德邦证券
Topsperity Securities

证券研究报告 | 行业专题
计算机
行业投资评级 | 优于大市（维持）
2024年8月20日

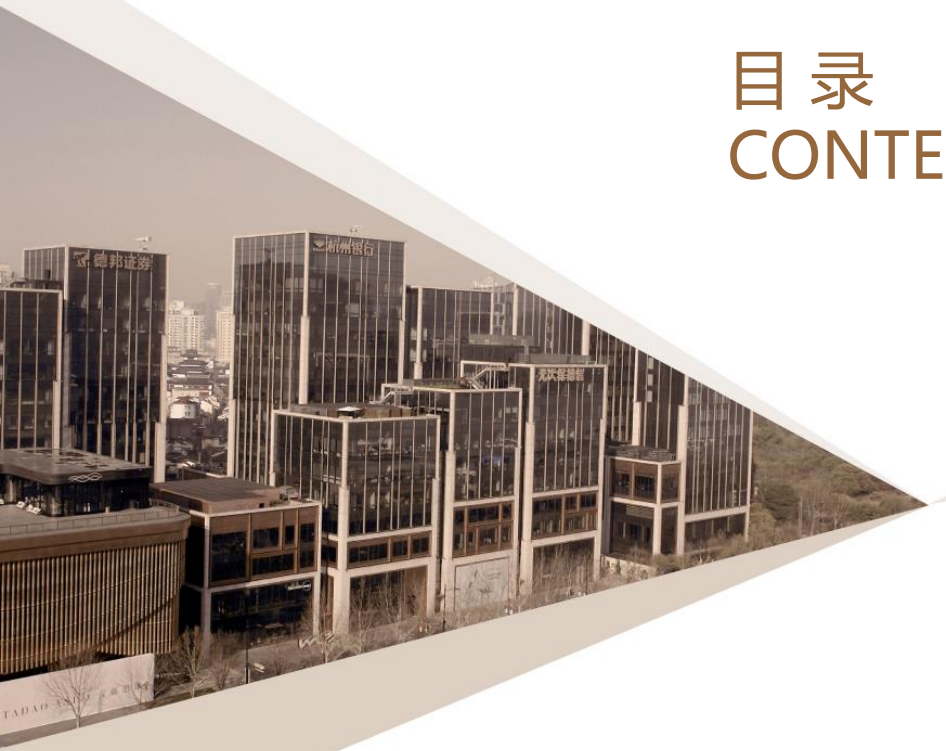
视频大模型奇点时刻加速到来

证券分析师
姓名：陈涵泊
资格编号：S0120524040004
邮箱：chenhb3@tebon.com.cn

研究助理
姓名：王思
邮箱：wangsi@tebon.com.cn

核心逻辑

- **视频大模型迎来Sora时刻，生产力工具蓄势待发。自Sora发布以来，国内外已有十多家公司发布或更新视频生成模型。**
  - ✓ 客观对比：与Sora差距缩小，抢占用户或为未来主线。国内外之间的差距正在逐步缩小，视频时长、分辨率等基础功能具有复制性，未来竞争或向抢占用户、提升粘性等方向迁移，从而需要保障生成质量更高的可用视频，使得视频一致性更高、文本指令遵循更准确、物理真实模拟能力更强。
  - ✓ 主观对比：视频质量提升显著，离物理世界模拟器仍有距离。文生视频领域，视频画面普遍清晰，而在动作幅度与物理还原度方面差异较大，我国Vidu、清影或已处于视频生成大模型第一梯队，生成时间相对其他模型倍数减少，且在动作幅度、物理还原度等复杂任务完成性较好；图生视频领域，整体效果优于文生视频，国内与国外差距进一步缩小。
  - ✓ 视频大模型具备商用潜力，下游应用正在储能。未来随着用户习惯的培育以及算力基础设施的完善，视频大模型的或者办公、广告、电影、游戏等多领域具有广阔的前景。
- **算法、算力、数据三要素，视频大模型通往AGI的必经之路。**
  - ✓ 算法：视频生成模型算法主要由基于SD逐帧生成和基于时空Patches两种范式构成，是效率与效益的比拼。前者模型更容易训练，而视频内容一致性要差一些，长视频训练存在困难；后者训练成本更高，但是生成视频的长度与一致性更容易得到保障。
  - ✓ 算力：以Sora为例，一定条件下测算，Sora训练算力需求是GPT-4的4.5倍，而推理算力需求接近GPT-4的400倍。
  - ✓ 数据：高质量数据是模型能力的保障，而用户数量或为开启模型迭代“数据飞轮”的关键。
- **投资建议：**建议关注（1）视频大模型厂商：科大讯飞、商汤、云从科技、格灵深瞳、拓尔思、昆仑万维等。（2）算力：海光信息、寒武纪、景嘉微、中科曙光、浪潮信息、工业富联、神州数码、拓维信息、四川长虹等。（3）接入大模型的应用标的：金山办公、万兴科技、福昕软件、虹软科技、彩讯股份、焦点科技、润达医疗、金证股份、泛微网络、金蝶国际等。
- **风险提示：**商业化落地不及预期；国内大模型在缺乏算力支持的情况下迭代速度放缓；国内大模型技术路线产生分歧。



