

《血佛》AI 短片 Demo

MiniMax H3 本地生成全流程复盘与 Prompt 手册

从小说片段 -> 设定图 -> 6 镜头生成 -> 尾帧连续 -> Premiere 成片

测试平台：RTX 4070 Laptop 8GB / 32GB RAM

输出：1248 x 704 / 24fps / 38.67s

参考性质：记录一次真实可复现的低显存 H3 视频生产流程，并保留踩坑。

2026.09

0. 这份手册记录什么

这不是“理论上可以”的方案，而是一条实际跑完并剪成 38.67 秒成片的工作流。目标是让别人能看懂：一段小说文本如何被拆成镜头、如何用一张设定图约束人物与世界、哪些镜头应该切、哪些镜头应该用尾帧续上，以及 8GB 显存机器上如何靠低显存工作流把 MiniMax H3 跑起来。

一句话结论

8GB 显存并不意味着 H3 “跑不了”。这次测试的关键是低显存注意力、分块 FF、DynamicVRAM/offload、分阶段加载，以及把长叙事拆成多个 5-7 秒镜头。代价是时间：单条生成大约 6.5-10 分钟，采样阶段最慢。

《血佛》Demo 的实际生产链



目录

- 1. 原文到镜头：先把剧情变成可生成事件
- 2. 测试环境、模型与 ComfyUI 低显存链路
- 3. 设定图：把人物、场景、道具一次性钉住
- 4. 六镜头结构与连续性策略
- 5. 六个镜头的成片表现
- 6. Prompt 设计方法
- 7. 完整 Prompt 附录
- 8. Premiere 后期与声音
- 9. 失败点与复盘
- 10. 一页复现清单

1. 原文到镜头：先把剧情变成可生成事件

原文核心事件（来自《血佛》第四章）

楚缘拿红蜡烛前去拜佛；蜡烛放到佛像下方后立刻自行燃烧；长廊里开始闪烁鬼影，所有鬼影都面向楚缘并露出与佛像相同的瘆人笑容；拜佛结束后红蜡烛熄灭，鬼影随之消失。

这段文字很适合做生成式视频 demo，因为事件链短、视觉标志强，而且天然能拆成“建立 -> 异变 -> 渐进揭示 -> 核心恐怖画面 -> 异象消失 -> 余悸”六段。真正重要的是：不要让一个 30-40 秒 Prompt 同时承担所有动作。让模型一次只完成一个镜头。

镜头 1	建立空间与人物：楚缘持红烛，望向佛堂。
镜头 2	仪式动作：跪拜、放置红烛、红烛自燃。
镜头 3	鬼影初现：远处先出现少量模糊人影，不揭脸。
镜头 4	核心 reveal：多名鬼影回望，统一“血佛笑容”。
镜头 5	异象结束：烛火熄灭，鬼影消失。
镜头 6	补一个情绪收尾：楚缘确认走廊空了，谨慎恢复并准备离开。



为什么镜头 6 是后加的？

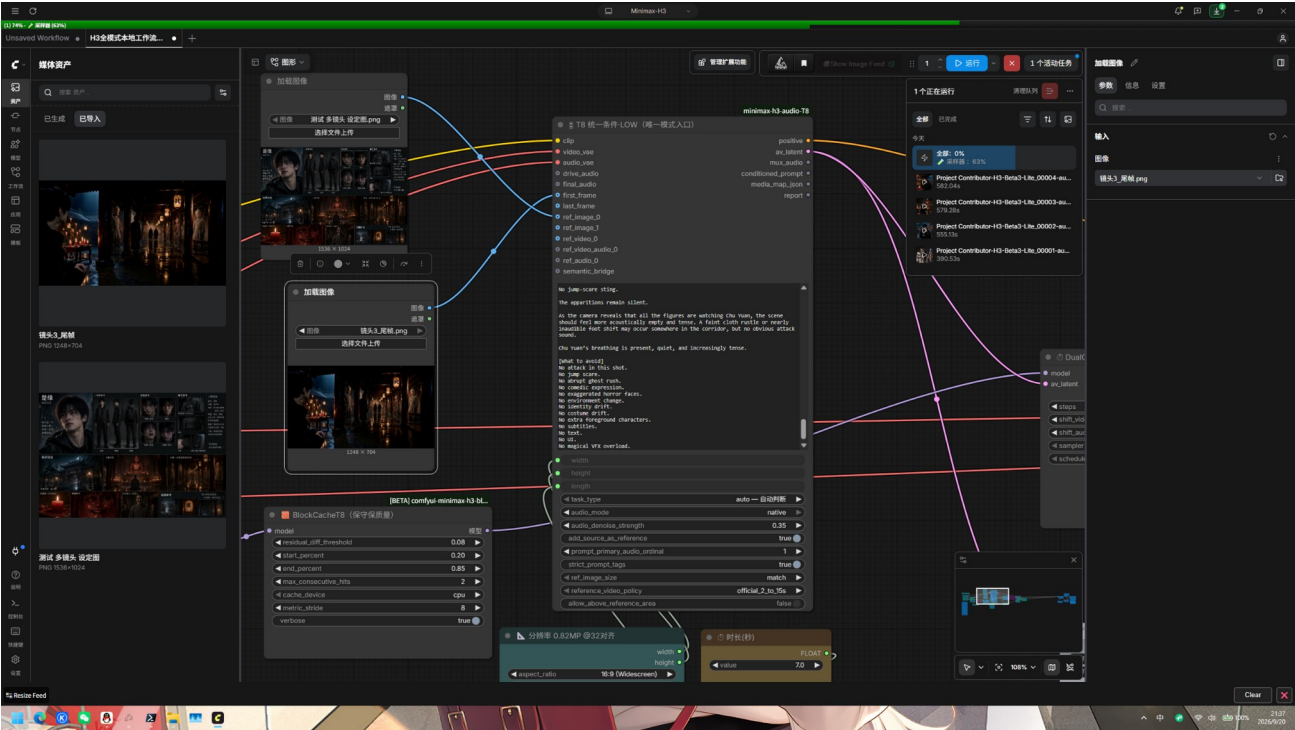
镜头 5 的生成结果停在“楚缘正准备起身”的中间状态。如果直接结束，情绪像被截断。补一个 4 秒左右的空走廊余悸镜头后，整段才真正“落地”。这也是生成视频常见情况：分镜表不是一次写死的，应该根据实际成片补镜头。

