

脚本 + 分镜 + 配音 + 画面 = 文生视频全流程深度解析

导语：

你是否曾想过，仅凭一段文字就能创作视频？随着AI的飞速发展，这个想法已经具有了落地可能性。本教程将带你一步步拆解从最初的文本创意到最终视频呈现的全过程。我们将深入探讨脚本创作、分镜设计、专业配音中AI的运用方法，展示AI驱动的画面生成与编辑，让你掌握文生视频的核心技能，开启AI视频制作之旅！

启程之前 —— 了解、规划与准备

0.1 明确目标

在开始之前，你需要确定视频制作的终极目标，这对把控整个流程中的AI产出非常重要。不同目标的视频会有不同的剧本创意、分镜表达、视觉风格、音效要求。你需要首先明确：

- 视频的核心目的是什么？（营销、教育、娱乐等）
- 你的目标观众是谁？他们的喜好和观看习惯？
- 视频要传递的关键信息是什么？
- 期望的视频风格？（严肃、幽默、科幻、温馨等）

例如：



为咖啡品牌创作的营销短片（20秒以内）。目标受众是年轻人。他们喜欢新潮事物。视频要传递品牌年轻化的调性，期望的视频风格是潮流、阳光、活泼的。



温馨的奇幻故事短片（3分钟以内）。目标受众是6-12岁儿童。喜欢能够产生共鸣的小故事。视频要传达友爱、互助等正向价值观。视频风格是温馨的插画风格。



一部玄幻短剧（15集）。目标受众是修仙体裁短剧的观众群体。他们熟悉玄幻世界观的一般形式，期待看到有新意的世界观塑造、角色塑造、故事发展。视频要以主角的成长、冒险为主线，也要有足够丰富的情感线。视频风格是3d玄幻风格，具有中式美学。

0.2 了解相关AI工具：

今天，许多AI工具可以帮助你完成视频制作，主要包括以下类型：

- 剧本写作AI（Gemini、豆包、Chatgpt等强大的大语言模型）：辅助剧本创作，可以生成角色、情节、场景、对白等，帮助设计视频的主要叙事内容
- 分镜绘制AI（豆包、SkyReels AI等）：将剧本内容拆分为分镜，视觉化预览故事表达，推敲镜头叙事的表达方式，在开始正式制作前，低成本预览、评估、修改分镜
- 静帧生成AI（Midjourney、Flux1.1 Kontext等图像生成AI工具）：生成特定风格的高质量静帧，以图生视频的方式，控制正式视频的内容、风格、视觉质量
- 视频生成AI（可灵、即梦、Wan2.1、FramePack等视频生成工具）：按分镜片段生成视频片段
- 语音合成AI与唇形匹配AI（可灵、即梦、GPT-SoVITS V2、D-ID、Heygen 等）：当视频中角色需要对话时，使用语音合成（TTS）类的AI工具合成语音文件，再通过唇形匹配功能将 语音+视频片段 作为输入进行合成。也可以选择直接将 文本+视频片段 作为输入，选择音色和情绪，生成带有语音的视频
- 音效AI（可灵、即梦、MMaudio等）：生成环境音、乐音等音效
- 基础剪辑工具（iMovie、剪映等）：进行最终剪辑，通过传统剪辑方式将视频片段串联在一起

0.3 时间规划与成本预算：

确定你的项目需要的制作周期、可支配的预算。对于新手个人用户来说，商业化付费工具的性能、效率、访问便捷性会略高于开源工具，请评估您的项目条件决定每一步采用的工具。以下演示中，每个步骤都将尽量推荐最合适的工具，也会对免费开源工具进行说明。

一 灵魂的基石 —— 剧本创作（Scripting）

1.1 什么是视频剧本？

视频剧本是视频故事的“骨架”，它定义了故事的“什么”（what）和“为什么”（why）。

- 与纯文学文本的区别：含场景描述、人物动作、对话、潜台词、音效指示等，强调视觉化和听觉化表达，需考虑画面、声音等呈现方式。
- 基本要素：
 - 场景（时间、地点）
 - 动作（人物行为）
 - 对话（人物语言）
 - 旁白（叙事声音）
 - 音效提示（如背景音乐、环境声）

1.2 剧本格式与规范

标准剧本格式特点：

- 使用专业格式（如Final Draft） ，明确区分动作、对话、角色名等。
- 简洁、清晰、便于拍摄执行。

1.3 AI辅助剧本创作：

大语言模型工具可以帮助我们辅助剧本创作，生成剧本创意，进而生成角色、情节、场景、对白等，帮助设计视频的主要叙事内容。

1.3.1 利用AI工具（如ChatGPT, Claude等）生成剧本创意/大纲

价值：快速得到剧本创意/大纲，或得到能够启发剧本创作的灵感

演示工具：Gemini

示例：

制作目标	Gemini输出
<p>你是一名资深编剧，请按以下视频创作需求，提出3条不同的剧本创意：</p> <p>为咖啡品牌创作的营销短片。目标受众是年轻人。他们喜欢新潮事物。视频要传递品牌年轻化的调性，期望的视频风格是潮流、阳光、活泼的。</p>	<div><div>1. 《咖啡漫游指南》</div><div><ul style="list-style-type: none">◦ 创意： 年轻人手持[品牌名]咖啡，探索城市中不为人知的小众打卡点（特色书店、艺术装置、秘密花园），咖啡是他的探索伙伴。◦ 风格： 探索、小清新、Vlog感。</div><div>2. 《暂停，然后“啡”起来》</div><div><ul style="list-style-type: none">◦ 创意： 年轻人学习/工作疲惫时，画面“卡顿”或变黑白，他喝下一口[品牌名]咖啡，瞬间世界恢复色彩和活力，效率倍增。◦ 风格： 趣味、对比、夸张。</div><div>3. 《咖啡星球奇遇记》</div><div><ul style="list-style-type: none">◦ 创意： 略带超现实。主角喝下咖啡后，仿佛进入一个由咖啡豆、咖啡杯、拉花图案构成的奇幻小宇宙，充满想象力。◦ 风格： 奇幻、可爱、动画结合。</div></div>
<p>你是一名资深编剧，请按以下视频创作需求，提出3条不同的剧本创意：</p> <p>温馨的奇幻故事短片（3分钟以内）。目标受众是6-12岁儿童。喜欢能够产生共鸣的小故事。视频要传达友爱、互助等正向价值观。视频风格是温馨的插画风格。</p>	<div><div>《分享味道的云朵》</div><div><ul style="list-style-type: none">• 故事梗概： 小狐狸烤了一个香喷喷的派，想分享给住在山那边的朋友小兔子，但路太远了。一朵路过的小云朵飘下来，小狐狸把派的味道“扇”给云朵。云朵吸饱了香味，飘到小兔子家上空，下了一场“闻起来像派”的香气雨。• 核心价值： 分享、友谊。• 奇幻设定： 能吸收和传递气味的云。• 视觉亮点： 云朵从白色变成温暖的派的顏色，散发着肉眼可见的香气粒子。</div></div>