目 录
CONTENTS

01 视频大模型迎来Sora时刻，生产力工具蓄势待发

1.1 国内外视频生成模型纷至沓来

- 根据APPSO微信公众号，自Sora发布以来，国内外已有**不少于十家公司**发布或更新视频生成模型。以7月为例：
- ✓ 7月31日，Runway宣布**Gen-3**可支持图生视频。用户可以使用任何图片作为视频生成的首帧，上传的图片既可以单独使用，也可以使用文本提示进行额外指导。
- ✓ 7月26日，**智谱AI版Sora清影发布**，人人可用、半分钟快速创作视频的时代已来。清影理论上仅需30秒即可完成6秒1440×960清晰度高精度视频的生成，展现出亮眼的推理速度，不仅具备高效的指令遵循能力，还具有内容的连贯性和调度灵活性。
- ✓ 7月24日，**爱诗科技将视频生成模型更新至PixVerse V2**，支持一键生成至多5段连续的视频内容，且片段之间会自动保持主体形象、画面风格和场景元素的一致性，视频效果再次提升。
- ✓ 7月17日，英国AI创企Haiper AI宣布**Haiper升级至v1.5**，时长延长到8秒，且提供视频延长、画质增强等功能。
- ✓ 7月6日，**智象未来在WAIC上发布了智象大模型2.0**，提供5、10、15秒三种视频生成时长，并增加文字嵌入生成、剧本多镜头视频生成、IP连贯一致性等能力。此外，智象支持视频增强至4K画质。

图表：清影AI视频效果展示

图表：PixVerse V2一次性生成多个一致性视频

图表：智象大模型2.0可增强生成4K画质视频

资料来源：智象未来官网，爱诗科技 Alsphere微信公众号，智谱清言官网等，德邦研究所

1.2 客观对比：与Sora差距缩小，抢占用户或为未来主线

- 目前，国内外大模型在视频时长、分辨率、画面比例切换等功能或性能指标均与Sora缩小差距，在部分功能已有赶超。
  - ✓ **生成类型：**主流大模型大多具有文生视频、图生视频功能。国外Sora、Haiper v1.5同时具备视频生视频功能，而我国清影视频生视频功能仅在demo中展示，尚未向用户开放。
  - ✓ **视频时长：**在Sora鲶鱼带动下，主流大模型视频时长大都达到5-10s级别，我国部分大模型在时长上处于第一梯队。例如，Vidu、Pixverse V2、可灵、Vimi等模型可通过视频延展等方式达到30-120s，进一步缩小与Sora差距，甚至实现赶超。
  - ✓ **分辨率：**之前的产品分辨率大多在1024*576左右，目前主流大模型以标清720p和高清1080p为主，我国Etna与智象大模型2.0可通过画质增强达到4K级别，赶超Sora的1080p。
  - ✓ **帧率：**部分厂商未公布帧率数据，目前以24或30fps为主，而我国的Etna的60fps处于领先地位，此前的产品帧率多为8-12fps。
  - ✓ **其他功能：**大部分模型已具备镜头运动、风格切换、画面比例切换等功能，提升视频生成质量与灵活性，国内Vidu和清影还可支持为视频配乐，生成视频更具想象力。
  - ✓ **商业模式：**目前大部分厂商采取免费方式吸引客户，而利用订阅模式开放更多功能等方式增强用户粘性。
- 综合来看，**我们认为国内外视频生成大模型之间的差距正在逐步缩小，视频时长、分辨率等基础功能具有复制性，未来竞争或向抢占用户、提升粘性等方向迁移，从而需要保障生成质量更高的可用视频，使得视频一致性更高、文本指令遵循更准确、物理真实模拟能力更强。**