2. 测试环境、模型与 ComfyUI 低显存链路

硬件	Intel i7-14700HX / RTX 4070 Laptop 8GB / 32GB DDR5-5600
ComfyUI	ComfyUI Desktop 独立实例: Minimax-H3
Python / Torch	Python 3.13.12 / PyTorch 2.12.1 + CUDA 13.0
主模型	10Eros_Max_h3_TURBO_ref2va_beta2_int8_convrot_skip_edges.safetensors
文本/视觉编码器	qwen3vl_32b_heretic_minimax_h3_nvfp4.safetensors, CLIP 类型必须选 minimax
视频 VAE	minimax_h3_video_vae_int8_convrot.safetensors
音频 VAE	minimax_h3_audio_vae_fp32.safetensors
模型文件总量	约 42GB (磁盘权重总量, 不等于运行时必须同时驻留 42GB 内存)
输出	1248 x 704 / 24fps / H.264 MP4 / CRF 19

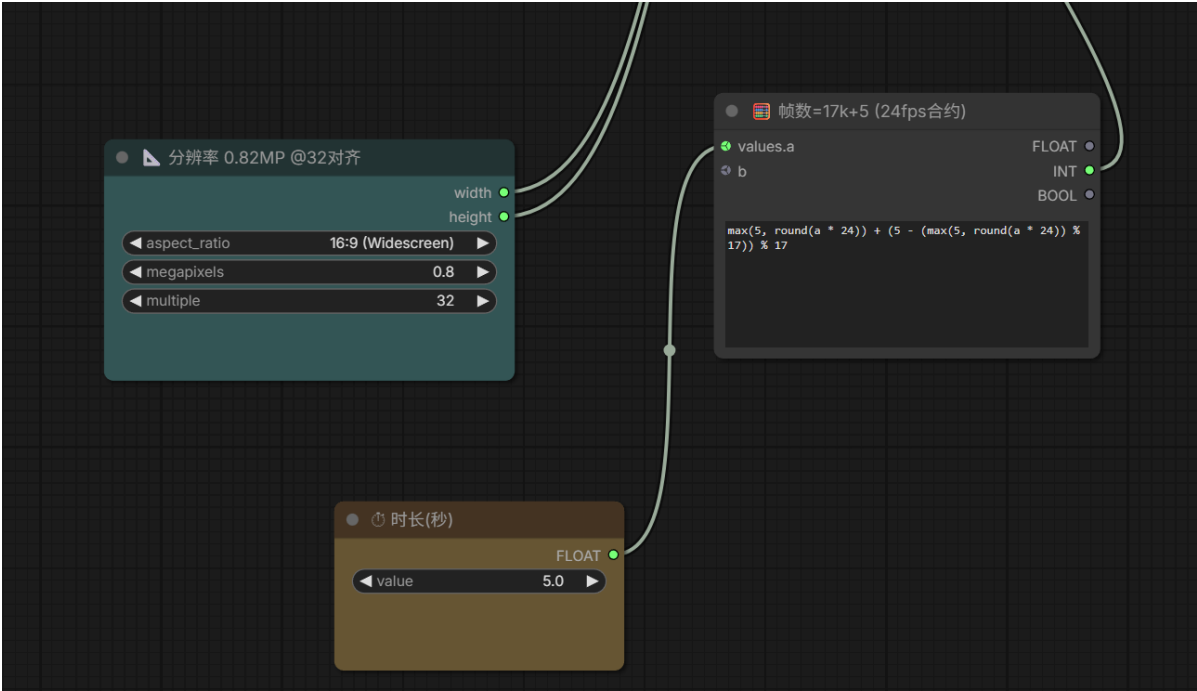
2.1 低显存工作流的关键

- Beta_3_Lite 使用纯 T8 骨架, DualClock steps=8。
- FP16 累计 + LowVRAM 注意力 (chunks=10) + 分块 FF (4/4096), 核心目的是降峰值显存, 而不是改变生成目标。
- 采样前、解码后、音频路各有 VRAM cleanup; 采样结束后模型可卸载, 显存不会一直占满。
- UNET 已内置 Turbo, 额外 Turbo LoRA 保持旁路。
- TDE 时序细节增强放在解码后, 作为零额外大模型的时序细化环节。



