



美图公司 (01357.HK)

买入 (首次评级)

公司深度研究

证券研究报告

AI 应用淘金系列一：

穿越周期的国产影像工具软件巨头

投资逻辑

- 我们认为公司不应受到“AI 吞噬软件”的叙事压制，并认为市场对美图系产品目前存在 4 大认知误区：1) 容错率误区：涉及人脸编辑、人物一致性生成的场景是存在标准答案的不容错场景，目前仅依靠多模态基模能力生成结果的稳定性较差。2) 审美活动需要提效误区：一个容易被忽视的用户洞察是 p 图本身是一件自我愉悦的事情，即使 AI 迭代后能够解决多数准确性问题，一键 p 图也会剥夺用户的体验感、参与感与成就感。3) 自然语言交互误区：从命令行到 GUI，功能按钮本身即为 Knowhow，自然语言交互的自由度与准确性反而低于美图系产品图形化面板的拖拉拽交互模式。4) 点状功能误区：美图系产品实际上是封装了成熟工作流的影像类工具箱，裁切→调整透视→滤镜→人像美容→贴纸→社交分享等功能场景环环相扣，能够对用户的深浅需求均做到 100% 的满足。公司影像与设计产品 18-24 年复合年均增长率/同比增速分别为 85.3%/45.2%，因此我们认为 AI 时代如美图系这类“功能厚度+国民认知”足够深的工具软件仍具备难以颠覆的价值。
- 我们认为公司长期成长空间广阔，出海+生产力工具有望启动第二增长曲线。公司 2008 年推出美图秀秀以来，先后尝试过平台转型流量变现（短视频平台增值服务、电商）、软件驱动硬件变现（美图手机、美容仪）等商业模式，最终回归影像工具软件的订阅化变现，订阅收入由 18 年的 0.5 亿元增长至 24 年的 20.9 亿元，其中 MAU/ARPU/付费率三项驱动因子 20-24 年分别增长约 2%/25%/519%。展望远期，我们认为“出海+生产力工具”有望贡献 MAU 未来增量，AI 增值功能+MAU 结构端变化有望增厚 ARPU，“功能厚度+社交属性”可将工具类 APP 付费率上限推升至 10+%。

盈利预测、估值和评级

- 我们预计公司 2025~2027 年营业收入分别为 40.21/50.12/62.45 亿元，经调整归母净利润分别为 9.63/12.21/16.09 亿元。采用市盈率法，给予公司 26 年 40 倍 PE (取经调整后的归母净利润) 估值，目标价 11.73 港元。首次覆盖，给予“买入”评级。

风险提示

- 影像编辑类工具软件被 AI 基模吞噬的风险；大厂入局导致竞争加剧的风险；手机厂商入局导致竞争加剧的风险。

计算机组

分析师：孟灿 (执业 S1130522050001)

mengcan@gjzq.com.cn

联系人：孙恺祈

sunkaiqi@gjzq.com.cn

市价 (港币)：7.790 元

目标价 (港币)：11.73 元



公司基本情况 (人民币)

项目	2023A	2024A	2025E	2026E	2027E
营业收入(百万元)	2,696	3,341	4,021	5,012	6,245
营业收入增长率	29.27%	23.93%	20.36%	24.64%	24.62%
经调整归母净利润(百万元)	368	586	963	1,221	1,609
经调整归母净利润增长率	/	59.16%	64.34%	26.73%	31.78%
摊薄每股收益(元)	0.09	0.14	0.23	0.29	0.39
每股经营性现金流净额	0.05	0.09	0.16	0.18	0.27
ROE(归属母公司)(摊薄)	9.21%	16.02%	14.66%	15.99%	17.72%
P/E	40.68	21.09	33.65	26.56	20.15
P/B	3.92	2.69	6.37	5.35	4.40

来源：公司年报、国金证券研究所



内容目录

一、为什么我们认为美图不会被 AI 吞噬？——从市场对美图的 4 个误区讲起	5
1.1 对“AI 吞噬软件”追问之后：AI 与应用间边界愈发清晰，关注 C 端软件错杀品种	5
1.2 容错率误区：涉及人脸编辑、人物一致性生成的场景是存在标答的不容错场景	6
1.3 审美活动需要提效误区：一个容易被忽视的用户洞察是 p 图本身是一件愉悦的事情	7
1.4 自然语言交互误区：从命令行到 GUI，功能按钮本身即为 Knowhow	9
1.5 点状功能误区：美图系产品实际上是封装了成熟工作流的影像类工具箱	9
二、如何看待美图的成长空间？——订阅收入的极限或在 100-225 亿元	12
2.1 商业模式：阅尽千帆归来仍是工具软件	12
2.2 MAU：“出海+生产力工具”贡献未来增量，背靠阿里 Gen AI 技术处于全球前列	16
2.3 ARPU：视频 > 图像，海外 > 国内，AI 增值功能有望增厚 ARPU	19
2.4 付费率：“功能厚度+社交属性”可将工具类 APP 付费率上限推升至 10+%	23
三、盈利预测与投资建议	24
3.1 盈利预测	24
3.2 投资建议及估值	26
四、风险提示	26
五、附录：多模态模型发展脉络及技术演变	27
5.1 多模态模型：从跨模态协同迈向原生融合，技术架构持续突破	27
5.2 多模态生成技术路线：跨越图像、音频与视频的融合创造	30
5.3 多模态理解技术路线：突破感知认知临界，AI 基座模型步入成熟期	34

图表目录

图表 1：“AI 吞噬软件”对 AI 应用板块产生叙事压制	5
图表 2：对“AI 吞噬软件”追问之后，众多 C 端软件仍具备较高护城河	6
图表 3：容错率误区：涉及人脸及人物一致性的场景实质是有标答不容错的	7
图表 4：摄像头功能在国际知名消费电子测评网站 DxOMark 上位居智能手机测评维度第一名	7
图表 5：自拍功能在国际知名消费电子测评网站 DxOMark 上位居智能手机测评维度第二名	7
图表 6：国际知名消费电子测评网站 DxOMark 对于摄像头功能的测评维度更注重照片及视频的光学还原度	8
图表 7：国际知名消费电子测评网站 DxOMark 对于自拍功能的测评维度更注重照片及视频的光学还原度	8
图表 8：第三方影像编辑工具能够提供“多样的美”，修图或为审美活动	8
图表 9：自然语言交互误区：在需要精细调整的场景，功能按钮界面自由度更高	9
图表 10：美图系产品实际上是封装了成熟工作流的影像类工具箱	9
图表 11：AI 自动消除功能比较	10
图表 12：AI 闪光灯功能比较	10



图表 13: AI 一键美颜功能比较.....	11
图表 14: AI 表情重塑功能比较.....	11
图表 15: AI 扩图功能比较.....	12
图表 16: 阅尽千帆归来仍是工具软件.....	13
图表 17: 以流量为中心（关注用户访问时间） VS 以产品为中心（关注用户体验）.....	14
图表 18: 公司影像与设计产品收入占比迅速提升.....	14
图表 19: 18-24 年影像与设计产品收入 CAGR=85.3%.....	14
图表 20: 20-24 年三因子模型主要依靠“付费率”贡献增长.....	15
图表 21: 公司产品矩阵三层架构清晰，发展侧重向生产力工具转移.....	15
图表 22: 公司应用产品矩阵涵盖生活场景应用与生产力工具两部分.....	16
图表 23: 公司国内生活场景应用 MAU 的成长上限或为 2.3 亿持有智能手机的 15-49 岁女性用户.....	16
图表 24: MAU 主要增量看“出海+生产力工具”.....	17
图表 25: 美图秀秀凭借 AI 合照功能荣获欧洲 14 个国家 APP Store 总榜第一.....	17
图表 26: 25 年 2 月美颜相机凭借“AI 换装”功能带动海外 MAU 快速提升.....	17
图表 27: 美图影像研究院聚焦人体技术、图像处理、视频编辑等核心技术处理的论文发表情况.....	18
图表 28: 视频生成权威评测 VBench 最新榜单中公司自研奇想智能大模型夺冠.....	18
图表 29: LMArena 图像编辑与文生视频榜单上阿里旗下模型位居全球 Top10.....	19
图表 30: Artificial Analysis 图像编辑榜单上阿里旗下模型位居全球第七.....	19
图表 31: Artificial Analysis 文生视频榜单上阿里旗下模型位居全球第五.....	19
图表 32: 国内主流图片编辑类工具 APP 会员价格及权益内容.....	20
图表 33: 国内主流图片编辑类工具 APP 年费约在 100-300 元.....	20
图表 34: 国外主流图片编辑类工具 APP 会员价格及权益内容.....	21
图表 35: 国外主流图片编辑类工具 APP 年费约在 40-140 美元（折合约 300-1000 元）.....	21
图表 36: 国内主流图片编辑类工具 APP 的 AI 功能普遍采取按使用量计费模式.....	21
图表 37: 国内主流视频编辑类工具 APP 会员价格及权益内容.....	22
图表 38: 国内主流视频编辑类工具 APP 年费约在 100-300 元.....	22
图表 39: 国内主流视频编辑类工具 APP 的 AI 功能普遍采取按使用量计费模式.....	23
图表 40: 具备社交属性的生活场景应用与生产力工具付费率可达 10+%，内容消费类 APP 付费率可达 20+%	24
图表 41: 公司订阅收入的极限或在 100-225 亿元之间.....	24
图表 42: 25/26 年有望实现约 40/50 亿元营收，同比增长 20%/25%.....	25
图表 43: 利润预测：2026 年有望维持综合毛利率 74%，经调整归母净利润 12.2 亿元.....	26
图表 44: 可比公司 26 年 PE 中位数接近 40X，PEG 接近 1.4X.....	26
图表 45: 一类典型的多模态模型结构包括编码器、连接器与大语言模型（LLM）.....	27
图表 46: 多模态模型关键技术演进历程（2019-2025 年）.....	28



图表 47: CLIP 联合训练图像编码器和文本编码器来预测一批 (图像, 文本) 训练示例的正确配对	28
图表 48: 通过将图像分割成网格并独立编码, 将 LLaVA-1.5 扩展到更高分辨率.....	29
图表 49: Veo 3 能够零样本解决大量以输入图像和文本提示指定的视觉任务.....	29
图表 50: 通过拼接或更通用的交叉注意力机制对潜在扩散模型 (LDMs) 进行条件约束	30
图表 51: 视觉自回归模型通用框架及核心组件	30
图表 52: TTS 的核心流程包含文本分析、声学模型和声码器三个关键环节.....	31
图表 53: 扩张因果卷积层堆叠的可视化.....	31
图表 54: Tacotron 2 系统架构框图	32
图表 55: VITS 训练流程与推理流程架构框图	32
图表 56: 语音合成技术的演进经历了从分阶段模型到端到端系统的转变.....	33
图表 57: 各类图像和视频转换为 visual patches 的潜在表示.....	33
图表 58: Gemini 模型支持将文本、图像、音频和视频的交错序列作为输入.....	34
图表 59: 主要多模态融合策略的基本流程.....	34
图表 60: 输入图像→CNN 提取特征→带注意力的 RNN (LSTM) →逐词生成描述	35
图表 61: 长短期循环卷积网络模型针对动作识别、图像描述和视频描述任务的特定实例化.....	35
图表 62: 整体的视觉问答框架.....	36
图表 63: 视觉常识推理 VCR 构建概述.....	36
图表 64: 跨模态检索一般框架.....	37
图表 65: 包含卷积神经网络 (CNNs) 和加权卷积神经网络 (WCNNs) 的双向提取结构	37



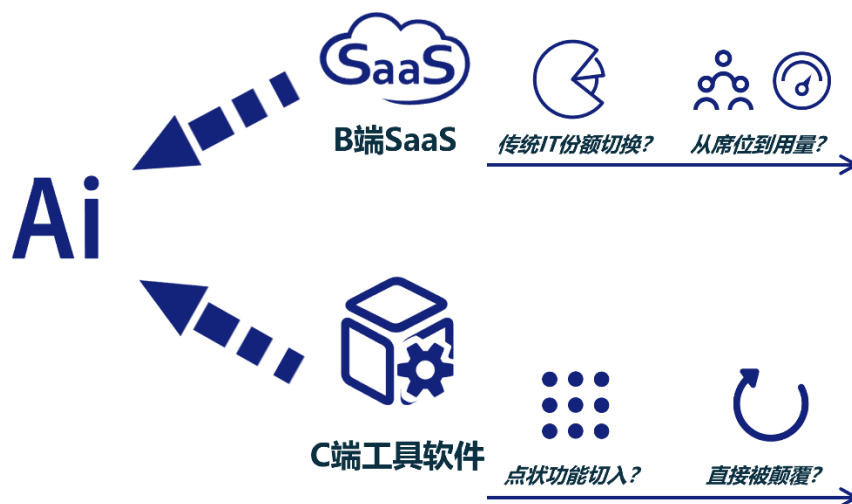
一、为什么我们认为美图不会被 AI 吞噬？——从市场对美图的 4 个误区讲起

1.1 对“AI 吞噬软件”追问之后：AI 与应用间边界愈发清晰，关注 C 端软件错杀品种

8 月以来的“AI 吞噬软件”对 AI 应用板块产生叙事压制。从 2025 年 8 月 11 日 Melius Research 下调 Adobe 评级开始，过去几个月以来，“AI is eating software”这一叙事始终压制着广义 AI 应用板块的股价表现，市场担心 AI 基模的能力会逐渐外溢到应用层，从而直接颠覆原本身处其位的软件厂商的产业卡位与商业模式。

- 对于 B 端 SaaS 软件而言，这种担心具体体现为，短期看 AI 在侵蚀客户的软件预算，在宏观经济相对难言强劲的环境下对客户的 IT 资本开支存在份额切换；长期看 AI 在侵蚀 SaaS 的商业模式，从 pay by account 到 pay by usage，有声音认为 SaaS 厂商的收入 TAM 与毛利率均可能受损于本轮 AI 浪潮。
- 对于 C 端工具软件而言，这种担心具体体现为，短期看 AI ChatBot 在一些点状功能上可以相对较好地满足消费者的需求（如问答、识图、AIGC 等）；长期看 AI 有可能重新整合 C 端入口，有声音认为非大厂 AI 玩家在算力、算法、数据、入口四要素当中均不占优，功能厚度较薄的效率型 C 端产品可能会被 AI 基模的衍生应用直接吞噬。

图表1：“AI 吞噬软件”对 AI 应用板块产生叙事压制



来源：阿里矢量图库，国金证券研究所

近期发生的一系列边际变化让 AI 与应用之间的边界愈发清晰。

2025 年 10 月 6 日，OpenAI 2025 DevDay 发布 Apps Inside ChatGPT，“抢占入口”野心昭然若揭，我们认为其对于帮助市场理解泛 C 应用与 AI 基模之间的关系也具有深刻意义。

- LLM-as-OS, ChatGPT 内置应用之后，用户可以在对话中被智能推荐到合适的应用，也可以直接通过自然语言在界面里唤起某款应用，上述“自然语言 + 应用接口”的模式，意味着 ChatGPT 本身成为了操作系统。
- 入口正在发生迁移，第一/二/三代入口门户网站/搜索引擎/超级 APP 依靠内容聚合/信息检索/信息分发&承载应用成为入口，9 月底 ChatGPT Pulse 版本的上线已展现出 OAI 希望推动 ChatBot 由被动接收 Prompt 给出结果的形式转为主动结合 Context 提供个性化信息推送；本次 ChatGPT 引入各类 APP (Booking.com, Canva, Coursera, Figma, Expedia, Spotify 及 Zillow) 之后，其一在技术层面有望成为操作系统的中间服务层，其二在商业层面也有望成为 AI 时代的应用商店，其背后的核心也许是依靠智能带来的“内容生成+应用分发”的能力，抢占 AI 时代的原生入口。
- 关注有分发价值的应用，我们认为，本次 ChatGPT 内置应用这一动作恰恰表达出尊重应用的态度，或者换言之——要关注“有分发价值的应用”：其一我们认为已对垂类工作流进行成熟封装的工具类软件（无论 2C2B，再强大的智能也无需在缺乏 knowhow 的背景下重复造轮子），其二是具有较强版权壁垒的内容消费类软件（无版权保护的内容极有可能由基模本身生成直接交付给用户），其三是已经聚合了较多供给侧服务资源的平台型应用（客户渠道资产较难短期重建）。

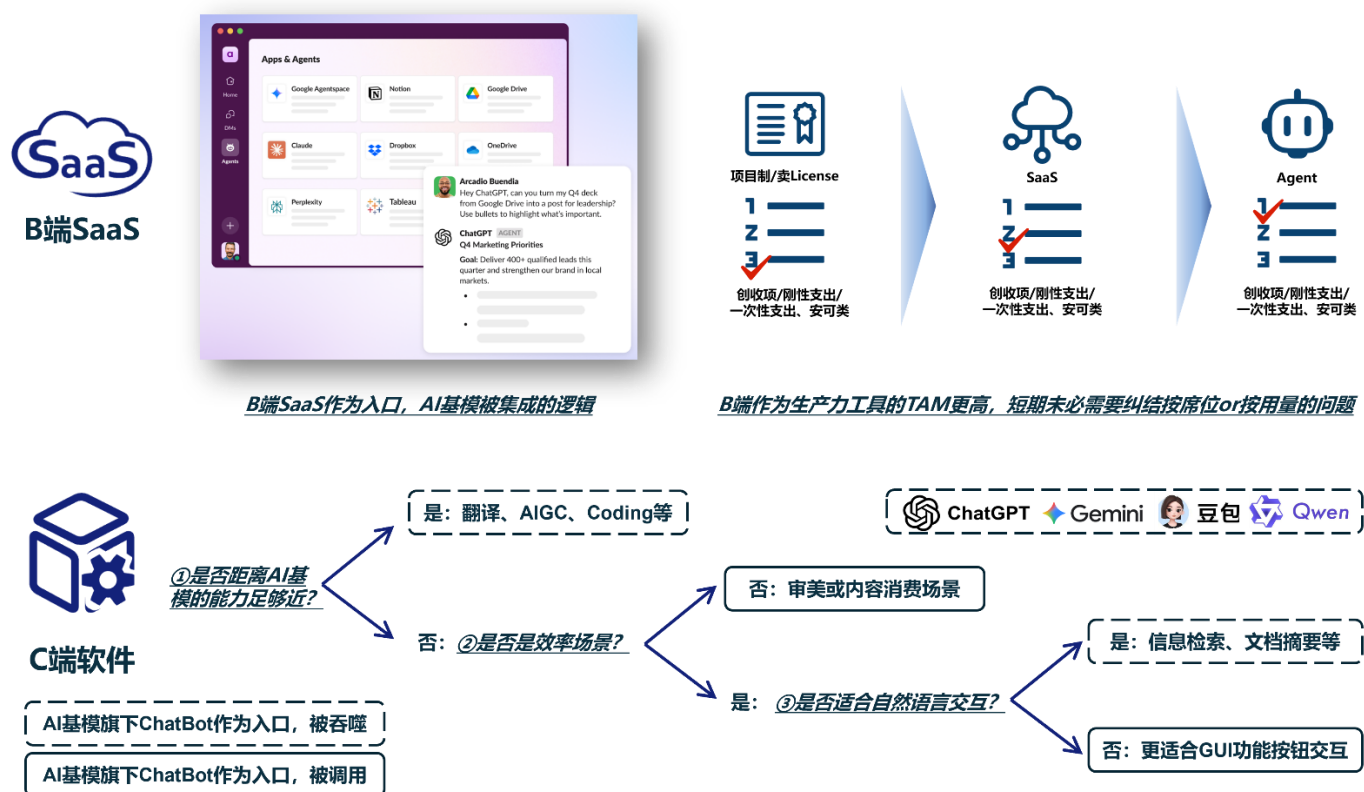


2025 年 10 月中旬，北美企业服务巨头 Salesforce 相继与 OpenAI、Anthropic、Google Gemini 达成或深化合作，合作形式几乎均以 Agentforce 集成外部基模能力为主，Salesforce 仍保留了面向企业级客户的服务入口。

- 与 OpenAI: Salesforce 将把 OpenAI 的推理、语音和多模态功能引入 Agentforce 360 平台；将 ChatGPT、Codex 引入 Salesforce 旗下的企业级办公协作平台 Slack；Salesforce 的应用程序也可通过 ChatGPT 访问，可将企业智能、CRM 数据和 Tableau 可视化直接带入 ChatGPT 对话中。
- 与 Anthropic: 引入 Claude 作为 Salesforce Agentforce 360 平台的基础模型；Salesforce 和 Anthropic 正在深度集成 Claude 和 Slack。
- 与 Gemini: 客户可以使用 Gemini 模型为 Salesforce 背后的 Atlas 推理引擎提供支持，并且可以利用 Gemini 的多模态智能来解决复杂的多步骤企业问题。由于推理引擎现在支持混合推理，Gemini 用户可以在 Salesforce 中部署 AI 代理。

