

视频生成模型之争

本报记者 宋婧

近日, Sora、Genie 等视频生成模型点燃了资本市场的热情。人工智能指数迎风大涨, 相关概念股接连涨停。据不完全统计, 近20家上市公司在各自的互动平台上披露了视频生成模型领域相关的业务情况。然而, 专家指出, 目前国内视频生成模型技术真正达到前沿水平的公司凤毛麟角, 多数公司只是在跟风炒作, 缺乏真正的技术储备和研发能力。

视频生成模型难在哪儿?

相较于文字和图片, 视频在多维信息表达、画面丰富性及动态性方面有更大优势。它可以结合文本、图像、声音及视觉效果, 在单一媒体中融合多种信息形式。从视频生成到文生视频、图生视频, 多模态的发展偏重于用更少的用户输入信息量实现更丰富的AI生成结果。

核心技术难点一方面在于数据, 文生视频需要大量的“文本-视频”配对数据, 且数据标注、清洗的工作量庞大; 另一方面在于逻辑, 视频是连续的多帧图像, 要有逻辑性, 而非简单的图片组合, 这会让模型的复杂度、计算难度和成本大幅提升。

“我们原本在文本、图像模型上就存在差距, 现在视频模型来了, 我们的差距也更大。”晨云磐盾信息技术有限公司总裁郝峻晟在接受《中国电子报》记者采访时坦言, “由于受到算力、显卡等多方面的限制, 国内人工智能企业在大模型领域的创新已经落后了一大步。而现在国外领先的大模型不再开源, 也不再公开核心技术细节, 因此, 所有的技术研发只能靠我们自己的了。大模型研发是一个系统性工程, 基础算法本身可能没什么差距, 但像数据的训练、清洗、标注、管理和能耗等细节才是真正拉开差距的地方。”

Sora、Genie 等视频生成模型的诞生离不开Runway ML的Gen-2、谷歌的Lumiere、Stable Video Diffusion等前沿技术或产品在前铺路。Sora的一个重要技术创新点在于其先将不同视频和图片数据压缩在一个低维空间中, 再分解成统一Patch作为训练大模型的基本单位, 这一改进使得文生视频更加逼真, 质量更高。

“所有的创新都是突发的, 具有偶然性, 无法被预测。但海外这些领先的视频生成模型至少已经帮助我们确定了技术方向。”郝峻晟说道。

除了技术突破, 应用赛道的选择亦非常重要。



要。赛迪顾问业务总监、软件与信息服务业研究中心总经理高丹在接受《中国电子报》记者采访时表示: “不管是ChatGPT还是Sora, 这类爆款产品都发迹于我们以往并没有过多关注的领域。而二者的共同点是都出现在数据积累比较多或者应用场景比较多的领域, 因此我认为爆款产品的出现首先是要选对赛道, 要关注数字化应用场景积累丰厚的领域。”

“另一个重要的点是企业要长期坚持, 并且有资本长期投入, 不追求短期效益。”高丹说道。这一点说起来容易, 真正能做到的企业寥寥无几。

谁将最先打造出“中国版Sora”?

在视频生成模型领域, 字节跳动被寄予厚望。“我比较看好字节跳动, 它本身在视频应用领域的积累就比较多。”郝峻晟对记者说道。他认为, 国内有丰富的视频数据集, 可供模型训练。要训练出类似Sora的视频生成模型其实并不难, 只要有充足的算力、显卡和数据, 再调高算法精度, 就能有效提升生成视频的质量。

根据公开信息, 2023年11月, 字节跳动发表视频生成研究成果PixelDance, 今年1月又发布了视频生成模型MagicVideo-V2。公开的实验评测数据显示, MagicVideo-V2生成的视频高清晰度、润滑度、连贯性、文本语义还原等比目前主流的文生视频模型Gen-2、Stable Video Diffusion、Pika 1.0等更出色。

近日, 字节跳动又悄然推出了一款名为Boximator的视频生成模型。不过, 字节跳动内部人士在回应外部关切时表示, Boximator目前仍是一个研究项目, 专注于视频生成领域中控制对象运动的技术方法。尽管其具有创新性和潜力, 但该项目尚未达到作为成熟产品推出的阶段。与此同时, 与国外领先的视频生成模型相比, Boximator在画面质量、保真率和视频时长等方面仍存在显著的差距。