	<p>《灰尘兔子的宝物》</p> <ul style="list-style-type: none">• 奇幻设定：床底下不起眼的灰尘团，其实是善良的“灰尘兔子”，它们会守护孩子们遗失的小零件。• 故事梗概：小杰最爱的玩具车坏了，少了一个小轮子，他怎么也找不到。晚上，他睡着后，一只由灰尘和绒毛组成的、圆滚滚的“灰尘兔子”从床底滚了出来，用小短腿把那个轮子推到了床边最显眼的地方。第二天早上，小杰惊喜地发现了轮子，并修好了玩具车。• 核心价值：失而复得的喜悦、发现看不见的帮助。 <p>《小手套的旅行》</p> <ul style="list-style-type: none">• 奇幻设定：丢失的手套会独自踏上旅程，帮助森林里的小动物，最后回到主人身边。• 故事梗概：莉莉在公园里不小心掉了一只手套。短片跟随这只手套的视角：它被风吹走，为一只小刺猬挡住了寒风；它掉在地上，成了一窝小田鼠温暖的床。第二天，莉莉回到公园寻找，发现手套挂在一个树枝上，里面还装着一颗小松果，像是小动物们留下的谢礼。• 核心价值：善良的传递、万物有灵。
<p>你是一名资深编剧，请按以下视频创作需求，提出3条不同的剧本创意：</p> <p>一部玄幻短剧。目标受众是修仙体裁短剧的观众群体。他们熟悉玄幻世界观的一般形式，期待看到有新意的世界观塑造、角色塑造、故事发展。视频要以主角的成长、冒险为主线，也要有足够丰富的情感线。视频风格是3d玄幻风格，具有中式美学。</p>	<ol style="list-style-type: none">1. 《罪血传人》<ul style="list-style-type: none">◦ 角色创新：主角是上古“罪仙”后裔，血脉中蕴含强大力量，但也背负着整个修真界追杀的诅咒。◦ 成长与冒险：主角在逃亡中觉醒血脉，努力证明“罪血”并非邪恶，并试图为先祖正名，揭露当年“罪仙”被陷害的真相。◦ 情感线：爱上了一位正道门派的“圣女”，两人身份对立，情感在正邪、道义与私情间备受考验。2. 《替身灵偶》<ul style="list-style-type: none">◦ 角色创新：主角本是一具为天才哥哥打造的“替死灵偶”，意外拥有自我意识后，渴望拥有真正的人生。◦ 成长与冒险：主角在模仿哥哥的过程中逐渐找到自我，并发现自己拥有独特的修炼天赋。最终要面对选择：继续做替身，还是为自己而活，哪怕与家族为敌。◦ 情感线：与一位知道他真实身份，却依旧鼓励他、爱护他的师妹产生感情，但哥哥也喜欢这位师妹。3. 《万界商行》<ul style="list-style-type: none">◦ 世界观创新与中式美学：主角经营着一家连接诸天万界的“商行”，贩卖各界法宝、丹药、情报。以“和气生财”为道，却也因此卷入万界纷争。

	<ul style="list-style-type: none">◦ 主角：一位精明能干，看似市侩，实则重情重义的年轻掌柜。◦ 成长与冒险：在经商过程中不断提升实力，拓展人脉，化解商业危机与修真实力危机。◦ 情感线：与一位来自异界的“顾客”，因一桩生意结缘，发展出跨越位面的爱恋，并为了对方不惜与强大势力对抗。
--	---

1.3.2 短视频的脚本初稿生成

在敲定剧本创意后，进一步生成完整的剧本初稿，方便对叙事逻辑进行预览、优化。

演示工具：Gemini

限制：对于较为长篇的视频故事，可能需要大语言模型分步拆解故事，先补全完整的剧本逻辑生成章节大纲，检查和优化整体情节，再细化生成每一章节之内的具体剧本。（例如上面提到的玄幻短剧创意）

示例：

我们选择较为短篇的儿童奇幻故事短片作为案例进行脚本初稿的生成，这里我们选择“《分享味道的云朵》”这个创意，让Gemini进行完整、格式规范的剧本生成。

提示词	Gemini输出
生成《分享味道的云朵》的完整剧本，使用中文，使用标准的Final Draft剧本格式。适合3分钟内的短片。	<p>[SCENE START]</p> <p>FADE IN:</p> <p>1. EXT. 小狐狸的家门口 - 白天</p> <p>阳光明媚，坐落在一片青翠草地上的小木屋。</p> <p>小狐狸（8岁）戴着一顶小小的厨师帽，用两只毛茸茸的手，小心翼翼地将一个金黄酥脆、热气腾腾的浆果派端到门口的木桩桌上。</p> <p>派的香甜气息以肉眼可见的、波浪形的【香气粒子】飘散在空气中。小狐狸深深吸了一口气，脸上露出满足的笑容。</p> <p>他拿起一块小木牌，上面写着“世界上最好吃的派！”。</p> <p>他望向远处，视线越过森林，落在了一座高耸的【青色山脉】上。他的笑容慢慢消失，耳朵也耷拉了下来。</p> <p><center>小狐狸</center></p> <p>（叹气）</p> <p>要是小兔子也能尝到就好了……可是山那么高，等我爬过去，派都凉了。</p> <p>他用小爪子托着下巴，满脸愁容。</p> <p>就在这时，一朵小小的、雪白蓬松的【小云朵】慢悠悠地飘了过来，高度刚好在小狐狸头顶不远处。它好奇地晃了晃身体，似乎被派的香味吸引了。</p>

小狐狸眼睛一亮，一个绝妙的主意在他脑中形成。他的耳朵重新竖了起来。

<center>小狐狸</center>

（对小云朵喊）

喂！小云朵，你可以帮我一个忙吗？

小云朵轻轻地上下浮动，像是在点头。

小狐狸兴奋地拿起一片巨大的荷叶，把它当作扇子。他站在派的旁边，使出全身力气，将派的香气朝着小云朵“扇”过去。

神奇的事情发生了。

那些金色的香气粒子，像被磁铁吸引一样，纷纷涌向小云朵。

小云朵的身体开始发生变化。它不再是纯白色，而是从中心开始，慢慢地、温柔地被染上了温暖的【橘黄色】，就像刚出炉的派的颜色。

很快，整朵云都变成了派的颜色，散发着柔和的光芒，看起来美味极了。

小云朵心满意足地晃了晃身体，再次向小狐狸“点点头”，然后轻快地升高，朝着青色山脉的方向飘去。

小狐狸在原地用力挥手，脸上是期待的笑容。

2. EXT. 天空/山脉 - 白天

【派味云朵】轻盈地飞越茂密的森林，越过高耸的山脉。从它的视角看下去，世界就像一个绿色的绒毯。

3. EXT. 小兔子的家门口 - 白天

山脉的另一边，小兔子（8岁）正在给她的胡萝卜菜园浇水。她看起来有点无聊，时不时抬头看看天空。

派味云朵正好飘到她的头顶上空。

小兔子好奇地看着这朵颜色奇怪的云。

<center>小兔子</center>

咦？这朵云怎么是黄色的？

派味云朵轻轻地“挤压”了一下自己的身体。

随即，一阵【细密的、闪着微光的香气雨】从云中洒落。雨丝没有打湿任何东西，它们只是纯粹的、看得见的香气。

香气雨落在小兔子的鼻尖上，落在她的胡萝卜上，落在她家门口的摇椅上。

小兔子闭上眼睛，用力地嗅了嗅。她的小鼻子抽动着。

惊喜的表情在她脸上绽放。

<center>小兔子</center>

（惊喜地）

这个味道……是小狐狸烤的浆果派！

她张开双臂，在香气雨中快乐地转圈。她能感觉到，这不仅仅是味道，更是朋友分享的温暖和心意。

4. EXT. 两人的家/天空 - 白天

【分屏画面】

左边：小狐狸坐在自家门口，看着派味云朵消失的方向，微笑着。

右边：小兔子在香气雨中开心地笑着，仿佛已经尝到了派的美味。

画面中央，那朵小云朵已经变回了白色，但看起来比之前更快乐、更蓬松。它满意地在蓝天中继续自己的旅程。

友谊的味道，跨越了山脉。

FADE OUT.