上述变化发生后，我们认为：1）“入口”的定位愈发明确。B 端仍以 SaaS 或软件厂商掌握入口及数据，AI 基模作为被集成的一方提供智能；C 端或渐渐被 AI 基模厂商（大厂 AI）掌握入口主导权，C 端软件存在被吞噬与被调用/分发两种结果。2）B 端作为生产工具的 TAM 更高。短期 IT 预算问题或对业绩存在扰动，但我们长期仍看好企业服务 Agent 化后的价值分成逻辑要强于按席位收费的 SaaS 逻辑。3）C 端应结合具体需求与场景分情况讨论，我们认为①距离 AI 基模能力越远、②越看重过程体验而非简单的结果、③越难以单纯依靠自然语言交互达成结果的场景，C 端软件被吞噬的风险越低，换言之，也越具备被 AI 基模旗下入口级应用调用与分发的价值。

图表2：对“AI 吞噬软件”追问之后，众多 C 端软件仍具备较高护城河



来源：Salesforce 官网，阿里矢量图库，国金证券研究所

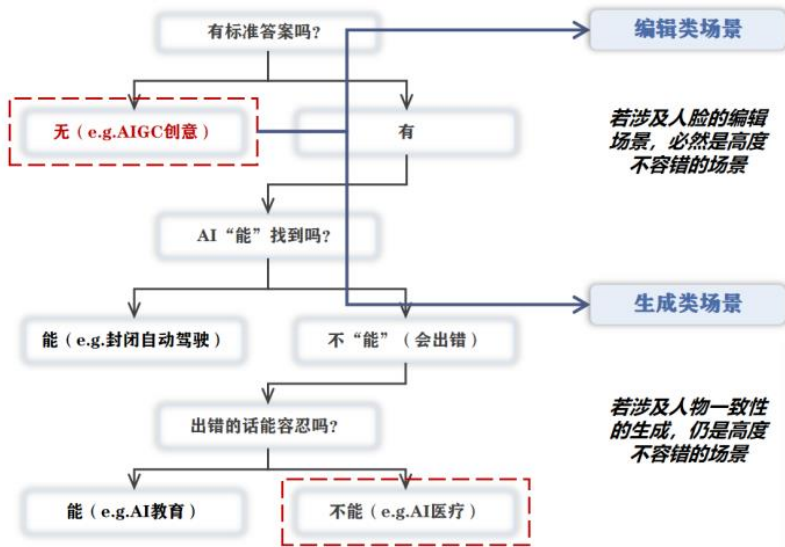
1.2 容错率误区：涉及人脸编辑、人物一致性生成的场景是存在标答的不容错场景

从 AI 落地的难易程度看，曾有声音误将美图系产品归类为 AIGC 类需求，也即“无标准答案、容忍幻觉”的场景，但实际上，美颜类 APP 所涉及的人脸编辑、人物一致性生成等场景，对结果的准确性与一致性要求极高——人脸的五官比例、动态表情、身份特征需严格契合用户审美预期，人物在生成或编辑过程中若出现五官扭曲、身份混淆等“幻觉”，将直接导致功能失效。这类场景并非天马行空的创意需求，而是存在隐性“标答”的精准化场景，AI 基模在当前技术阶段难以完全规避幻觉风险，因此美颜 APP 在人脸与人物一



致性场景的专业能力，构成了对 AI 基模吞噬的有效壁垒。

图表3：容错率误区：涉及人脸及人物一致性的场景实质是有标签不容错的



来源：RoboNeo, Gemini, 国金证券研究所

1.3 审美活动需要提效误区：一个容易被忽视的用户洞察是 p 图本身是一件愉悦的事情

许多年前，在 AI 吞噬论诞生之前，另一个美图系产品常被市场担忧的外界竞争是来自手机厂商自带美颜算法的入侵，但结果往往是单一的系统美颜较难满足用户多样化审美的需求，以及更重要的一点在于，众多女性用户在使用美图 p 图的过程本质是一种自我愉悦的审美活动，而非生产力效率场景。

根据国际知名消费电子测评网站 DxOMark 的智能手机维度排名，摄像头、自拍等拍摄相关指标在智能手机综合评分中占据核心地位，拍摄功能与硬件是智能手机的关键卖点之一。基于此，市场对公司美颜类 APP 是否会被上游手机厂商入侵的担忧具有现实依据：手机厂商可通过强化自身拍摄硬件与内置美颜功能，对独立美颜 APP 的生存空间形成潜在挤压，这一问题也成为美颜类应用赛道发展需重点探讨的方向。

图表4：摄像头功能在国际知名消费电子测评网站 DxOMark 上位居智能手机测评维度第一名

图表5：自拍功能在国际知名消费电子测评网站 DxOMark 上位居智能手机测评维度第二名

智能手机排名 摄像头分数									
排名	设备	发布价格	发布日期	摄像头	自拍	音频	屏幕	电池	
1.	Huawei Pura 80 Ultra	\$1099	2025年6月	175 vs	-	-	-	-	
2.	Oppe Find X8 Ultra	\$899	2025年4月	169 vs	-	-	-	-	
3.	Apple iPhone 17 Pro	\$1099	2025年9月	168 vs	-	-	-	-	
4.	Vivo X200 Ultra	\$1000	2025年4月	167 vs	-	-	149	-	
5.	Google Pixel 10 Pro XL	\$1299	2025年8月	163 vs	-	-	161	-	
=	Huawei Pura 70 Ultra	\$1499	2024年4月	163 vs	-	-	-	-	
7.	Apple iPhone 16 Pro Max	\$1099	2024年9月	161 vs	151	-	150	-	
8.	Google Pixel 9 Pro XL	\$1099	2024年8月	160 vs	148	-	158	-	
9.	Xiaomi 15 Ultra	\$1499	2025年3月	159 vs	-	-	150	-	
10.	Honor Magic6 Pro	\$1299	2024年2月	158	151	155	157	157	

智能手机排名 自拍分数									
排名	设备	发布价格	发布日期	摄像头	自拍	音频	屏幕	电池	
1.	Apple iPhone 16 Pro Max	\$1099	2024年9月	161 vs	151	-	150	-	
=	Honor Magic6 Pro	\$1299	2024年2月	158	151	155	157	157	
3.	Apple iPhone 15 Pro Max	\$1099	2023年9月	154	149	143	151	134	
=	Apple iPhone 15 Pro	\$999	2023年9月	154	149	142	151	114	
5.	Google Pixel 9 Pro XL	\$1099	2024年8月	160 vs	148	-	158	-	
6.	Apple iPhone 14 Pro Max	\$1099	2022年9月	146	145	142	146	133	
=	Apple iPhone 14 Pro	\$999	2022年9月	146	145	142	146	119	
=	Google Pixel 8 Pro	\$999	2023年10月	153	145	142	154	111	
=	Huawei Mate 50 Pro	\$1299	2022年10月	149	145	144	133	103	
10.	Apple iPhone 14 Plus	\$999	2022年10月	133	144	141	138	128	
=	Apple iPhone 14	\$799	2022年9月	133	144	141	138	112	

来源：DxOMark, 国金证券研究所

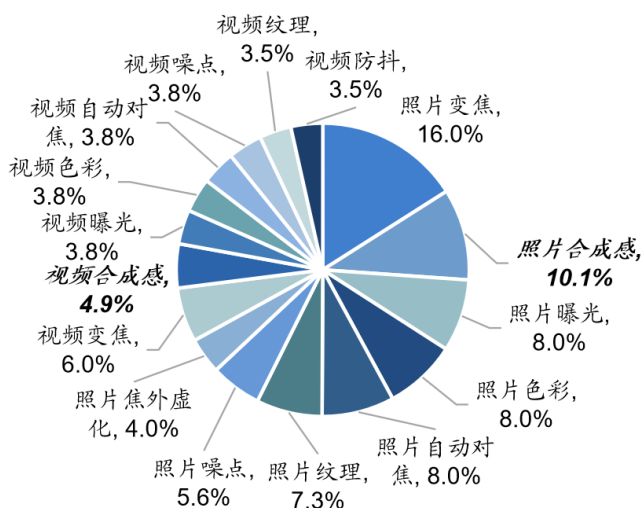
来源：DxOMark, 国金证券研究所

国际知名消费电子测评网站 DxOMark 对摄像头和自拍功能的测评维度，聚焦于照片及视频的光学还原度。测评涵盖照片变焦、合成感、曝光、色彩、自动对焦、纹理、噪点、焦外虚化，以及视频噪点、自动对焦、色彩、曝光、合成感、变焦、纹理、防抖等多个细分



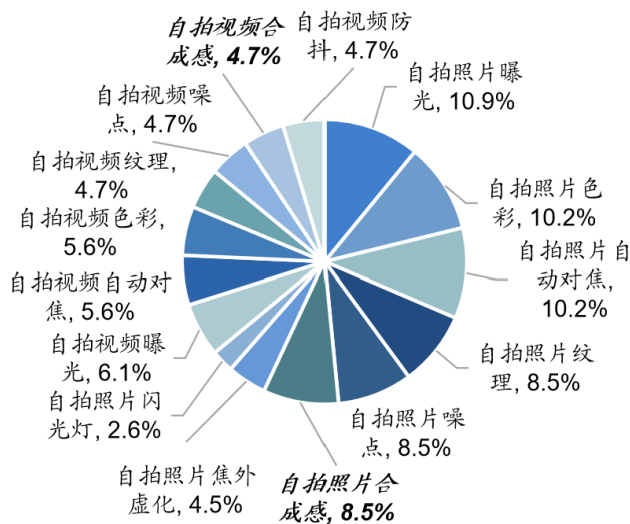
指标，重点考察拍摄功能对画面光学属性的还原能力，并未涉及美颜类功能的测评维度。这表明手机厂商在摄像头和自拍功能上的竞争重心在于光学层面的画质还原，与美图美颜 APP 提供的人像美化、风格化处理等功能存在明显差异，从而在一定程度上缓解了美图美颜 APP 被上游手机厂商功能入侵的担忧。

图表6：国际知名消费电子测评网站 DxOMark 对于摄像头功能的测评维度更注重照片及视频的光学还原度



来源：DxOMark，国金证券研究所

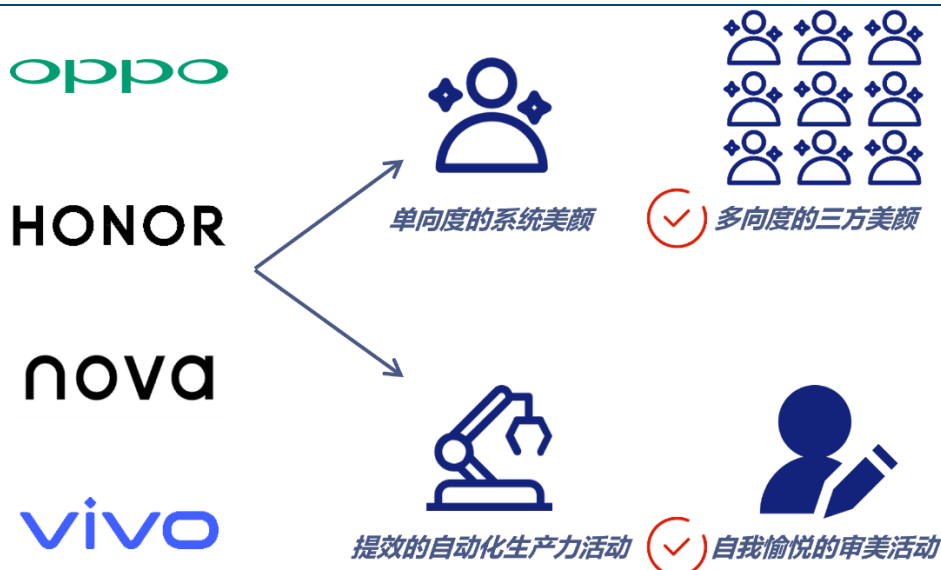
图表7：国际知名消费电子测评网站 DxOMark 对于自拍功能的测评维度更注重照片及视频的光学还原度



来源：DxOMark，国金证券研究所

美颜类 APP 在功能多样性与用户审美体验上形成对手机厂商功能入侵的有效壁垒。手机厂商内置的系统美颜属于单向度的预制化算法，适配性受限于光线环境、风格偏好、人种差异等，难以满足用户多样化美化需求，而美图美颜类 APP 提供多向度的美颜配方与滤镜插件，构建“大而全的美化百宝箱”，能充分适配不同场景下的个性化审美诉求；2025 年，清华大学心理学系学者 Hung, K. 等发表论文《数字自拍编辑揭示性别差异：处理偏见与生活满意度之间的关联》，实证研究发现，用户修图行为本质上是自我愉悦的审美活动。参与者对自身照片的满意度在经过编辑后得到显著提升，修图能够通过数字化塑造理想化自我形象，既满足了用户的个性化审美需求，又能带来切实的心理愉悦感。因此我们认为，用户修图行为本质上是自我愉悦的审美活动，而非单纯追求效率的生产力行为，修图过程中用户对审美创作的自主参与感具有不可替代性，手机厂商的自动化美颜功能难以替代第三方美颜 APP 所承载的“审美创作”价值。

图表8：第三方影像编辑工具能够提供“多样的美”，修图或为审美活动



来源：国产手机厂商官网，阿里矢量图库，国金证券研究所



1.4 自然语言交互误区：从命令行到 GUI，功能按钮本身即为 Knowhow

在人脸精细调整等需要精准控制的场景中，自然语言交互的自由度未必高于传统功能按钮界面。美颜类 APP 涉及人脸五官、轮廓、肤质等精细调整需求，对调整部位和调整程度的精准性要求极高。自然语言交互在描述这类精细需求时，可能会由于语义模糊而导致理解偏差；美颜类 APP 的功能按钮界面通过“拖拉拽”式的模块化设计，可对人脸各部位实现精准识别与分层调整，具备更高的操作自由度与精度。当用户能够用语言准确描述每一个精细需求点时，借助美颜 APP 的功能按钮界面往往能更快速、高效且准确地完成调整任务。因此，在需要精细调整的场景中，传统功能按钮界面的精准性与高效性构成了对 AI 基模自然语言交互模式的壁垒，美颜类 APP 的工具属性优势难以被替代。

图表9：自然语言交互误区：在需要精细调整的场景，功能按钮界面自由度更高

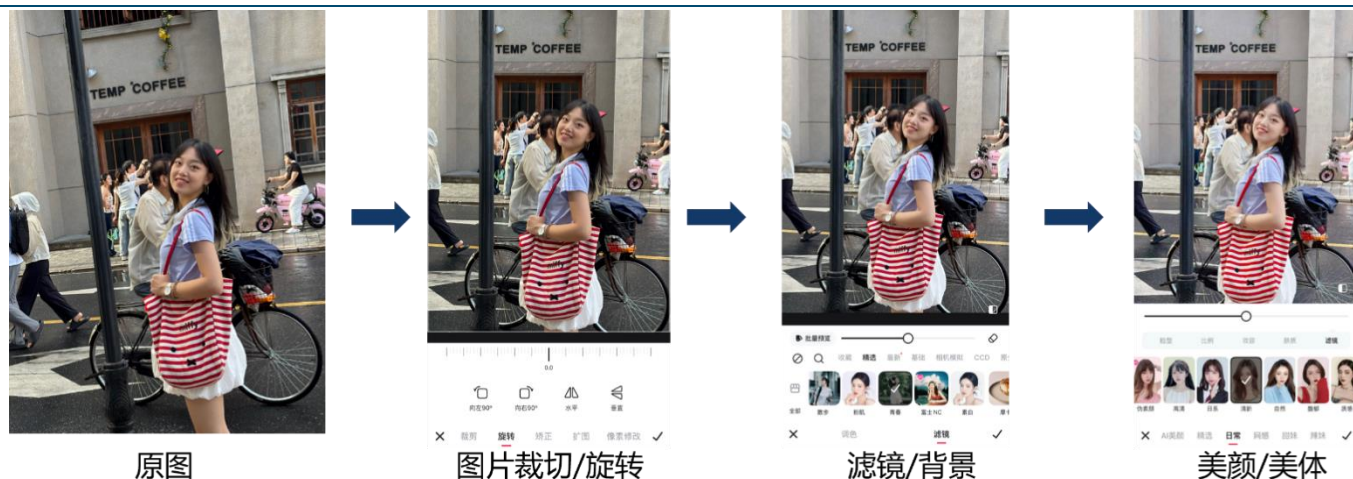


来源：RoboNeo，美图秀秀，国金证券研究所

1.5 点状功能误区：美图系产品实际上是封装了成熟工作流的影像类工具箱

AI 时代“功能厚度+国民认知”足够深的工具软件仍具备难以颠覆的价值。诚然，在我们的 AI 产品体验中，某些点状的生图、修图功能在豆包、ChatGPT、Gemini 的 ChatBot 中都已初步成熟，可以在“连续抽卡”之后满足部分用户点状的一次性浅需求（我们已在下文列示了 5 类典型的美图系付费 AI 功能与国内外代表性 AI ChatBot 生成的结果对比），但以美图秀秀为代表的美图系产品实际上是封装了成熟工作流的影像类工具箱，裁切→调整透视→滤镜→人像美容→贴纸→社交分享等功能场景环环相扣，能够对用户的深浅需求均做到 100% 的满足，且十数年来已经在全球范围内树立了极强的用户心智。

图表10：美图系产品实际上是封装了成熟工作流的影像类工具箱



来源：美图秀秀，国金证券研究所



图表11: AI 自动消除功能比较



来源: 美图秀秀, 豆包, Google Gemini2.5 Pro, 国金证券研究所

图表12: AI 闪光灯功能比较



来源: 美图秀秀, 豆包, Google Gemini2.5 Pro, 国金证券研究所



图表13: AI 一键美颜功能比较



来源: 美图秀秀, 豆包, Google Gemini2.5 Pro, 国金证券研究所

图表14: AI 表情重塑功能比较



来源: 美图秀秀, 豆包, Google Gemini2.5 Pro, 国金证券研究所



图表15: AI 扩图功能比较



来源: 美图秀秀, 豆包, Google Gemini2.5 Pro, 国金证券研究所

二、如何看待美图的成长空间？—订阅收入的极限或在 100-225 亿元

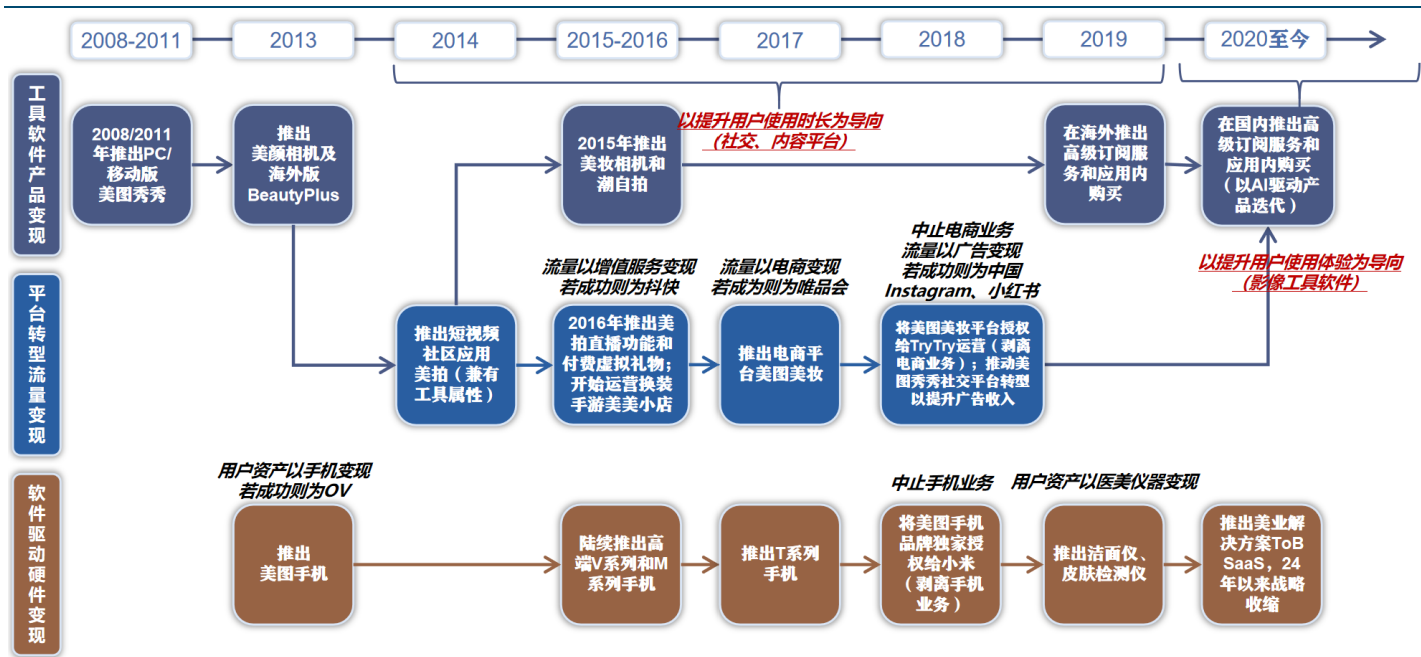
2.1 商业模式：阅尽千帆归来仍是工具软件

公司源起于影像工具软件开发，在经历美图手机、短视频社区应用、电商平台等业务尝试后，最终回归核心产品影像工具软件的订阅化转型：

- 工具软件产品变现：2008-2011 年，公司先后推出 PC 版和移动版的美图秀秀，迈出了影像工具软件领域的第一步。2013 年，公司推出美颜相机及海外版 BeautyPlus，进一步丰富影像工具产品矩阵，并于 2015 年推出美妆相机和潮自拍。2019 年，公司在海外推出高级订阅服务和应用内购买。2020 年至今，公司持续聚焦于核心业务，在国内推出高级订阅服务和应用内购买，并以 AI 驱动影响工具软件产品迭代。
- 平台转型流量变现：2014 年，公司推出兼具工具属性的短视频社区应用美拍，开启了在平台转型领域进行流量变现的探索。2016 年，美拍推出直播功能和付费虚拟礼物，并开始运营换装手游美美小店，探索流量以增值服务变现的路径。2017 年，公司推出电商平台美图美妆，转而尝试电商流量变现的模式。2018 年，公司中止电商业务，并将美妆平台授权给 TryTry 运营，同时推动美图秀秀社交平台转型，以提升广告收入。后续公司逐步退出平台转型相关业务，不再发展社交、电商类平台业务。
- 软件驱动硬件变现：公司于 2014 年推出美图手机，试图通过软件驱动用户资产以手机形式变现，并在 2015-2017 年陆续推出高端 V 系列、M 系列及 T 系列手机。2019 年，公司剥离手机业务，将美图手机品牌独家授权给小米，随后转向用户资产以医美仪器变现，推出洁面仪、皮肤检测仪，并进一步推出美业解决方案 ToB SaaS，但 24 年以来战略逐步收缩。