实际 ComfyUI 运行截图：设定图接 ref_image_0，连续镜头则把上一镜尾帧接到 first_frame。

2.2 分辨率与时长入口



工作中的分辨率、时长与帧数计算节点。最终导出实际为 1248x704。

aspect_ratio	16:9 (Widescreen)
megapixels	0.82（工作流选择器会对齐到实际尺寸）
multiple	32

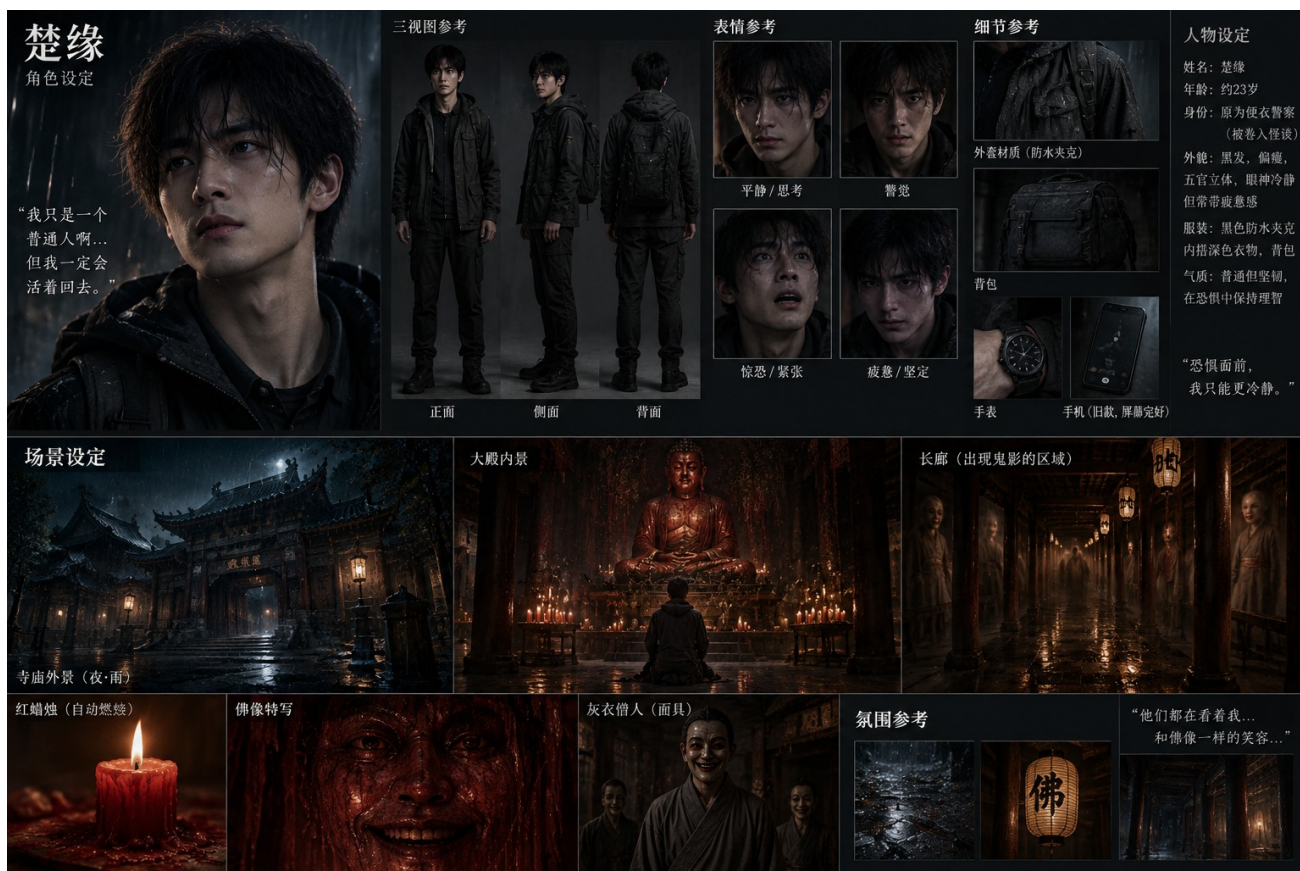
task_type	auto
audio_mode	native
audio_denoise_strength	0.35
参考图	add_source_as_reference = true; ref_image_size = match

安装时踩过的坑

SageAttention 最初报错，最终使用 sageattention 2.2.0 + cu130 对应 wheel，并安装 triton-windows 3.7.1.post27。
另一个关键点是 Qwen3-VL 的 CLIP Loader 类型：早期截图里曾误设为 yue2，正确值是 minimax。

3. 设定图：把人物、场景、道具一次性钉住

先做一张“制作设定板”，再让所有镜头共享它作为 ref_image_0。设定板同时承担人物身份、服装、寺庙视觉、血佛、长廊、鬼影和红烛道具的视觉锚点。这样每个镜头都不需要重新用文字从零定义世界。



实际使用的设定图：人物三视图、表情、服装、寺庙、血佛、长廊、鬼影与红烛。

3.1 设定图 Prompt 的职责

- 人物：楚缘约 23 岁、瘦削、黑色湿发、黑色防水夹克，保持写实真人感。
- 世界：夜雨寺庙、湿地反光、木柱、暖灯笼与冷蓝夜色并存。
- 核心恐怖物：暗红血佛，笑容克制但不舒服；灰衣僧/鬼影共享同一种笑容母题。
- 道具：红色仪式蜡烛。这里最后暴露出一个一致性问题：设定图是短粗胖蜡烛，但模型多次生成成细蜡烛。



设定图里的目标蜡烛：短、粗、宽圆柱



最终 Demo 镜头 1：蜡烛明显偏细，且状态也不完全符合最初设计

这次没有继续重跑的原因

这是 Demo，不是正式成片。为了修一根蜡烛继续烧十几二十分钟，收益已经低于成本。最终选择保留穿帮，并把它写进复盘：Prompt 约束不是万能的，关键道具如果必须严格一致，最好单独裁切成专用参考图。

4. 六镜头结构与连续性策略

镜头 1	约 0:00-0:06.58，独立生成；与镜头 2 直接 CUT。
镜头 2	约 0:06.58-0:13.17，独立生成；与镜头 3 直接 CUT。
镜头 3	约 0:13.17-0:19.75，独立起镜；结尾作为镜头 4 的 first_frame。
镜头 4	约 0:19.75-0:27.04，first_frame = 镜头 3 尾帧；结尾继续给镜头 5。
镜头 5	约 0:27.04-0:34.33，first_frame = 镜头 4 尾帧；结尾继续给镜头 6。
镜头 6	约 0:34.33-0:38.67，first_frame = 镜头 5 尾帧；情绪收尾。

连续性原则

切镜时不要为了“连续”硬接上一镜尾帧。镜头 1->2、2->3 本来就是机位变化，独立生成更自然；只有真正的连续动作段（3->4->5->6）才使用“上一镜尾帧 -> 下一镜 first_frame”。



镜头 3 尾帧



镜头 4 第一帧



镜头 4 尾帧



镜头 5 第一帧



镜头 5 尾帧



镜头 6 第一帧

剪辑时如果上一镜最后一帧与下一镜第一帧完全相同，要在 Premiere 里削掉下一镜开头的重复帧；如果 H3 在开头 hold 了几帧，就继续削到真正开始运动的位置。

5. 六个镜头的成片表现

镜头 1 | 建立



楚缘持烛进入仪式空间。实际成片里蜡烛已点燃，这是已知穿帮。

镜头 2 | 自燃



跪拜、放置红烛，暖光开始主导人物与血佛。

镜头 3 | 初现



长廊先出现少量、远距离、难辨身份的人形；不在这一镜揭脸。

镜头 4 | 百鬼回望



整段的核心 horror image：多层纵深的人影统一面向楚缘。

镜头 5 | 熄灭



暖光退出、鬼影消失，场景从“有人”突然变回空。

镜头 6 | 余悸



空走廊比鬼影本身更冷，人物慢慢恢复并准备撤离。

6. Prompt 设计方法

这次 Prompt 的核心不是堆“电影感”“8K”“大师镜头”之类形容词，而是把模型真正容易犯错的地方分区写清楚。生产 Prompt 基本都按下面这个顺序组织：

- Reference：明确设定图中哪一部分是人物、哪一部分是世界/道具参考，并要求忽略版式文字。
- Global constraints：锁身份、服装、环境、灯光逻辑和写实风格。
- Shot description：只写这一镜真正需要发生的动作，不跨镜头提前演。
- First/Last frame target：只在需要连续时写，告诉模型“从什么状态接、到什么状态停”。
- Performance constraints：恐怖片最容易过演，所以反复强调 restrained、no screaming、no exaggerated gestures。
- Ghost/prop constraints：对鬼影、蜡烛、血佛这种关键元素单独列红线。
- Camera constraints：慢推、慢移、禁止 whip pan / sudden zoom / shaky chaos。
- Audio：H3 只做 diegetic audio，明确 No background music；统一 BGM 放到 Premiere 再加。
- What to avoid：把最容易出现的错误再列一遍，尤其是文本、UI、身份漂移、怪物化、廉价透明鬼。

