

人工智能应用实践 开题报告

——人脸检测任务

学号：201720121206

班级：17 级人工智能班

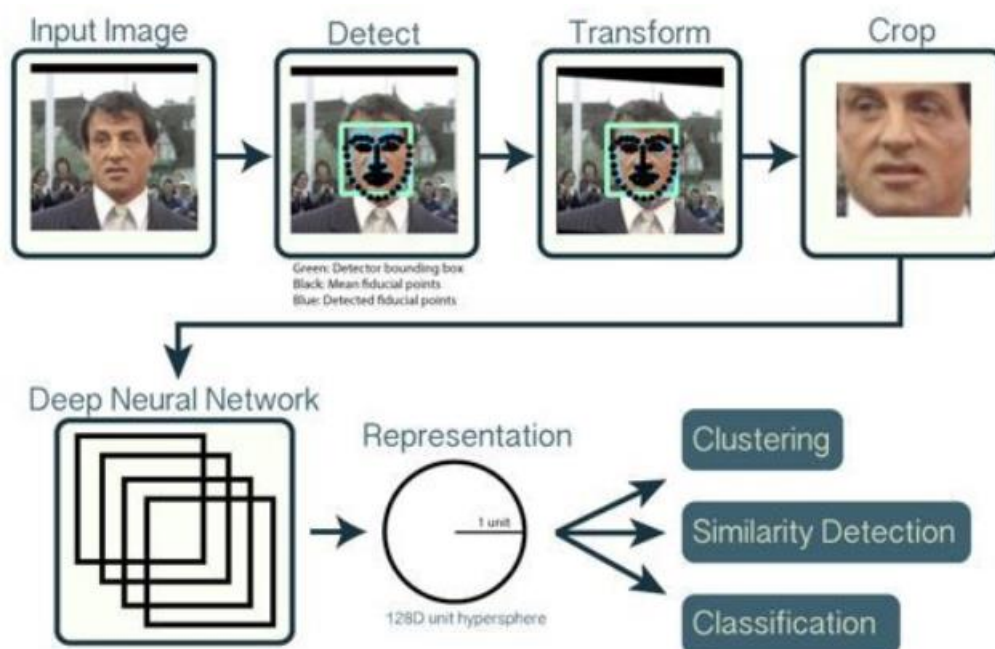
姓名：吕程

目录

1	研究课题	1
2	数据集	2
3	实现方法	3
3.1	早期方法	3
3.2	深度学习框架	3
4	开源代码	8
5	参考论文	9
6	预期成果	10

1 研究课题

人脸检测是机器视觉领域被深入研究的经典问题，在安防监控、人证比对、人机交互、社交等领域都有重要的应用价值。数码相机、智能手机等端上的设备已经大量使用人脸检测技术实现成像时对人脸的对焦、图集整理分类等功能，各种虚拟美颜相机也需要人脸检测技术定位人脸，然后才能根据人脸对齐的技术确定人脸皮肤、五官的范围然后进行美颜。在人脸识别的流程中，人脸检测是整个人脸识别算法的第一步。



2 数据集

LFW: 全名是 **Labeled Faces in the Wild**.这个数据集是人脸评估一定会用到的一个数据集,包含了来自 1680 的 13000 张人脸图,数据是从网上搜索来的。基本都是正脸。这个数据集也是最简单的,基本主流算法都能跑到 99%以上,貌似有 6 对 label 错了,所以最高正确率应该是 99.9%左右。这个都跑不到 99%的话别的数据集表现效果会更差。一般来说这个数据集是用来做人脸识别验证的。

CelebFaces: 总共包含 10177 个人的 202599 张图片,也是从搜索引擎上爬过来的,噪声不算多,适合作为训练集。同时这个数据对人脸有一些二元标签,比如是否微笑,是否戴帽子等。如果需要特定属性的人脸,也可以从中获取