图表16：阅尽千帆归来仍是工具软件



来源：公司官网，公司招股书，公司财报，国金证券研究所

公司商业模式逐步从以流量为中心转向以产品为中心，聚焦软件产品本身的价值挖掘与订阅制变现。

- 以流量为中心：公司早期以流量聚合与多元变现为核心商业逻辑，依托影像类产品构建用户基础并探索多路径变现。其核心流量入口为美图秀秀与美拍两大产品，依托于庞大的用户基数，公司布局四大方向进行流量变现：①硬件变现，推出不同系列的美图手机，实现用户资产的硬件端价值兑现；②广告变现，依托高用户关注度与访问时长，通过广告投放获取收入，用户访问时间越长则广告曝光效率越高，广告变现能力越强；③增值服务变现，在美拍平台上线直播功能与付费虚拟礼物功能，通过抽取用户打赏分成实现收入；④电商变现，通过美图秀秀、美拍等产品与化妆品品牌合作，推出虚拟试色、虚拟上妆等功能，尝试将用户流量直接转化为电商消费。
- 以产品为中心：探索与碰壁之后，公司商业模式逐步从流量中心化转向产品中心化，核心逻辑由“用户时长竞争”转变为“产品价值付费”，聚焦软件产品本身的价值挖掘与订阅制变现。在产品聚焦方面，公司专注于美图秀秀、美颜相机等生活场景类软件及相关生产力工具类应用，强化核心产品的功能体验与用户价值。在变现模式上，公司以订阅制为核心商业模式，其收入主要由月活跃用户数（MAU）、付费率及每付费用户平均收入（ARPU）三因子决定，即通过扩大用户基本盘、提升用户付费转化效率与单用户付费贡献，实现收入增长。商业模式的转变，不仅聚焦核心产品能力建设，也重塑了公司发展导向，从“广撒网式流量变现”转向“精细化产品价值运营”。



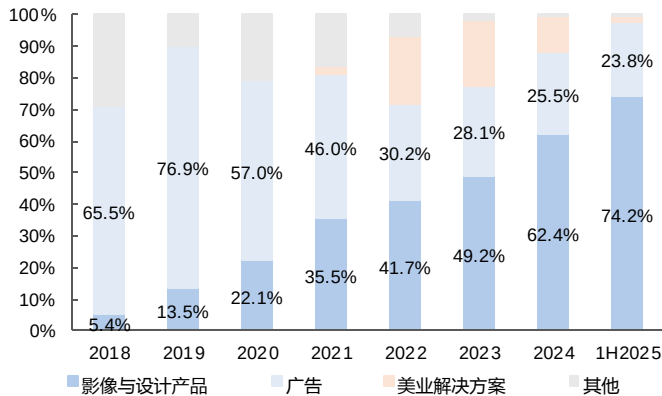
图表17: 以流量为中心 (关注用户访问时间) VS 以产品为中心 (关注用户体验)



来源: 公司招股书, 公司财报, 国金证券研究所

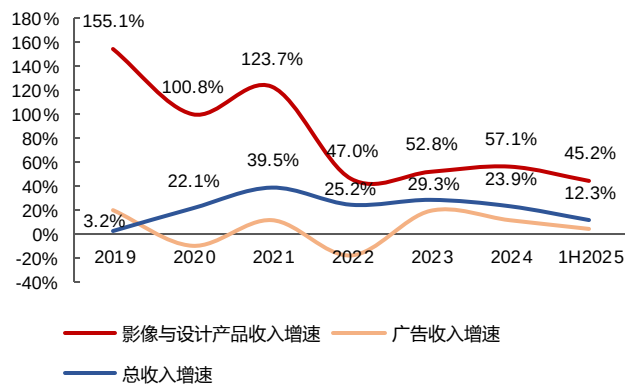
公司影像与设计产品收入占比迅速提升, 18-24 年业务收入占比由 5.4% 增长至 74.2%, 成为公司收入的核心支柱。影像与设计产品收入由 18 年的 0.5 亿元增长至 24 年的 20.9 亿元, 25H1 影像与设计产品收入为 13.5 亿元, 增长率为 45.2%。其中, 22/23/24 年影像与设计产品收入付费率分别为 2.3%/3.7%/4.7%, 增长率分别为 0.6pp/1.0pp/1.0pp。从分项收入增速分析, 影像与设计产品收入增速长期显著高于总收入增速, 18-24 年影像与设计产品收入复合年均增长率为 85.3%, 影像与设计产品成为拉动整体收入增长的关键引擎。

图表18: 公司影像与设计产品收入占比迅速提升



来源: 公司财报, 国金证券研究所 (注: 2022-1H2025 为公司财报实际披露结果, 2021 年影像与设计产品收入为 VIP 订阅+互联网增值服务, 2018-2020 年影像与设计产品收入为 VIP 订阅及影像 SaaS+互联网增值服务; 2021 年美业解决方案收入为 SaaS 及相关业务; 2018-2021 年的广告收入为在线广告)

图表19: 18-24 年影像与设计产品收入 CAGR=85.3%

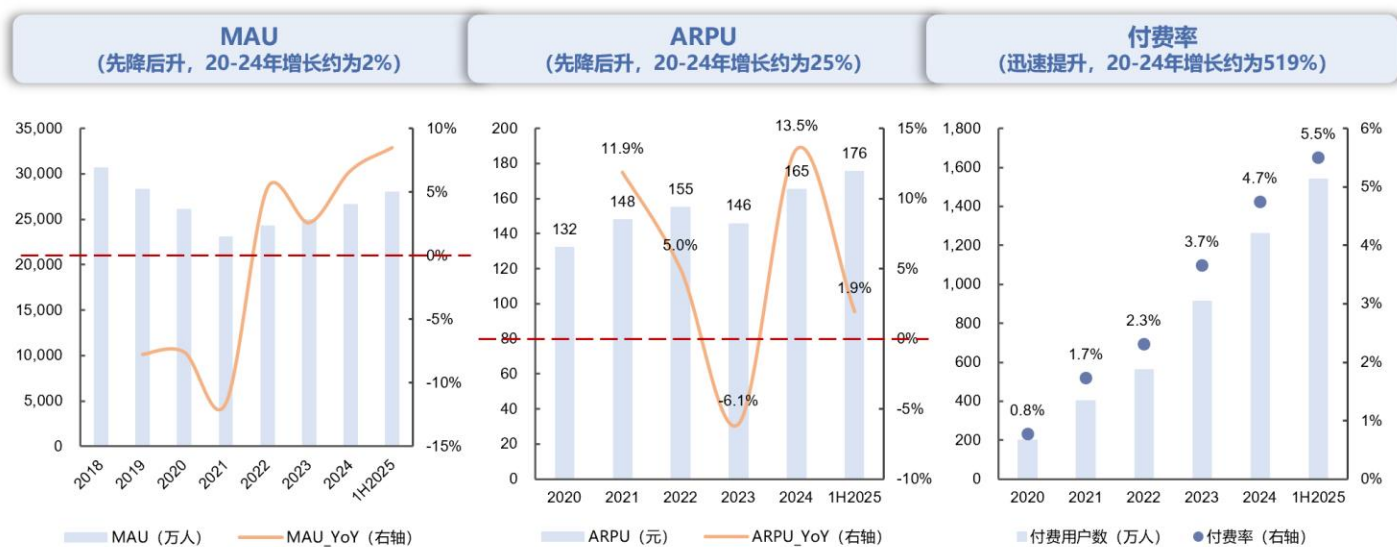


来源: 公司财报, 国金证券研究所 (注: 2022-1H2025 为公司财报实际披露结果, 2021 年影像与设计产品收入为 VIP 订阅+互联网增值服务, 2018-2020 年影像与设计产品收入为 VIP 订阅及影像 SaaS+互联网增值服务; 2021 年美业解决方案收入为 SaaS 及相关业务; 2018-2021 年的广告收入为在线广告)

公司影像与设计产品收入拆解为 MAU、付费渗透率和 ARPU 值的三因子模型。公司付费渗透率持续增长, MAU 与 ARPU 先降后升。20-24 年, MAU 增长约 2%, 整体规模波动且对收入拉动有限; ARPU 先降后升, 20-24 年增长约 25%, 呈波动态势, 贡献度不显著; 付费率迅速提升, 20-24 年增长约 519%, 成为影像与设计产品收入增长的核心拉动因素, 主导该业务收入的增长表现。



图表20: 20-24 年三因子模型主要依靠“付费率”贡献增长



来源: 公司财报, 国金证券研究所

公司三层业务架构清晰:

- 底层方面, 公司推出自研垂类大模型 **Miracle Vision**, 并广泛应用于旗下影像与设计产品, 助力电商、广告、游戏、动漫、影视五大行业, 同时, 公司与阿里巴巴集团进行大模型领域深度合作, 基于开源的通义万相首尾帧生视频、图生视频和文生视频等模型, 公司结合自身在影像领域积累的经验已训练多个垂直领域模型。
- 中间层设有美图 AI 开放平台与美图创意平台, 其中美图 AI 开放平台面向小 B 端用户, 通过插件化服务满足其特殊算法需求, 降低定制成本。技术层面, 美图 AI 开放平台依托于美图影像实验室 (MTlab), 聚焦人脸技术、人体技术、图像识别、图像处理、图像生成五大技术领域; 服务层面, 美图 AI 开放平台提供 Web API、Mobile SDK、定制解决方案等多种灵活服务形式。美图创意平台则助力创意资源整合, 为素材生产者提供创作空间和实现需求的生产工具, 支持 KOL 快速发布潮流创意内容, 在支持素材的制作发布的基础上, 助力应用层产品的创意内容供给。
- 应用层产品分为生活场景与生产力场景, 生活场景涵盖美图秀秀、美颜相机、Wink 等产品, 满足用户社交修图、视频处理等需求; 生产力场景涵盖开拍、美图云修、美图设计室等产品。整体而言, 公司从过去以生活场景社交需求为主, 逐渐向生活与生产力场景并重发展, 当前更侧重生产力工具场景的拓展。

图表21: 公司产品矩阵三层架构清晰, 发展侧重向生产力工具转移



来源: 公司微信公众平台, 国金证券研究所

公司构建了涵盖生活场景应用与生产力工具的产品矩阵。生活场景应用方面, 美图秀秀/



海外版 Airbrush 分别于 2011/2015 年推出，作为旗舰影像编辑应用，集成人像美容、视频美容等多元功能，海外版则为轻量化修图工具；美颜相机/海外版 BeautyPlus 于 2013 年推出，主打实时美颜自拍与滤镜直出；Wink 于 2022 年推出，作为 AI 视频编辑工具，具备视频画质修复、人像美容等功能，产品付费方式以订阅+单次付费为主。生产力工具领域，美图云修于 2021 年推出，为专业级批量修图 SaaS，服务摄影工作室，按张收费；美图设计室/海外版 X-Design 于 2022 年推出，面向电商场景提供 AI 设计服务；开拍于 2023 年推出，作为口播视频制作工具，支持 AI 脚本生成等功能；WHEE 于 2023 年推出，为多模态 AI 视觉创作平台；RoboNeo 于 2025 年推出，通过自然语言交互完成商业级修图。生产力工具付费方式涵盖按张收费、订阅、订阅+美豆充值等模式，并在付费模式上针对不同场景需求进行精细化设计。

图表22：公司应用产品矩阵涵盖生活场景应用与生产力工具两部分

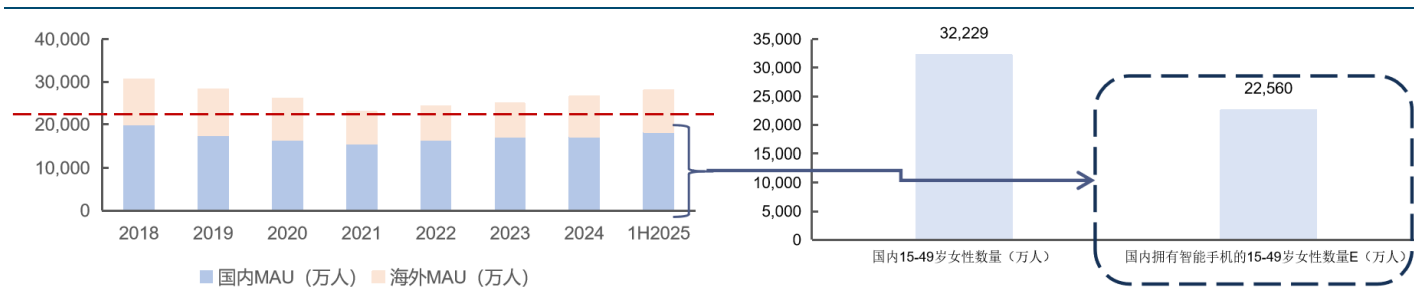
产品名称	推出时间	功能定位	付费方式
生活场景应用			
美图秀秀/海外版 Airbrush	2011 年/2015 年	美图秀秀：旗舰影像编辑应用，含人像美容、视频美容、美图配方、AI 修图、美图设计室等功能。 Airbrush：海外版修图工具，轻量化设计，支持一键人像美化。	订阅+单次付费
美颜相机/海外版 BeautyPlus	2013 年/2013 年	实时美颜自拍工具，主打滤镜直出与 AI 换装。	订阅+单次付费
Wink	2022 年	AI 视频编辑工具，其雏形来自美图秀秀视频美容模块，支持视频画质修复、视频人像美容等功能。	订阅+单次付费
生产力工具			
美图云修	2021 年	专业级批量修图 SaaS，服务摄影工作室。	按张收费，每张 0.21-0.33 元
美图设计室/海外版 X-Design	2022 年	面向电商场景的 AI 设计工具，其雏形来自美图秀秀设计室模块，支持商品图、模特图生成、海报设计、修图改图。	订阅
开拍	2023 年	口播视频制作工具，提供 AI 脚本生成、提词器、画质增强。	订阅+美豆充值
WHEE	2023 年	多模态 AI 视觉创作平台，支持文生图、图生图、图生视频、文生视频等功能。	订阅+美豆充值
RoboNeo	2025 年	AI Agent，通过自然语言交互完成商业级修图、品牌标识设计、室内设计图、营销视频、网页设计等。	目前免费

来源：公司招股书，公司财报，公司官网，公司微信公众平台，国金证券研究所

2.2 MAU：“出海+生产力工具”贡献未来增量，背靠阿里 GenAI 技术处于全球前列

公司国内 MAU 成长上限可从目标用户规模维度分析。国家统计局数据显示，国内 15-49 岁女性群体规模约 3.2 亿人，其中拥有智能手机的 15-49 岁女性数量约 2.3 亿人。公司国内 MAU 在 18-24 年在近 2 亿区间内小幅波动，已接近国内拥有智能手机的 15-49 岁女性群体规模，我们认为，在视频类工具取得大幅份额突破以前，公司国内生活场景用户 MAU 增长空间相对有限。

图表23：公司国内生活场景应用 MAU 的成长上限或为 2.3 亿持有智能手机的 15-49 岁女性用户



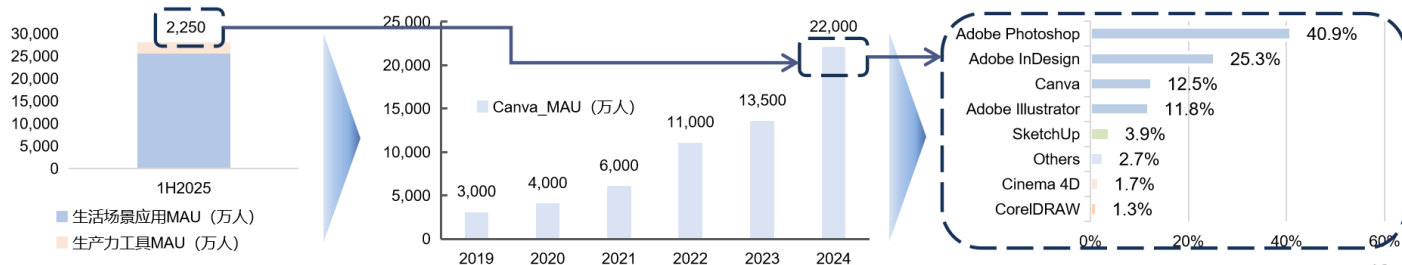
来源：国家统计局，公司财报，公司 IRPPT，国金证券研究所

公司 MAU 增量主要来源于出海业务拓展与生产力工具场景渗透两大方向。25 年上半年，公司生产力工具 MAU 为 0.2 亿人，对标海外设计工具巨头 Canva，其 24 年全球 MAU 已达 2.2 亿且增长势头未减，公司在生产力工具领域存在广阔的市场空间，可通过持续迭



代美图设计室、开拍等生产力产品，挖掘专业设计、口播视频制作等场景的用户需求，推动生产力工具板块 MAU 快速增长。从市场维度看，海外市场是关键增量场域。尽管国内生活场景用户增长空间趋于饱和，但全球范围内具备影像美化、设计创作需求的用户群体规模庞大，公司可以凭借美图秀秀海外版 Airbrush 等产品进行本土化布局，打开 MAU 出海增长的全新空间。

图表24: MAU 主要增量看“出海+生产力工具”



来源：国家统计局，公司财报，公司 IRPPT，国金证券研究所

公司以 AI 技术为核心驱动全球化布局，构建起覆盖多场景的产品矩阵并持续引爆全球市场。AI Agent 产品 RoboNeo 凭借自然语言与视觉指令跨模态解析能力，实现“图像精修-海报生成-视频特效”全流程自动化，上线后迅速登顶中国、西班牙 App Store 分类榜第一及越南 Google Play 分类榜第一，更跻身日本、韩国、英国等多国应用市场前列，累计斩获百万级月活用户；美图秀秀的 AI 合照功能支持用户上传照片后，通过自然语言指令生成虚拟合影并灵活调整风格，满足多元场景需求，2025 年 10 月一举拿下欧洲 14 国 App Store 总榜第一、28 国分类榜第一，此前更在俄罗斯、意大利、沙特阿拉伯等 12 国的 App Store 与 Google Play 摄影类榜单登顶，斩获柬埔寨及文莱 App Store 总榜第一。

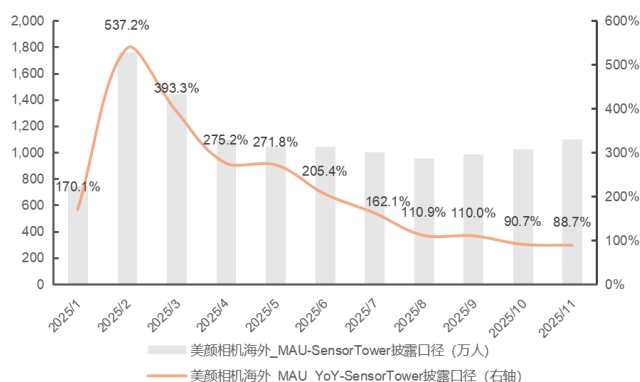
2025 年 2 月，美颜相机凭借创新“AI 换装”功能席卷东南亚，带动海外日活用户创历史新高并首次超越中国内地，单月新增 2300 万用户，泰国、越南、印尼单日新增均破百万，日、韩、美、英等高价值市场用户量同步激增，一举斩获新加坡、泰国等 12 国和地区 App Store 总榜第一；2025 年 6 月，Wink 连续 5 天占据印度尼西亚 App Store 总榜第一，成为当地现象级影像工具。美图将 B 端成熟的 AI 技术创新转化至 C 端场景，适配复杂用户场景，以设计师主导的美学洞察把控极致效果，同时依托阿里云的算力支持保障爆发式增长需求。