为什么 Prompt 很长？

因为这是多镜头连续制作，不是单条“好看就行”的视频。长 Prompt 的价值主要在一致性和边界控制。真正可以复用的部分应该提炼成模板；每一镜只改 Shot description / continuity / end state。

7. 完整 Prompt 附录

Prompt 来源说明

镜头 3、4、5 的文本直接从已生成 MP4 中保存的 ComfyUI workflow metadata 恢复，属于实际生产 Prompt。镜头 1、2 在 Demo 中出现蜡烛一致性问题，因此这里给的是复盘后的“推荐修正版”；镜头 6 则根据最终生产指令整理成可复用版。这样比把已经验证有问题的旧 Prompt 原样贴出来更适合作参考。

7.1 镜头 1 Prompt 推荐修正版

注意

最终 Demo 里镜头 1 的蜡烛仍然偏细且已经点燃；下面这版是针对该问题强化后的参考写法。

Use the provided reference image as the visual identity, prop-design, and world-design reference.

The male character shown in the portrait and three-view character sheet is Chu Yuan. Use those depictions as the identity reference. Ignore typography, labels, layout elements, and unrelated panels in the reference image.

[Critical prop reference]

Use the thick red ritual candle shown in the provided reference image as the exact visual reference for the candle prop. Match its short height, thick and wide cylindrical wax body, heavy visual weight, red color, and overall silhouette. Do not redesign it. Do not replace it with a thin taper candle, a long candle, or a generic slender candle.

[Global character constraints]

Maintain the same young Chinese male protagonist throughout the shot: Chu Yuan, about 23 years old, slim build, pale skin, black slightly messy short hair, tired but alert eyes, realistic facial features, wearing the same black waterproof jacket, dark inner shirt, and dark pants. Keep his face, hairstyle, body proportions, clothing silhouette, and overall identity consistent.

[World and style]

Dark ancient wooden temple at night, damp reflective floor, wooden pillars, cold blue-black ambient darkness, subtle mist, distant warm practical lantern light, realistic mainstream streaming-platform supernatural horror-drama quality. Cinematic live-action realism, natural human motion, realistic wet surfaces, shallow depth of field, controlled pacing.

[Shot description]

Interior, temple corridor, night. Chu Yuan stands in the dark corridor holding the exact thick red ritual candle from the reference image in his right hand near chest or waist height. The candle must be clearly visible and read as short, thick, wide, and heavy rather than slender. He faces toward the deeper Buddha hall in the distance.

The camera slowly pushes forward from a medium shot. Chu Yuan is cautious, quiet, and focused, as if something about the place is wrong. The corridor is empty. No apparitions are visible yet and no supernatural event has happened yet.

The candle must remain completely unlit from beginning to end. Do not show flame or glow from the candle. By the end, Chu Yuan still holds the unlit thick red candle while looking toward the Buddha hall, as though he is about to approach it.

[Performance]

Restrained, believable, slightly tense. No screaming, sudden panic, comedic acting, or exaggerated gestures.

[Camera]

Slow controlled push-in. No rapid cuts, whip pans, sudden zooms, or shaky action-camera movement. Preserve readable dark detail, wet reflections, natural lens behavior, and restrained low-saturation grading.

[Audio]

Diegetic audio only. No background music, soundtrack, or non-diegetic music. Faint distant rain, low wind through wooden architecture, occasional old wooden creaks, cloth movement, and quiet footsteps or body movement if appropriate. No dialogue.

[What to avoid]

No lit candle. No thin candle. No taper candle. No different candle design. No apparitions. No ghost reveal. No text, subtitles, UI, character identity drift, costume drift, or environment transformation.

7.2 镜头 2 Prompt 推荐修正版

Use the provided reference image as the visual identity, prop-design, and world-design reference. The male character shown in the portrait and three-view character sheet is Chu Yuan. Ignore typography, labels, layout elements, and unrelated panels.

[Critical prop continuity]

Use the thick red ritual candle shown in the reference image as the exact prop design. It is the same candle as shot 1: short, thick, wide cylindrical wax body, clearly red, heavy wax appearance, visibly thicker than a normal candle. Do not replace it with a thin taper candle or redesign it.

[Global constraints]

Maintain the same Chu Yuan identity, black waterproof jacket, dark clothing, body proportions, hairstyle, and realistic live-action appearance. Maintain the same ancient wooden Blood Buddha hall at night with wet reflective surfaces, cold blue-black ambience, warm practical light, subtle mist, and the same giant dark-red Blood Buddha statue.

[Shot description]

Interior, night, inside the Buddha hall. Chu Yuan is kneeling in front of the Blood Buddha altar in a medium-wide cinematic composition, seen from a three-quarter rear or side-rear angle.

At the beginning, he holds the exact same thick red ritual candle from the reference image. It is completely unlit. He slowly leans forward and carefully places it on the offering platform beneath the Blood Buddha.

For a brief beat, nothing happens. Then, without anyone touching it, this same thick candle ignites by itself. The flame starts small and stabilizes naturally. Warm red-orange candlelight gradually illuminates Chu Yuan's hand, sleeve, jawline, neck, and the lower portion of the Blood Buddha statue, while the rest of the hall remains cold and dark.

Chu Yuan's body subtly stiffens. He does not panic. He keeps his head lowered with only a small upward glance toward the statue. The camera slowly pushes in. By the end, the same thick red candle burns steadily on the altar, Chu Yuan remains kneeling, and the lower part of the Blood Buddha's face is faintly revealed. No apparitions appear yet.

[Performance]

Calm, restrained, psychologically grounded. No screaming, sudden standing, melodramatic gestures, or exaggerated shock.

[Blood Buddha]

Ancient, large, dark red, weathered, oppressive, motionless, with a subtle eerie smile. The statue does not move or transform.

[Camera]

Slow controlled push-in. No rapid cuts, shaky camera, whip pans, or sudden zooms. Keep cold ambient darkness plus warm candlelight coherent.