尽管字节跳动仍想保持低调, 但从字节跳动秘密组建AI产品研发团队、前抖音CEO张楠转战剪辑、谷歌顶尖科学家蒋路加入TikTok等举措来看, Sora的横空出世无疑大大加剧了字节跳动的焦虑。

一方面, Sora已经能生成60秒视频, 这

将直接冲击到剪辑的生存空间, 并且会改变抖音与TikTok的内容供给逻辑; 另一方面, Sora引领文生视频技术不断精进, 抖音及TikTok也将面临新技术带来的洗牌, 这也在一定程度上刺激字节跳动更加积极推进AI技术的投入。

“Sora的出现是技术、资本和数据综合的成果, 但是结合目前来看, 其更偏重于短视频的开发, 如果在多方要素基本相同的情况下, 我更看好前期相关数据积累较多的企业。”高丹分析道。

从国内厂商来看, 除了字节跳动, 百度在人工智能领域的布局较早, 无论是算力的充足、数据的丰富还是工程能力的先进程度, 都处于国内第一梯队。阿里、腾讯、科大讯飞等也是大模型竞争中的佼佼者。

“实际上, Sora打通的这条道路并不神秘, 总体上没有很多超预期的技术, 它是沿着大一统多模态大模型的既定路线, 在工程上取得的一项成果, 没有改变既定的技术范式和设计理念。”某大模型团队的工程副总裁表示。在他看来, Sora实现文生视频的大致思路和逻辑都已经在技术文档中公之于众了, 但

技术细节并未披露, 想要真正完成复现甚至超越, 仍然需要考验国内AI企业在算力、数据和工程能力等多方面的综合实力。

Sora面临的技术挑战仍不少

近日, 中信建投、国泰君安、申万宏源、招商证券等多家券商在研报中表示, Sora是人工智能发展进程中的“里程碑”, 预示AGI(通用人工智能)时代将加速到来, 众多行业将迎来颠覆式变革。而Sora出手即“王炸”, AI生成视频创业公司无疑将面临巨大压力。这也难怪Runway CEO瓦伦祖拉会在社交媒体上感叹: “Game On(游戏开始了)。”

“投资环境不同, 这也会对技术的创新产生一定的影响。”郝峻晟分析道。他表示, 如果可以给新技术一些耐心, 花十几年的时间持续投入研发, 做出来的产品、应用往往能给行业带来颠覆式的变化。“而国内很多投资方看到‘风’来了, 就会把资本投入进来, 然后‘风’走了, 投资就撤了。这对技术创新氛围的培育是不利的。”郝峻晟说道。

人工智能行业天使投资人郭涛也表达了类似的观点。他指出, 国内市场的特点和监管政策也可能影响企业的决策。国内企业要迎头赶上, 需要在技术研发、人才培养和市场洞察上下功夫, 同时还要考虑到国内外市场的差异性和合规性问题。

不过, 或许是考虑性能、安全等问题, Sora并未向公众开放, 目前还处于安全测试阶段, 仅向“红队测试人士”(针对潜在危险行为的测试)和少数创作者开放。据外媒预测, GPT-4经过6个月的测试后才正式向公众开放, 预计Sora或于8月向公众开放。

上海市人工智能行业协会秘书长钟俊浩分析指出, Sora面临的技术挑战仍有不少, 特别是与文本对话和图片生成相比, 训练成本高昂、高质量数据集的缺乏及视频描述的模糊性都将成Sora需要跨越的门槛。

“Sora的应用场景、商业前景比较明朗, 动漫游戏影视、短视频等都可以应用, 长视频的制作更加复杂, 还需要进一步开发, 同时模型训练成本也较高, 算力要求也较高, 商业落地还需要一段时间。”高丹表示。这也给国内厂商抢占市场留下了一段缓冲期。

郝峻晟表示, 国家层面大力建设智算中心, 加速完善新型算力基础设施, 将为我国人工智能发展提供肥沃的“土壤”。与此同时, 人工智能企业也要加快探索的步伐, 尤其是在垂直行业或者一些特定领域去做更多的积累与突破。

奋力谱写新型工业化发展新篇章



公益广告