图表25: 美图秀秀凭借 AI 合照功能荣获欧洲 14 个国家 APP Store 总榜第一

图表26: 25 年 2 月美颜相机凭借“AI 换装”功能带动海外 MAU 快速提升



来源：公司官网，国金证券研究所



来源：Sensor Tower，国金证券研究所（注：数据及信息涵盖 2025 年 1-11 月）

公司在多模态领域的布局以技术根基为核心，形成了“研究院引领、重学术站位”的体系。公司影像研究院（MT Lab）作为核心中枢，深耕计算机视觉、生成式 AI 等关键领域，不仅构建了覆盖人脸技术、图像生成等方向的多维度技术矩阵，还重点覆盖了交互式分割、3D 重建等多模态核心方向；在学术前沿站位上，美图影像研究院聚焦人体技术、图像处理、视频编辑等核心技术处理发表多篇学术成果，截至 2025 年 11 月已在人工智能领域顶级会议与顶级期刊上累计发表学术论文 58 篇。2025 年，美图影像研究院联合清华大学等高校完成的 5 篇多模态相关论文成功入选 CVPR2025，斩获 3 项 CVPR2024 国际人工智能顶会大奖，另有 2 项学术成果入选 AAAI2025，进一步凸显了其技术研究的前沿性与含金量。



图表27: 美图影像研究院聚焦人体技术、图像处理、视频编辑等核心技术处理的论文发表情况

时间	论文接收单位	具体技术成果
2025	CVPR2025	生成式 AI
	CVPR2025	交互式分割
	CVPR2025	3D 重建
	CVPR2025	生成式 AI
	CVPR2025	交互式分割
	AAAI2025	生成式 AI
	AAAI2025	超高分辨率抠图方法
2024	Electronics 2024	高清图像驱动视频编辑
	TPAMI2024	视频多人姿态估计
	ECCV2024	视频修复
	CVPR2024	3D 场景编辑
2023	NeurIPS2023	基于文生图模型的视频编辑
	ACM MM2023	人脸修复
	TCSVT2023	跨域少样本分类
	CVPR2023	视觉 Transformer 自注意力层改进
	TMM2023	视频多人姿态估计
2022	ACM MM2022	视频序列中的人类行为与关系
	ECCV2022	多层感知器类骨干网络
	CVPR2022	多人 3D 人体姿态估计
	AAAI2022	参考视频目标分割
	BIBM2022	图像自适应分割
	TMM2022	人体姿态估计的向心偏移预测

来源: 美图影像研究院官网, 国金证券研究所

公司自研 AI 视觉大模型 MiracleVision (奇想智能) 兼具显著技术与生态优势。该模型自 23 年 6 月内测并迭代至 4.0 版本, 为美图秀秀、美颜相机、Wink、美图设计室、WHEE、美图云修等知名影像与设计产品提供 AI 模型能力。根据视频生成权威评测框架 VBench 发布最新测试结果, 公司自研大模型以 85.31% 总分及 86.69% 视频生成质量分双获榜首, 一同入选的还有阿里旗下的通义万相, 海外近期爆火的 Luma AI 旗下的视频大模型。能精准理解语义生成流畅视频, 并针对性解决人像不真实等痛点; 同时为公司旗下多产品提供 AI 能力, 搭建起三层人工智能产品生态, 还融入美学认知赋能五大行业并推动中国传统文化数字化, 技术、生态与美学价值兼具。

图表28: 视频生成权威评测 VBench 最新榜单中公司自研奇想智能大模型夺冠

meitu



MiracleVision
奇想智能
懂调学的AI视觉大模型





4.0
MiracleVision
奇想智能



Model Name (clickable)	Sampled by	Evaluated by	Accessibility	Date	Total Score	Quality Score	Semantic Score
MiracleVision_V3	MV Team	MV Team		2025-01-21	85.31%	86.69%	79.76%
Wanx_2.1	Wanx Team	VBench Team	API	2025-01-08	84.70%	85.64%	80.90%
CausVid/2025-01-02_Sol	CausVid Team	VBench Team	Close Source	2025-01-02	84.27%	85.65%	78.75%
CausVid	CausVid Team	VBench Team	Close Source	2024-12-07	83.88%	85.21%	78.57%
StratoLens	StratoLensStud	StratoLensStudio Team		2024-11-22	83.78%	84.39%	81.34%
JT-CV-98	JiuTianCV Team	JiuTianCV Team		2024-11-18	83.69%	84.62%	79.97%
Luma	VBench Team	VBench Team	API	2025-01-14	83.61%	83.47%	84.17%
T2V-Turbo-v2	T2V-Turbo Team	T2V-Turbo Team		2024-10-02	83.52%	85.13%	77.12%
MiniMax-Video-01	VBench Team	VBench Team	API	2024-10-01	83.41%	84.85%	77.65%
STV (Apple)	Apple Team	VBench Team	Close Source	2024-12-19	83.35%	84.88%	77.57%
HunyuanVideo (Open-Source)	VBench Team	VBench Team	Open Source	2024-12-16	83.24%	85.09%	75.82%
Data-Juicer (2024-09-29_Tv	Data-Juicer Te	Data-Juicer Team		2024-09-23	82.53%	83.38%	79.13%
Gen-3 (2024-07)	VBench Team	VBench Team	API	2024-07-25	82.32%	84.11%	75.17%
Yichitect-2.0 (VEnhancer)	VBench Team	VBench Team	Open Source	2024-09-20	82.24%	83.54%	77.06%
CoVideoX1.5-58 (Se. SAT. or	VBench Team	VBench Team	Open Source	2024-11-15	82.17%	82.78%	79.76%
Data-Juicer (T2V-Turbo)	Data-Juicer Te	Data-Juicer Team	Open Source	2024-07-15	82.10%	83.14%	77.93%
Jimeng	VBench Team	VBench Team	API	2024-11-15	81.97%	83.29%	76.69%

来源: 公司微信公众平台, 国金证券研究所

公司背靠阿里巴巴, 在战略资源与技术能力上具备显著优势。2025 年 5 月, 公司与阿里巴巴签署 2.5 亿美元可转债战略协议, 同年 9 月携手阿里巴巴通义实验室, 结合通义万相



系列前沿 AI 模型为公司旗下产品赋能,为全球用户提供全新智能影像体验。根据 LMArena 图像编辑与文生视频榜单,阿里旗下的“qwen-image-edit”模型在图像编辑领域跻身全球前列,通义万相相关模型也位列文生视频领域全球 Top10。

图表29: LMArena 图像编辑与文生视频榜单上阿里旗下模型位居全球 Top10

美图公司与阿里巴巴达成战略合作并获得2.5亿美元投资

来源 美图公司 2025年05月20日 23:23 编辑

点击上方蓝字 美图公司 关注美图更多资讯

5月20日,美图公司(1357.HK)发布公告,宣布与阿里巴巴签署2.5亿美元可转债协议。同时,双方还将在电商平台、AI技术、云计算等领域展开战略合作。

公告显示,阿里巴巴以可转债的形式投资美图,本金总额为2.5亿美元(净筹资约2.496亿美元),为期限3年,年利率为1%。阿里巴巴可在发行日起至到期前5个工作日内行使转换权,按每股6.00港元的价格转换为美图公司股票。

美图携手通义万相:以AI影像黑科技重塑创作体验

来源 美图 2025年09月11日 11:30 编辑

点击上方蓝字 美图公司 关注美图更多资讯

美图公司与阿里巴巴通义实验室达成合作,通义万相系列前沿AI模型将助力美图旗下产品,为全球美图用户提供全新智能影像体验。

双方此次合作以技术驱动创新,将通义实验室在生成式AI领域的深厚积累与美图数亿月活用户的创作生态深度融合,聚焦短视频玩法及生产力场景,以精准、高效、高质量的特性打造优质AI功能,提升用户创作体验。

基于通义万相大模型,美图将持续提升垂类模型训练效果,通过拓展“AI动画”、“AI特效”等短视频玩法类场景的创作边界,打造“创意成片”等短视频生成功能,提升AI生产力工具在短视频创作领域的便捷性与效率,帮助美图秀秀、Wink、开拍、RoboNeo等产品打造产品力。同时,依托阿里巴巴通义实验室提供的专项技术支持,助力美图短视频产品迭代周期,为用户呈现优质效果与使用体验。

Image Edit 6 days ago

Rank (UB) ↑	Model ↓	Score ↓	Votes ↓
1	gemini-2.5-flash-image-previ...	1342	8,685,501
2	seedream-4-2k	1311	219,262
3	seedream-4-high-res-fal	1257	87,083
4	reve-v1	1226	2,799
4	seedream-4-fal	1211	150,056
6	flux-1-kontext-max	1194	339,004
7	qwen-image-edit	1182	829,191
7	flux-1-kontext-pro	1180	5,352,306
9	gpt-image-1	1159	2,110,165
9	flux-1-kontext-dev	1158	2,985,649

Text-to-Video 21 days ago

Rank (UB) ↑	Model ↓	Score ↓	Votes ↓
1	veo3-fast-audio	1397	11,720
1	veo3-audio	1395	8,318
3	veo3-fast	1293	6,562
3	veo3	1290	6,329
5	hailuo-02-pro	1256	6,437
6	hailuo-02-standard	1235	6,255
7	seedance-v1-pro	1213	7,860
8	klings-v2.1-master	1194	7,198
8	veo2	1190	5,663
10	wan-v2.2-a14b	1146	7,361

来源: 公司微信公众平台, LMArena, 国金证券研究所 (注: 数据及信息截止至 2025 年 10 月)

在人工智能模型的图像编辑与文生视频领域,阿里巴巴旗下模型展现出全球领先的技术实力。从 Artificial Analysis 的榜单可见,阿里旗下的“qwen-image-edit”模型/通义万相系列模型在图像编辑/文生视频赛道均位列全球 Top10。这为公司在 AI 图像编辑、视频生成等前沿领域的产品创新与技术突破提供了强大支撑,助力其在生产力工具与智能创作赛道持续强化竞争力

图表30: Artificial Analysis 图像编辑榜单上阿里旗下模型位居全球第七

Rank	Creator	Model	ELO	95% CI	Appearances	Release Date
1	Google	Gemini 2.5 Flash (Nano-Banana)	1,195	-10/+11	9,811	Aug 2025
2	ByteDance Seed	Seedream 4.0	1,191	-10/+12	8,086	Sept 2025
3	Reve	Reve V1	1,109	-12/+12	4,750	Oct 2025
4	Black Forest Labs	FLUX.1 Kontext [max]	1,105	-11/+11	6,264	May 2025
5	Black Forest Labs	FLUX.1 Kontext [pro]	1,090	-10/+10	8,858	May 2025
6	OpenAI	GPT-4o	1,087	-10/+10	8,468	Mar 2025
7	Alibaba	Qwen-Image-Edit	1,084	-9/+10	9,784	Aug 2025
8	ByteDance Seed	SeedEdit 3.0	1,083	-10/+11	7,951	Jun 2025
9	Black Forest Labs	FLUX.1 Kontext [dev]	1,004	-10/+11	8,905	Jun 2025
10	Google	Gemini 2.0 Flash Preview	1,000	+0/+0	8,643	May 2025

来源: Artificial Analysis, 国金证券研究所 (注: 数据及信息截止至 2025 年 10 月)

图表31: Artificial Analysis 文生视频榜单上阿里旗下模型位居全球第五

Rank	Creator	Model	ELO	95% CI	Appearances	Release Date
1	Kuaishou KlingAI	Kling 2.5 Turbo 1080p	1,246	-12/+13	3,829	Sept 2025
2	Google	Vevo 3 (No Audio)	1,229	-9/+9	8,780	Jul 2025
3	Luma Labs	Ray 3	1,219	-12/+11	4,162	Sept 2025
4	MiniMax	Hailuo 02 Standard	1,205	-10/+9	6,491	Jun 2025
5	Alibaba	Wan 2.5 Preview	1,190	-12/+13	3,797	Sept 2025
6	ByteDance	Waver 1.0	1,187	-7/+6	13,548	Aug 2025
7	Google	Vevo 3 Fast Preview (No Audio)	1,187	-7/+7	14,522	Jun 2025
8	MiniMax	Hailuo 02 Pro	1,185	-9/+10	6,685	Jun 2025
9	PixVerse	PixVerse V5	1,184	-10/+9	6,659	Aug 2025
10	ByteDance Seed	Seedance 1.0	1,174	-9/+10	6,126	Jun 2025

来源: Artificial Analysis, 国金证券研究所 (注: 数据及信息截止至 2025 年 10 月)

2.3 ARPU: 视频 > 图像, 海外 > 国内, AI 增值功能有望增厚 ARPU

从会员套餐的价格与权益维度来看,公司旗下的影像类 APP 呈现差异化布局策略。美图秀秀 SVIP 月度会员/季度会员价格为 38/108 元,权益涵盖海报模板商业使用、多端通用、无广告无水印及畅享美图秀秀 VIP 全部权益等;美图秀秀 VIP 连续包月/月度会员价格为 15/22 元,提供美化素材、视频美容、相机特效等权益;美颜相机 VIP 月度会员/季度会员价格为 25/60 元,包含精简广告、医美级美颜、正版授权 IP 等权益。



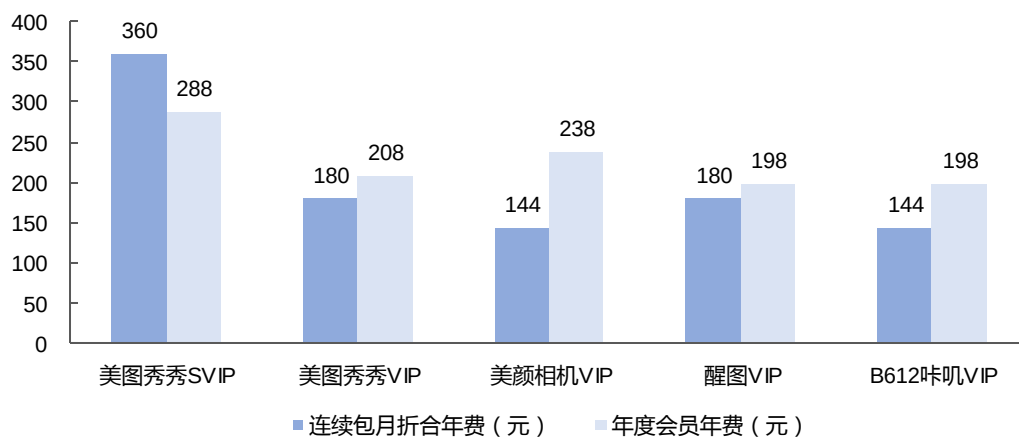
图表32: 国内主流图片编辑类工具APP会员价格及权益内容

套餐	美图秀秀 SVIP	美图秀秀 VIP	美颜相机 VIP	醒图 VIP	轻颜相机	Faceu 激萌	B612 咔叽 VIP
所属大厂	美图公司			字节跳动			亿睿科信息技术有限公司
会员价格	·连续包月:30 ·月度会员:38 ·连续包季:68 ·季度会员:108 ·连续包年:218 ·年度会员:288	·连续包月:15 ·月度会员:22 ·连续包年:128 ·年度会员:208	·连续包月:12 ·月度会员:25 ·连续包季:30 ·季度会员:60 ·连续包年:98 ·年度会员:238	·连续包月:15 ·月度会员:25 ·连续包年:118 ·年度会员:198	-	-	·连续包月:12 ·月度会员:25 ·连续包年:98 ·年度会员:198
权益内容	海报模板商业使用;会员多端通用;精美商业设计模板;每日更新设计模板;无广告无水印;畅享美图秀秀 VIP 全部权益	美化素材;视频美容;相机特效;高清人像;更多创意玩法等	精简广告;装饰素材;正版授权 IP;医美级美颜;视频创作神器等	海量素材;一件美化;面部重塑;面部丰盈;提拉去皱;画面调节;免广告;专属客服等	装饰素材;美颜滤镜;AI写真等	火爆贴纸;至美滤镜;生成漫画脸;视频拍摄等	超清画质;背景锁定;妆容滤镜;美体塑形;一键美颜;AI发型;AI写真等

来源: 美图秀秀 APP, 美颜相机 APP, 醒图 APP, 轻颜相机 APP, Faceu 激萌 APP, B612 咔叽 APP, 国金证券研究所

国内主流图片编辑类工具 APP 年费约 100-300 元, 公司 ARPU 约 160-170 元。公司旗下 APP 美图秀秀 SVIP 连续包年/年度会员价格为 218/288 元, 定价较高导致付费率偏低; 美图秀秀/美颜相机 VIP 年度会员分别为 208/238 元。国内竞品方面, 字节跳动旗下醒图 VIP 年度会员 198 元, 亿睿科旗下 B612 咔叽 VIP 年度会员 198 元, 与公司常规 VIP 定价区间趋同。整体来看, 公司常规类 VIP 与竞品在百元级年费区间形成直接竞争, 而高阶 SVIP 因定价较高, 付费率表现相对有限。

图表33: 国内主流图片编辑类工具APP年费约在100-300元



来源: 美图秀秀 APP, 美颜相机 APP, 醒图 APP, 轻颜相机 APP, Faceu 激萌 APP, B612 咔叽 APP, 国金证券研究所

从海外影像类 APP 的会员价格及权益维度看, 公司旗下 Airbrush 的会员价格分为多个梯度, 月度会员价格为 6.99/7.99/10.99 美元, 年度会员价格为 28.99/54.99 美元, 连续包年价格为 35.00/43.99 美元, 权益涵盖图像/视频编辑、在线化妆编辑、照片修饰等多元功能; Beautyplus 月度会员价格为 4.49/7.99/8.99 美元, 年度会员价格为 29.99/49.99 美元, 聚焦照片艺术风格转换、AI 滤镜及视频增强等特色权益。



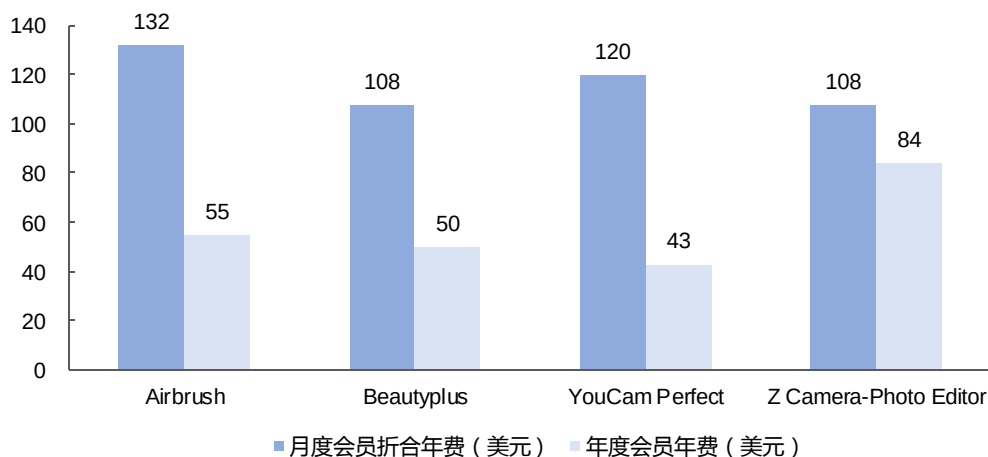
图表34: 国外主流图片编辑类工具APP 会员价格及权益内容

套餐	airbrush	beautyplus	YouCam Perfect	Z Camera-Photo Editor
所属大厂	美图公司		完美公司	Funnytech Corporation 趣科技公司
会员价格 (美元)	·月度会员:6.99/7.99/10.99 ·年度会员:28.99/54.99 ·连续包年:35.00/43.99	·月度会员:4.49/7.99/8.99 ·年度会员:29.99/49.99	·月度会员:5.99/8.49/9.99/32.99 ·年度会员:29.99/31.99/39.99/42.99	·周度会员:4.99/5.49/5.99/8.99 ·月度会员:12.99/19.99 ·年度会员:45.99/59.99/83.99
权益内容	图像/视频编辑;在线化妆编辑;照片修饰;移除物体/乘客;模糊背景;增强照片的锐利度和专业感;护照照片编辑器应用程序等	照片艺术风格转换;动漫、粘土、像素、蜡笔和卡通等 AI 滤镜; AI 照片编辑器; 增强视频; 提升图像质量等	物体移除/擦除; AI 换脸; AI 重拍; 瘦脸瘦身 增强照片效果 场景布局拼贴; 滤镜/贴纸等	最新贴纸和 AR 贴纸; 一键美化自拍照; 换脸功能; 时尚发型编辑器; 特殊拼贴模板; 化妆效果风格; 肌肉锻炼/ 3D 纹身效果等

来源: Airbrush App, BeautyplusApp, YouCam PerfectApp, Z Camera-Photo Editor App, Adapty, 国金证券研究所

国外主流图片编辑类工具 APP 年费约 300-1000 元。从定价维度看, 公司旗下海外图片编辑工具 Airbrush/BeautyPlus 年度会员为 55/50 美元; 国际竞品 YouCam Perfect/Z Camera-Photo Editor 年度会员为 43/84 美元。海外图片编辑类工具 APP 会员年费经汇率折算后显著高于国内水平。

