

更多AI工具可直接访问：<https://www.faxianai.com/>

混元：腾讯视频模型

原文链接：<https://mp.weixin.qq.com/s/kalNwoQP...>

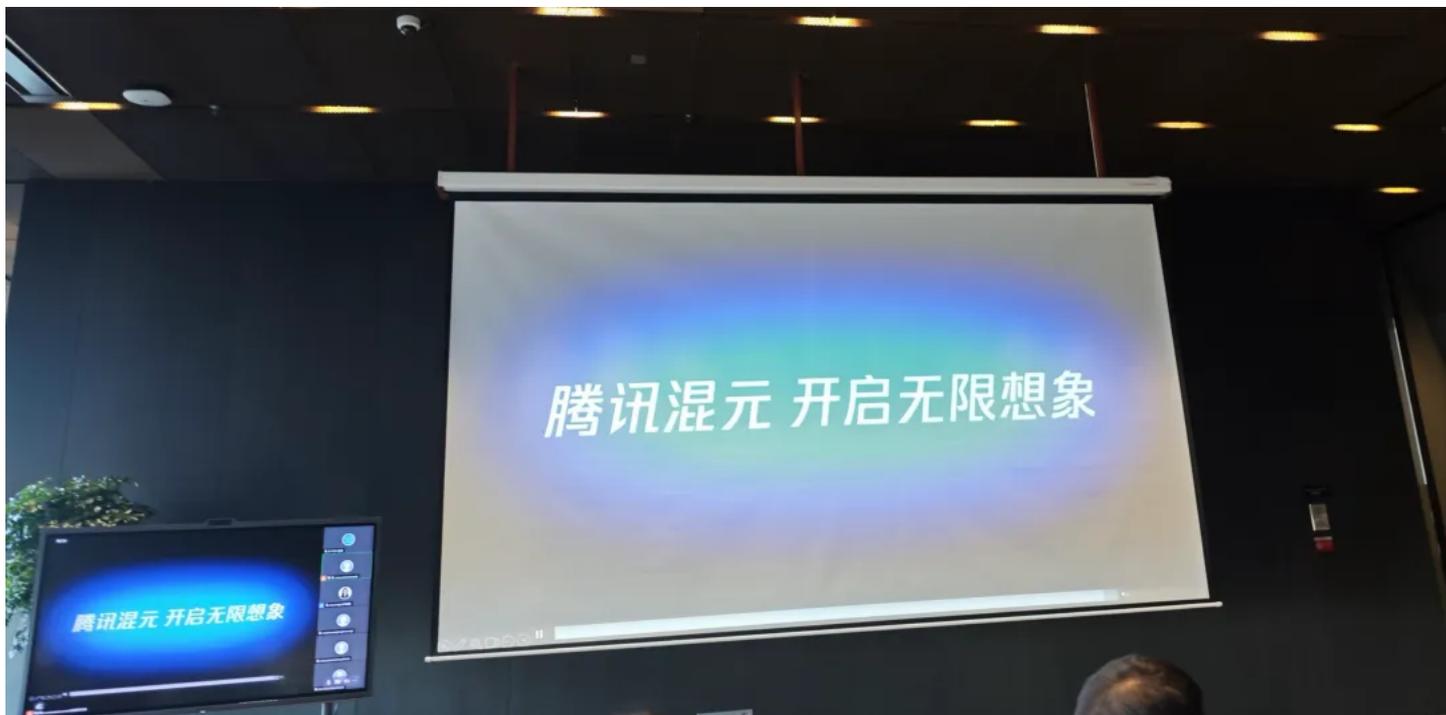
