**Fawkes图像隐身**

**如今强大的面部识别模型的激增对个人隐私构成了真正的威胁。面部识别系统在未经明确同意的情况下扫描了数百万公民。**

Fawkes实现图像隐身效果，此项目相对于大部分网民来说使用门槛较高，意在照片上传公共服务器前就实现隐身效果，其原理是对图像数据扰动来做到隐身，这让人不禁想到图像对抗算法deepfool，道理差不多，图像对抗算法实现暂且不谈，先就看fawkes的实验数据：

1. 部署：

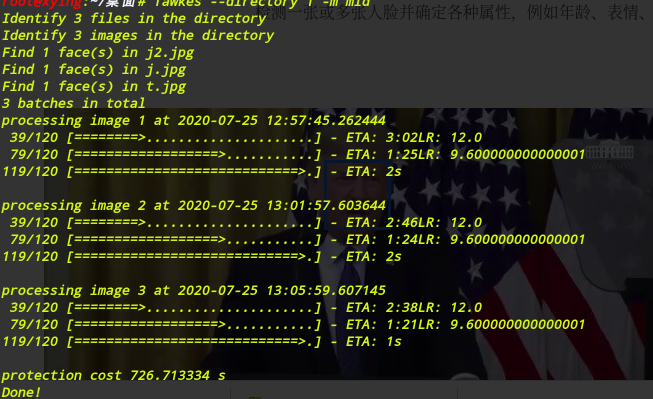
1、基于python3.5-python3.7环境

2、linux安装tensorflow-1.14.0 windows则为tensorflow-1.15.0此为目前语言环境各系统所能支持的最高版本，tensorflow最低不可低于1.12

二、大致原理：

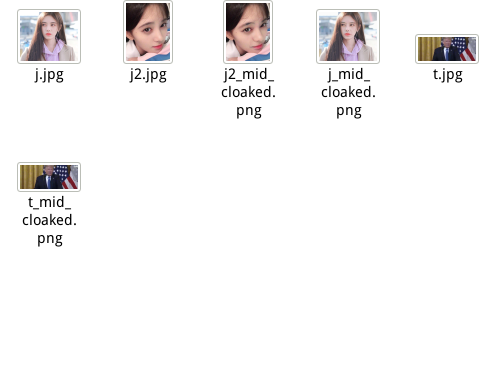
用法不再详细讲解，直接看结果，脚本处理数据按照目录，将目录内图像序列渲染，使用肉眼难以察觉的扰动(DSSIM≤0007),对图像的特征空间表示进行显著的改变。