[Audio]

Diegetic audio only. No background music or soundtrack. Temple room tone, distant rain, low wind, wooden creaks, cloth movement, a delicate ignition sound, subtle continuous flame, and faint controlled breathing.

[What to avoid]

No thin candle. No already-lit candle at the start. No extra people. No apparitions. No text, subtitles, UI, identity drift, environment transformation, or exaggerated magical VFX.

7.3 镜头 3 Prompt MP4 metadata 原始生产 Prompt

Use the provided reference image as the visual identity and world-design reference. The male character shown in the upper-left portrait and the three-view character sheet is Chu Yuan. Use those depictions as the identity reference. Ignore all typography, labels, layout elements, and unrelated panels in the reference image.

[Global constraints]

Maintain the same young Chinese male protagonist throughout the entire shot: Chu Yuan, about 23 years old, slim build, pale skin, black slightly messy short hair, tired but alert eyes, realistic facial features, wearing the same black waterproof jacket, dark inner shirt, and dark pants. Keep his face, hairstyle, body proportions, clothing silhouette, and overall identity consistent throughout the shot.

Maintain the same temple world established previously: an ancient wooden Buddha hall at night, damp floor and wooden surfaces, cold blue-black ambient darkness, subtle mist, restrained red-orange candlelight, and the same giant dark-red Blood Buddha statue. The environment must remain grounded and physically believable, with mainstream streaming-platform supernatural horror-drama realism.

The red candle is already burning at the beginning of this shot. The previous supernatural event has already happened. Do not show the candle igniting again.

Do not introduce obvious fantasy visual effects, glowing magic, exaggerated fog, surreal distortion, anime aesthetics, game-like rendering, subtitles, interface elements, or collage artifacts.

[Shot description]

Interior, night. Chu Yuan remains kneeling in front of the Blood Buddha altar after the red candle has ignited.

Begin with a quiet medium shot from slightly behind and to the side of Chu Yuan. The small red-orange candle flame burns steadily near him, softly illuminating part of his sleeve, jawline, and the lower portion of the Blood Buddha statue.

The camera slowly and deliberately begins to shift its attention away from the altar toward the long wooden corridor beside the Buddha hall.

At first, the corridor appears empty.

After a brief moment, a faint human-shaped silhouette becomes barely visible between two distant wooden pillars.

It should not suddenly materialize. It should emerge so subtly from the darkness that the audience is initially unsure whether something is actually there.

As the camera continues its slow movement, another indistinct human silhouette becomes visible farther down the corridor.

Then one or two more vague figures can be perceived deeper in the darkness.

The figures remain mostly still.

Their bodies should resemble human monks or human-shaped apparitions standing quietly in the corridor, partially concealed by wooden pillars and deep shadow. They should not yet be clearly identifiable. Their faces must remain mostly obscured in this shot.

The supernatural presence should increase gradually rather than appearing all at once.

Chu Yuan does not turn around to look directly at them.

He remains kneeling, but his body subtly stiffens as he senses something wrong through his peripheral vision. His breathing becomes slightly more controlled and tense. His shoulders tighten almost imperceptibly.

He must not make exaggerated gestures, scream, stand up, or directly confront the figures.

By the end of the shot, the audience should clearly understand that several human-shaped figures are now standing silently in the dark corridor watching Chu Yuan, but their exact faces and nature remain unrevealed.

This shot should end before the full reveal.

[First frame target]

If a first frame is provided, preserve its exact character identity, costume, Blood Buddha design, candle state, temple architecture, lighting direction, and general spatial layout.

The red candle must already be burning.

Chu Yuan should already be kneeling in front of the altar.

The corridor should initially appear empty or almost empty.

Do not repeat the candle ignition from the previous shot.

[Last frame target]

The final composition should contain Chu Yuan in the foreground or middle ground, still kneeling near the burning red candle, while several faint human-shaped silhouettes are now visible far down the adjacent corridor.

At least three separate figures should be perceptible at different depths of the corridor, partially hidden by pillars and darkness. Their faces must not yet be clearly visible.

They should feel motionless, unnatural, and silently attentive.

Do not fully reveal the Blood Buddha smile on their faces yet. That reveal belongs to the next shot.

End with unresolved visual tension, as though the audience has just realized that Chu Yuan is no longer alone.

[Performance constraints]

Chu Yuan remains restrained, cautious, and believable.

No screaming.

No sudden head turn.

No large gestures.
No exaggerated facial expression.
No standing up.
No direct eye contact with the apparitions.
Maintain consistent face, body, clothing, and hairstyle throughout.

[Ghost constraints]
The apparitions must remain human-shaped and restrained.
No floating monster bodies.
No violent movement.
No crawling.
No attacking.
No grotesque transformations.
No obvious transparent CGI ghosts.
No full facial reveal yet.
No excessive number of figures.
They should feel like silent people who have somehow always been standing in the darkness.

[Camera and visual constraints]
Cinematic live-action realism.
Mainstream streaming-platform psychological and supernatural horror quality.
Use a slow, controlled camera drift or gentle lateral movement from Chu Yuan toward the corridor.
No rapid cuts.
No whip pans.
No sudden zoom.
No shaky action-camera movement.
Preserve spatial continuity between the altar, Chu Yuan, and the corridor.
Lighting remains motivated by the small warm candle flame and cold ambient night light.
Deep shadows should retain texture and detail rather than becoming crushed black areas.
Use restrained contrast, realistic wet wooden textures, realistic skin and fabric, subtle depth of field, natural lens behavior, and low saturation.

[Audio constraint]
Diegetic audio only.
No background music.
No soundtrack.
No non-diegetic music.

[Audio]
Maintain continuous night temple ambience: distant rain outside, faint wind passing through the wooden structure, very subtle room tone, occasional old wooden creaks, and the quiet burning sound of the red candle.
Chu Yuan's breathing is faint and controlled.
As the silhouettes gradually become visible, do not add a jump-scare sound.
Instead, allow the environmental sound to become slightly quieter and more spatially unsettling, with perhaps a distant soft wooden footstep or cloth movement whose source cannot be identified.
The apparitions themselves remain silent.

[What to avoid]
No candle ignition sequence.
No dialogue.
No background music.
No jump scare.
No attacking ghosts.
No clearly visible ghost faces yet.
No extra foreground characters.
No sudden lighting changes.
No character identity drift.
No costume changes.
No environment transformation.
No text.
No subtitles.
No UI.