[SCENE END]

1.3.3 AI辅助润色、改写和风格迁移

另外，除了帮助编写剧本，大语言模型也可以在润色、改写和风格迁移方面提升编剧的效率。比如：

- AI辅助润色 (AI-Assisted Polishing)

例如：“让主角的语言更加口语化”、“检查剧本是否符合行业标准的格式”

- AI辅助改写 (AI-Assisted Rewriting)

例如：“让这个场景的节奏更快”、“将主角的内心独白改写为旁观者的客观描述”、“如果角色A在这里选择了B而不是C，后续会怎样发展？”

- AI辅助风格迁移 (AI-Assisted Style Transfer)

例如：“用王家卫的风格改写这段对话”、“将一个爱情故事的片段改写成悬疑惊悚风格”、“将现代口语对话改写成具有民国时期韵味的台词”

二 视觉的蓝图 —— 分镜设计 (Storyboarding)

2.1 为什么需要分镜？

剧本是“写”故事的，而分镜是“画”故事的。分镜故事版详细描绘了每个镜头，定义了故事将“如何被看见”（how）。分镜的作用包括：

- 可视化叙事，方便预览，或与合作伙伴进行探讨。
- 预演镜头，发现潜在问题。
- 作为草稿指导后续最终画面的生成。

2.2 分镜的核心要素：

- **画面构图：** 景别（远、全、中、近、特）、角度（平、仰、俯）、主体位置。

- **镜头运动：**推、拉、摇、移、跟等。
- **角色动作与表情：**关键帧如何捕捉关键动作/表情。
- **文字说明：**对话、旁白、音效、时长、镜头编号等。

2.3 AI辅助生成分镜脚本、分镜参考图

通过AI工具，我们可以快速生成分镜脚本、静态分镜图片，达到快速预览分镜组合效果，在开始正式制作前，低成本预览、评估、修改分镜设计。

这里会用到两类AI：

- **大语言模型：**用于将之前的剧本拆解为分镜脚本，以文字的方式描述每个具体的镜头内容。角色形象和制作风格（包括视觉/音效风格）也可以由大模型根据剧本进行初步定义（但特别需要人工进行审美评判，灵活修改或丰富）。
- **具有文生图功能的图片生成模型：**用于将分镜脚本转化为静态分镜图（非正式版），在开始正式制作前，低成本预览、评估、修改分镜

接下来，我们将按顺序演示以下步骤：生成角色形象描述、生成制作建议（视、听风格）、生成分镜脚本、生成角色参考图、生成分镜关键帧预览图、时间轴上的分镜预览

工具：Gemini、midjourney、Flux、豆包、任意视频编辑工具（如即梦-故事创作功能、iMovie、AE）

■ 生成角色形象描述

提示词	Gemini输出
为剧本《分享味道的云朵》生成所有角色的形象描述，简洁概括。	<p>《分享味道的云朵》角色形象描述</p> <p>1. 小狐狸 (Little Fox)</p> <ul style="list-style-type: none">● 外形: 一只大约8岁的拟人化小狐狸。拥有橘红色、毛茸茸的蓬松尾巴和柔软的皮毛。他的眼睛大而明亮，充满好奇心和创造力。耳朵非常灵敏，能根据情绪耷拉或竖起。● 穿着: 总是戴着一顶对他来说略显宽松的白色厨师帽，这是他作为“小厨神”的骄傲标志。● 性格: 热情、善良、动手能力强。热爱烘焙和发明，把与朋友分享美食看作是最大的快乐。他是一个积极的思考者和问题解决者。 <p>2. 小兔子 (Little Rabbit)</p> <ul style="list-style-type: none">● 外形: 一只同样8岁左右的拟人化小兔子。皮毛是温柔的米白色或浅灰色，有一对长长的、能表达情绪的耳朵。她的眼神文静而温柔。● 穿着: 通常穿着一条朴素的、带有田园风格的碎花小围裙或连衣裙，显得干净利落，符合她喜欢照料花园的设定。

	<ul style="list-style-type: none">• 性格: 文静、心思细腻、懂得感恩。她享受宁静的田园生活，珍视与小狐狸的友谊。她能从微小的事物中感受到巨大的幸福。 <h3>3. 小云朵 (The Little Cloud)</h3> <ul style="list-style-type: none">• 外形: 一朵拥有自我意识、充满灵性的云。它比普通的云朵要小巧，质感像棉花糖一样蓬松柔软。<ul style="list-style-type: none">◦ 初始形态: 纯净的雪白色。◦ 派味形态: 在吸收了派的香气后，会从中心向外温柔地渐变成温暖的、如同烤派一般的橘黄色，并散发着肉眼可见的柔和光晕和香气粒子。• 性格: 充满好奇心、乐于助人、天真烂漫。它不会说话，但通过轻轻浮动、晃动身体、变换形态来表达“点头”、“同意”和快乐等情绪，是连接友谊的魔法信使。
--	---

■ 生成制作建议（视、听风格）

提示词	Gemini输出
为剧本《分享味道的云朵》提供整体制作建议，包括：风格、色调、音效	<h2>《分享味道的云朵》整体制作建议</h2> <p>**核心定位:** 一部治愈系、无对白（或极少对白）的奇幻童话动画短片。</p> <hr/> <h3>1. 视觉风格 (Visual Style)</h3> <p>**总基调:** 温暖、柔软、富有手作感。</p> <ul style="list-style-type: none">• 动画风格建议:<ul style="list-style-type: none">◦ 3D定格动画质感 (3D Stop-motion Feel): 即使使用3D软件制作，也应追求类似《小羊肖恩》或Laika工作室作品（如《久保与二弦琴》）的“手作质感”。角色的毛发、衣服的纹理、场景中的木纹和草地都应带有轻微的不完美和触感，避免过度光滑的塑料感。◦ 毛毡或黏土质感: 可以将角色设计成毛毡玩偶或黏土公仔的质感，这会立刻拉近与观众的距离，让整个世界看起来像一个可触摸的微缩景观。• 角色设计:<ul style="list-style-type: none">◦ 小狐狸: 动作灵活，表情丰富。重点突出他竖起和耷拉的耳朵，以及充满创造力的大眼睛。厨师帽可以做得稍微大一点，显得俏皮可爱。◦ 小兔子: 动作文静优雅。她的长耳朵是表达情绪的关键，比如在闻到香味时会惊喜地弹跳一下。◦ 小云朵: 设计得像一团有生命的、会呼吸的棉花糖。它的形态变化要流畅、柔软，避免僵硬的变形。• 节奏与运镜:

- **节奏舒缓：**整体节奏应是宁静而舒缓的，给观众足够的时间去感受情绪和环境。在小狐狸灵光一闪、小兔子闻到香味等关键时刻，可以通过稍快的剪辑和特写来制造节奏变化。
- **运镜温柔：**多使用平稳的推、拉、摇镜头。航拍云朵飞行的镜头要轻快、自由，营造开阔感。在表现“香气雨”等奇幻场景时，可使用慢镜头和旋转镜头，最大化其梦幻效果。

2. 色彩基调 (Color Palette)

