

音视频生成模型横评报告

1. 评测信息

1.1 评测背景与目标

这次测试聚焦图声视频生成：输入文本 Prompt 和参考图，让 5 个模型各自生成视频，再从参考图一致性、Prompt 一致性、音画同步性、画面质量、音频质量、运动质量、表演可信度七个维度打分。目的是看看这些模型在不同难度下的表现差异，重点对比它们对 Prompt、参考图、画面、运动、音频、音画同步及表演可信度的理解和还原能力。

1.2 评测信息概览

项目	内容
评测任务	真人剧场景音视频模型横评
评测模型	即梦 seedance2.0、谷歌 omni flash、可灵 Kling-V3-Omni、快乐马 happyhorse1.0、万相 wan2.7
数据规模	24 条 Prompt（低/中/高三档各 8 条），最终评分表 120 条模型记录
评分方式	1-5 分制，逐条逐维度打分后取平均
评分范围	覆盖低/中/高三档，每档 8 条，5 个模型逐条评

1.3 数据集设计

难度	特点	考察重点
低难度	无对白	画面、参考图、运动、配乐、环境声与整体氛围
中难度	日常对话	台词内容、音画同步、自然表达和音频稳定性
高难度	多对白、复杂调度或复杂情绪	多角色稳定、音色稳定、动作音效、口型同步和情绪层次

每档 8 条 Prompt，每条 Prompt 下 5 个模型视频，共 120 条模型级评分记录。

1.4 评分流程

- 逐条逐维度打分 → 复核确认 → 汇总统计 → 输出报告
- 最终评分表涵盖 120 条模型记录，记录模型、数据 ID、难度和视频分数
- 复核：分数偏极端或拿不准的样本会反复回看，重点确认音画同步性、音频质量（类型适配）、表演可信度等主观维度

2. 评测标准

2.1 评分维度

维度	核心含义	细分维度
参考图一致性	人物、服装、道具、场景与参考图是否稳定对应	首帧参考图一致；时序一致性（主体前后一致性）；风格前后一致性
Prompt 一致性	主体、场景、动作、镜头、风格和负面要求是否按 Prompt 执行	动作指令；场景描述；镜头描述
音画同步性	口型、动作音效、环境声与画面事件是否同步	口型同步；对白时序同步；动作音效同步；环境声源同步；配乐节奏同步
画面质量	清晰度、稳定性、真实感、崩坏/闪烁/AI 瑕疵控制	基础质量；细节丰富度；材质逼真度；光影协调度
音频质量	清晰度、噪声、爆音、混响、断裂、音量平衡；对白、环境声、动作音效、配乐类型是否符合 Prompt	人声清晰度；噪声与失真控制；响度平衡；声场空间合理性；音频类型适配度
运动质量	动作链、手物交互、多人调度、镜头运动和物理效果是否自然	结构准确性；运动合理性；叙事流畅性（因果链、情绪链、动作链、镜头链）
表演可信度	表情、肢体、情绪、对白等表演是否自然可信	微表情自然度；眼神交互合理性；肢体表演自然度；情绪表达准确性；情绪变化连贯性；对白口吻适配度；多人对戏反应真实性

3. 评测结果

3.1 整体排名

排名	模型	低难度	中难度	高难度	平均分
----	----	-----	-----	-----	-----

1	即梦 seedance2.0	4.12	4.15	4.57	4.28
2	谷歌 omni flash	4.29	3.53	3.77	3.86
3	可灵 Kling-V3-Omni	3.95	3.76	3.50	3.74
4	快乐马 happyhorse1.0	3.94	3.71	3.41	3.69
5	万相 wan2.7	4.05	3.50	3.16	3.57

即梦 seedance2.0 以 4.28 分位列平均分第一；万相 wan2.7 平均分最低。

3.2 分难度表现

难度	第1名	第2名	第3名	第4名	第5名
低难度	谷歌 4.29	即梦 4.12	万相 4.05	可灵 3.95	快乐马 3.94
中难度	即梦 4.15	可灵 3.76	快乐马 3.71	谷歌 3.53	万相 3.50
高难度	即梦 4.57	谷歌 3.77	可灵 3.50	快乐马 3.41	万相 3.16

几点观察：

- 低难度最优：谷歌 omni flash，无对白视觉展示和环境声氛围突出
- 中难度最优：即梦 seedance2.0，与第二名拉开约 0.39 分
- 高难度最优：即梦 seedance2.0，得分 4.57，在高难度下依然表现稳定
- 谷歌波动最大：低难度第1 → 中难度第4 → 高难度第2
- 万相难度敏感最强：低难度第3 → 中/高难度第5
- 即梦随难度上升仍保持高分，万相随难度上升降幅明显

4. 评测结论

4.1 各模型表现一览

即梦 seedance2.0（平均分第1，4.28 分）

维度	分析
强项	

	音频质量（音色一致性 5.00、综合音质 4.83，人声清晰稳定）；音画同步性 5.00（口型/对白时序/动作音效同步俱佳）；复杂多角色调度下稳定性最好
短板	音频质量·音频类型适配度偏弱（原音频内容准确性 3.67）；Prompt 一致性 3.75（动作指令/场景描述执行有偏差）；运动质量 4.00（结构准确但叙事流畅性一般）
推荐场景	中高难度、多对白、多角色、音画同步要求高的视频
不推荐场景	仅追求低难度视觉展示且需控制成本时