原创 数字生命卡兹克 数字生命卡兹克 2024年12月03日 15:05 北京

腾讯悄悄开源混元版「Sora」，这就是开源领域的No.1。

今天，人又在腾讯混元发布会的现场。

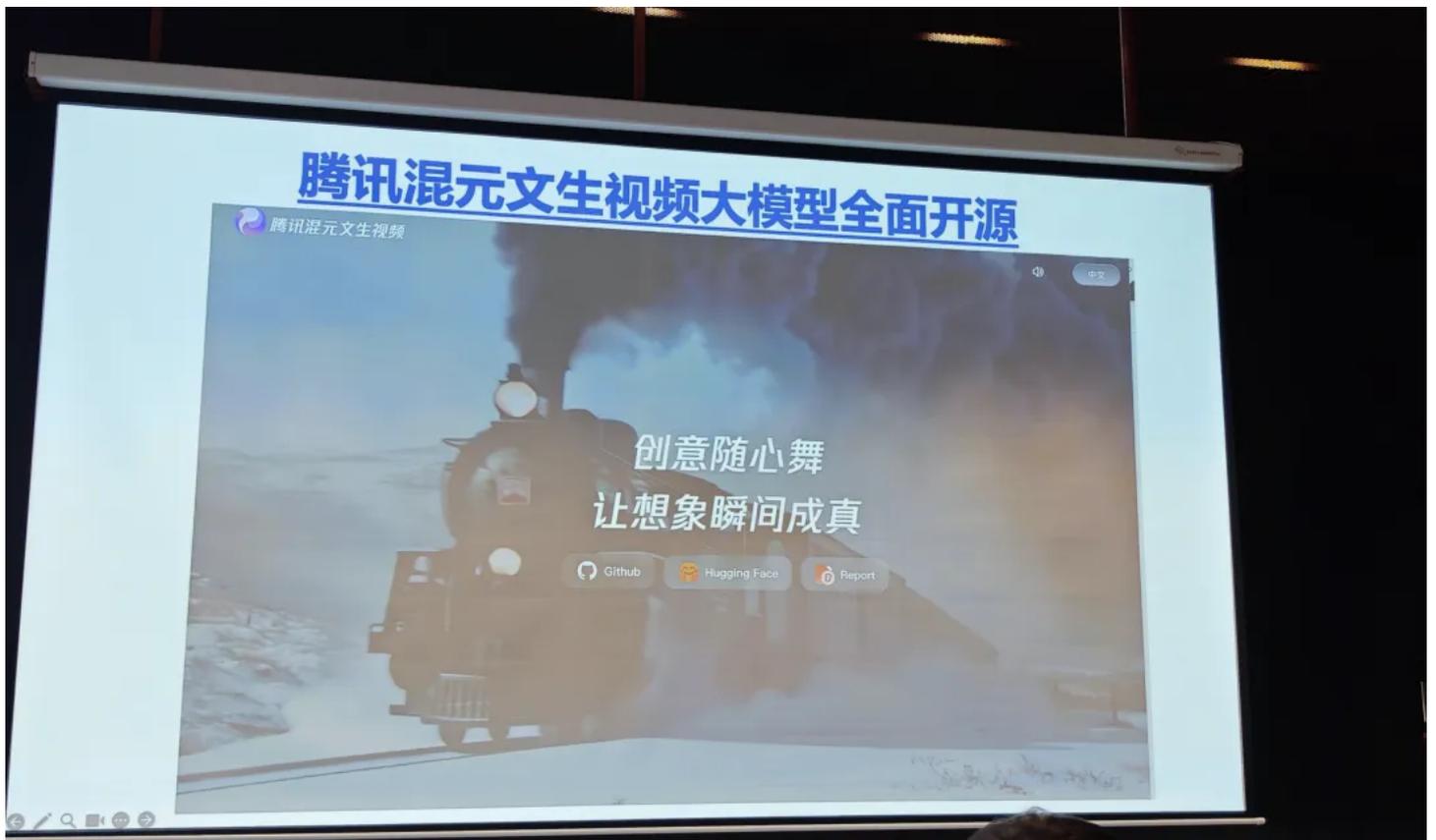
上个月5号，他们宣布开源大语言模型混元Large和3D大模型 Hunyuan3D-1.0。