****总基调：** **高饱和度、低对比度的暖色调，如同午后阳光下的绘本。

- ****主色调：** **以【枫糖色】、【奶油黄】和【青草绿】为基础，构建一个温暖、安全、充满生机的世界。
- **场景用色：**
 - **小狐狸的世界：** 以暖色为主。他的木屋是温暖的棕色，周围环境是明快的橘色和黄色调，象征着他的热情与创造力。他烤的派必须是画面中最诱人的【金黄色】。
 - **小兔子的世界：** 以柔和的冷暖色调结合。她的菜园是【薄荷绿】和【胡萝卜橙】，远处的天空是【婴儿蓝】，营造一种宁静、恬淡的田园氛围。
 - **天空与山脉：** 天空是清澈的【天蓝色】，山脉则是带有童话感的【青色】或【薰衣草紫色】，与地面形成柔和的层次感。
- **魔法元素用色（核心视觉亮点）：**
 - **香气粒子：** 设计成【半透明的金色光点】，带有微微的拖尾效果，使其在空气中的流动轨迹清晰可见。
 - **云朵变色：** 从纯白平滑过渡到【温暖的、如同蜂蜜或烤派内馅的橘黄色】。这个过程要非常柔和，带有光晕效果，让云朵本身像一个发光体。
 - **香气雨：** 这不是水，而是【闪烁的金色粒子雨】。在阳光下应该有水晶般的折射光芒，强调其非现实的、魔法般的特质。

3. 音效与配乐 (Sound Design & Music)

****总基调：** **轻快、灵动、治愈人心。

- **配乐 (Music):**
 - **乐器选择：** 以原声乐器为主，营造自然、质朴的感觉。**尤克里里、木吉他、钢琴、铁琴 (Glockenspiel)** 和弦乐的拨奏 (Pizzicato) 是绝佳选择。长笛或单簧管可以用来点缀，增加童话色彩。
 - **主题旋律：** 创作一段简单、抓耳、温暖的主题旋律。这段旋律可以根据情节进行变奏：
 - 小狐狸烘焙时 -> **轻快、跳跃**
 - 小狐狸发愁时 -> **放慢速度，变为小调，略带伤感**
 - 云朵出现并传递香味时 -> **旋律变得悠扬、充满希望和奇幻感**
 - 小兔子闻到香味时 -> **主题旋律以最欢快、最完整的方式奏响，推向高潮**

	<ul style="list-style-type: none">● 音效 (Sound Effects - SFX):<ul style="list-style-type: none">○ 核心原则: 音效要“拟人化”和“可爱化”，比现实声音更夸张、更悦耳。○ 关键音效设计:<ul style="list-style-type: none">■ 派的香味: 当香气粒子飘散时，可以配上轻微的、如同【小风铃】或【气泡破裂】的清脆声音，将嗅觉感受转化为听觉。■ 云朵吸收香味: 一种轻柔的【“咻——”的吸入声】，伴随着温暖的嗡鸣声，表示它正在“吃”味道。■ 云朵点头: 可以用【果冻晃动】的Q弹声音来表现。■ 香气雨: 绝对不能用真实的雨声。应该使用【细碎的水晶或玻璃珠滚落】的声音，叮叮咚咚，清脆悦耳，强调其魔法特质。■ 角色动作音: 小狐狸扇动荷叶是“呼呼”的风声，小兔子的脚步声是轻快的“嗒嗒”声。他们的惊叹、开心等情绪，可以用可爱的语气词（如“咦？”、“哇！”）或清脆的笑声来表达，即使没有台词。
--	--

■ 生成分镜脚本

提示词	Gemini输出
根据剧本《分享味道的云朵》决定每段情节合适的时长/镜头数量，以专业的镜头组合和运镜方式，生成分镜脚本。不要使用表格。	<p>《分享味道的云朵》分镜脚本</p> <p>总时长: 约2分30秒</p> <p>核心视觉: 温馨、奇幻、治愈</p> <hr/> <p>场景1: 小狐狸的家门口 (预计时长: 70秒)</p> <p>镜头 1.1</p> <p>时长: 5秒</p> <p>景别: 大全景 (ELS)</p> <p>镜头内容: 阳光灿烂的早晨。一个坐落在青翠草地上的可爱小木屋。远景是连绵的青色山脉，画面宁静美好。</p> <p>运镜: 固定镜头。</p> <p>音效: 鸟鸣，微风声。</p> <p>镜头 1.2</p> <p>时长: 6秒</p> <p>景别: 中景 (MS)</p> <p>镜头内容: 小狐狸戴着厨师帽，骄傲地将一个金黄色的浆果派从木屋里端出来，放在门口的木桩上。派散发着肉眼可见的、波浪形的金色【香气粒子】。</p> <p>运镜: 从木屋门口【跟拍】小狐狸走出来。</p> <p>音效: 轻快的背景音乐起。</p> <p>镜头 1.3</p>

时长：4秒

景别：特写 (CU)

镜头内容：派的特写。热气蒸腾，浆果酱鲜红欲滴，派皮酥脆金黄。香气粒子更明显地从派上飘起。

运镜：固定镜头，焦点清晰。

音效：“滋滋”的热气声。

镜头 1.4

时长：8秒

景别：过肩镜头 (OTS)

镜头内容：从小狐狸的肩膀后方看去，他满足地吸了一口气，然后望向远处的青色山脉。他的表情由满足变为失落，耳朵耷拉下来。

运镜：镜头缓慢【向前推】，越过他的肩膀，聚焦到远处的山脉上。

音效：音乐变得稍显忧伤。小狐狸轻轻的叹气声。

镜头 1.5

时长：7秒

景别：仰拍中景 (Low Angle MS)

镜头内容：小狐狸托着下巴发愁。此时，一朵小而蓬松的白云悠悠地飘入画面上方，它好奇地晃了晃。

运镜：固定镜头。

音效：一阵轻柔的风铃声，暗示云朵的出现。

镜头 1.6

时长：5秒

景别：特写 (CU)

镜头内容：小狐狸的脸。他先是注意到云朵，然后眼睛一亮，仿佛想到了什么，耳朵“噌”地一下竖了起来。

运镜：快速【拉近】到他眼睛的特写，捕捉“灵光一闪”的瞬间。

音效：一个清脆的“叮”声。

镜头 1.7

时长：8秒

景别：全景 (WS)

镜头内容：小狐狸兴奋地拿起一片巨大的荷叶，对着派用力扇动。金色的香气粒子被扇向空中的小云朵。

运镜：固定镜头，展现完整的动作关系。

音效：扇动荷叶的“呼呼”声，背景音乐重归活泼。

镜头 1.8

时长：10秒

景别：中景特写 (MCU)

镜头内容：聚焦于小云朵。金色的香气粒子像有生命一样涌入云朵。云朵的中心开始被染上温暖的橘黄色，并逐渐向外扩散。

运镜：镜头缓慢【推进】，强调色彩变化的细节。

视觉特效：云朵颜色变化的柔光特效。

镜头 1.9

时长：7秒

景别：全景 (WS)

镜头内容：整朵云已经变成了派的颜色，散发着柔和的光晕。它在空中满意地晃动，然后向更高处飘去，朝山脉方向飞去。小狐狸在地面上用力挥手告别。

运镜：镜头【向上摇 (TILT UP)】，跟随“派味云朵”远去。

音效：音乐达到一个小高潮，充满希望感。

场景2：天空/山脉 (预计时长: 10秒)

镜头 2.1

时长：10秒

景别：航拍跟拍 (Aerial Tracking Shot)

镜头内容：“派味云朵”轻快地飞越茂密的森林和高耸的山脊。下方景物飞速后退，展现旅途的轻快和辽阔。

运镜：侧面【跟拍】云朵，模拟与它一同飞行。

音效：悠扬、开阔的过场音乐，风声。

场景3：小兔子的家门口 (预计时长: 55秒)

镜头 3.1

时长：5秒

景别：全景 (WS)

镜头内容：山脉的另一边，小兔子的家门口。这里有一个可爱的胡萝卜菜园。小兔子正在无聊地给胡萝卜浇水。

运镜：固定镜头。

音效：浇水声，音乐转为平缓。

镜头 3.2

时长：7秒

景别：仰拍 (Low Angle)

