https://www.douyin.com/video/7336849049473125682

# 标题:未找到标题  
## 关键字: 未找到关键字  
## 作者: 严伯钧  
## 视频ASR文本:  
 都看到了吧啊 open ai 开年放了个大招年初期就来了个 solo 啊文字生成视频这视频效果如果是真的不是剪辑的那真的是太炸裂了啊感觉谷歌呀真的是被 open ai 偷家偷的很彻底啊 china gbd 的底层技术 transformer 架构呢最早就是谷歌发的论文 那这一次这个 sora 背后的技术你以为就不是谷歌的这个论文吗其实也是啊 transformer 架构呢是基于一篇叫做 attention is all you need 的论文 sora 背后的技术叫做这个 diffusion transformer model 啊猜猜这篇论文是谁发的也是谷歌呀就这一篇啊 photorealistic textured image diffusion models with deep language understanding 啊这篇论文啊可是珍藏那就熟练的在 archive 里面点这个 link 把论文放到 t x y z 里面 研读一下很快就能搞明白 t s i z 啊最近还出了个新功能就是它给你回答的内容里面啊有 reference 了啊就点一下这里这个配 啊就直接会出现在原文里面给你标注出来他这个回答的内容啊是依据原文里哪一段给你提炼的反正呢我很快就通过原论文看懂了这个纹身图纹身视频的原理了这里面那个关键就叫做这个 diffuser model 啊这个中文直接翻译的话呢叫做扩散模型这个 diffuser 的过程其实就是模拟物理学当中的扩散过程 你想象一杯清水我往这个清水里面滴一滴墨水然后呢这一滴墨水就扩散到整杯水里面清水呢就会被染色了但是墨水在清水里扩散的这个过程啊这个路径啊其实是随机的并且跟很多参数有关比方温度比方里面的这个离子分布等等 defuser model 的原理就跟这个很像我要怎么通过文字生成一张图片呢哎先说图片啊再说视频这里面就涉及到了一个 diffuser model 的神经网络了啊这个事情要做到当然是要靠训练但训练的东西是什么呢其实是路径那又是什么路径呢其实是一张图片被 完全模糊掉的路径这个训练过程是这样的比方我有一堆图片这些图片呢我是打好语言标签的例如这张图片我就给他一个描述叫做这是 t x y z 的 logo 啊好了那下一步是什么呢我就要给这张图片加噪音 怎么加呢用高斯分布的噪音这张图片呢是由像素点构成的我每次加噪音啊都是以这张图片的每个像素点的颜色码作为中间值我们知道像素点的颜色码就是 rgb 码以这个码为中间值给它随机的加减一个值这个值呢是随机的但是这个随机值的概率分布满足高斯分布 我加了一轮噪音以后啊这个图就跟原来的图有一点不一样了这个就好像你往清水里面滴一滴墨水这个墨水会慢慢扩散然后呢我就不断的一轮轮的给这张图片加噪音直到什么程度呢直到这个噪音多到这张图片我已经看不出来是什么图了然后我把整个加噪音的过程啊这里会有 n 张 张图片我的 n 张图片都塞进神经网络这样的话呢神经网络就会学习学习到的是什么呢其实这张图片从完全模糊到清晰所经历的路径这个路径啊其实是个矢量的点的集合这样的话呢我这个训练过程就完成了实际情况我不可能只训练一张图片我可能有几千一张图片用来训练全是打标签呢 好了我训练完了这个神经网络我要怎么去通过文字生成图片呢这其实就是把刚才那个过程给它反过来当我跟 ai 说给我生成一个 t x y z 的 logo 他做的事情啊其实就是根据羽翼在矢量空间里面找到这条对应 t x y z logo 的这条路径 然后再反推刚才的过程就可以生成图片了但我们训练的图片集是非常巨大的这里面就出现创造力了例如我跟 ai 说给我生成一张长着两只头的鸡的照片那很显然我们训练的图片集里面不一定有两个头的鸡那这种创造力要怎么实现呢这个就跟 open ai 的这个文字处理师 不开了你跟 ai 说一句话他要很好的把你这句话进行 embedding 变成矢量然后在矢量空间里面找到对应的路径再进行一个加权的平均使得加权出来的路径贴合你的文字的意思这就是为什么 open ai 的这个 sora 效果好跟它这个自然语言处理的能力啊是分不开的 必须是对文字的理解到位才能保证生成的图片跟文字的含义是贴合的好了这是图片那视频要怎么办呢其实对于 ai 来说啊图片跟视频的区别其实是差一个维度那就是时间的维度 数据的类型多了一维因为视频呢就是一秒二十四张图片我把二十四张图片放在一块其实就是每个像素点不光有一个 rgb 码而是加了时间轴的 rgb 码然后它对应的也是一个矢量 方法论上跟上面的图片的 case 呢其实没有太多的区别那为什么 sora 这个效果那么好呢主要还是它文字的白领做的很好然后就是数据的资源好数据又多质量又高是最关键的 科学上的这个东西啊其实就还是 diffusion transformer 但具体执行过程中啊肯定有很多技术的细节在里面这这里就不清楚了但原理就是这个原理这就又要说到这个谷歌被 open ai 偷家的事了本来这些技术啊都是谷歌发明的为啥谷歌没有整出来呢 这个呢我就有比较一手的信息了啊主要原因大概有两个第一呢就是谷歌这样的大公司啊各种合规数据的隐私啊各种规矩条条框框很多这就导致数据的效率不够高简单来说呢就是公司大了盯着的人多干起事来啊束手束脚第二呢就是谷歌这种上市公司啊内部做一些事情组织很大有很多决策环节 例如你要做这个纹身图啊纹身视频动画你就要有足够的硬件资源去分配到这个事情上那么大公司就会要求你的内部创新啊用了那么多的资源最好是基于谷歌现有的产品去进行打造的这样才能带来经济的利益啊例如我就听说了当年第一个发明了啊纹身视频的技术路径的研究员说当时 他已经搞出来这个技术路径了但是谷歌愣是让他花了两年时间在想这个东西怎么改到油管上总之呢就是大公司不够灵活第二有产业压力有变现压力就导致了 这个被 open ai 给偷了家了啊 anyways 啊 sora 这个炸裂的效果我需要思考一下之后聊聊会有哪些行业会受到它的严重冲击啊听没听懂都点个赞呗