CASIA-WebFace: 该数据集是从 IMBb 网站上搜集来的,含 10K 个人的 500K 张图片。同时做了相似度聚类来去掉一部分噪声。**CAISA-WebFace** 的数据集源和 **IMDb-Face** 是一样的,不过因为数据清洗的原因,会比 **IMDb-Face** 少一些图片。噪声不算特别多,适合作为训练数据。

VGG-Face: 来自 2622 个人的 2 百万张图片。每个人大概要 2000+图片,跟 **MS-Celeb-1M** 有很多重叠的地方(因为都是从搜索引擎来的),这个数据集经常作为训练模型的数据,噪声比较小,相对来说能训练出比较好的结果。

3 实现方法

3.1 早期方法

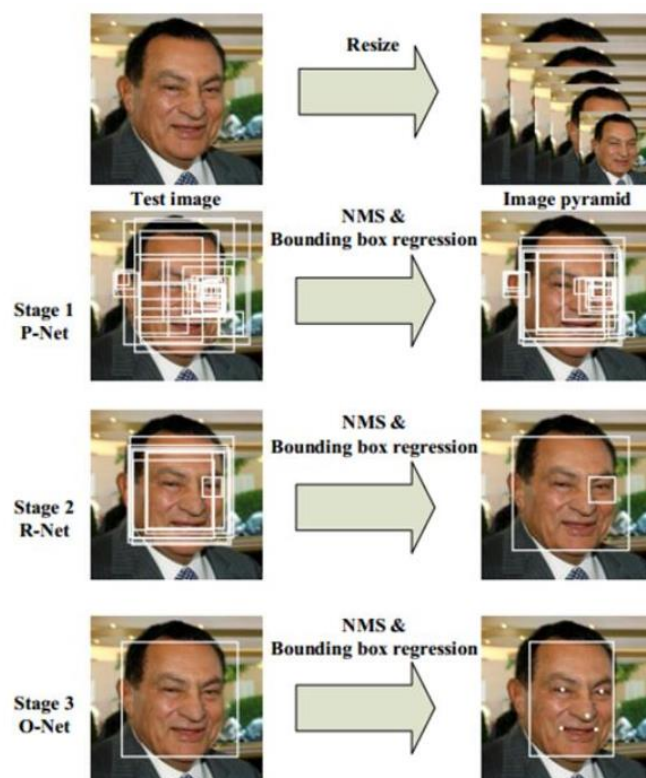
早期的人脸检测算法使用了模板匹配技术，即用一个人脸模板图像与被检测图像中的各个位置进行匹配，确定这个位置处是否有人脸；此后机器学习算法被用于该问题，包括神经网络，支持向量机等。以上都是针对图像中某个区域进行人脸-非人脸二分类的判别。

3.2 深度学习框架

MTCNN (Multi-task convolutional neural networks) 人脸对齐

该 MTCNN 算法出自深圳先进技术研究院，乔宇老师组，是 2016 年的 ECCV

MTCNN 顾名思义是多任务的一个方法，它将人脸区域检测和人脸关键点检测放在了一起，同 Cascade CNN 一样也是基于 cascade 的框架，但是整体思路更加巧妙合理，MTCNN 总体来说分为三个部分：PNet、RNet 和 ONet，如下图所示：