镜头内容：从小兔子的视角向上看，她注意到天上那朵颜色奇特的“派味云朵”。

运镜：固定镜头。

音效：小兔子好奇的“咦？”一声。

镜头 3.3

时长：8秒

景别：中景 (MS)

镜头内容：“派味云朵”飘到小兔子头顶正上方，它像海绵一样轻轻“挤压”了一下自己。

运镜：固定镜头。

音效：一声轻柔的、类似水滴的音效。

镜头 3.4

时长：10秒

景别：慢动作特写 (Slow-motion CU)

镜头内容：闪着微光的、半透明的【香气雨】缓缓落下。雨丝落在小兔子的鼻尖上，落在胡萝卜的叶子上，但没有打湿任何东西。画面唯美梦幻。

运镜：镜头缓慢【向下摇 (TILT DOWN)】，跟随香气雨的飘落。

音效：梦幻般的、闪烁的音效，背景音乐变得温柔。

镜头 3.5

时长：6秒

景别：大特写 (ECU)

镜头内容：小兔子的脸。她闭上眼睛，小鼻子用力地、可爱地抽动着，感受这奇妙的香气。

运镜：固定镜头。

音效：满足的吸气声。

镜头 3.6

时长：5秒

景别：特写 (CU)

镜头内容：小兔子猛地睁开眼睛，惊喜万分。

运镜：固定镜头。

音效：音乐上扬，小兔子惊喜的独白：“是小狐狸烤的浆果派！”

镜头 3.7

时长：14秒

景别：全景 (WS)

镜头内容：小兔子张开双臂，在金色的香气雨中快乐地转圈，脸上洋溢着幸福的笑容。

运镜：【弧形运镜 (ARC SHOT)】，围绕着小兔子旋转，增强动感和幸福感。

音效：欢快的音乐和兔子清脆的笑声。

场景4：两人的家/天空 (预计时长: 15秒)

镜头 4.1

时长：10秒

景别：分屏画面 (Split Screen)

镜头内容：

- **左屏：**中景。小狐狸坐在自家门口，看着远方，欣慰地微笑着。
- **右屏：**中景。小兔子仍在香气中微笑，仿佛能感受到朋友的心意。

运镜：两个画面均为固定镜头。

音效：主题音乐温柔地响起，将两个场景的情感连接起来。

镜头 4.2

时长：5秒

景别：大全景 (ELS)

镜头内容：分屏画面向上【划像】消失，露出湛蓝的天空。那朵小云朵已经变回了纯白色，心满意足地、轻快地继续飘向远方。

运镜：固定镜头，画面宁静而悠远。

音效：音乐缓缓收尾，留下最后的余音。

FADE OUT.

为下一步生成分镜关键帧参考图提供角色形象参考、画面风格参考。这里我们推荐将Midjourney（需要付费，访问方式有自学成本）作为首选，因为其在角色形象设计和风格表现方面，都更容易达到最佳审美效果。

Midjourney官网：<https://www.midjourney.com/>

当然，其他文生图工具也可以作为替代，比如开源的Flux。优点包括：开源模型，可本地部署；支持comfyui生态；在魔搭平台可以找到免费体验的space。


魔搭space链接：<https://www.modelscope.cn/aigc/imageGeneration?tab=default>

🦋 角色参考图将不仅用于本节的下一步“生成分镜关键帧预览图”，还将用于最终视频制作中，决定最终视频内容中的角色形象/画面风格。提高这张图片的美学质量是非常重要的。

演示：

提示词	Midjourney输出
3D stop-motion animation style, with felt and clay textures, bright and soft colors. The picture depicts a peaceful pastoral scene: a green hillside with a bright blue sky and a cozy cabin at the foot of the mountain. In the center of the foreground are three cute characters of the same size: the little fox is an anthropomorphic orange-red fox about 8 years old, with a big fluffy tail, a white chef's hat, a floral apron, and big, bright eyes. The little rabbit, with off-white fur, long ears, and a pink and white striped dress, stands quietly and gently beside the fox; floating on the side is the little cloud,	<div><div><div><div><div><div></div><div></div></div></div><div><div><div></div><div></div></div></div><div><div><div></div><div></div></div></div></div></div><div><div>Midjourney Bot</div><div>3D stop-motion animation style, with felt and clay textures, bright and soft colors. The picture depicts a peaceful pastoral scene: a green hillside with a bright blue sky and a cozy cabin at the foot of the mountain. In the center of the foreground are three cute characters of the same size: the little fox is an anthropomorphic orange-red fox about 8 years old, with a big fluffy tail, a white chef's hat, a floral apron, and big, bright eyes. The little rabbit, with off-white fur, long ears, and a pink and white striped dress, stands quietly and gently beside the fox; floating on the side is the little cloud, a small cotton candy-textured cloud, pure white with no facial features or limbs. The overall layout of the picture is balanced, the atmosphere is warm and cheerful, and it has a handmade texture and natural soft light effect. --ar 16:9 --v 6.1 --relax --upsample by @POLYH33 (relax)</div></div><div></div></div> <div></div>


<p>a small cotton candy-textured cloud, pure white with no facial features or limbs. The overall layout of the picture is balanced, the atmosphere is warm and cheerful, and it has a handmade texture and natural soft light effect. -- ar 16:9</p>	
--	--

提示词	Flux输出
<p>Felt and clay textures. Grassland background. An orange fox in a white chef's hat and floral apron, an off-white bunny in a pink and white striped dress, and a floating cotton candy texture cloud.</p>	

■ 生成分镜关键帧预览图

豆包可以一键批量生成生成所有分镜的首尾帧、关键帧，这种批量生成的能力极大提高了效率，避免仅仅生成分镜帧的预览版本，就耗费过多时间，非常适合快速预览故事的分镜内容。能够在正式开始投入时间和成本制作之前，快速判断目前的分镜表达是否合适，并进行调整，这对后续的制作会很有帮助。

注意：由于形象和风格一致性的控制能力不足，经常产生不准确的画面内容，这里的关键帧参考图只起到大致预览作用，而不能直接用于图生视频。

提示词	豆包输出
<p>为奇幻短片《分享味道的云朵》生成每个镜头的首尾帧和核心帧，总计：19个镜头，预计总时长约2分30秒。画面为横屏，比例为16:9。与参考图中的角色形象保持一致，风格保持一致。</p>	<p>请点击图片查看</p> 

+{角色形象描述}（简要概括）

+{制作风格/色调}（简要概括）

+{分镜脚本}

+角色参考图



1. 项目概述	2. 角色设定	3. 场景设定	4. 镜头语言	5. 分镜脚本	6. 制作流程	7. 后期制作	8. 成片展示
1.1 项目背景	1.2 角色设定	1.3 场景设定	1.4 镜头语言	1.5 分镜脚本	1.6 制作流程	1.7 后期制作	1.8 成片展示
2.1 角色设定	2.2 角色设定	2.3 角色设定	2.4 角色设定	2.5 角色设定	2.6 角色设定	2.7 角色设定	2.8 角色设定
3.1 场景设定	3.2 场景设定	3.3 场景设定	3.4 场景设定	3.5 场景设定	3.6 场景设定	3.7 场景设定	3.8 场景设定
4.1 镜头语言	4.2 镜头语言	4.3 镜头语言	4.4 镜头语言	4.5 镜头语言	4.6 镜头语言	4.7 镜头语言	4.8 镜头语言
5.1 分镜脚本	5.2 分镜脚本	5.3 分镜脚本	5.4 分镜脚本	5.5 分镜脚本	5.6 分镜脚本	5.7 分镜脚本	5.8 分镜脚本
6.1 制作流程	6.2 制作流程	6.3 制作流程	6.4 制作流程	6.5 制作流程	6.6 制作流程	6.7 制作流程	6.8 制作流程
7.1 后期制作	7.2 后期制作	7.3 后期制作	7.4 后期制作	7.5 后期制作	7.6 后期制作	7.7 后期制作	7.8 后期制作
8.1 成片展示	8.2 成片展示	8.3 成片展示	8.4 成片展示	8.5 成片展示	8.6 成片展示	8.7 成片展示	8.8 成片展示

