各位老师好，我是数媒的徐嘉宁，我的答辩题目是基于生成式对抗网络的角色形象生成项目的实现。

我将由四个方面展开讲述。

随着深度学习技术的成熟，人工智能在人们日常生活中越来越常见，甚至在艺术方面都有了一席之地。2014年….

相比于传统的深度学习模型，生成对抗网络的优势在于其使用了一对神经网络模型。

生成模型的输入是随机噪声，输出是一张图片，判别模型的输入是来自数据集的图片或者生成模型产生的图片，输出则是一个标量，用于评判图片质量的好坏。

因此该问题就转化成一个求最小极大值的问题，对于判别模型来说，它要尽可能将生成模型的图判别为0；对于生成模型来说，它要尽可能让判别模型输出1.

DCGAN相比于GAN采用了三个改进，其中最重要的是它引入了全卷积网络这一点。

DCGAN模型构成如下，生成模型目标是生成图片，因此可以通过逆卷积操作不断扩大图像尺寸；判别模型目标是输出一个数，因此可以使用卷积操作不断缩小图像尺寸。

StyleGAN3模型构成如下，很明显，它要比DCGAN模型多了很多隐层，结构非常复杂，大大提高了神经网络的性能。

训练过程可视化通过visdom框架，在界面中左边是生成模型生成的假图，右侧是来自数据集的真图，每20次迭代更新一次。右侧是每次迭代时输出的损失值。

结果的可视化基于flask框架，两个标签页可以一键生成相应的生成图片，都设置了一些可以控制图片生成效果的参数。

DCGAN的训练过程如图所示，可以看出，随着迭代的进行，图片中颜色逐渐丰满，人物的轮廓逐渐清晰，300次之后图片已经相对完善了，到500次迭代后程序运行结束，等同最终效果。

噪声均值和方差都与图片的一致性和质量有关，均值越大，图片的一致性就越高（长得都一样了），同时图片质量越低。

方差越大，则图片一致性越弱，大家都非常不一样，而且颜色更加饱和，但数值过大图片质量也会降低。

StyleGAN3生成效果就好非常多，毕竟训练时长是DCGAN的几倍。最终的结果可以通过输入一个随机种子号得到，每个种子对应了一张图片，只要是整数就行。生成的效果……

Psi截断类似于DCGAN中的噪声均值和方差，都与图片一致性有关，psi值越大，图片则越不一致，越有个体特征。当psi截断=0时，无论随机种子设置为多少，其生成图像永远为第一张图，大家都没区别了。

这个旋转与我们常识中的概念不太一样，是StyleGAN3为了调整某些图片的细节改进的，因此有些时候微调可以改善图片质量，数值过大则会破坏图片。

参考StyleGAN3的优化方案以及其他GANs的优化策略，我对DCGAN也做了一些优化工作。DCGAN在生成生成自然风景图时更占优势，风景图中的事物拥有自然痕迹，即，其画面时连续的、可以模糊的、可变的，例如，山高一点还是山，云扭曲一点还是云；但人物画面则不是，图像不清晰就会很难看，眼睛移动了一点位置就会很奇怪，因此对应自然痕迹，人工痕迹需要的是细节的、精确的、固定的表达。我认为，影响“人工痕迹”主要有三点：……

第一点，使用更加纯净的数据集和数据清洗工作。最开始使用的数据集是原项目自带的数据集，里面的图片大小不一，尺寸都比较小，而且有些背景很不明显，这样很影响神经网络的学习。后来我在kaggle上找到了一份更加干净、规范的数据集，并且用脚本剔除了一些打不开的、错误的图片，整个数据集的情况都好了不少。

第二点是参数的调整，先讲一下两个参数……我认为原项目中的学习率设置过大了，因此改小后重新训练。

第三点是预训练模型，解释一下。例如将迭代的总次数延长成1000次，可以提升图片质量。但是更好的做法是，先训练500次，导出模型，再以这个模型为初始模型，叠加500次迭代。效果会更加好一点。

如果我们拥有无限庞大的数据集，确实没有必要创造一个生成模型。现实是，我们的数据集是有限的，然而有限数据集的特征可以组合出近乎无限的高维特征，例如头发的颜色加上眼睛的颜色，光是这两个特征就可以组成数百种可能。作为人类，要认真地创造一个蓝头发加红眼睛的角色就可能需要耗费许多时间成本，而生成模型允许我们仅改变一个参数，就能实现一个角色的诞生。因此，GAN能够成为节省时间的利器。

GAN帮助我们训练出来的模型并不是顽固的，即使我们使用同一个模型，设置同样的属性，生成的结果也会有微小的不同，这是顽固的数据集无法做到的

图展现了GAN与图像连贯性之间的联系，这让人不禁想到另一个场景——动画，而现在确实也有人将GAN技术应用在动画生成中，实际上在引言中提到的Crypko网站也在进行同样的项目。因此从创新和交叉领域的角度来说，GAN技术是相当有意义的。