7.4 镜头 4 Prompt MP4 metadata 原始生产 Prompt

Use the provided reference image as the visual identity and world-design reference. The male character shown in the upper-left portrait and the three-view character sheet is Chu Yuan. Use those depictions as the identity reference. Ignore all typography, labels, layout elements, and unrelated panels in the reference image.

[Global constraints]

Maintain the same young Chinese male protagonist throughout the entire shot: Chu Yuan, about 23 years old, slim build, pale skin, black slightly messy short hair, tired but alert eyes, realistic facial features, wearing the same black waterproof jacket, dark inner shirt, and dark pants. Keep his face, hairstyle, body proportions, clothing silhouette, and overall identity consistent from beginning to end.

Maintain the same temple world established previously: a dark ancient wooden Buddha hall and corridor at night, damp reflective floor, wooden pillars, lanterns, cold blue-black ambient darkness, subtle mist, and warm red-orange candlelight. The same Blood Buddha atmosphere must persist even if the statue is not always fully visible in frame. Keep the environment grounded, realistic, and suitable for a mainstream streaming-platform supernatural horror drama.

Preserve the continuity of the provided first frame. The shot must begin from that exact visual state: Chu Yuan is in the foreground/left side area, the corridor is already visible, and several human-shaped figures are already present in the distance. Use this starting image faithfully and continue the scene naturally from it.

No stylized anime look, no game-like rendering, no exaggerated fantasy magic, no surreal distortion, no collage artifacts, no subtitles, and no UI elements.

[Shot description]

Interior, night. This is the key reveal shot.

At the start of the shot, Chu Yuan remains in the foreground, tense and silent, still oriented toward the corridor through peripheral awareness rather than fully turning his body around. The corridor already contains several still human-shaped figures in the distance.

The camera slowly and deliberately pushes forward and slightly shifts laterally, revealing more of the corridor space and making the hidden figures clearer.

As the shot progresses, more human-shaped figures become visible at different depths of the corridor, partially emerging from behind the wooden pillars and dim lantern light. The number of figures should gradually increase, but in a restrained and believable way. They must not pop in abruptly.

These figures now become identifiable as monk-like apparitions or silent human-like presences standing motionless in the corridor.

Most importantly, as the reveal continues, it becomes clear that all of them are facing Chu Yuan.

Their faces should now gradually become readable: each face carries the same faintly eerie, unnatural smile associated with the Blood Buddha. The smiles should feel deeply unsettling, restrained, and inhumanly uniform rather than exaggerated or grotesque.

The audience should clearly understand that these figures are all watching Chu Yuan in silence.

Chu Yuan does not scream or jump up. Instead, his body becomes subtly more rigid. His breathing grows tighter. He remains restrained, frozen by dread, aware that something is horribly wrong. He must not make exaggerated movements.

By the end of the shot, the corridor contains multiple silent monk-like apparitions at different distances, all turned toward Chu Yuan, all with variations of the same Blood Buddha-like smile, creating a powerful sense of oppressive supernatural attention.

This is the main horror-image shot of the sequence. It should feel iconic, cinematic, and deeply unsettling.

[First frame continuity]

Preserve the starting composition provided by the first frame:

- Chu Yuan remains in the foreground / left side
- The candle-lit side table and temple lighting logic remain coherent
- The corridor layout, pillars, lanterns, and reflective floor remain spatially consistent
- The distant figures already present in the first frame must continue naturally into the shot

Do not restart the scene. Continue from this exact moment.

[End-state target]

By the end of the shot, several apparitions should be clearly visible throughout the corridor, standing at different depths and partially framed by pillars and lanterns.

At least some of their faces should now be readable.

They are all turned toward Chu Yuan.

Their expressions share the same eerie smile motif associated with the Blood Buddha.

The shot should end on a strong tableau of silent supernatural attention focused on Chu Yuan.

Do not resolve the threat yet.

Do not make them attack yet.

Do not make them disappear yet.

[Performance constraints]

Chu Yuan remains restrained, believable, and controlled.

No screaming.

No running.

No standing up suddenly.

No exaggerated gestures.

No dramatic head whipping.
 No melodramatic acting.
 He should communicate fear through stillness, breath, and subtle body tension.
 Maintain stable character identity, stable costume, stable hairstyle, and stable body proportions throughout.

[Ghost / apparition constraints]
 The apparitions should remain mostly still.
 No attacking.
 No rushing forward.
 No floating wildly.
 No distorted monster transformations.
 No excessive transparency.
 No cheap CGI ghost look.
 They should resemble monk-like human figures or silent temple presences standing in darkness.
 Their faces should now become more visible than in the previous shot, but still naturalistic within the horror style.
 The shared smile must be uncanny, restrained, and disturbing, not cartoonish.

[Camera and visual constraints]
 Cinematic live-action realism.
 Mainstream streaming-platform psychological supernatural horror quality.
 Use a slow controlled push-in and/or subtle lateral camera drift.
 No rapid cuts.
 No shaky camera chaos.
 No whip pans.
 No sudden zooms.
 Preserve strong spatial continuity and visual coherence between Chu Yuan, the corridor, the pillars, the lanterns, and the apparitions.
 Lighting remains motivated by practical sources: warm lanterns and candlelight contrasted with cold ambient night illumination. Deep shadows should retain texture and detail. Wet wooden and stone surfaces should have realistic reflective highlights. Keep colors restrained and cinematic, with low saturation and strong tonal control.

[Audio constraint]
 Diegetic audio only.
 No background music.
 No soundtrack.
 No non-diegetic music.

[Audio]
 Maintain continuous night temple ambience: distant rain outside, faint wind through wooden architecture, subtle room tone, occasional old wood creaks, and faint candle burning sound if still audible from the foreground area.
 No jump-scare sting.
 The apparitions remain silent.
 As the camera reveals that all the figures are watching Chu Yuan, the scene should feel more acoustically empty and tense. A faint cloth rustle or nearly inaudible foot shift may occur somewhere in the corridor, but no obvious attack sound.
 Chu Yuan's breathing is present, quiet, and increasingly tense.

[What to avoid]
 No attack in this shot.
 No jump scare.
 No abrupt ghost rush.
 No comedic expression.
 No exaggerated horror faces.
 No environment change.
 No identity drift.
 No costume drift.
 No extra foreground characters.
 No subtitles.
 No text.
 No UI.
 No magical VFX overload.