图表35: 国外主流图片编辑类工具 APP 年费约在 40-140 美元(折合约 300-1000 元)



来源: Airbrush App, BeautyplusApp, YouCam PerfectApp, Z Camera-Photo Editor App, Adapty, 国金证券研究所

主流图片编辑类工具 APP 的 AI 功能普遍采用按使用量计费模式, 存在增量付费空间。以美图秀秀、轻颜、B612 咔叽等产品为例, 其 AI 功能在图片美化处理、形象照制作等场景中, 通过单次付费、VIP/SVIP 权益内的功能使用限制等模式实现商业化。按使用量计费的策略, 在订阅制基础上, 能够进一步挖掘用户对 AI 功能的付费潜力, 丰富收入结构, 为 ARPU 增长开辟新路径。

图表36: 国内主流图片编辑类工具APP 的AI 功能普遍采取按使用量计费模式

产品	分类	图片美化/处理	形象照制作	AI	
美图秀秀 美颜相机	具体功能	滤镜/贴纸	写真 证件照 形象照	AI 绘画	图生图、文生图、头像制作、线稿上色、涂鸦生图
		美颜/美妆		AI 设计	AI 海报、AI 写真、AI Logo、百变 AI 头像、AI 绘画、AI 捏捏特效、AI 消除
		拼图		其他	AI 动漫、AI 扩图、AI 图文、AI 文案、AI 换背景、AI 模特
	收费模式	VIP 权益:部分滤镜贴纸 美妆特效	单次付费:写真/证件照:8.8 元/12.8 元/次 形象照:6.9 元/次, 10.9 元/3 次, 25.9 元/10 次	单次付费:百变 AI 头像, 1 张 50 美豆(1 元为 7 美豆) SVIP 权益:AI 设计功能, 部分功能限时免费 VIP 权益:AI 扩图(每日免费 3 次)	



轻颜	具体功能	滤镜/贴纸、风格拍摄		AI 生图	AI 写真
	收费模式	-		单次付费:9.9元/10次	
B612 咔叽	具体功能	滤镜/贴纸、风格拍摄		AI 设计	AI 漫画脸生成
	收费模式	-			
妙鸭相机	具体功能	高清图质、发型设计、面部修复、精修写真	写真 证件照 艺术照	AI 消除、AI 修脸、色彩测试	
	收费模式	单次付费	单次付费:9.9元/张	专家模式:9.9元/次	
醒图	具体功能	美颜/美妆、拼图、批量修图、超清画质、抠图	写真	AI 写真、AI 消除、AI 替换、AI 商品图	
	收费模式	VIP 权益:部分滤镜贴纸 美妆特效, 超清画质	单次付费:11.8元/张	VIP 权益:AI 写真免费生成	

来源: 美图秀秀 APP, 美颜相机 APP, 醒图 APP, 轻颜相机 APP, B612 咔叽 APP, 国金证券研究所

从国内视频编辑类 APP 的会员价格及权益维度看, 公司旗下的 Wink APP 连续包月/连续包年/年度会员价格分别为 18/168/218 元, 权益涵盖图像/画质修复、AI 美容、人像增强、身材美型等多元功能。

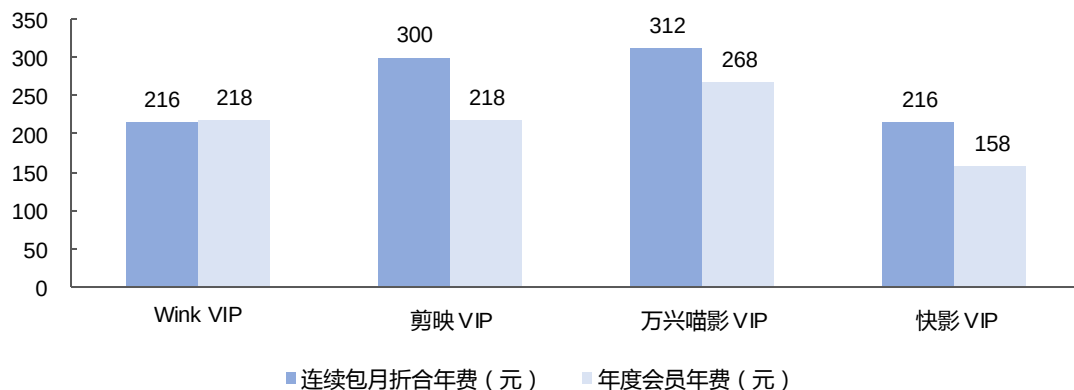
图表37: 国内主流视频编辑类工具 APP 会员价格及权益内容

套餐	Wink VIP	剪映 VIP	万兴喵影 VIP	快影 VIP
所属大厂	美图公司	字节跳动	万兴科技	快手科技
会员价格	连续包月:18 连续包年:168 年度会员:218	连续包月:25 月度会员:30 连续包年:188 年度会员:218	连续包季:78 连续包年:178 年度会员:268	连续包月:18 月度会员:30 连续包年:128 年度会员:158
权益内容	全能修复; 画质修复; AI 美容; 人像增强; 身材美型等	精选模板自由剪辑; 智能识别加字幕; 视频画质感升级; 人像精修美颜塑 形; 轻量 AI 创作	自定义水印; 动态贴纸; 千首经典音 乐; 质感滤镜; 不限视频时长; 多条 轨道视频; 无广告; 资源持续更新	会员专属素材; 精选付费模板; 解锁 模板草稿; 会员专属配音; 会员专属 装扮; 5G 云空间; 更多会员 AI 玩法

来源: WinkAPP, 剪映 APP, 万兴喵影 APP, 快影 APP, 国金证券研究所

国内主流视频编辑类工具 APP 年费约在 200-300 元。公司 Wink/字节跳动剪映/万兴科技万兴喵影/快手科技快影 VIP 年度会员费分别为 218/218/268/158 元。国内主流视频编辑类工具 APP 通过多元会员套餐, 满足用户付费需求, 形成了差异化的定价策略以适配不同用户的消费能力与使用需求。

图表38: 国内主流视频编辑类工具 APP 年费约在 100-300 元



来源: WinkAPP, 剪映 APP, 万兴喵影 APP, 快影 APP, 国金证券研究所



国内主流图片编辑类工具 APP 的 AI 功能普遍采用按使用量计费模式，为 ARPU 的提升提供新路径。以 Wink、剪映、万兴喵影、快影等产品为例，其 AI 功能在视频生成/编辑、图片制作等场景中，通过单次付费、VIP 权益内的功能使用限制（如免费次数、积分兑换）等模式实现商业化。这种按使用量计费的策略，在订阅制基础上，通过 AI 功能的单点付费环节，可进一步挖掘用户付费潜力，丰富收入结构，从而推动 ARPU 的提升。

图表39：国内主流视频编辑类工具APP的AI功能普遍采取按使用量计费模式

产品	分类	视频生成/编辑	图片制作	AI	
Wink	具体功能	视频美容、视频剪辑、画质修复	跳转至美图秀秀	AI 视频	AI 消除、AI 扩展、AI 超清、AI 美容、AI 动图、AI 绘画、AI 动漫
	收费模式	VIP 权益:部分滤镜贴纸视频特效，人像增强		VIP 权益:AI 动漫免费生成 10 秒，AI 扩图、视频修复等含免费次数 单次付费:AI 动漫 30 秒/12 元，60 秒/20 元	
剪映	具体功能	美颜美体、视频美化、提词拍摄、超清视频、视频变速	视频转拼图、超清图片、智能抠图、图片编辑	AI 视频	AI 图文成片、AI 故事成片、AI 剪视频、数字人、AI 剪口播
	收费模式	VIP 权益:部分滤镜贴纸视频特效，美颜美体功能	VIP 权益:部分智能抠图效果，AI 换背景、画质提升功能	单次付费：根据视频长度消耗对应 AI 积分（VIP 免费享有 1200 积分/月）	
万兴喵影	具体功能	创建视频、视频剪辑、添加字幕、视频变速、平面追踪、多机位剪辑、屏幕录制、无损剪切	-	AI 剪辑	智能扣人像、视频降噪、精准仿色、智能人脸马赛克、AI 图像、AI 音效、AI 视频翻译、高光卡点
	收费模式	VIP 权益:部分滤镜贴纸视频特效，水印编辑功能		单次付费:根据转换难度进行积分兑换（VIP 享有 3000 积分/年）	
快影	具体功能	一键出片、混剪视频、音乐 MV、游戏大片	图文编辑	AI 视频	AI 故事成片、AI 小说转漫画、AI 创作、AI 数字人口播、AI 文案成片
	收费模式	VIP 权益:部分滤镜贴纸视频特效编辑功能	VIP 权益:部分滤镜贴纸视频特效、美颜美体功能	VIP 权益:AI 数字人、小说转漫画功能畅享	

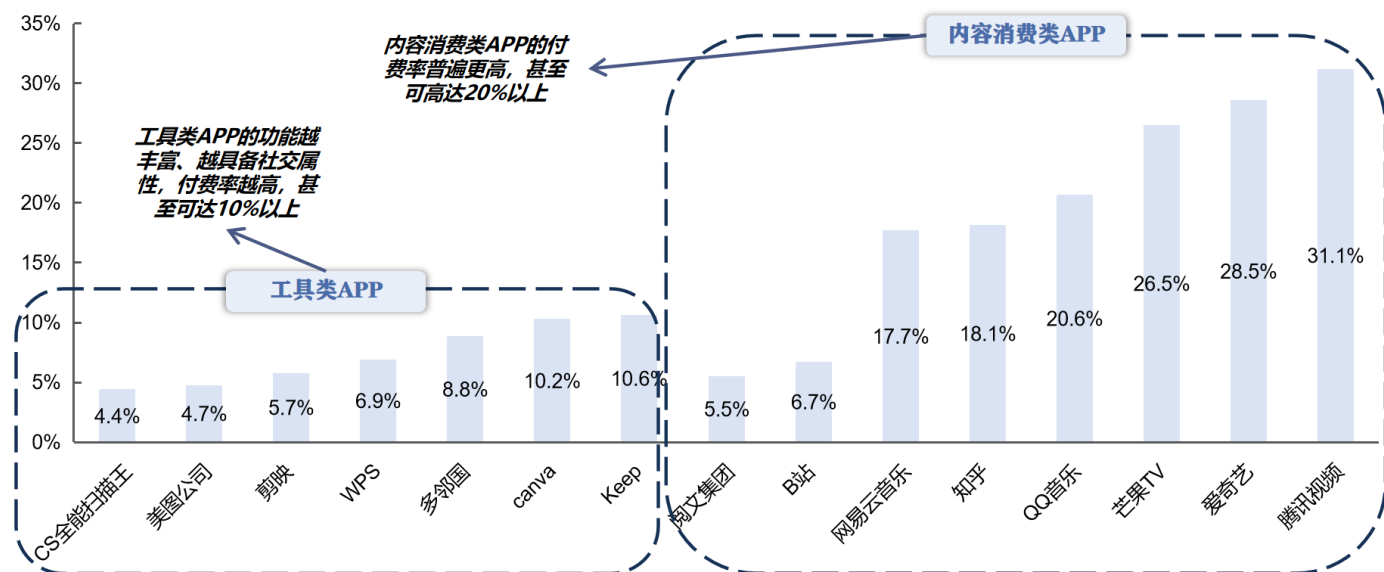
来源：美图秀秀 APP，美颜相机 APP，醒图 APP，轻颜相机 APP，B612 咔叽 APP，国金证券研究所

2.4 付费率：“功能厚度+社交属性”可将工具类 APP 付费率上限推升至 10+%

工具类与内容消费类 APP 付费率表现差异显著，工具类 APP 付费率普遍低于内容消费类 APP。内容消费类 APP 付费率整体可达 20+%，腾讯视频/爱奇艺/芒果 TV 付费率分别为 31.1%/28.5%/26.5%；工具类 APP 付费率整体可达 10+%，功能越丰富、具备社交属性的产品付费率越高，Keep/Canva/多邻国付费率分别为 10.6%/10.2%/8.8%，公司付费率仅为 4.7%。基于此，公司生活场景应用与生产力工具的付费率存在较大提升空间，未来可通过强化功能丰富度与社交属性，推动付费率向工具类 APP 付费率上限靠拢。



图表40：具备社交属性的生活场景应用与生产力工具付费率可达10+%，内容消费类APP付费率可达20+%



来源：各APP界面，各APP背后上市公司财报，界面新闻微信公众平台，上海证券报，Demandsage官网，国金证券研究所

公司订阅收入极限或在100-225亿元区间。1H2025，公司旗下产品MAU合计2.8亿人，付费率约为5.5%，ARPU值约在176元。我们假设，公司未来将拥有2亿国内生活场景应用、2亿海外生活场景应用及1亿生产力工具构成的中长期MAU，以此为锚对ARPU和付费率进行双因子敏感性分析。从ARPU维度，假设国内生活场景应用/海外生活场景应用/生产力工具ARPU分别为150/300/250元，按权重加权得ARPU预测值为230元；付费率维度，假设国内生活场景应用/海外生活场景应用/生产力工具付费率分别为10%/15%/13%，按权重加权得付费率预测值为13%。根据三因子模型，远期订阅收入约为150亿元。根据ARPU和付费率的变动，订阅收入极限或在100-225亿元之间。

图表41：公司订阅收入的极限或在100-225亿元之间

付费率/ARPU	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300
8.0%	60	64	68	72	76	80	84	88	92	96	100	104	108	112	116	120
8.5%	64	68	72	77	81	85	89	94	98	102	106	111	115	119	123	128
9.0%	68	72	77	81	86	90	95	99	104	108	113	117	122	126	131	135
10.0%	75	80	85	90	95	100	105	110	115	120	125	130	135	140	145	150
10.5%	79	84	89	95	100	105	110	116	121	126	131	137	142	147	152	158
11.0%	83	88	94	99	105	110	116	121	127	132	138	143	149	154	160	165
11.5%	86	92	98	104	109	115	121	127	132	138	144	150	155	161	167	173
12.0%	90	96	102	108	114	120	126	132	138	144	150	156	162	168	174	180
12.5%	94	100	106	113	119	125	131	138	144	150	156	163	169	175	181	188
13.0%	98	104	111	117	124	130	137	143	150	156	163	169	176	182	189	195
13.5%	101	108	115	122	128	135	142	149	155	162	169	176	182	189	196	203
14.5%	109	116	123	131	138	145	152	160	167	174	181	189	196	203	210	218
15.0%	113	120	128	135	143	150	158	165	173	180	188	195	203	210	218	225

来源：国金证券研究所

三、盈利预测与投资建议

3.1 盈利预测

订阅收入：公司订阅收入=MAU*付费渗透率*ARPU，我们预计25/26/27年公司订阅收入



分别为 30.3/39.8/51.7 亿元，同比增速为 45.3%/31.5%/29.9%。其中：

1) MAU: 公司 1H25 MAU 约为 2.8 亿，考虑到公司出海业务的持续拓展及生产力工具领域存在的广阔市场空间，我们预计 25/26/27 年公司 MAU 维持稳步增长，分别为 3.0/3.3/3.5 亿人。

2) 付费渗透率: 公司 1H25 的付费率约为 5.5%，考虑到公司在生活场景应用与生产力工具的付费率存在较大提升空间，我们预计 25/26/27 年公司付费率仍将维持提升趋势，分别为 6.0%/7.0%/8.0%。

3) ARPU: 公司 1H25 的 ARPU 值为 176 元/人，考虑公司在订阅制基础上进一步挖掘用户对 AI 功能的付费潜力，我们预计 25/26/27 年公司 ARPU 值有望持续提升，分别为 170/175/185 元。

美业解决方案收入: 美业解决方案业务属于非核心战略业务且毛利率较低，预计未来 3 年保持收缩趋势。我们预计 25/26/27 年公司美业解决方案收入分别为 0.8/0.7/0.7 亿元，同比增速为 -80.0%/-10.0%/-5.0%。

广告收入: 公司持续优化程序化广告运营，同时优化广告质量、投放精准度，同时将更重视产品订阅化收入转向，预计广告收入未来 3 年保持持续小幅增长态势。我们预计 25/26/27 年公司广告收入分别为 9.0/9.4/9.9 亿元，同比增速为 5.0%/5.0%/5.0%。

其他收入: 其他业务主要为与“生产力和全球化战略”无关的传统业务，预计未来 3 年保持稳定态势。我们预计 25/26/27 年公司其他收入分别为 0.2/0.2/0.2 亿元，同比增速为 0%/0%/0%。

图表42: 25/26 年有望实现约40/50 亿元营收，同比增长 20%/25%

	2021A	2022A	2023A	1H2024	2024A	1H2025	2025E	2026E	2027E
营业总收入 (亿元)	16.7	20.9	27.0	16.2	33.4	18.2	40.2	50.1	62.5
YOY	39.5%	25.2%	29.3%	28.6%	23.9%	12.3%	20.4%	24.6%	24.6%
影像与设计产品收入 (亿元)	5.9	8.7	13.3	9.3	20.9	13.5	30.3	39.8	51.7
YOY	123.7%	47.0%	52.8%	/	57.1%	45.2%	45.3%	31.5%	29.9%
MAU (万人)	23,064	24,288	24,916	25,773	26,583	28,000	29,713	32,525	34,963
MAU_YOY	-11.6%	5.3%	2.6%	/	6.7%	8.6%	11.8%	9.5%	7.5%
其中: 国内	15,388	16,312	17,147	17,300	17,132	18,200	18,845	20,353	21,574
其中: 海外	7,676	7,976	7,768	8,473	9,451	9,800	10,868	12,172	13,389
ARPU (元/人)	148	155	146	172	165	176	170	175	185
ARPU_YOY	11.9%	5.0%	-6.1%	/	13.5%	1.9%	2.8%	2.9%	5.7%
付费用户渗透率	1.7%	2.3%	3.7%	4.2%	4.7%	5.5%	6.0%	7.0%	8.0%
付费用户渗透率_YOY	1.0pct	0.6pct	1.4pct	/	1.1pct	1.3pct	1.3pct	1.0pct	1.0pct
美业解决方案 (亿元)	0.4	4.4	5.7	2.7	3.8	0.3	0.8	0.7	0.7
YOY		1036.7%	29.1%	/	-32.4%	-88.9%	-80.0%	-10.0%	-5.0%
广告 (亿元)	7.7	6.3	7.6	4.1	8.5	4.3	9.0	9.4	9.9
YOY	12.5%	-17.8%	20.5%	/	12.5%	5.0%	5.0%	5.0%	5.0%
其他 (亿元)	2.7	1.5	0.4	0.1	0.2	0.1	0.2	0.2	0.2
YOY	8.5%	-46.1%	29.3%	/	-57.7%	-12.4%	0.0%	0.0%	0.0%

来源: 公司财报, 国金证券研究所 (注: 表格中标黄部分为预测值。)

毛利率: 综合考虑影像与设计产品订阅收入占比提升的规模效应与高毛利广告业务的占比降低，预计 25/26/27 年公司综合毛利率将维持在 74.0%/74.0%/74.0%。

费用率方面，我们认为：

1) 研发费用率: 公司持续保持研发投入，Gen AI 相关的费用小幅增加，同时受益规模效应费用率小幅下降。我们预计 25/26/27 年公司研发费率分别为 25.0%/23.0%/22.0%。

2) 销售及营销费用率: 公司国外市场推广主要用于生活场景应用，而中国内地市场推广增长主要用于推广生产力工具。预计影像与设计产品国内外营销开支同比上升，销售费用



率维持在较高比率。我们预计 25/26/27 年公司销售费率分别为 16.0%/16.0%/16.0%。

3) 行政费用率: 公司收入快速增长, 规模效应下管理费用率小幅下降。我们预计 25/26/27 年公司管理费率分别为 11.0%/10.0%/9.3%。

利润端, 我们预计 25/26/27 年归母净利润分别为 8.6/11.2/15.1 亿元, 经调整归母净利润分别为 9.6/12.2/16.1 亿元。

图表43: 利润预测: 2026 年有望维持综合毛利率 74%, 经调整归母净利润 12.2 亿元

	2021A	2022A	2023A	1H2024	2024A	1H2025	2025E	2026E	2027E
综合毛利率	67.5%	56.9%	61.4%	64.9%	68.7%	73.6%	74.0%	74.0%	74.0%
销售及营销费用率	19.3%	19.3%	15.9%	12.6%	14.5%	16.1%	16.0%	16.0%	16.0%
行政费用率	15.9%	13.0%	11.2%	12.0%	12.0%	12.0%	11.0%	10.0%	9.3%
研发费用率	28.1%	23.6%	26.2%	27.3%	27.3%	24.8%	25.0%	23.0%	22.0%
归母净利润 (亿元)	-0.4	0.9	3.8	3.0	8.1	4.0	8.6	11.2	15.1
归母净利率	-2.7%	4.5%	14.0%	18.7%	24.1%	21.8%	21.5%	22.4%	24.2%
经调整归母净利润 (亿元)	0.9	-0.3	3.7	2.7	5.9	4.7	9.6	12.2	16.1
YoY	39.7%			80.4%	59.2%	71.3%	64.3%	26.7%	31.8%
经调整归母净利率	5.1%	-1.4%	13.7%	16.8%	17.5%	25.7%	24.0%	24.4%	25.8%

