尾：三维文字“THE FUTURE IS NOW”冲出画面

② 技术参数要求：

分辨率：1920×1080 30fps

风格：赛博朋克+学术风混合（权重比 4:6）

启用“动态运镜”增强模式

具体操作步骤如下：

（1）登录 Liblib AI 创作平台，进入“文生视频”模块。

（2）在参数设置界面配置相关参数如图 3-30 所示。

基础模型：F.l.safetensors

提示词引导：7.2

VAE：自动匹配



图 3-30 文生视频参数配置界面

(3) 生成候选版本后, 选择版本进行优化: 使用“时序编辑器”微调粒子动画速度、添加校方提供的品牌色滤镜 (RGB 15, 82, 186)、输出为带 Alpha 通道的 MOV 格式, 如图 3-31 所示。

传统视频制作中, 素材转视频需要复杂的剪辑合成。Liblib AI 的图生视频功能如同一位精通蒙太奇的语言学家: 当输入星空照片时, 它能演绎出银河旋转坍塌成 DNA 链的哲学隐喻; 当处理建筑效果图时, 又会自动生成穿梭于钢结构间的粒子轨迹; 更惊艳的是其“风格迁移”能力——将水墨画的笔触转化为墨滴晕染开场的动态过程, 让静态作品获得二次生命。



图 3-31 视频生成效果

林同学在完成文生视频后, 继续使用校园摄影作品生成辅助视频。

提示词 (图生视频部分基于 VIDU 模型):

① 基于提供的 10 张校园摄影素材, 生成符合要求的动态视频: 每张图片持续 4 秒 (含 2 秒动态过渡)、统一添加“数据流光”边缘特效、对含人物的照片启用“动态重绘” (保护肖像不变形)。

② 增强效果要求: 图书馆照片生成书本翻页粒子化效果、实验设备照片添加 AR 标注浮动窗口、运动场景同步速度线特效。

具体操作步骤如下:

(1) 上传图片至“图生视频”工作区, 设置动态参数:

变焦幅度: $\pm 15\%$

旋转角度: 5 度

透视畸变: 关闭

(2) 启用高级功能 (如图 3-32 所示), 加载“科技感特效包”预设、设置关键帧插值模式: Bezier 曲线、开启智能避障 (防止文字重叠)。

(3) 最终生成成果包含主视频文件、独立特效素材包、自动生成的分镜脚本文档。

本案例展示了 AI 视频生成技术如何颠覆传统制作流程: 将原本需要专业团队两周的工作量压缩至单人 8 小时内完成, 同时实现传统手段难以企及的特效质量。实践表明, 掌握“视觉提示词工程”与“动态风格控制”能力将成为数字内容创作的新门槛, 这种通过语言精确操控视觉元素的生产方式, 正在重新定义“导演”的职能边界。作为新时代的创作者, 应当以 AI 为“数字场记”, 将更多精力投入创意构思而非技术实现, 从而释放更大的艺术潜能。