1.2 客观对比：与Sora差距缩小，抢占用户或为未来主线

图表：国内外主流视频生成大模型性能对比

	公司	产品/模型	推出时间	生成类型	生成时长	分辨率	帧率	其他功能	价格	是否可用
国外	Open AI	Sora	2月16日	文生视频 图生视频 视频生视频	60s	1920*1080	-	比例切换，时长可延展、镜头运动、真实世界模拟、世界交互等	-	否
国外	Stability AI	Stable Video	2月21日	文生视频 图生视频	4s	1024*576	24fps	比例切换、风格选择、镜头移动	免费可用，积分付费：500积分/10美元、3000积分/50美元	是
国外	Luma AI	Dream Machine	6月13日	文生视频 图生视频	5s(可延长至10s)	1360*752	24fps	视频延长	免费可用，月付费版：23.99/51.99/79.99/399.99美元	是
国外	Runway	Gen 3	6月17日	文生视频 图生视频	5/10s	1280*720	-	镜头运动、比例切换、风格选择、导演模式	15美元/月、144美元/月	是
国外	Haiper AI	Haiper v1.5	7月17日	文生视频 图生视频 视频生视频	2/4/8s	1280*720(可增强至1080p)	24fps	比例切换、视频延长、画质增强	免费可用，月付费版：10/30，年付费8折	是
国内	七火山科技	Etna	3月7日	文生视频	8-15s	最高3840*2160	60fps	-	-	否
国内	生数科技	Vidu	4月27日	文生视频 图生视频	4/8s（理论32s）	1920*1080	-	风格切换，支持配乐	免费可用，月付费版：9.99/29.99/99.99元，年付费8折	是
国内	字节	即梦	5月9日	文生视频 图生视频	3/6/9/12s	1280*720	8fps	镜头移动、比例切换、视频延长、补帧、对口型、画质增强、运动速度	免费可用，年付费版：659/1899/5199元	是
国内	快手	可灵	6月6日	文生视频 图生视频	5/10s（理论120 s）	1280*720	30fps	比例切换	免费可用，月付费版：66/266/666元	是
国内	商汤	Vimi	7月4日	图生视频	60s	-	-	可控人物、多种方式控制、风格切换	-	是
国内	智象未来	智象大模型2.0	7月6日	文生视频 图生视频	5/10/15s（商业化分钟级）	1024*576（可增强至4K）	24fps	比例切换、反向提示词、镜头运动、4K增强	月付费版：9.9/39.9/129.9/389.9元	是
国内	爱诗科技	Pixverse V2	7月24日	文生视频 图生视频	5/8s（可延展5倍）	1920*1080	-	视频延长（一键生成至多5段连续的视频内容）、镜头运动	免费可用，月付费版：4/24/48美元	是
国内	智谱AI	清影	7月26日	文生视频 图生视频 视频生视频未开放	6s	1440*960	-	镜头移动、风格选择、支持配乐、情感氛围选择	免费可用，付费版：5元/天，199/年	是

资料来源：各公司官网，上海科技微信公众号，智东西微信公众号，APPSO微信公众号等，德邦研究所

1.2 客观对比：与Sora差距缩小，抢占用户或为未来主线

- 2024年7月31日，中文专用的多层次文生视频基准测评AIGVBench-T2V发布更新。
- ✓ **Gen-3在综合得分和多项指标中表现最佳，智谱清影和快手可灵紧随其后，位于视频生成大模型第一梯队。**
- ✓ **国内模型在高难度任务中表现强劲。**国内模型如智谱华章的智谱清影（75.24）、爱诗科技的PixVerse V2（75.29）、字节跳动的Dreamina即梦（75.80）在高难度任务中表现优异，分别位列国内第三、第二和第一。此外，快手的可灵网页版（73.13）和可灵（70.98）也表现不俗，进入国内前五。这显示了国内模型在高难度任务处理上的强劲实力和竞争力。

图表：SuperCLUE中文专用的多层次文生视频基准测评AIGVBench-T2V测评结果

模型名称	所属机构	综合得分	视频感官质量	文本指令遵循能力	物理真实模拟能力	高难度任务分数	测评时间
Gen-3	Runway	**79.2**	79.03	**87.08**	**71.5**	**80.92**	8月1日
智谱清影	智谱华章	**75.08**	71.19	**92.79**	61.76	75.24	8月1日
可灵网页版	快手	**75.02**	73.04	**89.75**	62.28	73.13	8月1日
PixVerse V2	爱诗科技	**73.32**	74.36	86.06	59.55	75.29	8月1日
即梦	字节跳动	**72.99**	**80.31**	78.81	60.21	75.8	7月2日
可灵	快手	**71.89**	77.77	71.63	**66.25**	70.98	7月2日
Luma	Luma AI	**70.89**	75.16	68.75	**69.45**	69.97	7月2日
PixVerse	爱诗科技	**70.18**	**82.55**	69.87	58.1	70.64	7月2日
WHEE	美图	**66.92**	**82.7**	64.32	53.94	66.04	7月2日
Pixeling	智象未来	**66.04**	71.22	70.52	56.34	68.19	7月2日
Pika Art	Pika	**63.95**	71.75	63.16	56.33	63.95	7月2日
星火绘镜	科大讯飞	**61.55**	72.07	57.8	56.73	61.55	7月2日
Gen-2	Runway	**58**	65.33	56.28	51.56	58	7月2日
Vega AI	右脑科技	**57.22**	68.57	49.38	53.07	57.22	7月2日

资料来源：SuperCLUE，CLUE中文语言理解测评基准微信公众号，德邦研究所

1.3.1 文生视频：画面普遍清晰，动作幅度物理还原差异大

德邦证券 Topsperity Securities

➢ 我们对主流视频生成大模型在相同prompt下进行测试。

图表：Sora视频结果（20s）

图表：Gen-3视频结果（10s）

图表：Dream Machine视频结果（5s）

图表：Haiper v1.5视频结果（4s）

图表：Vidu视频结果（4s）

图表：可灵视频结果（5s）

图表：Pixverse V2视频结果（9s）

图表：清影视频结果（6s）

注：中文prompt：“镜头跟随一辆带有黑色车顶行李架的白色老式SUV，它在陡峭的山坡上一条被松树环绕的陡峭土路上加速行驶，轮胎扬起灰尘，阳光照射在SUV上行驶土路，给整个场景投射出温暖的光芒。土路缓缓地蜿蜒延伸至远方，看不到其他汽车或车辆。道路两旁都是红杉树，零星散落着一片片绿意。从后面看，这辆车轻松地沿着曲线行驶，看起来就像是在崎岖的地形上行驶。土路周围是陡峭的丘陵和山脉，上面是清澈的蓝天和缕缕云彩。”英文prompt：“The camera follows behind a white vintage SUV with a black roof rack as it speeds up a steep dirt road surrounded by pine trees on a steep mountain slope, dust kicks up from it’s tires, the sunlight shines on the SUV as it speeds along the dirt road, casting a warm glow over the scene. The dirt road curves gently into the distance, with no other cars or vehicles in sight. The trees on either side of the road are redwoods, with patches of greenery scattered throughout. The car is seen from the rear following the curve with ease, making it seem as if it is on a rugged drive through the rugged terrain. The dirt road itself is surrounded by steep hills and mountains, with a clear blue sky above with wispy clouds.”