来源: 公司财报, 国金证券研究所 (注: ①表格中标黄部分为预测值; ②经调整部分主要系加密资产处置损益与以股权支付的费用。)

3.2 投资建议及估值

目前, H 股市场中尚无与公司主要产品完全相同的上市公司, 考虑到公司所处细分行业为影像与设计类软件, 主要涵盖拍摄、修图修视频、社交分享、商业设计、视觉创作、商业摄影、专业视频编辑等功能的生活场景与生产力场景工具, 因此我们选取了奥多比、万兴科技、合合信息、金山办公、KEEP 作为公司的同行业可比公司。以 26 年业绩为锚, 行业可比公司的 PE 中位数接近 40 倍, PEG 中位数接近 2 倍。

综合考虑上述五家公司的估值水平, 若参考行业可比公司 40X PE, 公司对应估值 490 亿元人民币, 目标价 11.73 港元, 给予“买入”评级。

图表44: 可比公司 26 年 PE 中位数接近 40X, PEG 接近 1.4X

证券代码	证券简称	市值 (亿元)	净利润 (亿元)					PE					PEG		
			2023A	2024A	2025E	2026E	2027E	2023A	2024A	2025E	2026E	2027E	2025E	2026E	2027E
ADBE.O	奥多比	9,268	386.0	399.6	510.4	572.7	660.0	33.1	32.0	18.2	16.2	14.0	0.65	1.33	0.92
300624.SZ	万兴科技	149	0.9	-1.6	0.0	1.4	1.6	151.1	-	451.2	151.6	93.6	-	(0.05)	4.55
688615.SH	合合信息	273	3.2	4.0	4.7	5.5	6.8	/	50.5	45.6	37.5	32.0	2.71	2.04	1.37
688111.SH	金山办公	1,416	13.2	16.5	18.5	21.5	25.0	110.8	80.5	76.7	66.0	56.6	6.25	4.06	3.43
3650.HK	KEEP	19	11.1	-5.3	0.1	0.7	1.5	6.0	(5.3)	167.6	28.0	12.9	(1.64)	0.06	0.11
中位数								71.9	50.5	61.1	37.5	32.0	2.71	2.04	1.37
1357.HK	美图公司	324	3.7	5.9	9.6	12.2	16.1	40.7	21.1	33.7	26.6	20.2	0.52	0.99	0.63

来源: 公司财报, 国金证券研究所 (注: ①截至 2025 年 11 月 21 日收盘; ②万兴科技、合合信息、金山办公盈利预测来自国金计算机团队, 奥多比、KEEP 盈利预测来自 iFinD; ③2024 年可比公司 PE 中位数已剔除 KEEP 的负值, 2025 年 PE 中位数已剔除万兴科技的 451X, 2025、2026 年 PEG 中位数已剔除万兴科技、KEEP 的负值。)

四、风险提示

- 影像编辑类工具软件被 AI 基模吞噬的风险。



全球多模态基模能力日趋成熟，伴随 NanoBanana、Sora 等爆款“基模即应用”出圈，或存在多模态 AI 基模吞噬应用的风险；对此我们认为市场存在“容错率误区”与“自然语言交互误区”，封装成熟工作流的工具软件仍具有其价值。

■ 大厂入局导致竞争加剧的风险。

近期国内大厂 AI 进展迅速，或增强大厂竞品的竞争优势；对此我们认为公司核心优势在于十数年积累的数据与审美 knowhow，且有自研 AI 垂类模型，并能够借助全球第一梯队的阿里基模能力赋能旗下产品。

■ 手机厂商入局导致竞争加剧的风险。

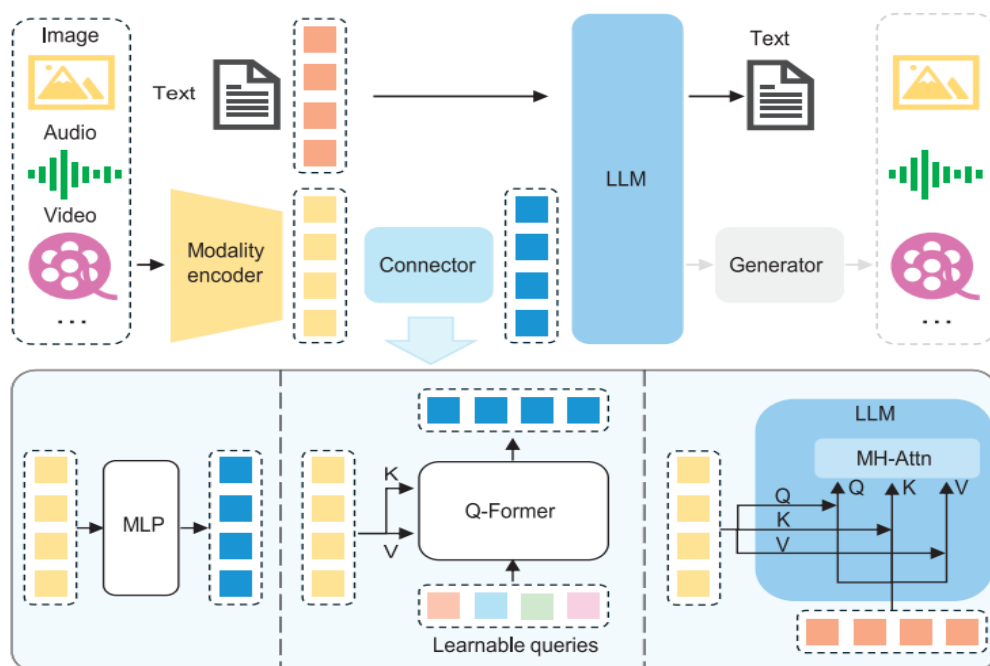
拍照修图功能为智能手机核心卖点之一，若智能手机厂商投入资源结合 AI 开发适配其硬件的影像设计类工具，或对美图系产品的市场份额产生扰动。

五、附录：多模态模型发展脉络及技术演变

5.1 多模态模型：从跨模态协同迈向原生融合，技术架构持续突破

多模态模型是处理图像、音频、视频等非文本模态数据，融合多模态信息实现理解与生成的 AI 模型，其核心架构包含编码器、连接器与大语言模型（LLM），还可选择性附加生成器以输出非文本模态内容。其中，编码器负责将图像、音频或视频等输入转化为模型可处理的特征；连接器作为关键中间模块，分为基于投影、基于查询与基于融合三类，以适配 LLM 的语义理解逻辑；LLM 则承担信息整合、推理及文本生成的核心任务，附加的生成器可进一步拓展其输出范围，实现图像、音频等非文本模态内容的生成。

图表45：一类典型的多模态模型结构包括编码器、连接器与大语言模型（LLM）

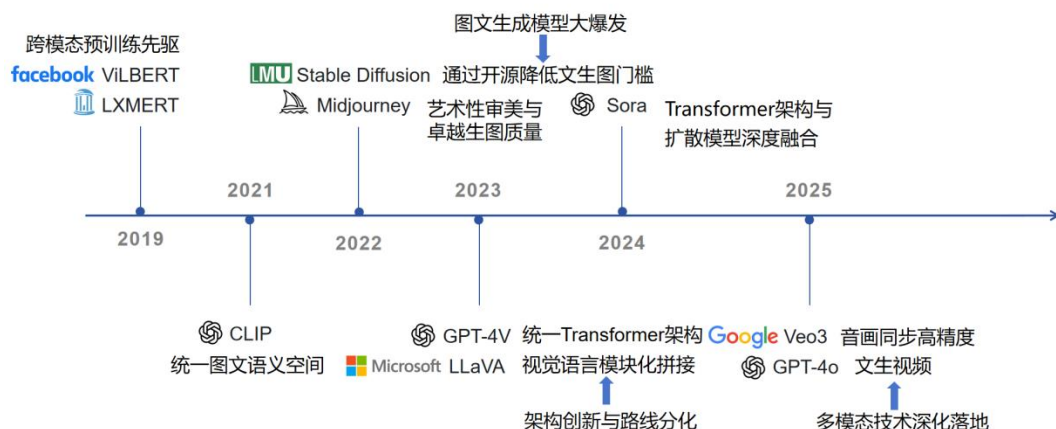


来源：Liang Z 等《A survey of multimodal large language models》(2024)，国金证券研究所

多模态模型的发展脉络包含奠基探索、架构创新与技术分化、统一融合与模态泛化三个阶段。2019-2022 年，以 ViLBERT、LXMERT 等跨模态预训练模型为开端，CLIP 实现图像与文本在统一语义空间的映射，由 Stable Diffusion 和 Midjourney 引爆图文生成应用的爆发式增长。2023 年，随着技术能力的成熟，模型架构迎来创新与路线分化。GPT-4V 探索了端到端的统一 Transformer 架构，LLaVA 则采用“模块化拼接”策略，形成两条并行的技术探索路径。2024 年至今，技术路线开始重新走向融合，Sora 通过将 Transformer 与扩散模型深度结合，将生成能力从静态图像泛化至复杂的动态视频。Google Veo3 和 GPT-4o 等模型聚焦于音画同步和高精度生成，标志着多模态技术正迈向更为全面、深入的应用落地新纪元。



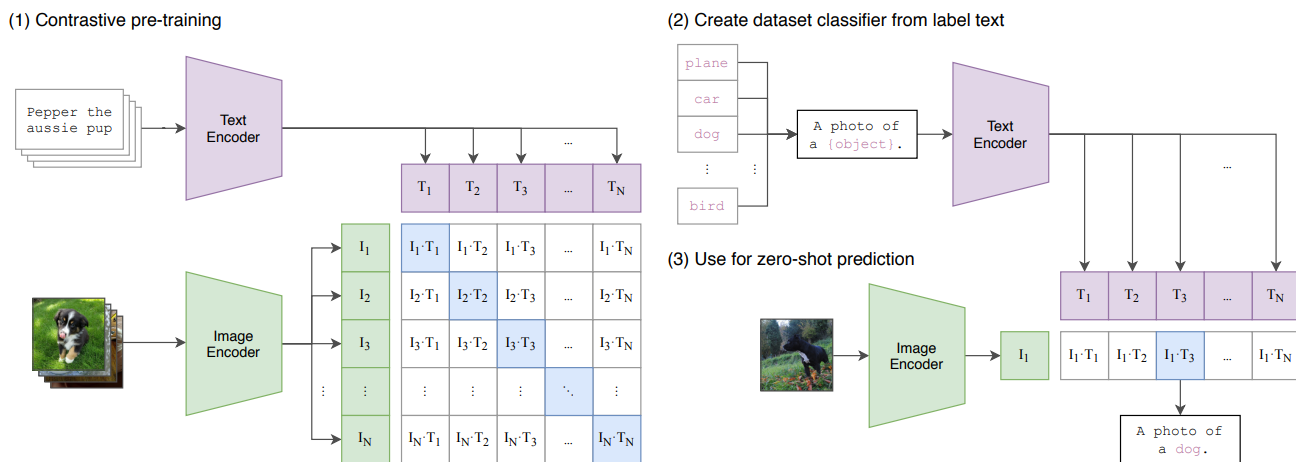
图表46：多模态模型关键技术演进历程（2019-2025 年）



来源：Hao Tan 等《LXMERT: Learning Cross-Modality Encoder Representations from Transformers》(2019), Jiasen Lu 等《Visual Instruction Tuning》(2019), Alec Radford 等《Learning Transferable Visual Models From Natural Language Supervision》(2021), Stable Diffusion 官网, Midjourney 官网, OpenAI GPT-4V 系统卡, Haotian Liu 等《Visual Instruction Tuning》(2023), Sora 官网, OpenAI GPT-4o 系统卡, Thaddäus Wiedemer 等《Video models are zero-shot learners and reasoners》(2025, 百度百科, 国金证券研究所)

多模态模型的崛起，依托于神经网络架构的底层突破，更源于“跨模态信息协同”的技术创新。在多模态大模型的起源探索阶段，佐治亚理工学院、俄勒冈州立大学与 Facebook 联合推出的 ViLBERT 和北卡罗来纳大学教堂山分校推出的 LXMERT 是多模态预训练的先驱模型，上述模型开创性地采用 Transformer 架构实现视觉与语言的联合预训练，首次从技术层面验证了跨模态信息深度交互的可行性；2021 年 OpenAI 推出的 CLIP 不再满足于简单的特征拼接，通过 40 亿规模“文本-图像”的对比预训练，实现了图像与文本在统一语义空间的映射，突破性达成零样本图像分类能力，彻底打开了多模态大模型“以自然语言理解图像、以图像反哺语言语义”的双向交互大门，成为多模态领域具有里程碑意义的标志性研究。在 CLIP 构建的强大图文理解能力之上，生成模型迎来了大爆发。2022 年 8 月 CompVis 正式发布 Stable Diffusion，通过开源极大降低了高质量文生图（AIGC）的门槛，引发全球性创作热潮；而 2022 年 11 月 Midjourney 发布的 Midjourney V4 模型，则凭借其对复杂、艺术性提示语的深刻理解和卓越的图像生成质量，创造富有想象力和风格化的艺术作品。

图表47：CLIP 联合训练图像编码器和文本编码器来预测一批（图像，文本）训练示例的正确配对



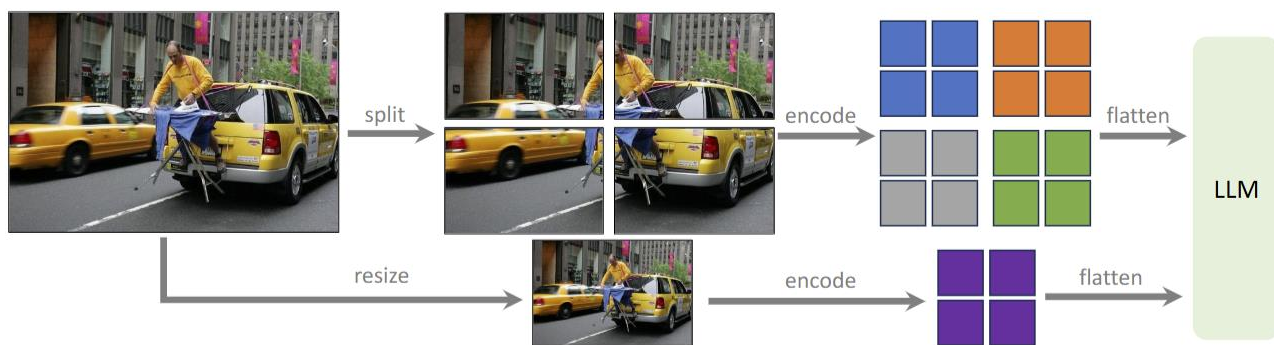
来源：Yanxu Mao 等《From LLMs to MLLMs to Agents: A Survey of Emerging Paradigms in Jailbreak Attacks and Defenses within LLM Ecosystem》(2015), 国金证券研究所

2023 年是架构创新与路线分化的关键节点，随着基础能力的成熟，业界开始探索通往更强大多模态智能的不同路径，形成了两条主流技术路线。Open AI 推出的 GPT-4V 凭借“统一 Transformer 架构”，摒弃传统“视觉编码器+语言解码器”的拼接模式，将视觉与语



言 Token 在同一注意力矩阵中交互，旨在构建一个原生多模态的大脑，从而实现从像素级细节到高级语义概念的全面理解。LLaVA 则采用“模块化拼接”策略，以 CLIP-VT 作为视觉编码器、Vicuna-LLM 作为语言解码器，通过“投影层对齐预训练+指令微调”的两阶段训练策略，将图像分割成网格并独立编码以提高分辨率，仅新增不到 5% 的参数数量就实现了完整的视觉对话能力，为多模态大模型的低成本研发与开源生态建设提供了典型范式。在架构探索的同时，Google Gemini 将重点放在了模态的广度上，支持文本、图像、音频、视频、代码等多种模态的输入和处理，Gemini 不再局限于图文，而是尝试理解视频中的物理规律，或结合音频与文本生成内容，将多模态交互推向了更复杂的现实场景。

图表48：通过将图像分割成网格并独立编码，将 LLaVA-1.5 扩展到更高分辨率

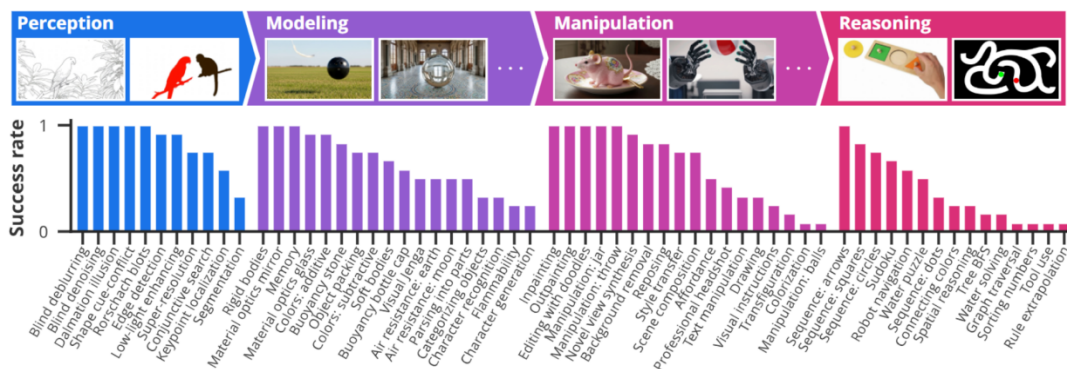


来源：Haotian Liu 等《Improved Baselines with Visual Instruction Tuning》(2024)，国金证券研究所

2024 年至今，分化的技术路线开始重新交汇，业界的目标愈发明确：打造能够端到端处理一切模态、实现无缝人机交互的终极模型，多模态大模型技术进入统一融合与全模态泛化阶段。

- 视频生成与物理世界模拟：2024 年 2 月，OpenAI 推出的 Sora 模型将 Transformer 架构与扩散模型深度融合，进一步强化了文本到视频和图像生成的功能，可生成长达 60 秒的 1080P 高清视频，并能模拟复杂场景和物理规律，实现了从静态图像生成到动态世界模拟的飞跃。
- 多模态原生交互：2024 年 5 月，OpenAI 推出的 GPT-4o 重点解决了 GPT-4V 在效率、精度与多模态兼容性上的痛点，真正实现了文本、音频、图像、视频的多模态原生处理，通过单一神经网络端到端支持多模态任意组合的输入输出，标志着多模态从“单模态交互”向“全场景智能交互”的跨越。
- 多模态技术的深化落地：2025 年 1 月，阿里通义千问发布的 Qwen2.5-VL 模型，作为多模态大模型“高清细节理解”的代表，其专注于攻克高清图像的细节理解难题，能够无损解析 4K 医疗影像、卫星图等专业图像。2025 年 5 月，Google 发布 Veo3 模型则在强大的视频生成能力之上，原生集成了音频生成，实现了高达 99.8% 的音画同步精度，并具备了媲美专业工具的视觉推理能力。在解决通用交互能力后，多模态模型向专业化、高精度、高保真方向演进，旨在解决现实世界中更具挑战性的问题。

图表49：Veo 3 能够零样本解决大量以输入图像和文本提示指定的视觉任务



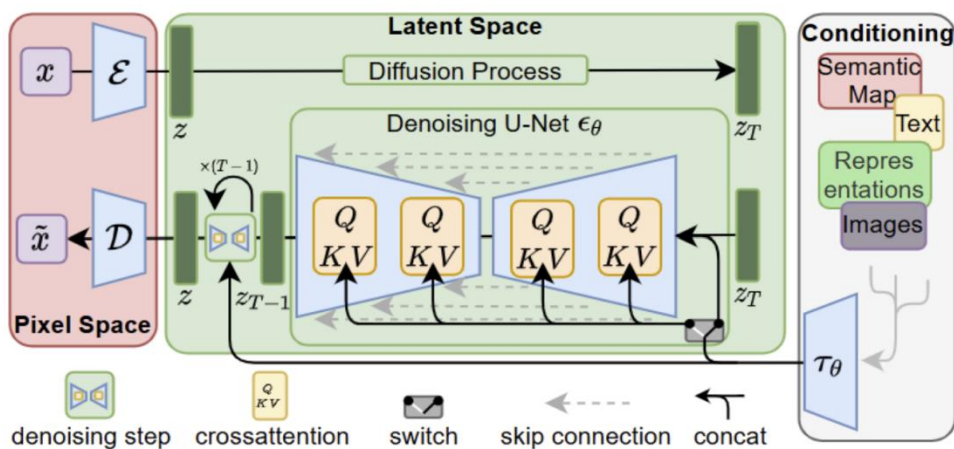
来源：Thaddäus Wiedemer 等《Video models are zero-shot learners and reasoners》(2025)，国金证券研究所