7.5 镜头 5 Prompt MP4 metadata 原始生产 Prompt

Use the provided reference image as the visual identity and world-design reference. The male character shown in the upper-left portrait and the three-view character sheet is Chu Yuan. Use those depictions as the identity reference. Ignore all typography, labels, layout elements, and unrelated panels in the reference image.

[Global constraints]

Maintain the same young Chinese male protagonist throughout the entire shot: Chu Yuan, about 23 years old, slim build, pale skin, black slightly messy short hair, tired but alert eyes, realistic facial features, wearing the same black waterproof jacket, dark inner shirt, and dark pants. Keep his face, hairstyle, body proportions, clothing silhouette, and overall identity consistent from beginning to end.

Maintain the same temple world established previously: a dark ancient wooden Buddha hall and corridor at night, damp reflective floor, wooden pillars, lanterns, cold blue-black ambient darkness, subtle mist, and warm candle-lit practical light. Preserve the same spatial continuity, horror-drama realism, and mainstream streaming-platform visual quality.

Preserve the continuity of the provided first frame. The shot must begin from that exact visual state: Chu Yuan is in the foreground, tense and frightened; several monk-like apparitions with eerie smiles are visible in the corridor; the temple space, lighting direction, and wet reflective surfaces remain coherent.

No anime stylization, no game-like rendering, no surreal fantasy distortion, no magical VFX overload, no UI, no text, and no collage artifacts.

[Shot description]

Interior, night. This is the resolution shot of the sequence.

At the beginning of the shot, Chu Yuan remains frozen in fear, facing the corridor from the foreground while several smiling monk-like apparitions stand at different depths in the hall and corridor.

For a brief moment, the oppressive tableau holds.

Then the candlelight that has been sustaining the warm illumination suddenly goes out.

The extinguishing of the candle must feel abrupt but natural, as though the ritual state has been broken.

As the candle goes out, the warm red-orange light disappears from the scene, leaving the environment noticeably colder, darker, and emptier.

At nearly the same moment, the apparitions vanish.

They should not run away or explode or transform. Instead, they should disappear in a restrained supernatural manner — either dropping out of visibility almost instantly with the loss of light, or seeming to no longer be there when the darkness settles.

The vanishing should feel eerie and silent, not flashy.

Chu Yuan reacts with a restrained physical shock: a small inhale, a stiffening of the shoulders, a sudden loss of certainty. He does not scream.

After the apparitions are gone, he remains in the now much darker temple space, breathing tensely.

He can either remain frozen for a beat, or begin a subtle motion of rising or preparing to leave, but do not turn this into a running escape. The ending should preserve dread rather than action.

By the end of the shot, the corridor is empty again. The supernatural presence has vanished, but the unease remains. Chu Yuan is left in darkness and silence, visibly affected by what he has just witnessed.

[First frame continuity]

Preserve the first-frame setup exactly:

- Chu Yuan in the foreground
- Corridor and pillars in the same layout
- Apparitions already visible at different depths
- Warm practical candle lighting and temple ambience still present at the start

Do not restart the reveal.

Continue directly from the end of the previous shot.

[End-state target]

By the final moments of the shot:

- The warm candlelight is gone
- The apparitions have disappeared
- The corridor is empty
- The environment is darker, colder, and more hollow
- Chu Yuan remains alone, tense, breathing, and visibly unsettled

The ending should feel like a haunted silence after a supernatural event, not like a victory or release.

[Performance constraints]

Chu Yuan remains believable, restrained, and psychologically grounded.

No screaming.

No overacting.

No melodramatic gestures.

No exaggerated collapse.

No sudden full-speed running.

Use subtle performance: tightening breath, frozen posture, a slight flinch, a small attempt to recover composure, or the beginning of a cautious movement to leave.

Maintain stable identity, stable costume, stable hairstyle, and stable body proportions.

[Ghost / apparition constraints]

The apparitions are already present at the start of the shot.

They must disappear during the shot.

Do not make them attack.

Do not make them move aggressively.

Do not let them morph into monsters.

Do not create flashy dissolves, horror CGI bursts, or smoke explosions.

Their disappearance should feel unnaturally quiet and deeply unsettling.

[Lighting constraints]

Lighting continuity is crucial.

At the start, preserve the mixed lighting established previously: warm candle/practical light combined with cold night ambience.

When the candle goes out, the warm component should drop sharply, leaving mostly cold blue-black night tones and very weak practical spill.

The lighting change must make the disappearance of the apparitions feel more terrifying.

Keep realistic shadow behavior, realistic contrast, wet reflections on the floor, and detailed dark textures.

[Camera and visual constraints]

Cinematic live-action realism.

Mainstream streaming-platform supernatural horror quality.

Use a slow, controlled camera hold or a very subtle continuation of movement from the previous shot.

No rapid cuts.

No shaky camera chaos.

No sudden zoom.

No whip pan.

This is a dread-and-release shot. Let the space breathe.

Keep the composition readable and elegant.

[Audio constraint]

Diegetic audio only.

No background music.

No soundtrack.

No non-diegetic music.

[Audio]

Maintain night temple ambience: distant rain outside, faint wind through wooden architecture, subtle room tone, and old wood creaks.

At the beginning, the scene still contains the quiet sound of candle flame and Chu Yuan's tense breathing.

When the candle goes out, include a small extinguishing sound and let the soundscape suddenly feel emptier.

The apparitions remain silent before vanishing.

After they disappear, emphasize the cold emptiness of the temple: faint rain, far wind, room tone, and Chu Yuan's controlled breathing.

No jump-scare sting.

[What to avoid]

No attack.

No chase.

No jump scare.

No dramatic monster transformation.

No exaggerated VFX.

No screaming performance.

No character identity drift.

No costume drift.

No environmental reset.

No subtitles.

No text.

No UI.

7.6 镜头 6 Prompt 最终生产指令整理版

Use the provided reference image as the visual identity and world-design reference for Chu Yuan and the temple. Preserve the provided first frame faithfully if one is supplied.

[Continuity]

This shot begins directly from the exact end state of shot 5. The apparitions are already gone. The corridor is empty. The candle has gone out. The warm ritual light has collapsed, leaving cold blue-black night ambience and weak practical spill. Chu Yuan is alone and visibly affected by what just happened.

[Shot description]

Interior, night. Hold on Chu Yuan in the foreground with the empty temple corridor behind him. He takes one tense, controlled breath and cautiously finishes rising or recovering from his half-risen posture. His eyes search the corridor, trying to confirm that the figures are truly gone.