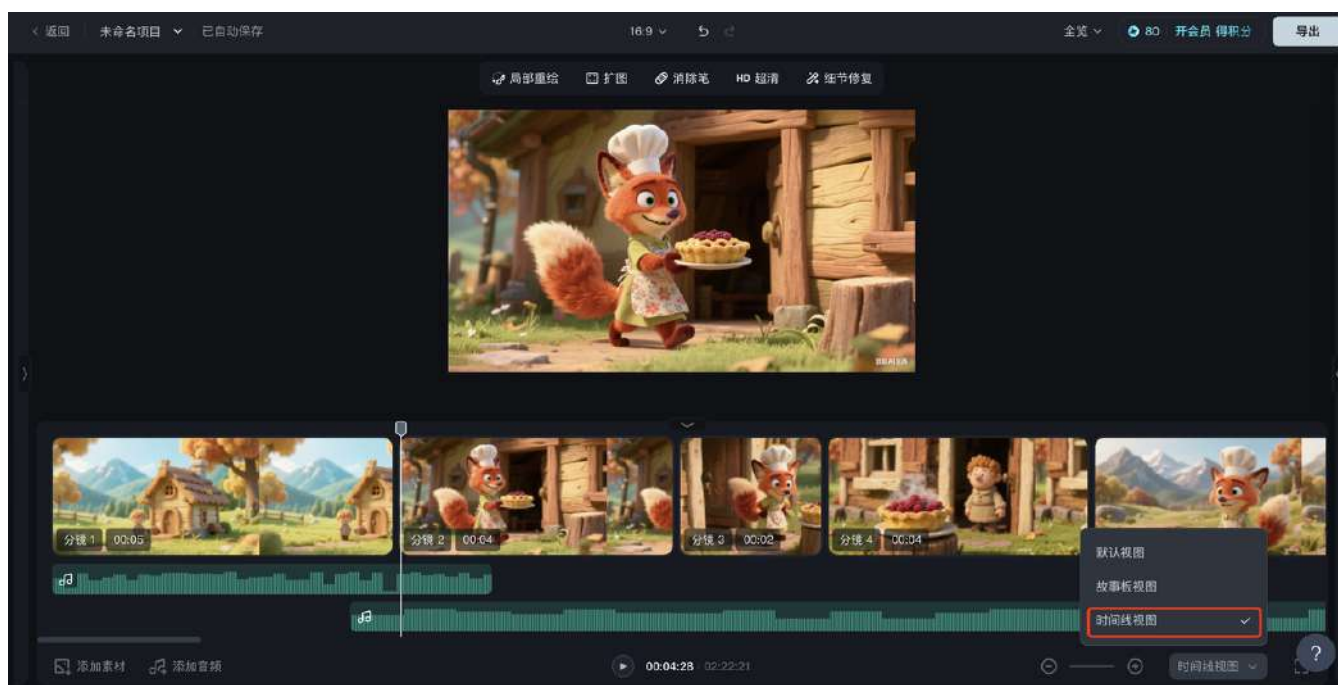


■ 时间轴上的分镜预览：

选择一些分镜关键帧按分镜脚本中的时间计划串联，配上临时音轨，即可快速地大致预览视频的时间节奏，以便进行必要的计划调整（例如对部分冗长镜头的时间计划进行缩短）。临时的配乐和音效可以自开源音效网站如FMA、Freesound等下载。（素材来源已在文末注明）



这里的音效/配乐仅作预览。实际上，视频声音的处理是一个非常独立的专业领域，充满各种复杂的技巧。限于作者的知识面有限，本文下一节中会展开讲解一些基础概念，但这仍然是比较粗糙、简化的。





通过预览视频，我们可以记下一些对某些镜头内容、配乐选择、配乐时间分配等进行调整的想法。

到这里，我们已经完成了故事的视觉化预览，接下来即将进入正式版视频的制作流程。

三 视觉的盛宴 —— 画面生成与编辑 (Visual Generation & Editing)

3.1 视频编辑的基本流程：

首先，在开始生成各类最终素材之前，需要了解整个视频编辑过程的基本流程和其中需要的素材类型：

- 导入素材 (AI生成的图像/视频片段, 配音, BGM, SFX)。
- 时间线操作：将图像/视频片段按分镜顺序排列，与配音对齐。
- 转场效果：自然过渡。
- 文本与图形：添加字幕、标题、信息图表。
- 调色与校正：统一画面风格。
- 节奏控制：根据配音和BGM调整画面时长和剪辑点。

镜头片段的生成是整个过程的起点，形成整个视频的画面内容，至关重要。

3.2 了解AI视频片段生成的必要功能：

对于生成不同时长、不同内容的视频素材，AI视频工具中最至关重要的功能有：

- **图生视频：** 以静态图片为初始帧生成，起到绝对控制画面内容/风格的作用

- **首尾帧：**以两张图片分别作为首尾帧，除内容外进一步控制更精确的画面变化过程（如更精确的动作过渡或运镜过程），或进行两个镜头间的完全连接。
- **时长控制：**精确到至少秒级的视频时长控制
- **视频延长：**对于一次生成达不到的时长，需要对较短镜头进行延长
- **视频高清化：**提高画面清晰度，修复质量

3.3 为镜头生成进行的帧图片生成：

上面我们提到了图生视频和首尾帧功能的重要性，那么如何得到这些静态帧图片呢？

Flux1.1 Kontext是进行这些图片生成的完美工具，且目前几乎是无可替代的。其图片编辑功能（选择max模型）可以保持稳定的角色/场景/风格一致性，同时修改画面的内容，适合生成最终版分镜的静态帧。使用上一节中生成的**角色参考图**作为起始图片，直接使用提示词改变画面内容。

Flux1.1 Kontext官网-图像编辑：[FLUX Playground - Black Forest Labs](#)

Liblib在线 workflow：[kontext 图像编辑 | 黑森林官方版-workflow-WildboarA6-LiblibAI](#)

在上一节的“分镜关键帧预览生成”部分，其实已经得到了各镜头中首帧、关键帧、尾帧的提示词，只需要在提示词中再加入保持角色形象、画面风格、场景布局一致性的要求，就可以直接命令Flux1.1 Kontext生成。根据结果，可以再进行灵活地提示词调整，充分控制单帧内容。

镜头1.2

- 首帧：3D定格动画质感，毛毡或黏土质感，高饱和度、低对比度暖色调。中景，小狐狸（一只大约8岁的拟人化小狐狸，拥有橘红色、毛茸茸的蓬松尾巴和柔软的皮毛，眼睛大而明亮，耳朵灵敏，戴着一顶略显宽松的白色厨师帽，穿着碎花围裙）骄傲地从木屋里端出一个金黄色的浆果派，开始向门口的木桩走去。
- 核心帧：小狐狸将浆果派放在门口的木桩上，派散发着肉眼可见的、波浪形的金色【香气粒子】。
- 尾帧：小狐狸站在木桩旁，看着派，表情骄傲。



各镜头中首帧、关键帧、尾帧的提示词（来自豆包）





Edit > Keeping the image style and character image unchanged, generate: The little fox with a chef's hat takes the first step out of the wooden house, ready to come out with a ...