仅仅一个月时间，他们又从深圳奔赴北京，邀请了一些老朋友，又开了一次私密的闭门会。



而这一次的项目，就是被N多人期待了很久的，腾讯混元视频生成模型。

同样，现场宣布，直接，开源。



腾讯也活成了，马斯克心中，那个OpenAI的模样。

聊聊这个腾讯混元的AI视频模型，我已经先行测试了一周，跑了几百个case。

先说结论：**偏科战神，强的部分强到没边，弱的地方也急需优化，但是瑕不掩瑜，综合来看，闭源模型中排在T1附近，开源AI视频中，无可争议的T0。**

开源地址：<https://github.com/Tencent/HunyuanVideo>

普通用户也可以去 腾讯元宝APP，进入AI应用，就能看到这个AI视频了。



可能普通用户需要资格申请，但是以腾讯的速度，应该非常快，看了群友的反馈，有的申请了不到一会就拿到了体验资格。

我先放几个我跑的Case，再来细说。



很有意思，很特别的模型。

如果让我来总结混元的3个特点，那就是：

超强的真实质感，很强的语义理解，可以切换镜头。

一个一个来说。

一. 超强的真实质感

说实话，混元的真实感，是非常惊艳到我的。

有点当年Flux出来的时候，那种既视感。

那种光影，那种质感，那种色彩，就是会让你感觉，这是现实场景，就是很真。

而且，几乎没有任何的AI视频中的柳絮抖动的那种质感，稳的一笔。

比如：**两个女人面对面哭泣。**



你就感觉，像在看电视剧。

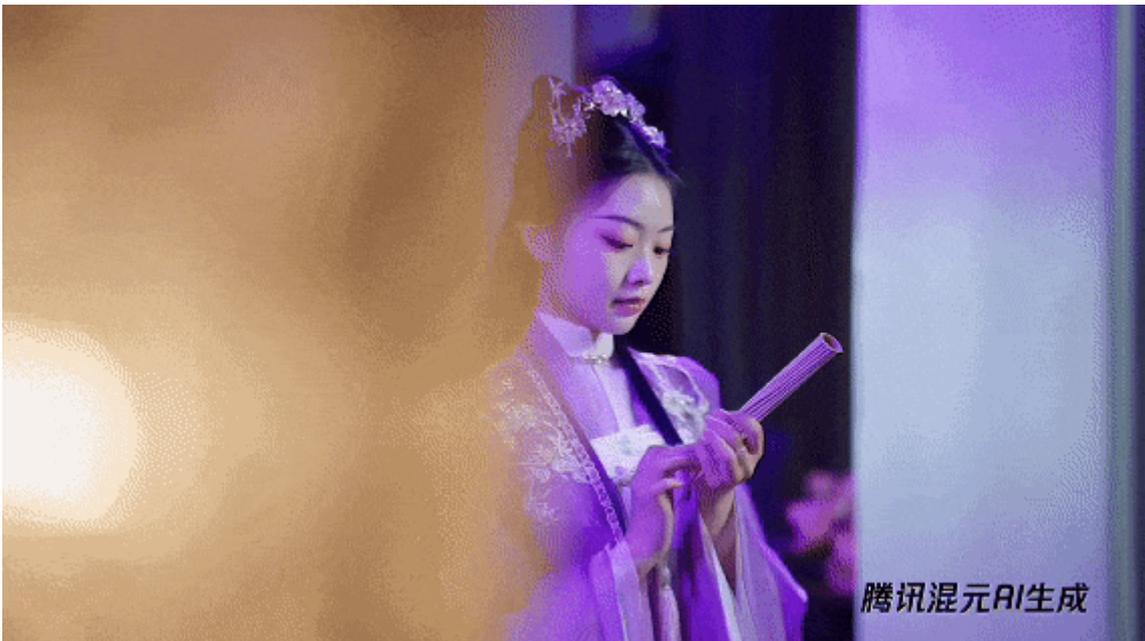
下属跟老板打架也是。



这光影和稳定性，实在是强到有点飞起了。

除了现代的戏，古装也是真实到不可思议。

比如：**长焦特写，一位紫衣宫女在月下绣花。银光洒在锦缎上，针线穿梭。背景是雕花窗棂，氛围静谧优雅。**



我愿称之为，最具有“电视剧质感”的AI视频大模型。

第一次，能看到，做中国古装，服化道和真实质感，好到如此程度的。

不只是国人，外国人的真实质感，也极强。



像拍的，不像生成的，实话。

但是，但是来了。

混元AI视频的写实质感，可以拉满，我可以说不，就是现在的No.1。

但是，我开头说它个超级偏科战神，就是因为，它如果不做超写实，做一些偏幻想的风格，或者2D、3D之类的，那个审美，真的有点丑==

这就仿佛一个学生跟你说，他语文吊炸天，150分考了149分，然后你问他英语多少，他支支吾吾的跟你说：不告诉你。。

另外有一点，是需要狠狠夸混元的。

就是跑过AI视频的人都知道，一旦人物的脸，在画面上占比一小，很多都会糊脸，五官是一坨，看不清。

但是我用混元跑的几百个case里，有不少是脸很小的，但是几乎都非常准确，甚至还是动嘴型说话。可以看下面这个例子。



开门的那个女人，脸没糊，那么小的面积，你都能看到她的神情，看到他的嘴形，这个点，太牛逼了。

二.很强的语义理解

混元的语义理解，是我觉得在所有的AI视频大模型里，都能排到前列的。

看这个Prompt:

一只银渐层在游乐园里奔跑，跳到一个小女孩的怀里。



看着简单，但是其实蛮多坑。

银渐层、游乐园、奔跑、跳、小女孩、怀里。

这些个关键词，其实都不好理解，更别提跑着跑着跳到小女孩怀里这种操作了。

首先要准确识别出银渐层这个特定品种的猫，还得理解它在游乐园这个复杂场景中的运动轨迹。

更难的是，模型需要精准捕捉从奔跑跳跃的动作转换，还要准确把握跳入怀中这个互动场景的空间关系。

这个能完美还原，就挺牛逼的。

还有这个case:

45度俯拍，一位紫衣女修在竹林中抚琴，琴音化作七彩音符在空中飘荡。翠竹摇曳，月光如水。



虽然这个七彩音符加的吧，总会让我想起一些非常古早的渐变PPT，但是咱们忽略审美的事，你会发现，混元都给你还原出来了，45度俯拍、紫衣、抚琴。

还有那句最重要的：琴音化作七彩音符在空中飘荡。

你如果用其他AI视频都把这个Prompt跑一遍，你就知道，能精准的出现七彩音符这事，有多难了。

还有一个我超级喜欢的case：

星系边缘，宇宙战舰引爆反物质引擎。能量涟漪以光速扩散，撕裂周围星体。



这个镜头，是我幻想过，在科幻片里看到的一幕。

而现在，是由AI，给我精准的还原出来了。

就是我心里的，奇点爆炸。

当然，在我测的几百个case中，并不是说，什么语义他都能理解。

翻车的也有很多，但是整体成功率，还是相比于其他我用过的，高出不少。但是带来的负面也很明显，有的视频，那个审美真的挺差。

语义理解和审美之间，其实一直需要找一个平衡点，语义理解强，就代表着可控性很强，也代表着用户对AI的干涉非常强，一旦用户自己随心所欲的去创作，也可能会让出来的美感，没有保障。

典型的就Midjourney，它的语义理解，我觉得到现在为止，都可以说是一坨屎，包括可控性也是一坨，ControlNet说了一年了，到现在也没加。

而他们给出的理由，就是如果让用户对AI的生成干涉过多，就会影响审美，让出来的东西变丑，作为一家有审美洁癖的公司，这个是他们不可接受的。

所以就一直拖一直拖。

混元的AI视频，走到了Midjourney的反面，给用户的自由度很高，但是这时候就又吃用户自己对于画面的理解了，换句话说，就是：

上限很高，下限同样，也极低。

三.可以切换镜头

混元这个，跟PixelDance是目前全世界唯二，可以自主控制切镜头的。

比如：**广角镜头，破碎的镜子碎片漂浮在空中，每片镜子里都映照出不同的时空画面。画面中央是一位银发时空术士，他穿着暗纹长袍，双手编织时空之线。**



我其实没有写类似于镜头切换的关键词。

但是混元在理解完我的Prompt后，给我切镜了。而切镜完的效果，确实张力好一些。

最重要的是，人物、场景几乎完全一致。这点，就很酷。

还有士兵看信这个case。

战场废墟中，一位士兵跪地，颤抖着手打开战友留下的信。然后镜头切换到信件的特写。远处炮火映红天际，烟尘在空中飘散。战争片风格。



这次我是主动说了“镜头切换”这个词，混元理解的非常好，场景、人物一致性也堪称完美。这一点，我非常喜欢。

写在最后

AI视频这个行业，我一直觉得对于创业者或者小公司来说，不是特别友好。

不友好的点在于，获取高质量数据的难度，相比于文本、图像啥的，太高了。

大厂啊，护城河还是太高、太深了。

最关键的是，腾讯用这个自己深厚的家底，先把这个模型的v1版本做完了，然后，直接开源，免费送。

这尼玛，谁顶得住。

目前混元AI视频模型，只支持文生视频，不过他们说图生视频也很快会上线了，马上就做完了。

文本大模型、AI绘图大模型、3D生成大模型，再加上这次的AI视频大模型。

如果再来一个AI声音模型，腾讯就是，真正的全系开源了。

只能说，腾讯对于自己的市场定位和核心竞争优势，也有着极度明确的认知。

腾讯的城堡，还在向天空挺进。

以上，既然看到这里了，如果觉得不错，随手点个赞、在看、转发三连吧，如果想第一时间收到推送，也可以给我个星标★~谢谢你看我的文章，我们，下次再见。

>/ 作者：卡兹克

>/ 投稿或爆料，请联系邮箱：wzglyay@gmail.com