深度学习测试作业实验报告

代码结构

1. model.py 训练模型model.hdfs，主要参考了 CSDN 博客：https://blog.csdn.net/qq_20989105/article/details/82760815, 训练出来准确率为 0.8734
2. ssim.py 用于判断攻击样本相似度，参考 ssim 定义实现
3. attack.py 包含 generate()外部接口，遍历图片集生成攻击样本并计算时间和 ssim，get_label()返回图片标签，change_ pixel()改变像素点，random_attack()循环攻击保留相似度最高的图片，security_attack()是一个保底攻击，load_data()加载数据

算法详解

本来想采用fgsm算法，但是论文没看太弄懂，写了几天也没很好的实现思路，最后就用简单的修改像素点实现了。随机定位1个像素点，随机修改为某个颜色。
另外为了避免攻击失败的情况，还设置了一个保底攻击算法 security_attack()，添加了 10 个白色噪音。
循环 10 次所得图片的 SSIM 值大概为 0.051，一个用例耗时约为 5.6s.

个人感受

第一次接触机器学习，最开始对于作业题目无处下手，看了很多博客后才稍有思路，自己的方法比较简单，跑出来结果也不高，尝试了好几种修改像素点的方法但循环 10 次的 SSIM 都在 0.05 左右，循环 1500 次能跑到 0.2，但是电脑跑不完 1000 个用例，而且就算是 0.2 也是一个很低的数值，可能一开始的方向就错了，后来也想试着去用一些现成算法但是在实现上没什么思路，只好作罢。除此之外再环境配置上也花了较多时间，tensorflow 老是在版本兼容上出问题。

参考资料

1. Little Programmer 《TensorFlow之tf.keras的基础分类》https://blog.csdn.net/qq_20989105/article/details/82760815