资料来源：各公司官网，德邦研究所整理（注：除Sora外，以上测试结果均为我们在各公司官网的测评，选用第一次结果，未进行抽卡；国内模型采用中文prompt，国外模型采用英文prompt，其中PixVerse V2由于主要面对海外客户，也采用英文）

1.3.1 文生视频：画面普遍清晰，动作幅度物理还原差异大

- 纵向来看，Sora发布之后几个月，国内外视频大模型生成效果提升显著，表现为：1）得益于分辨率的提升，视频画面**清晰度普遍提升**；2）模型**语义理解和一致性表现较好**，视频能较好地理解prompt内容，并能记住画面中出现的内容，前后保持连贯与一致。然而，生成视频依然存在不足：1）实测时间差异较大，**存在排队行为从而导致等待时间较长**，Dream Machine生成时间接近半小时，影响用户体验；2）动作幅度依然存在提升空间，生成的视频的策略偏向于小幅度运动，复杂动作较少，从而导致视频的稳定性和流畅度可能存在问题；3）生成视频部分未能很好还原实际物理世界，运动状态、光影灰尘等效果表现不佳。
- 横向来看，我国**Vidu、清影或已处于视频生成大模型第一梯队**，生成时间相对其他模型倍数减少，且在动作幅度、物理还原度等复杂任务完成性上与Sora的差距进一步缩小。

图表：主流视频大模型文生视频评价

	说明	Sora	Gen-3	Dream Machine	Haiper v1.5	Vidu	可灵	Pixverse V2	清影
生成时间	输入提示后得到结果的时间	-	1m54s	26m	3m37s	30s	5m14s	6m	1m25s
分辨率	清晰与否	高	高	中	高	高	高	中	高
可控性	运镜幅度	高	低	高	低	高	低	高	高
动作幅度	视频画面前后变化大小、动作幅度大小	高	低	高	低	高	低	高	中
语意理解	提示词是否能被完整准确地理解和表达	高	中	高	中	高	中	中	高
一致性	视频内容的前后连贯性和一致性，场景转换是否平滑，各元素之间是否协调统一	高	中	中	低	高	中	中	高
稳定性	是否会出现画面变形、撕裂或其他异常现象	高	中	低	低	高	中	低	高
流畅度	动作是否自然连贯，整体观感是否顺畅	高	低	高	低	高	低	低	高
物理还原度	光影效果、液体流动是否自然，交互行为是否符合物理规律	高	中	低	低	高	中	低	中
综合评价	-	整体对语义及物理现实还原度高，视频在较大幅度变化依然可以保持画面一致性与流畅性。	生成速度较快、画质较高，然而整体动作幅度和流畅度较低。	整体动作幅度较大、流畅度较高，然而生成时间较长，并且存在画面变形、非物理现实现象出现。	整体动作幅度较小、流畅度稳定性较低，存在变形与非物理画面。	生成速度最快，在较短的视频长度下依然可以达到Sora级别的动作幅度与流畅度，语义理解和物理还原度均较高。	优点是画面质量较高；缺点是生成时间较长、动作幅度较小、流畅度欠缺等。	优点是动作幅度大，视频前后画面差异大，但在一致性物理还原度等方面存在不足。	整体语言理解较强，画面一致性和稳定性高，然而动作幅度适中，存在一些非物理现象。

注：以上是根据前文prompt生成结果的对比，不同prompt结果或有差异。

资料来源：各公司官网，德邦研究所整理（注：在不同时间点调用模型用户拥挤度存在差异，以上生成时间与测评效果仅作参考；低中高效果递进变好；以上结果为研究员根据实际使用体验得到的评价；生成时间是研究员的个人计时）

1.3.2 图生视频：整体效果优于文生视频，国内向国外看齐

德邦证券 Topsperity Securities

➢ 我们对主流视频生成大模型在相同图片与prompt下进行测试。

图表：Sora视频结果（8s）

图表：Gen-3视频结果（10s）

图表：Dream Machine视频结果（5s）

图表：Haiper v1.5视频结果（4s）

图表：Vidu视频结果（4s）

图表：可灵视频结果（5s）

图表：Pixverse V2视频结果（9s）

图表：清影视频结果（6s）

注：中文prompt：“平面设计风格的怪物插图，描绘了一个多样化的怪物家族。这个群体包括一只毛茸茸的棕色怪物、一只长着触角的黑色光滑的怪物、一只斑点的绿色怪物和一只小小的有波尔卡圆点的怪物，它们都在一个有趣的环境中互动。”英文prompt：“Monster Illustration in flat design style of a diverse family of monsters. The group includes a furry brown monster, a sleek black monster with antennas, a spotted green monster, and a tiny polka-dotted monster, all interacting in a playful environment.”