角色图+提示词修改画面内容，精确控制（来自Flux1.1 Kontext）

以镜头1.2的首帧为例，输入和输出如下：

提示词	Flux1.1 Kontext输出
<p>Keeping the image style and character image unchanged, generate: The little fox with a chef's hat takes the first step out of the wooden house, ready to come out with a golden berry pie. Medium view.</p> <p>（中文翻译：保持图片风格和人物形象不变，生成：戴着厨师帽的小狐狸迈出了木屋的第一步，准备带着金黄色的浆果派走出来。中景。）</p> <p>+角色参考图</p> 	

所有镜头的静态帧如下：

Scene1	Scene2	Scene3
--------	--------	--------

镜头1.1

首帧



镜头1.2

首帧、关键帧



镜头1.3

仅首帧



镜头1.4

首帧、核心帧、尾帧



镜头2.1

首帧



镜头3.1

首帧



镜头3.2

首帧



镜头3.3

首帧



镜头3.4

首帧



镜头3.5

首帧



镜头1.5:
首帧、尾帧



镜头1.6
首帧



镜头1.7
首帧、尾帧



镜头3.6
首帧、尾帧



镜头3.7
首帧



镜头4.1
首帧



镜头4.2
首帧、尾帧



镜头1.8

首帧、尾帧



镜头1.9

首帧



3.4 视频片段生成：

本文选择了FramePack、Wan2.1作为视频生成的主要工具。作为开源模型，他们在时长控制、视频延长方面比多数付费AI视频工具（如可灵、即梦）都更加自由，可以达到秒级/帧数级控制。同

时，他们的视频质量在多数简单的画面运动（简单主题、常见动作、常见运镜等）的表现上足够过关。当然，开源无法有时无法实现最佳效果，例如极其快速地运动、非常自然的表情变化等。必要时请灵活尝试更好的付费工具（如Veo3、可灵、即梦）

本文使用了魔搭平台上的space演示，但也推荐在Liblib等平台寻找他们的Comfyui工作流进行本地部署，会具有更多的自由度。

- FramePack: [FramePack图生视频](#) 、 [FramePack_F1视频延长](#)
- Wan2.1: [wan2.1图生视频+首尾帧](#)

单帧生视频演示：



镜头 3.1

时长： 5秒

景别： 全景 (WS)

镜头内容： 山脉的另一边，小兔子的家门口。这里有一个可爱的胡萝卜菜园。小兔子正在无聊地给胡萝卜浇水。

运镜： 固定镜头。

音效： 浇水声，音乐转为平缓。

单帧图生视频，只需设定时长、帧频，用提示词简洁准确的描述视频内容。

输入	输出
<div></div> <div>工具：FramePack-图生视频</div> <div>Total Video Length (Seconds)（时长）：5s</div> <div>Frames Per Second（帧频）：24</div> <div>Prompt（提示词）： The little rabbit is holding a green watering can and watering the carrots.</div>	<div></div>

首尾帧生成、视频延长演示：



镜头 1.7

时长： 8秒

景别： 全景 (WS)

镜头内容： 小狐狸兴奋地拿起一片巨大的荷叶，对着派用力扇动。金色的香气粒子被扇向空中的小云朵。

运镜： 固定镜头，展现完整的动作关系。

音效： 扇动荷叶的“呼呼”声，背景音乐重归活泼。

镜头1.7中，需要展现云朵吸入香气粒子的过程和颜色变化，因此使用了首尾帧。

由于首尾帧允许生成的总时长有限，因此首先生成5s的局部片段，再使用视频延长功能。为了留有剪辑余地，延长了5s（2s剪辑余量）。

输入	输出
<div></div> <div>工具：wan2.1-首尾帧功能</div> <div>提示词：</div> <div>镜头内容： 小狐狸站立在固定位置，拿起一片巨大的荷叶对着派快速地挥舞。少量金色的发光粒子从派上向上升起，被扇向空中的小云朵。云朵慢慢发出金色柔光。运镜： 固定镜头。</div> <div>时长： 5s</div>	<div></div> <div>工具：FramePack-视频延长功能</div>

Prompt（提示词）：

The little fox stands in a fixed position, picks up a huge lotus leaf and waves it quickly at the pie. A small amount of golden luminous particles rise from the pie and are fanned towards the small clouds in the sky. The clouds slowly emit a soft golden light. Camera movement: Fixed lens.

Additional Video Length to Generate (Seconds)(延长时间)：3s



四 声音的魅力 —— 配音与音效 (Voiceover & Sound)

4.1 声音——独立的叙事工具

声音是如今所有类型的影视视频中不可获取的一部分，是非常主要的叙事工具。如果在观看电影时关闭声音，你将明确意识到声音对情绪、节奏和信息传达有多么巨大的影响。

声音不是画面的附属品，而是独立的叙事语言。

4.2 声音的分类

按照功能不同，影视视频中的声音可以分为3个大类：

- 对白 (Dialogue): 故事的骨架。它的首要任务是“清晰”。
- 音乐 (Music): 情绪的催化剂。分为“片中角色能听到的音乐” (有源音乐) 和“只有观众能听到的配乐” (无源音乐)。
- 音效 (Sound Effects / SFX): 世界的血肉。它构建了环境的真实感与超现实的想象力。

音效还可以分为：

- 拟音：模拟画面中人物与物体互动产生的声音，如脚步声、衣物摩擦声、开关门声。它赋予了角色“重量感”和“质感”，是真实感的核心。
- 环境音：特定场景中持续存在的背景声，如城市街道的嘈杂、森林的鸟鸣风声、安静房间的空调声。它能瞬间告诉观众“我们在哪”，并奠定场景的情绪基调（宁静、紧张、喧闹等）。
- 硬音效：所有非拟音、非环境的标志性声音。小到电话铃声，大到宇宙飞船引擎、爆炸、魔法光波。

4.3 音效设计工作流程

音效师负责从剧本出发，构思、创造并组合所有非对白、非音乐的声音，让世界“听起来”是真的。

- 剧本解读与声音定位 (Spotting Session)：与导演一起观看影片，逐个镜头讨论需要哪些声音、声音的风格和情绪，这是整个工作的“设计图纸”。
- 声音资产的整理与规划：根据声音定位，规划需要录制哪些声音，需要从音效库购买或寻找哪些素材。
- 对白处理/剪辑/替换：保证故事清晰；清理现场录音的杂音、选择最佳的录音版本；在录音棚里让演员重录无法使用的现场对白。
- 拟音录制：拟音师在拟音棚观看画面，使用各种道具，同步“表演”出画面中的声音并录制下来。
- 环境音搭建：分层搭建，像画画一样，铺上基础层（如风声），再叠加细节层（远处的狗叫、偶尔的汽车声），创造立体感。
- 硬音效录制或合成：从庞大的音效库中寻找合适的声音，针对特定声音进行实地录音（如录制真实的汽车关门声）；或创造现实中不存在的声音，例如，“激光剑”的声音可能是由电动机声和干冰的滋滋声叠加处理而成。
- 混音：预混 (Pre-Dubbing)阶段将成百上千条音轨，先分类别（如所有脚步声、所有环境音）进行初步混合，形成“分轨” (Stems)；终混 (Final Mix)阶段将对白、音乐、音效的“分轨”进行最终的艺术化平衡，输出我们在影院听到的最终版本 (如5.1环绕声或全景声)。

4.4 AI赋能的声音设计新范式

在AI为主的视频制作流程中，声音与画面一样，可以完全从0开始生成，不必依赖于现场录制或特定的合成技术。

- AI生成对白

对白声音的生成可以分为两个步骤：文生语音、唇形匹配。一些AI视频工具中也可以实现在生成视频后，一次性完成这两个步骤（例如：可灵）

文生语音：以一段参考音频和对白台词为基础，生成对白音频

开源工具：GPT-SoVITS-V2 [GPT-SoVITS V2 在线语音生成（中日英韩粤）](#)