5.2 多模态生成技术路线：跨越图像、音频与视频的融合创造

在图像生成技术领域，扩散模型通过“正向扩散加反向去噪”的机制实现图像生成：先对初始图像逐步添加随机噪声使其演变为纯粹随机噪声，模型则学习反向去噪以从随机噪声重构清晰图像。其生成流程为，利用文本编码模型将输入文本转化为嵌入向量，以此作为条件输入送入 U-Net 去噪网络，经 U-Net 逐步去噪最终生成清晰图像。Open AI 于 2021 年 1 月推出的 DALL-E 以 CLIP 模型对齐文本与图像空间，在训练阶段将文本特征向量转化为图像特征向量，decoder 融合 U-Net 与 Transformer 实现加噪-去噪的图像还原，奠定了多模态文生图的技术基础。Latent Diffusion Models (LDM) 在潜在空间执行扩散，通过拼接或更通用的交叉注意力机制对潜在扩散模型进行条件约束，大幅降低计算成本；Diffusion Transformers (DiT) 以 Transformer 替代 U-Net，提升生成质量；多条件控制方式如 CLIP 引导、分类器无关引导，增强了生成可控性。扩散模型优势在于生成图像质量高、细节丰富且条件控制灵活，但生成速度慢，需较多采样步骤，计算成本较高。

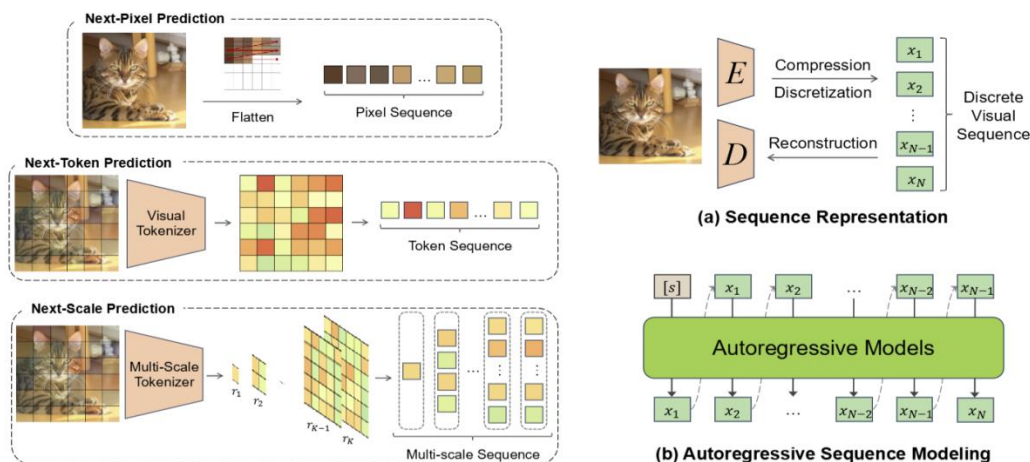
图表50：通过拼接或更通用的交叉注意力机制对潜在扩散模型（LDMs）进行条件约束



来源：Robin Rombach 等《High-Resolution Image Synthesis with Latent Diffusion Models》(2022)，国金证券研究所

在图像生成领域，自回归模型以“基于已有内容逐步预测后续内容”为核心思想，将图像生成转化为序列预测问题，主要分为像素级预测（如 2016 年 PixelRNN 模型支持逐像素生成但计算成本高）、分词级预测（如 2017 年 VQ-VAE 模型先将图像压缩为离散 token 再用 Transformer 建模）、多 token 预测（如 2024 年视觉自回归模型 VAR 模型通过分层预测多个 token 提升效率）三类；尽管当前扩散模型在图像质量上更具优势，但自回归模型因与大语言模型（LLM）结构一致，在多模态统一框架构建中更受青睐，为图像与文本的联合建模提供了天然的架构兼容性。

图表51：视觉自回归模型通用框架及核心组件



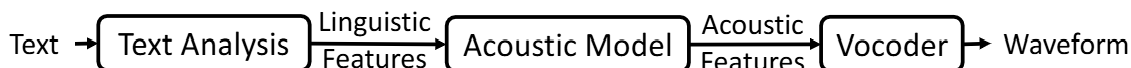
来源：Jing Xiong 等《Autoregressive Models in Vision: A Survey》(2025)，国金证券研究所

在音频生成技术领域，三段级联语音技术（ASR+LLM+TTS）是重要的技术路径，其中



TTS（文本转语音）的核心流程包含文本分析、声学模型和声码器三个关键环节：文本分析模块将输入文本转换为音素或语言特征，声学模型基于这些语言特征生成声学特征，最后声码器将声学特征转化为波形输出。但三段级联的架构存在延迟高、非文本信息和上下文信息丢失等问题，且各模块单独训练易导致级联后损失增大。为此，端到端 TTS 模型成为发展方向，以 GPT-4o 为代表的端到端模型具备延迟低、泛化性强、精确度高的优势，但可控性不足。从技术演进逻辑看，TTS 端到端的关键在于简化语音特征表示，提升生成效率与拟真度，未来还需解决与大语言模型（LLM）骨干的连接问题；而 ASR 端到端的关键突破则在于摒弃 CTC 时序对齐机制，通过注意力机制使解码器在生成文本时自适应关注输入语音的不同部分，隐式解决序列对齐问题，即便存在冗余也不影响效果。

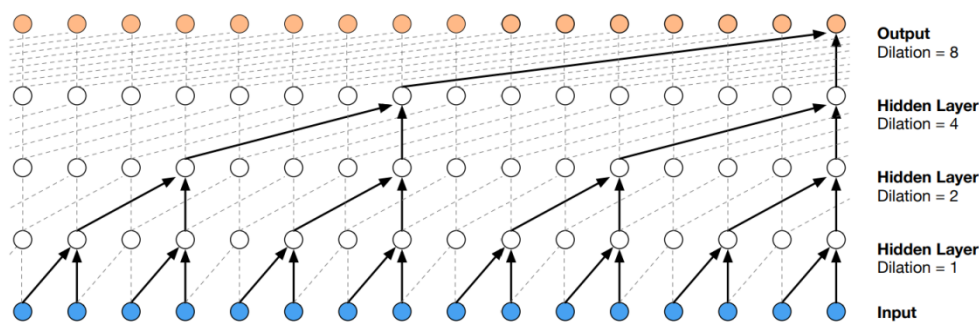
图表52: TTS 的核心流程包含文本分析、声学模型和声码器三个关键环节



来源：Xu Tan 等《A Survey on Neural Speech Synthesis》(2021)，国金证券研究所

在端到端音频生成领域，2016 年 DeepMind 提出的 WaveNet 开创了生成式模型的先河，其首次实现直接从原始音频波形生成语音，核心创新在于采用扩张因果卷积，可捕捉长达数千个时间步的上下文依赖，从而生成高保真音频。这种方法极大地提升了音频生成的质量，使其接近真实人声，克服了传统方法中常见的音质不自然和断裂感问题。同时，引入 μ -law 量化技术将 16 位音频压缩为 256 个离散值，大幅提升音频生成质量并降低计算需求，还使得其能够在更高效的硬件上运行。虽存在自回归速度慢的局限，但为后续研究奠定了基础。

图表53: 扩张因果卷积层堆叠的可视化

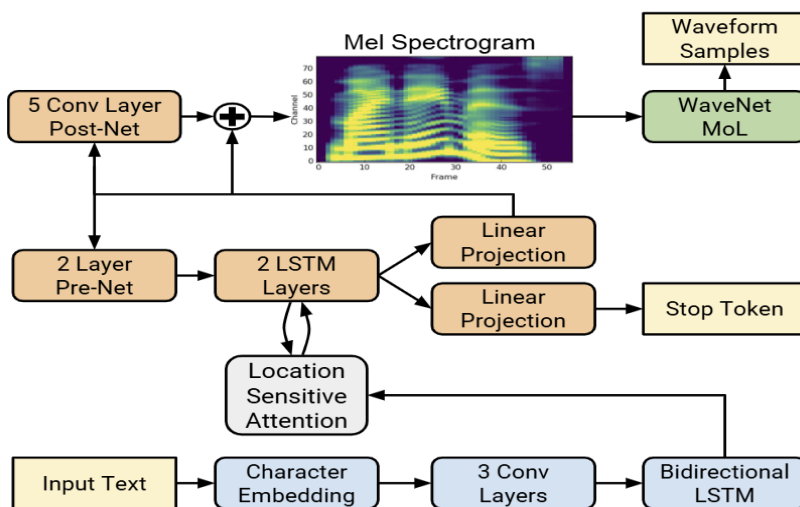


来源：A Van Den Oord 等《WAVENET: A Generative Model For Raw Audio》(2016)，国金证券研究所

2017 年 Google 团队提出的 Tacotron 2 端到端合成简化了语音合成的流程，推动 TTS 系统的端到端发展。该流程采用了两阶段架构：第一阶段是序列到序列（Seq2Seq）模型，它将输入的文本直接映射为梅尔频谱图。梅尔频谱图是一种通过模拟人耳听觉特性来增强低频信号并抑制高频噪声的频率表示形式，能够大大提升合成语音的自然度；在第二阶段，改进的 WaveNet 声码器来将梅尔频谱图转化为最终的音频波形。与传统的 Griffin-Lim 算法相比，WaveNet 声码器能够通过预测 16 位音频样本的混合逻辑分布来减少相位估计误差，从而生成更为精细、逼真的语音。不再需要依赖手工标注的语言学特征（如音高、时长等），并通过动态停止机制自适应地控制语音的长度，进一步提升了模型的灵活性和应用性。



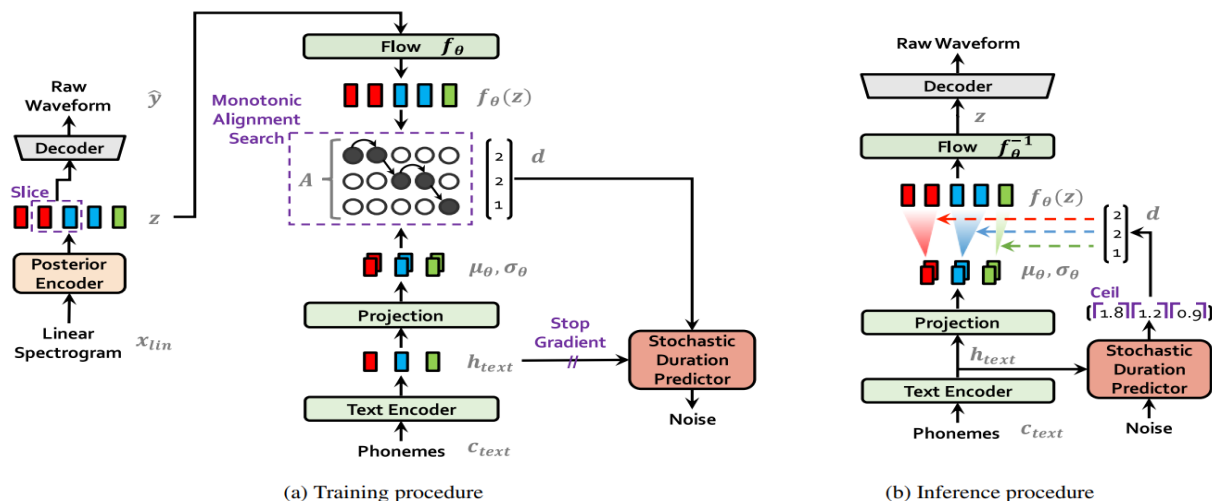
图表54: Tacotron 2 系统架构框图



来源: Jonathan Shen 等《Natural TTS Synthesis By Conditioning Wavenet On Mel Spectrogram Predictions》(2018), 国金证券研究所

2021 年提出的单阶段多样化 VITS, 首次通过将条件变分自编码器 (VAE) 和对抗训练相结合, 实现了单阶段端到端并行生成。VITS 的核心优势在于能够同时生成高质量音频, 并对语音的时长、风格等细节进行控制, 让语音合成更加灵活多样。该模型采用变分推断, 结合标准化流和对抗训练过程, 提高了生成建模的表现能力, 还通过随机时长预测器, 从输入文本中合成具有不同节奏的语音。

图表55: VITS 训练流程与推理流程架构框图



来源: Jaehyeon Kim 等《Conditional Variational Autoencoder with Adversarial Learning for End-to-End Text-to-Speech》(2021), 国金证券研究所

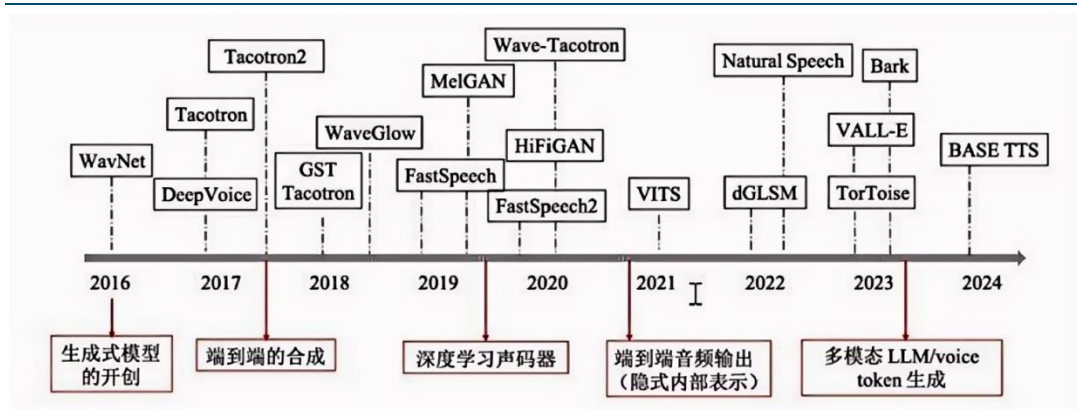
语音合成技术的演进经历了从分阶段模型到端到端系统的转变:

- 生成方式上, WaveNet 依赖自回归逐点生成, 虽然能够实现极高的保真度, 但其生成速度较慢。Tacotron2 则通过两阶段架构, 在保证语音质量的同时提高了生成效率。而 VITS 则引入了变分推断与对抗学习, 实现了单阶段端到端并行生成, 不仅提升了速度, 也减少了计算复杂度, 成为当前速度最快的技术。
- 在自然度与多样性方面, WaveNet 受限于固定条件输入, 因此其生成的语音多样性较低。Tacotron2 通过梅尔频谱作为中间表示, 增强了语音的自然度, 但语音的多样性仍然受到外部控制的影响。而 VITS 则通过随机时长预测和隐变量建模, 使得语音的节奏和风格可以实现灵活的调控, 大大提升了多样性与自然度的表现。
- 在应用场景方面, WaveNet 因其高保真性, 广泛应用于语音和音乐合成。Tacotron2 则因其稳定性和端到端的设计, 成为工业级 TTS 系统的首选技术。VITS 则在实时交



互和个性化语音生成中展现了巨大的潜力，尤其是在需要快速响应和个性化定制的场景中具有广泛的应用前景。

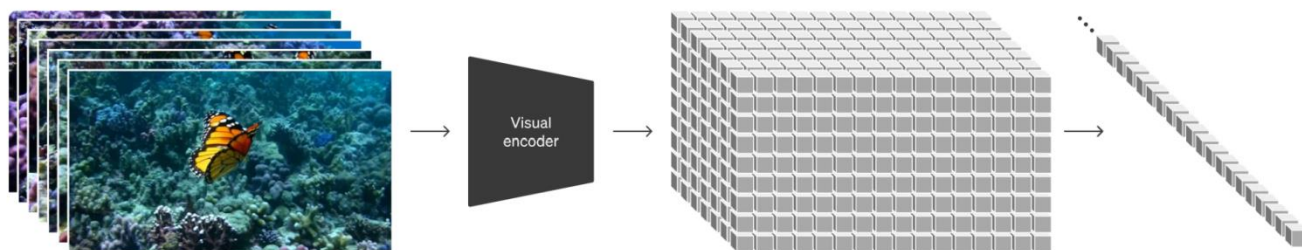
图表56：语音合成技术的演进经历了从分阶段模型到端到端系统的转变



来源：华南理工大学智能算法研究中心，国金证券研究所

在视频生成技术领域，现有研究主要采用循环神经网络（受长度限制）、生成对抗网络（多样性不足且生成视频短）、自回归 Transformer（生成速度慢且累计误差大）及扩散模型等方法，且多局限于特定视觉数据、较短时长或固定尺寸视频。而 Sora 作为通用视觉模型，突破了这些限制，可生成不同时长、宽高比和分辨率的视频及图像，其技术核心在于：一方面采用视频压缩神经网络将原始视频转换为时间和空间压缩的潜在向量表示，在该空间中完成训练与生成，再通过解码器（如 AVE）映射回像素空间；另一方面借鉴大语言模型的 token 化思路，实现“文本 token”与“视频 patch”的模态对齐，直接理解文本指令生成对应内容。同时，ViViT 模型提出的 4 种时空 patch 信息聚合策略（时间-空间联合注意力、双编码器架构、分解自注意力、级联注意力）为其时空建模提供了技术参考。当前，该领域聚焦于稳定性、可控性、叙事性三大核心要素，需突破长时视频逻辑一致性、细粒度生成可控性及多模态交互与世界模型融合等难点。

图表57：各类图像和视频转换为 visual patches 的潜在表示

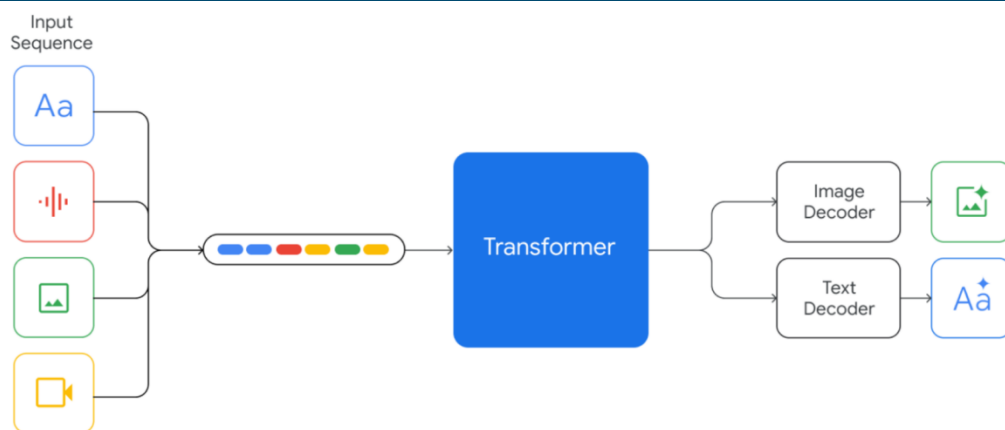


来源：OpenAI 官网，国金证券研究所

在视频生成领域，自回归技术基于时序依赖性建模视频生成，通过将视频按时间维度离散化为帧或帧级子块，构建有序视觉时序序列，再通过因果注意力掩码等机制施加严格因果约束，最终模型依托对海量真实视频时序规律（如动作演化、光影变迁）的学习，从初始时序单元开始逐一生成后续单元，逐步构建完整视频序列，以此保障生成内容的时序连贯性与逻辑一致性。Google 分别于 2023 年 12 月推出 Gemini 系列、2025 年 5 月推出 Veo3 模型，Gemini 打破传统多模态模型“编码器拼接”局限，将视频帧、文本、音频统一编码为 Token 序列输入纯 Transformer 解码器，通过时空 Transformer（ST-Transformer）优化计算复杂度，在保证 4K 画质的同时支持数分钟长视频生成，还能结合历史帧与文本指令动态调整内容、维持场景一致性；Veo3 则采用“自回归动力学模型+扩散模型”混合架构，自回归模块生成符合物理规律的动态序列，扩散模块优化细节，依托“帧链（CoF）推理”实现零样本泛化，且支持首尾帧控制等功能，适配影视特效、科学模拟等场景需求。



图表58: Gemini 模型支持将文本、图像、音频和视频的交错序列作为输入

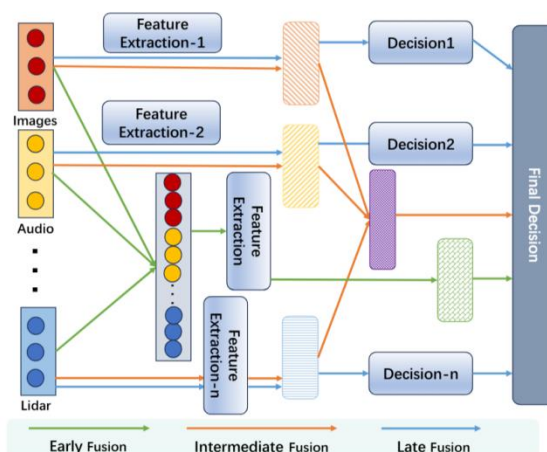


来源: Gemini Team 等《Gemini: A Family of Highly Capable Multimodal Models》(2025), 国金证券研究所