资料来源：各公司官网，德邦研究所整理（注：除Sora外，以上测试结果均为我们在个公司官网的测评，选用第一次结果，未进行抽卡；国内模型采用中文prompt，国外模型采用英文prompt，其中PixVerse V2由于主要面对海外客户，也采用英文）

1.3.2 图生视频：整体效果优于文生视频，国内向国外看齐

德邦证券 Topsperity Securities

- 整体而言，图生视频效果优于文生视频，而在动作幅度、物理还原度提升空间依旧较大。
- 国内视频生成效果向国外龙头模型看齐，Vidu、清影和Sora、Gen-3差距或在缩小。

图表：主流视频大模型图生视频评价

	说明	Sora	Gen-3	Dream Machine	Haiper v1.5	Vidu	可灵	Pixverse V2	清影
生成时间	输入提示后得到结果的时间	-	1m54s	1h+	3m28s	30s	4m17s	5m40s	12m53s
分辨率	清晰与否	高	高	高	高	高	高	高	高
动作幅度	视频画面前后变化大小、动作幅度大小	高	中	低	低	高	低	中	高
语义/图意理解	提示词是否能被完整准确地理解和表达	中	高	高	中	高	中	高	高
一致性	视频内容的前后连贯性和一致性，场景转换是否平滑，各元素之间是否协调统一	高	高	低	低	高	中	高	高
稳定性	是否会出现画面变形、撕裂或其他异常现象	高	高	中	低	高	中	高	高
流畅度	动作是否自然连贯，整体观感是否顺畅	高	高	低	中	高	中	中	高
物理还原度	光影效果、液体流动是否自然，交互行为是否符合物理规律	高	高	低	低	高	中	中	中
综合评价	-	未能很好呈现“交流”语义理解，其他方面整体较好。	部分形象动作幅度较小其他方面整体较好。	整体不够连贯，流畅性不足。	整体动作幅度较小、存在变形等异常现象。	整体效果较好，且存在一些创新性画面。	动作幅度较小、流畅度和一致性欠缺。	动作幅度、流畅度等方面存在提升空间。	动作幅度虽然大，但是内容较为单调。

注：以上是根据前文prompt+单一图片生成结果的对比，不同prompt与图片的组合结果或有差异；以上结果为研究员根据实际使用体验得到的评价；生成时间是研究员的个人计时。

资料来源：各公司官网，德邦研究所整理（注：在不同时间点调用模型用户拥挤度存在差异，以上生成时间与测评效果仅作参考；低中高效果递进变好；清影文生视频与图生视频生成时间差异大主要系研究员测试时图生视频排队时间较长，剔除排队影响，二者时间类似）

1.4 视频大模型具备商用潜力，下游应用正在储能

- **由前文推断，我们认为，主流视频大模型已经实现了不错的时长和稳定一致性，“翻车”现象大幅减少，生成的视频不再是简单的动图和“PPT式”变化，下一步迭代的重点方向是动作幅度和物理模拟能力。**
- **视频大模型的成熟奠定了AIGC应用普及的基础，在垂直领域具有广阔的应用场景和市场价值，向用户开放正在初步验证商用潜力。**未来随着用户习惯的培育以及算力基础设施的完善，视频大模型或在办公、广告、电影、游戏等多领域具有广阔的前景。

✓ 4月15日，全球多媒体巨头Adobe在官网宣布，将Sora、Pika、Runway等集成在视频剪辑软件Premiere Pro中。在发布短片中，PR展现出在视频中添加物体、消除物体以及生成视频片段等能力。通过AI驱动的音频功能已普遍可用，可使音频的编辑更快、更轻松、更直观。

✓ 视频大模型在短剧市场潜力已被验证。据智东西微信公众号，截至7月底至少有8部AI短剧可以成为产业发展的关键节点。当月，国内首部AIGC奇观剧《山海奇镜之劈波斩浪》短剧播出，十余人的创作团队取代传统百人规模，制作周期从通常的3-6个月缩短到了2个月，成本达到传统制作流程的1/4以下，大大缩短制作周期和成本，验证视频大模型在短剧的商用潜力。

图表：Adobe的PR实现在视频中添加物体、消除物体以及生成视频片段

图表：2024年引起关注的已播或待播AI短剧情况

剧名	题材	主推出方	首播时间	状态
《中国神话》	玄幻	央视	3月22日	已完结
《英雄》	历史	央视	6月28日	已完结
《爱永无终止》	伦理	央视	6月28日	已完结
《奇幻专卖店》	科幻	央视	6月28日	已完结
《三星堆：未来启示录》	科幻	抖音	7月8日	已完结
《觉醒》	科幻	悟空AI	7月9日	更新中
《山海奇镜之劈波斩浪》	玄幻	快手	7月13日	已完结
《因AI求真》	公益	上海广电	7月23日	更新中

