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  
##   
  
以下是补全标点符号和修订错别字的文本：  
  
都看到了吧，啊？Open AI 开年放了个大招，年初就来了个 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 偷家偷得很彻底。China GBD 的底层技术，Transformer 架构，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”。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 TX、YZ 里面，研读一下，很快就能搞明白 TX、YZ。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。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 TX、YZ 的 logo。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，我就不断地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 TX、YZ 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 TX、YZ 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只头的鸡。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。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，