处理需要一定时间：



中等级别数据保护处理平均每张图像耗费一分钟多，越高级别保护则处理时长越多

处理完成后生成对应图片。



我们使用中级防护处理后的图像数据先做试验

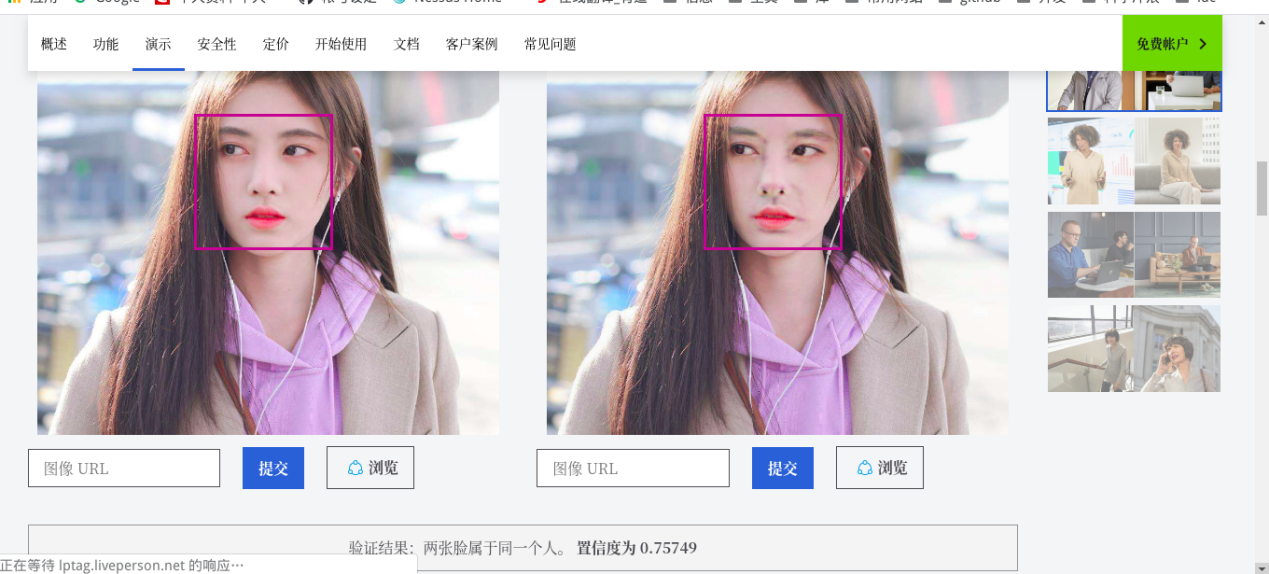
微软Azure人脸识别结果：

首先上传原图：





可以看到，微软面部检测都十分详细，我们在看生成后照片的相似度：





属于同一个人，我们在将扰动后的图像上传到面部检测：





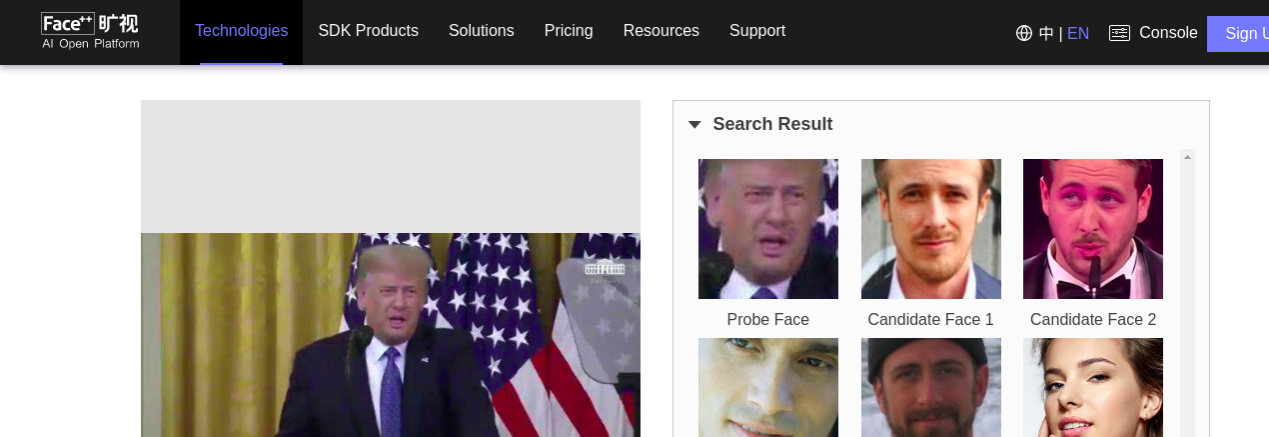
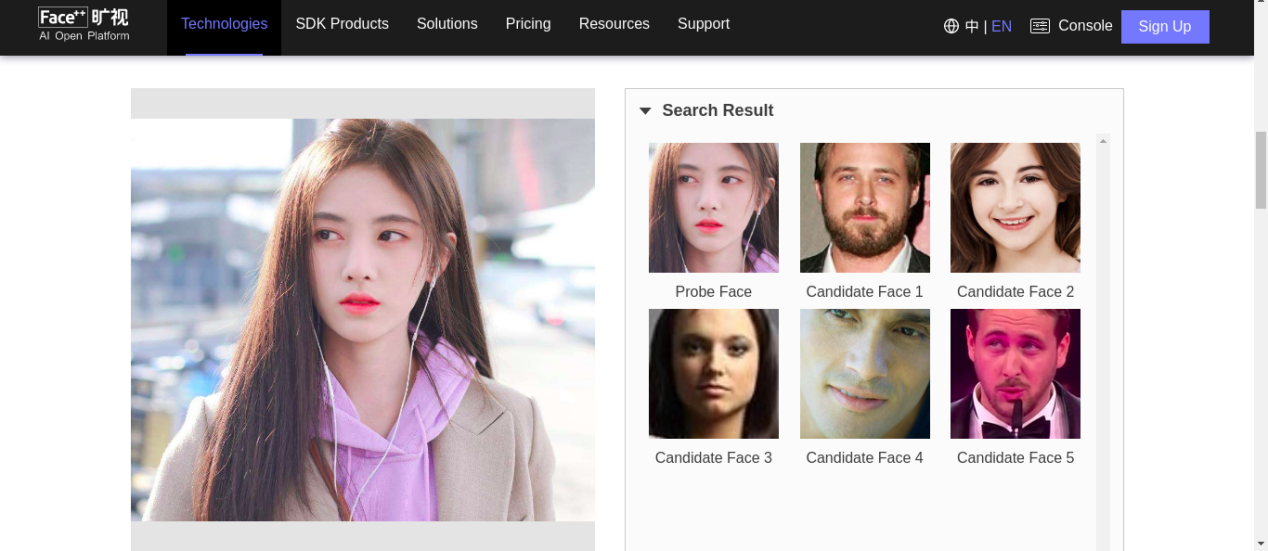
可以看到，AI识别出现了错误。当然怎么渲染也不会合成到那么大的：