资料来源：Adobe官网，智东西微信公众号，德邦研究所

02 算法、算力、数据三要素，视频大模型通过AGI的必经之路

1.1 算法：Transform与U-net，效益与效率的比拼

1.2 算力：视频生成训推算力需求指数级增长

1.3 数据：质量决定模型能力，用户激发模型迭代的潜能

1.1 算法：Transform与U-net，效益与效率的比拼

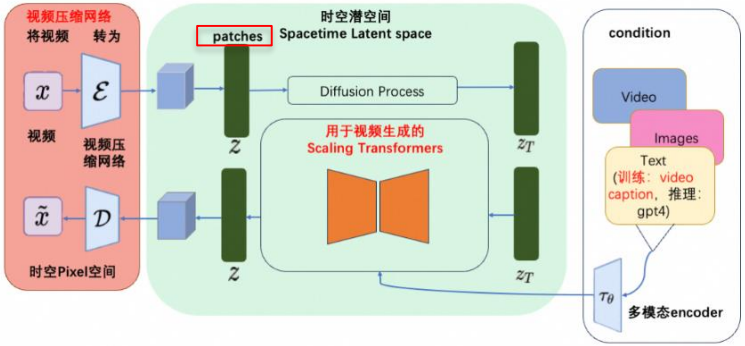
- **目前，视频生成模型算法主要由基于SD逐帧生成和基于时空Patches两种范式构成。**前者是以SD模型作为初始条件，将其转化为视频生成模型，模型架构以U-Net为主，模型代表为Stable Video Diffusion；后者则是从头开始视频训练，将视频压缩为时空Patches，后通过transfomer机制生成视频，模型架构以DiT或U-ViT为主，模型代表为Sora、Vidu等。
- **两种架构是效率与效益的比拼。**基于SD逐帧生成的模型，模型更容易训练，然而生成的视频内容一致性要差一些，长视频生成存在困难；基于时空Patches生成的架构，训练成本更高，但是生成视频的长度与一致性更容易得到保障。

图表：主流视频生成模型架构优劣势对比

视频生成范式	基于SD逐帧生成	基于时空Patches生成	
模型架构	U-Net	DiT	U-ViT
模型名称	Pika、Gen-2、Stable Video Diffusion、MagicVideo-V2等	Sora、清影、Pixverse V2、可灵	Vidu
模型优点	以SD作为初始化，模型更容易训练，训练成本可控	视频内容的一致性有保证，可以生成长视频	
模型缺点	视频内容的一致性要差一些，长视频生成有困难	整个模型需要从头训练，训练成本很高	

图表：MagicVideo-V2 SD模型范式：采用插帧的方式利用

图表：Sora模型范式猜想：基于时空Patches，利用Transformer架构

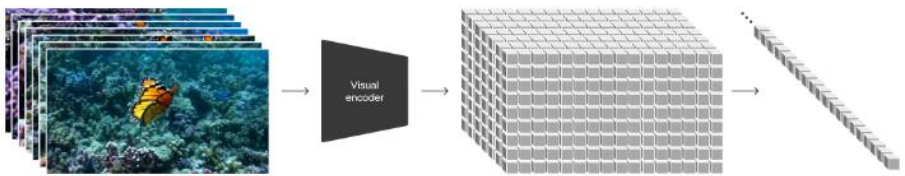


资料来源：Bytedance Inc.《MagicVideo-V2: Multi-Stage High-Aesthetic Video Generation》，智源社区微信公众号，魔搭ModelScope社区微信公众号等，德邦研究所

1.2 算力：视频生成训推算力需求指数级增长

- 相比于文字等单模态，图片、视频等多模态所包含的信息更多，计算复杂度显著提高，训练该类大模型所需算力需求更高。以Sora为例，一定条件下测算，**Sora训练算力需求是GPT-4的4.5倍，而推理算力需求达到了GPT-4的390.5多倍**。
- Sora此类大模型可一次性生成1分钟视频，并具备较高的稳定性、文字理解能力等，已成为实用生产力工具，有望掀起新一轮的内容创作革命，对后续的推理算力要求更高。
- **训练侧：**Sora算力规模或达到9.8万ZFLOPS，为GPT-4的4.5倍，大致需要9096张H100卡训练180天，系训练90天GPT-4的2倍多。
- ✓ **GPT-4：**根据semianalysis的测算，GPT-4总共包含了1.8万亿参数，采用了13万亿token的训练数据集。
- ✓ **Sora：**大约具有30亿参数规模，训练数据或达到百万亿级patches。在训练过程中，Sora通过将视频压缩到低维潜在空间，并借鉴了LLM中将文本信息转化为token的思路，针对视频训练视觉patch。DiT论文为Sora基础之一，根据其第一作者谢赛宁推算，Sora参数量约30亿。就训练数据而言，为方便计算，我们假设Sora采用了Runway发布的文生视频模型Gen-2数据集的十倍，即24亿张图片和6400万视频片段（假设单片段长度为1分钟）。图片和视频分辨率假设为高清图像（1920*1080），借鉴谷歌论文，假设压缩到16*16（像素）潜在空间；视频长度假设为1分钟，帧率为30FPS；同时视频帧率被压缩到潜在空间中，借鉴DiT论文，假设压缩系数为8。经过测算，Sora的训练数据规模或达到136.08万亿patches。