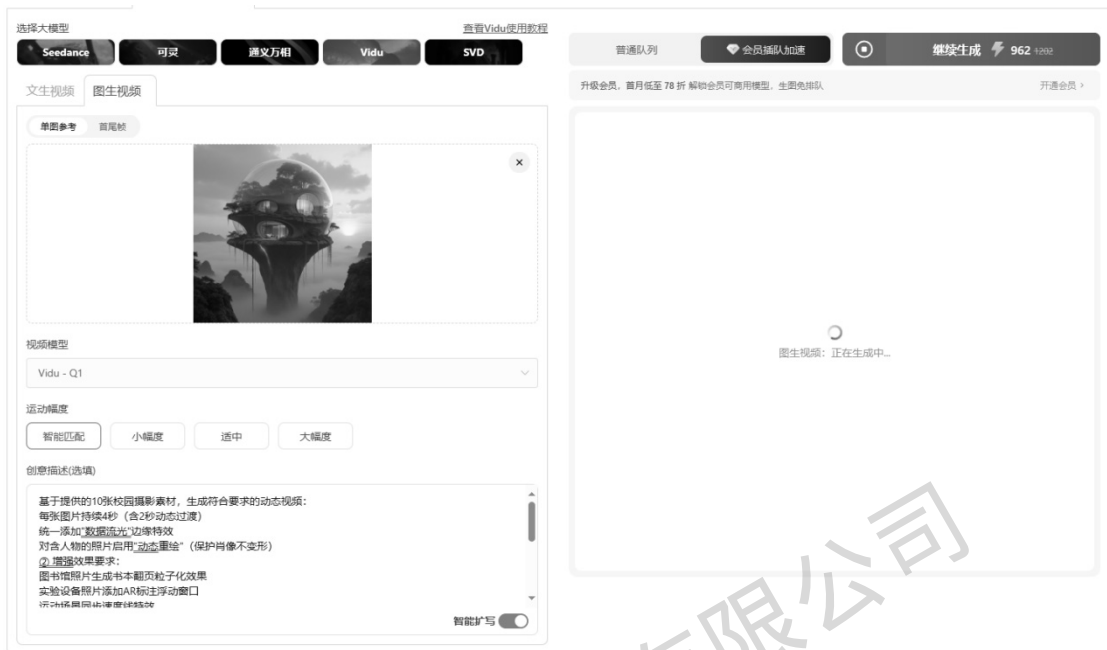


图 3-32 图生视频高级设置界面

项目考核与评价

一、选择题

- 在 AI 图像生成技术中，Checkpoint 模型与 LoRA 模型的核心区别在于（ ）。
 - Checkpoint 模型仅用于艺术创作，LoRA 模型仅用于商业设计
 - Checkpoint 是全参数大模型，LoRA 是轻量化微调模块
 - Checkpoint 模型体积更小，LoRA 模型体积更大
 - Checkpoint 需依赖 LoRA 才能工作，反之则不然
- 文生图技术的工作流程中，负责将文本描述转化为机器可理解的向量表示的阶段是（ ）。
 - 像素重建阶段
 - 文本编码阶段
 - 迭代去噪阶段
 - 潜空间映射阶段
- 与文生图技术相比，图生图技术的显著特点是（ ）。
 - 完全不依赖文本描述
 - 从随机噪声开始生成图像
 - 以参考图像为起点进行迭代去噪
 - 无法实现风格迁移功能
- 在 AI 音乐生成领域，Suno 平台与 Udio 平台的主要差异在于（ ）。
 - Suno 更擅长生成完整结构化音乐，Udio 更侧重风格融合创新
 - Suno 仅支持英文输入，Udio 仅支持中文输入
 - Suno 生成的音乐无法导出，Udio 生成的音乐可自由导出
 - Suno 属于第一代 AI 音乐系统，Udio 属于第二代
- 文生视频技术中，“运镜”的核心作用是（ ）。
 - 定义画面的色彩风格
 - 控制镜头的运动方式与视角变化

C. 决定视频的时长与分辨率

D. 生成视频的背景音乐与音效

二、填空题

1. 扩散模型在图像生成中的核心原理是通过多阶段降噪过程，将（ ）逐步转化为目标图像。

2. 图生图技术中，输入图像会先被送入（ ）压缩至潜空间，再与文本条件共同引导迭代去噪过程。

3. AI 音乐生成技术的发展经历了三代，其中第三代系统的特点是结合了（ ）和（ ）的优势，兼顾音乐结构合理性与音色表现力。

4. 在视频制作中，“景别”是指画面中主体的取景范围，常见的景别包括特写、近景、中景、全景和（ ）。

5. 多模态融合技术在视频生成中的核心价值是整合文本、图像、音频等不同模态信息，生成（ ）的视频内容。

三、简答题

1. 简述 Checkpoint 模型与 LoRA 模型在图像生成中的功能定位差异。

2. 请说明文生图技术中迭代去噪过程的基本原理，并列举两种常用的采样算法。

3. 图生图融合技术在商业设计中有哪些典型应用？请结合具体场景说明。

4. 分析 AI 视频生成技术当前面临的主要挑战（至少列举两点）。

5. 结合实际应用场景，说明“参考主体生成视频”技术的实用价值。