5.3 多模态理解技术路线: 突破感知认知临界, AI 基座模型步入成熟期

多模态学习是一种通过整合多种数据模态(如文本、图像、音频、视频等)来提升模型对复杂信息的理解能力的技术,多模态学习的两个核心问题是多模态融合和跨模态对齐。多模态融合指通过整合不同模态的数据来提高模型的感知和理解能力。在多模态融合领域,表征学习通过从原始数据中自动提取各模态有效特征,生成稳定全面的多模态表征,其本质近似于向量化操作。从数据处理的层次角度划分,多模态融合可分为数据级融合、特征级融合和目标级融合。跨模态对齐是确保不同模态数据准确对应、为融合提供可靠基础的关键环节,主要分为显式对齐和隐式对齐两大类。显式对齐直接建立不同模态间的对应关系,涵盖无监督对齐和监督对齐;隐式对齐不直接建立对应关系,而是通过模型内部机制实现跨模态对齐,包括注意力对齐和语义对齐。

图表59: 主要多模态融合策略的基本流程

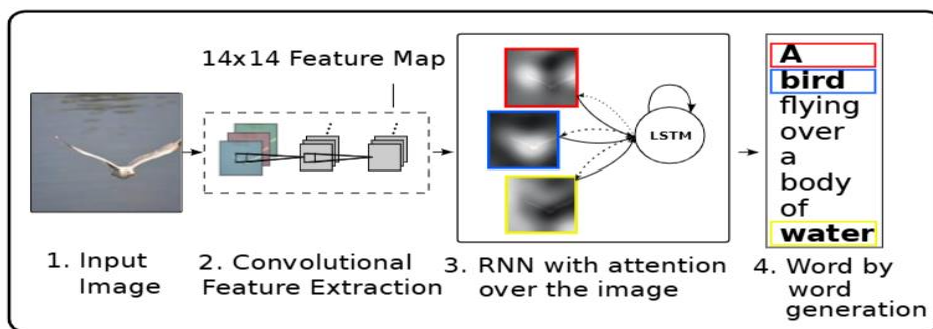


来源: Xiaofeng Han 等《Multimodal Fusion and Vision-Language Models: A Survey for Robot Vision》(2025), 国金证券研究所

传统方法下,多模态理解主要应用于图像描述、视频描述及视觉问答,多模态推理主要应用于视觉常识推理、跨模态检索任务。其中,图像描述旨在让计算机根据给定图像自动生成描述性文字。受机器翻译领域中编码器-解码器(Encoder-Decoder)模型的启发,图像描述可以通过端到端的学习方法直接实现图像和描述句子之间的映射,将图像描述过程转化成为图像到描述的“翻译”过程。这一过程通常使用卷积神经网络(CNN)来提取图像中的视觉特征,如颜色、纹理、形状等;然后利用循环神经网络或其变体或Transformer架构来捕捉这些特征之间的时序依赖关系,并生成对应的自然语言描述。在生成描述的过程中,还可能采用注意力机制来增强模型对图像关键区域的关注度,从而提高描述的准确性和相关性。注意力机制并非简单地将输入图像编码成一个固定的特征向量,而是通过引入上下文向量,对每个时间步的解码过程进行动态调整,以此增强图像区域与生成单词之间的相关性,从而捕捉并表达更多的图像语义细节。



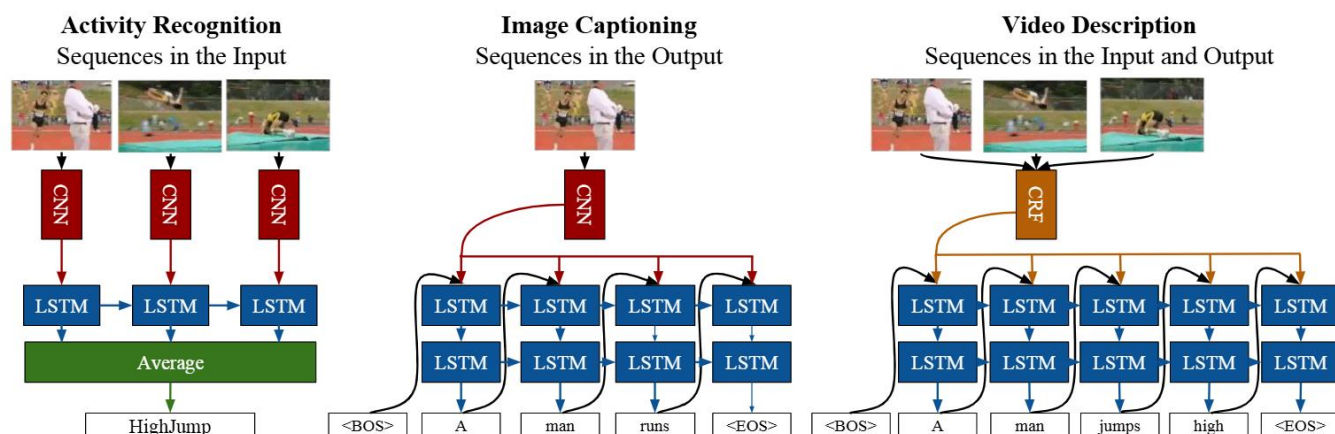
图表60: 输入图像→CNN 提取特征→带注意力的RNN (LSTM) →逐词生成描述



来源: Kelvin Xu 等《Show, Attend and Tell: Neural Image Caption Generation with Visual Attention》(2015), 国金证券研究所

视频描述指通过机器自动生成视频内容的描述语句的技术,旨在将视频中的视觉和听觉信息转化为易于理解的自然语言文本,从而帮助用户快速了解视频内容。通过卷积神经网络 (CNN) 提取视频帧中的关键视觉特征,结合循环神经网络 (RNN) 及其变体捕捉时序信息,再融合音频特征和文本信息等多模态数据,最后利用序列到序列 (Seq2Seq) 模型和注意力机制将这些特征转化为自然、准确的语言描述。可以应用视频定位 (Video Localization) 技术来精准标记视频关键元素或事件的位置及时间点。借助目标检测与跟踪,实时定位物体、人物;时间标注则记录关键事件或场景变换的具体时刻,两者协同工作,实现对视频内容的深度解析。

图表61: 长短期循环卷积网络模型针对动作识别、图像描述和视频描述任务的特定实例化

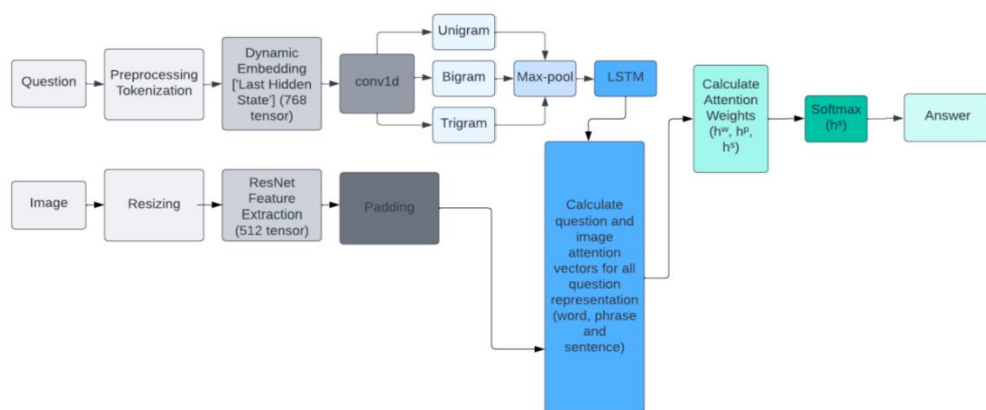


来源: Jeff Donahue 等《Long-term Recurrent Convolutional Networks for Visual Recognition and Description》(2015), 国金证券研究所

视觉问答指系统接收一张图像和一个关于这张图像的自然语言问题作为输入,经过系统处理后,输出一个准确的自然语言答案。为了实现视觉问答,需要通过预训练的卷积神经网络提取图像特征,利用自然语言处理技术转化问题为特征向量,进而采用联合嵌入、注意力机制或多模态融合等方法将图像与问题特征有效结合,并在特征融合的基础上,运用分类器或生成式模型生成与问题紧密相关的自然语言答案。



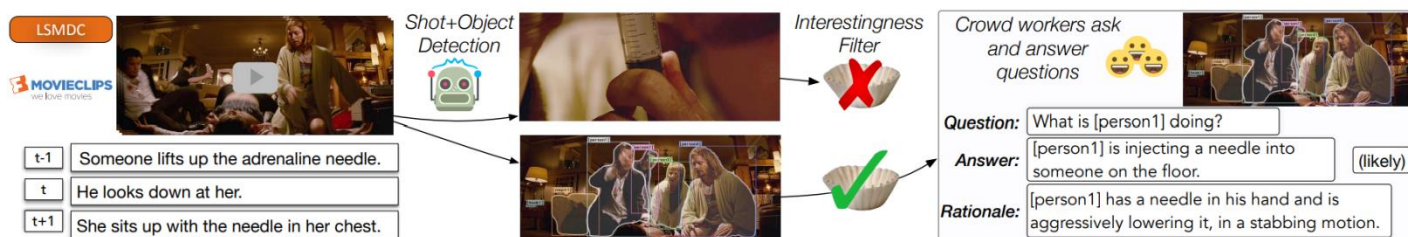
图表62: 整体的视觉问答框架



来源: Stanislaw Antol 等《VQA: Visual Question Answering》(2015), 国金证券研究所

视觉常识推理指需要在理解文本的基础上结合图片信息, 基于常识进行推理。给定一张图片、图中一系列有标签的 **bounding box**, VCR 实际上包含两个子任务: 根据问题选择答案; 根据问题和答案进行推理, 解释为什么选择该答案。VCR 数据集由大量的“图片-问答”对组成, 主要考察模型对跨模态的语义理解和常识推理能力。知识图谱的基本组成单位是“实体-关系-实体”三元组, 以及实体及其相关属性一值对, 实体间通过关系相互联结, 构成网状的知识结构。基于知识图谱的推理通常依赖于神经网络模型来捕捉图谱中的结构信息, 这些模型通过前向传播过程进行推理预测。例如, 采用基于图神经网络(GNN)的推理方法, 特别是基于注意力机制的图卷积神经网络, 该方法通过计算实体之间的相似度并为它们分配相应的权重, 从而推断出实体之间的关系。

图表63: 视觉常识推理 VCR 构建概述

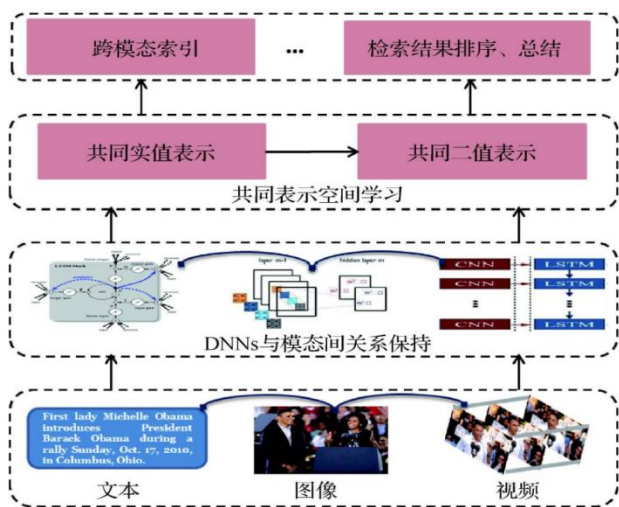


来源: Rowan Zellers 等《From Recognition to Cognition: Visual Commonsense Reasoning》(2019), 国金证券研究所

跨模态检索指通过整合文本、图像、音频、视频等异构模态数据的语义关联, 实现以一种模态为查询检索其他模态相关内容的技术, 典型场景包括“文搜图”“图搜音频”“视频搜文本”等。该过程主要依赖“特征提取-模态对齐-相似度计算”三步核心原理: 首先针对不同模态采用专属编码器提取特征, 如用 CNN 提取图像视觉特征、用 WCNN 或 BERT 提取文本语义特征、用梅尔频谱与 CNN 结合提取音频特征; 随后通过联合嵌入技术将多模态特征映射至共享语义空间, 早期方法如 DBRLM 模型通过双向结构强化匹配样本的相似度, 现代大模型则借助对比学习实现更精准的跨模态对齐; 最终通过余弦相似度等度量方式计算查询与候选样本的语义距离, 返回关联度最高的结果。

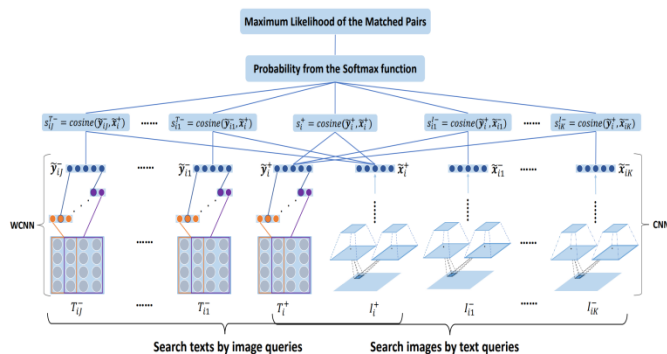


图表64：跨模态检索一般框架



来源：尹奇跃等《基于深度学习的跨模态检索综述》(2021)，国金证券研究所

图表65：包含卷积神经网络（CNNs）和加权卷积神经网络（WCNNs）的双向提取结构



来源：Yonghao He 等《Cross-Modal Retrieval via Deep and Bidirectional Representation Learning》(2016)，国金证券研究所


附录：三张报表预测摘要
损益表(人民币 百万)

	2022A	2023A	2024A	2025E	2026E	2027E
主营业务收入	2,085	2,696	3,341	4,021	5,012	6,245
增长率	25.2%	29.3%	23.9%	20.4%	24.6%	24.6%
主营业务成本	898	1,040	1,047	1,045	1,303	1,624
%销售收入	43.1%	38.6%	31.3%	26.0%	26.0%	26.0%
毛利	1,187	1,656	2,294	2,975	3,709	4,622
%销售收入	56.9%	61.4%	68.7%	74.0%	74.0%	74.0%
营业税金及附加	0	0	0	0	0	0
%销售收入	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
销售费用	403	428	483	643	802	968
%销售收入	19.3%	15.9%	14.5%	16.0%	16.0%	15.5%
管理费用	272	301	400	422	516	625
%销售收入	13.0%	11.2%	12.0%	10.5%	10.3%	10.0%
研发费用	586	635	911	1,025	1,228	1,468
%销售收入	28.1%	23.6%	27.3%	25.5%	24.5%	23.5%
息税前利润 (EBIT)	168	396	779	895	1,176	1,590
%销售收入	8.1%	14.7%	23.3%	22.3%	23.5%	25.5%
财务费用	-7	-42	-44	-16	-20	-36
%销售收入	-0.3%	-1.6%	-1.3%	-0.4%	-0.4%	-0.6%
投资收益	-13	-19	-24	-20	-20	-20
%税前利润	-7.5%	-4.4%	-2.9%	-2.2%	-1.6%	-1.2%
营业利润	-74	272	497	874	1,145	1,547
营业利润率	-3.6%	10.1%	14.9%	21.7%	22.9%	24.8%
营业外收支						
税前利润	175	438	823	911	1,196	1,625
利润率	8.4%	16.3%	24.6%	22.6%	23.9%	26.0%
所得税	156	72	17	55	60	81
所得税率	89.2%	16.4%	2.0%	6.0%	5.0%	5.0%
净利润	19	366	806	856	1,136	1,544
少数股东损益	-75	-12	1	9	11	15
归属于母公司的净利润	94	378	805	847	1,125	1,529
净利率	4.5%	14.0%	24.1%	21.1%	22.4%	24.5%

现金流量表(人民币 百万)

	2022A	2023A	2024A	2025E	2026E	2027E
净利润	94	378	805	847	1,125	1,529
少数股东损益	-75	-12	1	9	11	15
非现金支出	-6	-34	-243	-2	-7	-14
非经营收益						
营运资金变动	83	18	99	-114	-26	-33
经营活动现金净流	234	413	746	874	1,242	1,635
资本开支	-51	-56	-53	-200	-150	-150
投资	-7	-672	118	-411	-300	-300
其他	41	83	-58	8	30	43
投资活动现金净流	-17	-645	7	-603	-420	-407
股权募资	0	1	0	0	0	0
债权募资	-37	115	49	50	50	50
其他	0	-198	-153	-10	-12	-14
筹资活动现金净流	-37	-81	-104	40	38	36
现金净流量	208	-306	661	309	859	1,263

资产负债表(人民币 百万)

	2022A	2023A	2024A	2025E	2026E	2027E
货币资金	947	641	1,301	1,610	2,469	3,732
应收款项	1,016	1,287	1,303	1,765	2,200	2,741
存货	26	54	73	73	90	113
其他流动资产	468	787	1,850	2,082	2,232	2,393
流动资产	2,456	2,769	4,528	5,530	6,991	8,979
%总资产	49.1%	48.0%	64.3%	66.5%	70.0%	73.8%
长期投资	124	122	97	197	197	197
固定资产	441	464	470	584	635	679
%总资产	8.8%	8.0%	6.7%	7.0%	6.4%	5.6%
无形资产	688	851	572	515	463	417
非流动资产	2,549	2,998	2,519	2,790	2,989	3,186
%总资产	50.9%	52.0%	35.7%	33.5%	30.0%	26.2%
资产总计	5,004	5,767	7,047	8,319	9,980	12,166
短期借款	10	15	267	317	367	417
应付款项	135	141	179	203	253	316
其他流动负债	924	1,109	1,374	1,729	2,155	2,686
流动负债	1,069	1,264	1,820	2,249	2,775	3,418
长期贷款	0	135	0	0	0	0
其他长期负债	226	254	194	184	184	184
负债	1,295	1,653	2,014	2,433	2,959	3,602
普通股股东权益	3,770	4,109	5,027	5,873	6,996	8,523
其中：股本	0	0	0	0	0	0
未分配利润	4,145	4,463	5,357	6,202	7,326	8,853
少数股东权益	-60	5	6	14	26	41
负债股东权益合计	5,004	5,767	7,047	8,319	9,980	12,166

比率分析

	2022A	2023A	2024A	2025E	2026E	2027E
每股指标						
每股收益	0.02	0.09	0.18	0.19	0.25	0.33
每股净资产	0.85	0.92	1.10	1.29	1.53	1.87
每股经营现金净流	0.05	0.09	0.16	0.19	0.27	0.36
每股股利	0.00	0.00	0.00	0.00	0.00	0.00
回报率						
净资产收益率	2.50%	9.21%	16.02%	14.43%	16.08%	17.94%
总资产收益率	1.88%	6.56%	11.43%	10.18%	11.27%	12.57%
投入资本收益率	0.49%	7.77%	14.40%	13.56%	15.12%	16.82%
增长率						
主营业务收入增长率	25.17%	29.27%	23.93%	20.36%	24.64%	24.62%
EBIT 增长率	1118.21%	135.82%	96.54%	14.89%	31.41%	35.22%
净利润增长率	311.49%	301.83%	112.84%	5.23%	32.75%	35.91%
总资产增长率	5.41%	15.24%	22.19%	18.06%	19.96%	21.90%
资产管理能力						
应收账款周转天数	61.1	49.3	42.8	43.2	45.1	45.1
存货周转天数	6.1	13.7	21.9	25.1	22.5	22.5
应付账款周转天数	48.0	47.8	55.1	65.9	63.1	63.1
固定资产周转天数	71.2	60.5	50.3	47.2	43.8	37.9
偿债能力						
净负债/股东权益	-19.16%	-5.75%	-16.71%	-18.86%	-27.33%	-36.57%
EBIT 利息保障倍数	18.8	82.7	108.2	87.7	98.4	116.0
资产负债率	25.87%	28.67%	28.58%	29.24%	29.64%	29.60%

来源：公司年报、国金证券研究所


市场中相关报告评级比率分析

日期	一周内	一月内	二月内	三月内	六月内
买入	1	3	6	12	35
增持	0	0	0	1	0
中性	0	0	0	0	0
减持	0	0	0	0	0
评分	1.00	1.00	1.00	1.08	1.00

来源：聚源数据

市场中相关报告评级比率分析说明：

市场中相关报告投资建议为“买入”得 1 分，为“增持”得 2 分，为“中性”得 3 分，为“减持”得 4 分，之后平均计算得出最终评分，作为市场平均投资建议的参考。

最终评分与平均投资建议对照：

1.00 = 买入； 1.01~2.0=增持； 2.01~3.0=中性
 3.01~4.0=减持

投资评级的说明：

买入：预期未来 6—12 个月内上涨幅度在 15%以上；

增持：预期未来 6—12 个月内上涨幅度在 5%—15%；

中性：预期未来 6—12 个月内变动幅度在 -5%—5%；

减持：预期未来 6—12 个月内下跌幅度在 5%以上。



特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海	北京	深圳
电话：021-80234211	电话：010-85950438	电话：0755-86695353
邮箱：researchsh@gjzq.com.cn	邮箱：researchbj@gjzq.com.cn	邮箱：researchsz@gjzq.com.cn
邮编：201204	邮编：100005	邮编：518000
地址：上海浦东新区芳甸路 1088 号 紫竹国际大厦 5 楼	地址：北京市东城区建国内大街 26 号 新闻大厦 8 层南侧	地址：深圳市福田区金田路 2028 号皇岗商务中心 18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究