图表：Sora将视频数据转化为patch的过程



图表：Sora训练数据规模测算

变量	单张图片	1秒视频
分辨率	1920*1080	1920*1080
压缩空间	16*16	16*16
帧率（FPS）		30
帧率压缩系数		8
单位patch（token）	8100	30375

Sora训练数据集	图片规模（亿张）	视频规模（万视频片段）
	24	6400
patch数量（T token）	19.44	116.64
Sora训练patch总数量（T token）	**136.08**	

注：假设一个视频片段为60s；Sora训练数据集规模与视频片段时长为研究员自行假设，此处仅为示意性测算，可能与实际情况存在一定程度的差异，具体数据以各公司公布的数据为准）

资料来源：OpenAI官网，Runway《Structure and Content-Guided Video Synthesis with Diffusion Models》，Google《AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE》等，德邦研究所测算

1.2 算力：视频生成训推算力需求指数级增长

- **训练侧：**Sora算力规模或达到9.8万ZFLOPS，为GPT-4的4.5倍，大致需要9096张H100卡训练180天，系训练90天GPT-4的2倍多。
  - ✓ 迭代次数假设：由于扩散模型在去噪降噪过程中需要多次迭代，参考Stable Diffusion 30~50次的步数，我们假设Sora迭代了40次；算力利用率假设：根据semianalysis，由于大量的故障导致训练需要重新启动的原因，GPT-4的算力利用率大概为32%~36%，我们假设Sora和GPT-4算力利用率均为35%；训练时间假设：根据semianalysis，GPT-4训练了90~100天，由于多模态模型计算更为复杂，我们假设GPT-4和Sora分别训练了90 / 180天。经过测算，我们发现Sora训练算力或达到9.8万ZFLOPS，系GPT-4的4.5倍；Sora需要9096张H100加速卡训练180天，系训练90天的GPT-4所需卡数的2.2倍。

图表：Sora和GPT-4训练算力需求对比

变量	Sora	GPT-4
参数规模（N，B）	3	1800
激活参数规模（N，B）	3	280
训练数据量（D，T token）	136.1	13.0
单次训练计算量（ZFLOPS）	2449.4	21840.0
迭代次数	40	1
训练总计算（ZFLOPS）	**97977.6**	**21840.0**
算力利用率	35%	35%
算力需求对比（以GPT-4为基准）	**4.5**	**1**
训练时间假设（Days）	180	90
H100单卡FP16算力（TFLOPS）	1979	1979
所需H100数量（张）	**9096**	**4055**
加速卡需求对比（以GPT-4为基准）	**2.2**	**1**

注：假设Sora和GPT-4训练时间分别是180、90天；该假设为研究员自行假设，此处仅为示意性测算，可能与实际情况存在一定程度的差异，具体数据以各公司公布的数据为准）

资料来源：semianalysis官网，腾讯技术工程微信公众号，英伟达官网，OpenAI《Scaling Laws for Neural Language Models》等，德邦研究所测算

1.2 算力：视频生成训推算力需求指数级增长

- **推理侧：**Sora生成1分钟视频算力规模达到437.4PFLOPS，系生成2k token GPT-4的390.5倍；Sora在60秒响应时间需要10.5张H100，系响应时间为10秒GPT-4的65.1倍。
- ✓ 输出token假设：假设Sora和GPT-4分别生成1分钟视频 / 2k token，1分钟视频大致为182万token；算力利用率假设：我们假设Sora和GPT-4算力利用率均为35%；推理响应时间假设：我们假设Sora生成一分钟视频需要响应60s，而GPT-4生成2k token需要响应10s。经过测算，我们发现Sora生成1分钟视频算力达到437.4PFLOPS，系生成2k token GPT-4的390.5倍；在此基础上，Sora需要10.5张H100加速卡响应60s，系响应10s的GPT-4所需卡数的65.1倍。

图表：Sora和GPT-4推理算力需求对比

变量	Sora	GPT-4
参数规模（N，B）	3	1800
激活参数规模（N，B）	3	280
输出token数（D，K）	1822.5	2
单次推理计算量（TFLOPS）	10935.0	1120.0
迭代次数	40	1
推理总计算（PFLOPS）	**437.4**	**1.1**
算力利用率	35%	35%
算力需求对比（以GPT-4为基准）	**390.5**	**1**
推理响应时间假设（S）	60	10
H100单卡FP16算力（TFLOPS）	1979	1979
所需H100数量（张）	**10.5**	**0.2**
加速卡需求对比（以GPT-4为基准）	**65.1**	**1**

注：假设一次推理Sora输入1分钟视频，而GPT-4一次对话是输出1500-2000个字，假设是2000tokens；假设Sora生成一分钟视频需要响应60s，而GPT-4生成2k token需要响应10s；该假设为研究员自行假设，此处仅为示意性测算，可能与实际情况存在一定程度的差异，具体数据以各公司公布的数据为准）

资料来源：semianalysis官网，腾讯技术工程微信公众号，英伟达官网，OpenAI《Scaling Laws for Neural Language Models》等，德邦研究所测算