谷歌 omni flash（平均分第2，3.86 分）

维度	分析
强项	音频质量（音色一致性 4.25、综合音质 4.63，低难度氛围突出）；音画同步性 4.81（口型/配乐节奏同步好）；低难度表现第1
短板	参考图一致性 2.84（首帧/风格对应弱）；音频质量·音频类型适配度 3.46；运动质量 3.50（结构准确性与运动合理性偏弱）；中难度明显下滑
推荐场景	低难度、无对白、科幻/视觉展示类任务
不推荐场景	中等复杂对白、强参考图稳定性任务

可灵 Kling-V3-Omni（平均分第3，3.74 分）

维度	分析
强项	音频质量（音色一致性 5.00、综合音质 4.59）；音画同步性 4.25
短板	运动质量 3.04（结构准确性与运动合理性均弱）；Prompt 一致性 3.17（动作指令/场景描述执行偏差）；音频质量·音频类型适配度 3.29
推荐场景	中等复杂度、音频稳定性要求适中的任务
不推荐场景	高难度动作链和多角色调度任务

快乐马 happyhorse1.0（平均分第4，3.69 分）

维度	分析
强项	音频质量（音色一致性 4.88、综合音质 4.63）；音画同步性 3.94

短板	Prompt 一致性 3.00（动作指令执行差）；运动质量 3.25（结构准确性弱）；表演可信度 3.30（原自然度与情感表达，情绪表达准确性/肢体表演自然度不足）
推荐场景	低到中等复杂度、基础可用性要求的任务
不推荐场景	高难度多对白、多角色、高同步要求任务

万相 wan2.7（平均分第5，3.57 分）

维度	分析
强项	音频质量（音色一致性 4.62、综合音质 4.42）；音画同步性 3.81
短板	运动质量 2.54（结构准确性最弱）；画面质量 2.92（基础质量/材质逼真度不足）；Prompt 一致性 3.21；随难度上升降幅最大
推荐场景	低难度、简单场景、参考图要求明确的任务
不推荐场景	多角色、多对白、复杂音画同步任务

4.2 典型问题

问题类型	表现
指令理解失败	高难度下弱势模型 Prompt 一致性下降，复杂主体、动作链和多角色关系易失控
参考图偏离	中高难度多角色场景中，人物身份、服装、构图和主体稳定性易低分
画面和运动问题	打斗、递物、采访、多人走位等手物交互和调度场景，运动质量低分主要来源
音频类型适配度错误	有对白任务中，台词缺失、动作音效缺失、环境声类型不匹配是主要 badcase 类型
音频质量/音画同步问题	高难度多角色对白中更明显，影响音色一致性和音画同步性
情绪表达不足	中高难度要求克制、紧张、动摇、纪实等细腻情绪时，弱势模型易出现机械或错位表达

4.3 总体排名

- 平均分最优：即梦 seedance2.0，三档平均分 4.28
- 低难度最优：谷歌 omni flash，适合无对白视觉展示和氛围类任务
- 高难度表现最好：即梦 seedance2.0，在复杂音画同步和多角色任务中优势明显

- **波动最大**：谷歌 omni flash，低难度强、中难度明显下滑、高难度回升
- **对难度变化最敏感**：万相 wan2.7，从低难度第3 降至中/高难度第5
- **中间梯队差距小**：可灵与快乐马分差较小，需更多样本确认稳定差异

5. 具体 case

Goodcase

数据 ID	难度	模型	亮点维度	Prompt 摘要	结论
古装6	高	即梦	参考图一致性 5、画面质量 5、运动质量 5、音频质量（类型适配） 5	真人古装动作剧，阴天宗门山门前，多人场景	高难度动作链、音频质量（类型适配）、音画同步性多项满分，代表即梦在复杂多人场景中的强项
科幻1	低	谷歌	Prompt 一致性 5、参考图一致性 5、画面质量 5、运动质量 5	真人科幻电影，年轻女研究员在未来实验舱内研究能量体	低难度视觉展示场景均分最高，谷歌无对白科幻展示任务的正向样例
科幻5	高	即梦	参考图一致性 5、画面质量 5、音频质量（类型适配） 5、音频质量（综合音质） 5	时间穿越实验失控，环形装置高速运转，受伤研究员与年轻女性的互动	高难度科幻特效与音频质量维度稳定，说明即梦在高难度下依然稳定

Badcase

数据 ID	难度	模型	低分维度	Prompt 摘要	结论
古装4	中	谷歌	Prompt 一致性 1、画面质量 1、运动质量 1、表演可信度 1	真人古装剧，酒楼雅座，年轻公子与夫人的对话	多维标为 1 分，集中暴露 Prompt 一致性、画面质量、运动质量和表演可信度问题，反映谷歌中难度下滑
都市5	高	万相	Prompt 一致性 1、参考图一致性 1、运动质量 1、表演可信度 1	都市场景，昏暗出租屋内，警探、嫌疑人与证人三角对峙	参考图一致性 1 分，复杂多角色都市场景中参考图稳定性严重不足
年代6	高	万相			

			音频质量（音色）1、 Prompt 一致性 2、画面质量 2、运动质量 2	九十年代老工厂 车间，工人围听 改革消息，中年 女工紧攥工牌	音频质量（综合音质）2、 音频质量（音色）1、音画 同步性 2，代表万相在高难度多角色音频稳定性上的核心短板
古装6	高	谷歌	Prompt 一致性 1、画面质量 2、运动质量 2	与即梦 goodcase 同一 Prompt	同 Prompt 下与即梦形成鲜明对比，Prompt 一致性/画面质量/运动质量/音频质量（类型适配）均受拖累
科幻4	中	万相	画面质量 1、音频质量（类型适配） 2、Prompt 一致性 3	科幻对白场景， 男研究员与女科学家的情绪碰撞	中难度对白任务中表演可信度均低，体现万相在复杂表达中的不稳定

目评测集

(注：内容由 AI 生成，请谨慎参考)