He does not run. He does not shout. After a short beat, he makes a small backward step or begins a restrained turn as though preparing to withdraw from the hall. The camera remains slow and controlled, using a gentle hold or subtle drift rather than a dramatic move.

The empty corridor should now feel more disturbing than when it contained the apparitions. Wet floor reflections, old wooden pillars, distant lanterns, rain ambience, and negative space should carry the tension. Nothing supernatural reappears. No new threat is revealed.

End on lingering dread rather than relief: Chu Yuan is alive and the corridor is empty, but the audience should feel that the event has not truly ended.

[Performance]

Restrained, believable, exhausted tension. Fear is communicated through breath, eye movement, stiffness, and cautious recovery. No screaming, collapse, exaggerated gestures, or full-speed escape.

[Camera and lighting]

Cinematic live-action realism. Slow controlled hold or very subtle drift. Preserve cold post-extinguish lighting and readable shadow detail. No sudden zoom, whip pan, or shaky camera.

[Audio]

Diegetic audio only. No background music, soundtrack, or non-diegetic music. Distant rain, faint wind, wooden creaks, room tone, clothing movement, footsteps only if he shifts position, and controlled breathing. No jump-scare sting.

[What to avoid]

No ghosts reappearing. No new monsters. No attack. No magical effects. No warm candlelight returning. No text, subtitles, UI, identity drift, or environment transformation.

8. Premiere 后期与声音

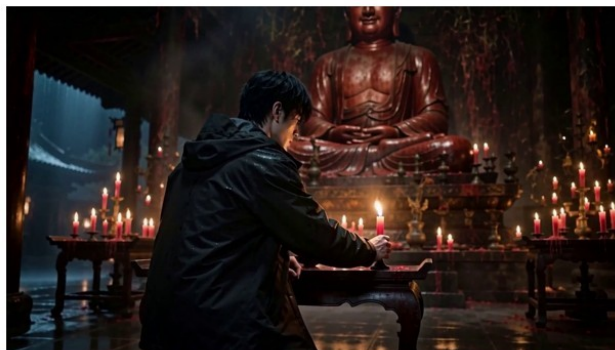
时间线	镜头 1 CUT 镜头 2 CUT 镜头 3 -> 镜头 4 -> 镜头 5 -> 镜头 6
分割快捷键	Ctrl + K（播放头位置切当前目标轨道）
画面处理	连续镜头边界去掉重复首帧；必要时削掉开头几帧静止 hold。
H3 原声	保留雨声、风、木头、呼吸、烛火、衣料等现场声。
全局 BGM	不要在每个 H3 Prompt 里生成 BGM；在 Premiere 上铺一条连续 BGM。
声音衔接	镜头边界对雨声/房间底噪做短交叉淡化，避免每个生成片段的环境声硬断。

8.1 声音情绪曲线

- 镜头 1：BGM 几乎听不见，让雨夜寺庙先成立。
- 镜头 2：红烛自燃后逐渐抬低频/仪式感。
- 镜头 3-4：鬼影出现到“百鬼回望”达到最高压迫。
- 镜头 5：蜡烛熄灭时把 BGM 的低频和密度迅速抽走。
- 镜头 6：只留尾音 + 雨声/空间底噪，让空走廊继续不舒服。



shot_01



shot_02



shot_03



shot_04



shot_05



shot_06

最终成片六镜头代表帧。

9. 失败点与复盘：哪些值得修，哪些 Demo 可以放过

关键道具漂移	设定图是短粗红烛，模型多次生成细烛。结论：关键道具需要单独参考图，单靠文字重复 “short/thick/fat” 仍可能不够。
状态穿帮	镜头 1 理论上应未点燃，最终 Demo 却已经有火焰。正式制作必须重跑；Demo 则可用剪辑弱化。
生成文字	灯笼/牌匾文字会出现无意义字符。除非文字是剧情信息，否则尽量放在背景、浅景深或暗部。
鬼影阶段感	镜头 3 不应太早把脸揭清；把“初现”和“完整 reveal”分成两镜，效果明显更稳。
连续镜头	不是所有相邻镜头都需要 first_frame。切镜就应该允许重新构图。
过度优化	一个 39 秒 Demo 的价值是验证流程。若某个小穿帮需要额外几十分钟，先看它是否真的会破坏叙事，再决定是否重跑。

最重要的经验

先把“完整可看”做出来，再决定哪里值得烧算力。生成阶段盯单帧时很刺眼的问题，放进完整叙事以后往往会被注意力重新分配。这个 Demo 里的蜡烛就是典型例子。

9.1 本机性能观察

- 最早一条约 5.17 秒 / 124 帧的样本，总耗时约 390.53 秒（6 分 30 秒）。
- 后续 6 秒左右镜头大约 9 分 15 秒 - 9 分 40 秒；7 秒镜头约 9 - 10 分钟。
- 采样阶段明显是最慢环节；采样结束后解码、TDE、视频合成相对很快。
- 8GB 显存能够跑的根本原因是低显存工作流 + RAM/offload，而不是把约 42GB 权重同时塞进显存。

10. 一页复现清单

- 准备一段 20-60 秒、视觉事件明确的原文，不要一上来做长剧情。
- 先拆成 5-7 秒的镜头；明确哪些是 CUT，哪些需要连续。
- 做一张统一设定图：人物三视图 + 表情 + 服装 + 场景 + 关键道具 + 核心怪物/气氛。
- 所有镜头把设定图接到 ref_image_0；不要为了连续把上一整段视频塞进 ref_video_0。
- 真正连续的镜头，用上一镜尾帧 PNG 接下一镜 first_frame。
- 16:9、约 0.82MP、multiple=32、24fps；低显存机器优先从 5-6 秒开始试。
- Prompt 分区写：Reference / Global / Shot / Continuity / Performance / Camera / Audio / Avoid。
- H3 只生成 diegetic audio；全局 BGM、连续雨声与最终混音放到 Premiere。
- 剪连续镜头时删掉重复首帧；每个 CUT 镜头允许重新构图。
- 正式成片前重点复查：人物身份、关键道具、灯光方向、鬼影数量、文字伪影、重复帧、音频硬断。

最终结果

本次 Demo 最终成片：38.67 秒，1248 x 704，24fps。最有价值的验证不是“某一帧多漂亮”，而是：本地 H3 在 8GB 显存机器上已经能够完成带人物一致性、参考图、原生音频、连续镜头和后期剪辑的一段完整叙事。

END | 建议把这套流程当作模板，而不是把参数当作永远不变的答案。