1.3 数据：质量决定模型能力，用户激发模型迭代的潜能

- **训练数据的规模和质量是视频生成模型的重要考虑因素。**
- ✓ 一方面，互联网数据是训练数据的重要来源，然而互联网视频质量普遍偏低，难于满足训练需求，**数据的筛选能力成为大模型厂商的重要竞争标准**。例如，快手大模型团队构建了较为完备的标签体系，可以精细化的筛选训练数据，或对训练数据的分布进行调整。
- ✓ 另一方面，**提高训练视频的文本描述性，能够显著提升视频生成模型的文本指令响应能力。**例如，Sora建立在过去DALL·E3和GPT模型的研究基础之上，构建视频re-captioning，为视觉训练数据生成高度描述性的字幕，使得模型具有强大的语言理解能力。
- ✓ 此外，对于模型的升级改良，**用户的涌入有望提升模型迭代速度的斜率**。我们认为，用户生成的数据与反馈能够更有效地转化为高质量数据，用来调整模型的升级方向，训练出更符合用户需求的模型，再通过吸引更多的用户开启模型迭代的“数据飞轮”。以Midjourney为例，Midjourney V5版本是文生图历史上的一个关键临界点，正式从“玩具”蜕变成了生产力工具，而这一次产品能力的突破，带来的是用户大规模涌入，数据飞轮开始转动，效果日新月异。

图表：Midjourney迭代提升生成效果

图表：Sora根据文本说明生成高质量视频

03 投资建议

3. 投资建议

- 建议关注（1）视频大模型厂商：科大讯飞、商汤、云从科技、格灵深瞳、拓尔思、昆仑万维等。（2）算力：海光信息、寒武纪、景嘉微、中科曙光、浪潮信息、工业富联、神州数码、拓维信息、四川长虹等。（3）接入大模型的应用标的：金山办公、万兴科技、福昕软件、虹软科技、彩讯股份、焦点科技、润达医疗、金证股份、泛微网络、金蝶国际等。

04 风险提示

4. 风险提示

- 商业化落地不及预期：海内外视频大模型发展仍未成熟，商业模式还在探索期，未来仍存在不确定性；
- 国内大模型在缺乏算力支持的情况下迭代速度放缓：美国对国内AI算力硬件采取严格封锁措施，国内在缺乏先进GPU的情况下，大模型迭代速度可能放缓；
- 国内大模型技术路线产生分歧：国内视频大模型厂商数量众多，他们或都采取差异化的技术路线维持自身竞争力，但不利于集中力量攻克大模型发展难题。

信息披露

分析师与研究助理简介

陈涵泊：德邦证券计算机行业首席分析师，上海交通大学信息安全本科，电子与通信工程硕士，曾任职于中信证券研究部、天风证券研究所，多年计算机行业研究经验，具备成熟的计算机研究框架、自上而下产业前瞻视野，云计算领域深入研究。2022-2023年新财富最佳分析师入围（团队），2023年新浪金麒麟最佳分析师第五名（团队）。

王思：德邦证券计算机行业研究助理，湖南大学金融学学士、武汉大学金融学硕士，主要覆盖AI大模型、工业软件、网安等方向。

投资评级说明

	类 别	评 级	说 明
1. 投资评级的比较和评级标准： 以报告发布后的6个月内的市场表现为比较标准，报告发布日后6个月内的公司股价（或行业指数）的涨跌幅相对同期市场基准指数的涨跌幅；	股票投资评级	买入	相对强于市场表现20%以上；
		增持	相对强于市场表现5%~20%；
		中性	相对市场表现在-5%~+5%之间波动；
		减持	相对弱于市场表现5%以下。
2. 市场基准指数的比较标准： A股市场以上证综指或深证成指为基准；香港市场以恒生指数为基准；美国市场以标普500或纳斯达克综合指数为基准。	行业投资评级	优于大市	预期行业整体回报高于基准指数整体水平10%以上；
		中性	预期行业整体回报介于基准指数整体水平-10%与10%之间；
		弱于大市	预期行业整体回报低于基准指数整体水平10%以下。

免责声明

分析师声明：本人具有中国证券业协会授予的证券投资咨询执业资格，以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，本报告所采用的数据和信息均来自市场公开信息，本人对这些信息的准确性或完整性不做任何保证，也不保证所包含的信息和建议不会发生任何变更。报告中的信息和意见仅供参考。本人过去不曾与、现在不与、未来也将不会因本报告中的具体推荐意见或观点而直接或间接收任何形式的补偿，分析结论不受任何第三方的授意或影响，特此声明。

法律声明：

本报告仅供德邦证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。市场有风险，投资需谨慎。本报告所载的信息、材料及结论只提供特定客户作参考，不构成投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需要。客户应考虑本报告中的任何意见或建议是否符合其特定状况。在法律许可的情况下，德邦证券及其所属关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供投资银行服务或其他服务。本报告仅向特定客户传送，未经德邦证券研究所书面授权，本研究报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

如欲引用或转载本文内容，务必联络德邦证券研究所并获得许可，并需注明出处为德邦证券研究所，且不得对本文进行有悖原意的引用和删改。

根据中国证监会核发的经营证券业务许可，德邦证券股份有限公司的经营范围包括证券投资咨询业务。

德邦证券
Topsperity Securities

德邦证券股份有限公司

地 址：上海市中山东二路600号外滩金融中心N1幢9层

电 话： +86 21 68761616　　传 真： +86 21 68767880

400-8888-128