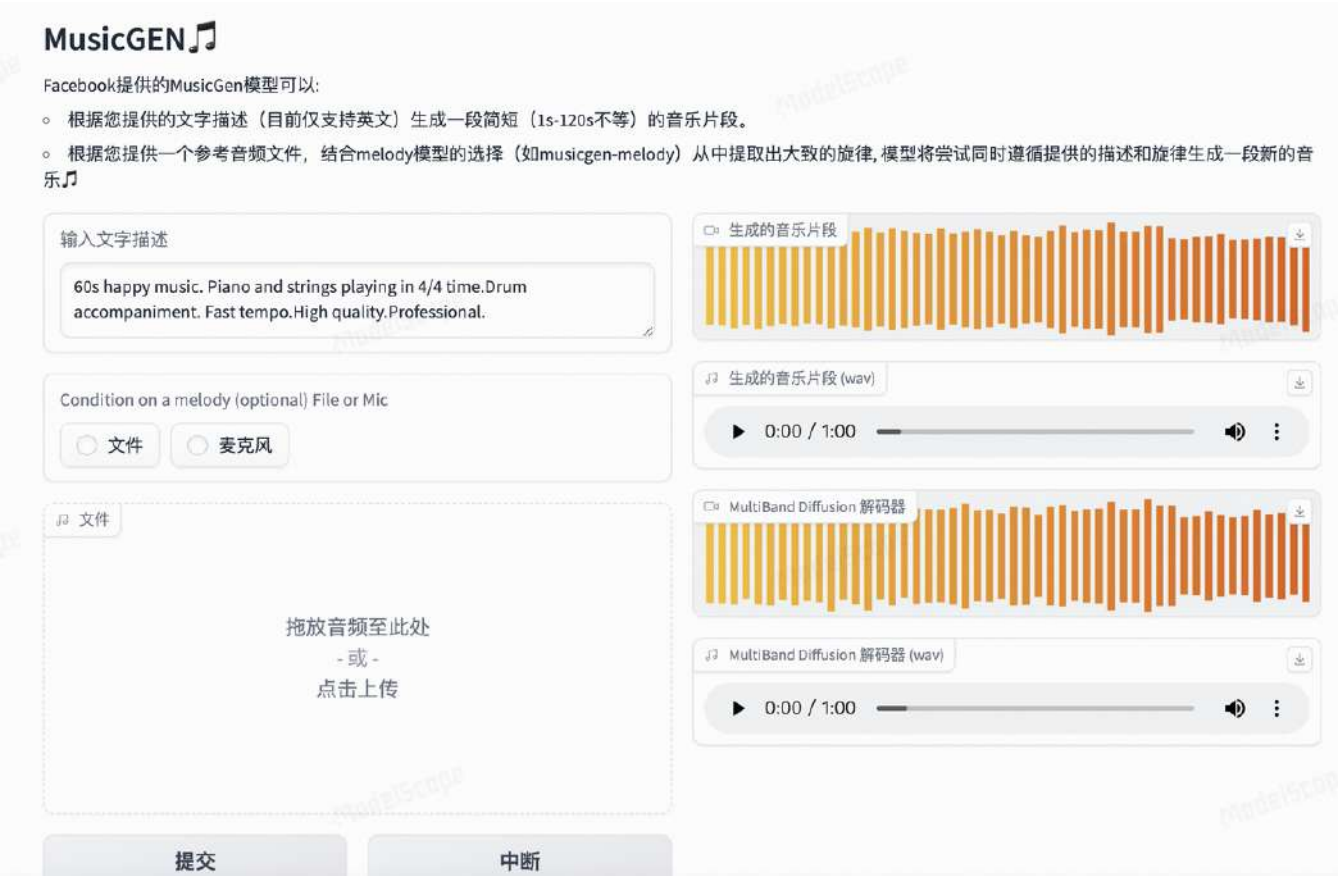
唇形匹配（仅限用于人/面部类人的角色）：将对白音频与视频片段结合，生成唇形与对白音频一致的新视频

开源工具：LatentSync [对口型](#)

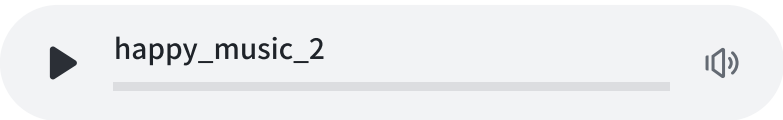
■ AI生成配乐

文生音乐：以一段音乐描述提示词为基础，生成特定时长的音乐

开源工具：[MusicGen: 让AI作曲触手可及](#)



演示：

输入	MusicGen输出
60s happy music. Piano and strings playing in 4/4 time. Drum accompaniment. Fast tempo. High quality. Professional. 音频时长：60s	 A screenshot of an audio player interface. It features a play button on the left, the filename "happy_music_2" in the center, and a volume icon on the right. Below the filename is a progress bar.



■ AI生成音效

文生音效、视频生音效功能能够解决大部分的音效生成问题。

开源工具：[SFX-Sound-magic](#)、[MMAudio](#)

- 拟音：一般需要音画同步，采用 视频生成音效功能更加适合。将

演示：

输入	SFX-Sound-magic
<div data-bbox="244 282 675 533"></div> <div data-bbox="244 566 595 672"><p>Prompt:</p><p>The sound of waving a fan.</p></div> <div data-bbox="244 752 475 857"><p>Negative Prompt:</p><p>music.</p></div>	<div data-bbox="707 282 1473 719"></div>

- 环境音：

判断某段环境音的视频片段时长，用提示词文生音效即可。

- 硬音效：

判断某个音效时长，用提示词文生音效即可。

- 混音与剪辑

完成所有声音资产的生成后，即可将画面和不同音轨进行剪辑，然后将对白、音乐、所有音效元素按重要性、远近、大小进行平衡与调和。最终，让观众能听清重要的，感受到该感受的，最终形成一个和谐统一、富有层次感的声音体验。

最终效果：



结语：

到此为止，我们已经完成了AI文生视频全流程的学习，你可以开始尝试制作自己的第一步全AI短片了。

AI在影视流程中的应用其实还有很多类型可以探索，例如高清化视频（Topaz Video AI）、智能混音（iZotope Neutron (Mix Assistant)）、对白智能降噪（Adobe Podcast Enhance）等。非常期待您能解锁更多花式应用AI的方法

附录：

临时音效：

来自：<https://freesound.org/>

鸟鸣与风声：Birdsong Wind Blowing.wav by totalcult -- <https://freesound.org/s/388641/> --
License: Creative Commons 0

风铃声：Wind Chimes by harrietniamh -- <https://freesound.org/s/271509/> -- License: Attribution 4.0

山谷风声：Highland Winds FX by Evil_Ear_Recordings -- <https://freesound.org/s/523967/> --
License: Creative Commons 0

水滴声：Water Drop With Large Bouncy Reverb Tail by 8bitmyketison --
<https://freesound.org/s/701071/> -- License: Creative Commons 0

吸气呼气: inhale/exhale.wav by superEGsonic -- <https://freesound.org/s/273979/> -- License: Creative Commons 0

临时配乐

来自: <https://freemusicarchive.org/>

木吉他: **Relaxing Acoustic Guitar**by Greg Kirkelie

悲伤的钢琴: **Sad**by Lite Saturation

快乐的木吉他: **Happy Acoustic Guitar Background**by BlackTrendMusic

琶音: **Piano Arpeggio - I - Variation I.mp3**by Gregor Quendel

欢乐的钢琴: **My Happy Dance**by Xennial