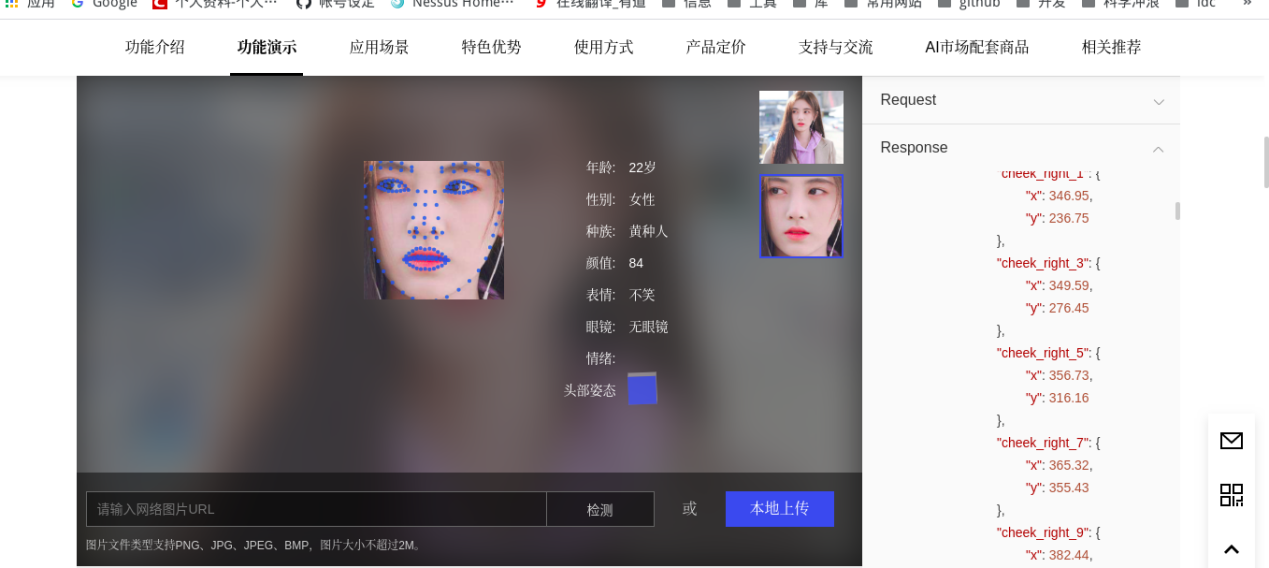
中级别防护旷视实验结果：

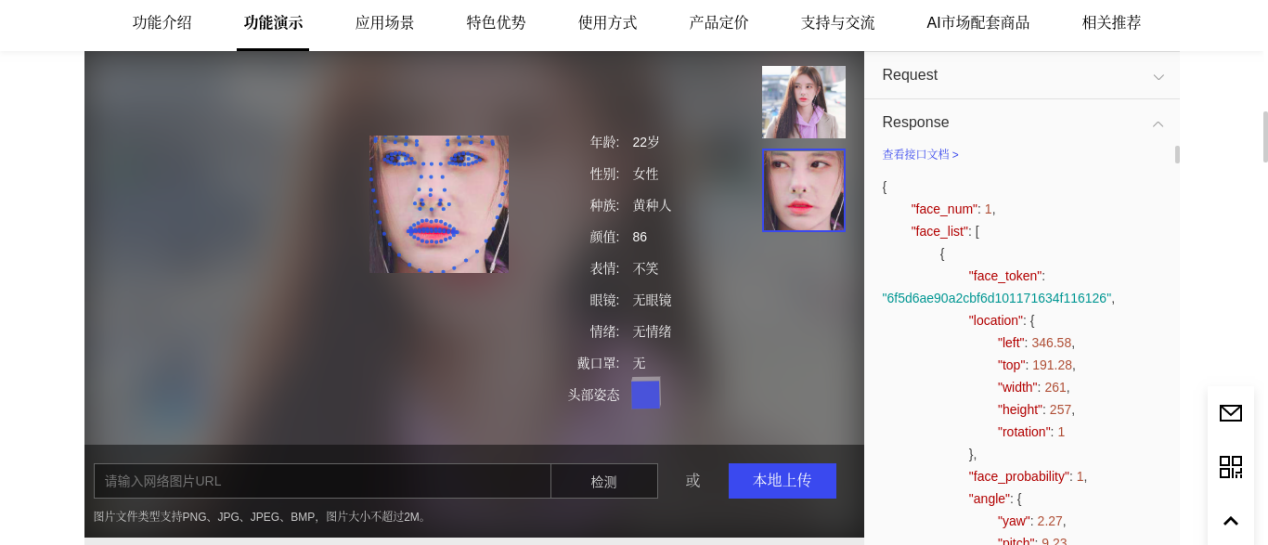
节省时间不做对比，直接上传生成后的



并没有对旷视AI造成影响，只不过多让它花费了点时间识别而已。

百度AI亦如此：





在花费一些时间来等高级别防护对抗样本

后续高保护级别样本：

可影响AI判断，但变化较大，就：