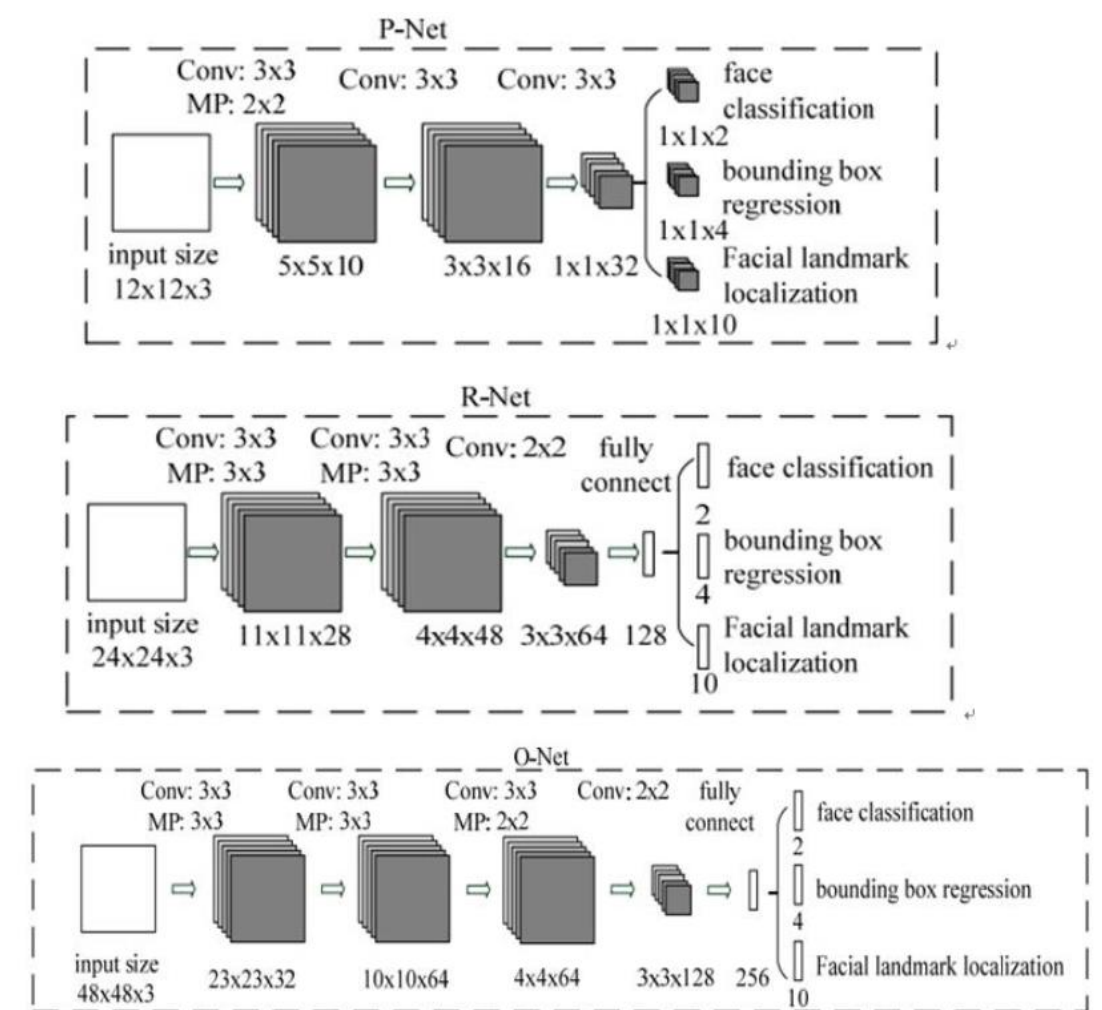
Proposal Network (P-Net): 该网络结构主要获得了人脸区域的候选窗口和边界框的回归向量。并用该边界框做回归，对候选窗口进行校准，然后通过非极大值抑制（NMS）来合并高度重叠的候选框。

Refine Network (R-Net): 该网络结构还是通过边界框回归和 NMS 来去掉那些 false-positive 区域。

只是由于该网络结构和 P-Net 网络结构有差异，多了一个全连接层，所以会取得更好的抑制 false-positive 的作用。

Output Network (O-Net): 该层比 R-Net 层又多了一层卷基层，所以处理的结果会更加精细。作用和 R-Net 层作用一样。但是该层对人脸区域进行了更多的监督，同时还会输出 5 个地标（landmark）。

详细的网络结构如下图所示：



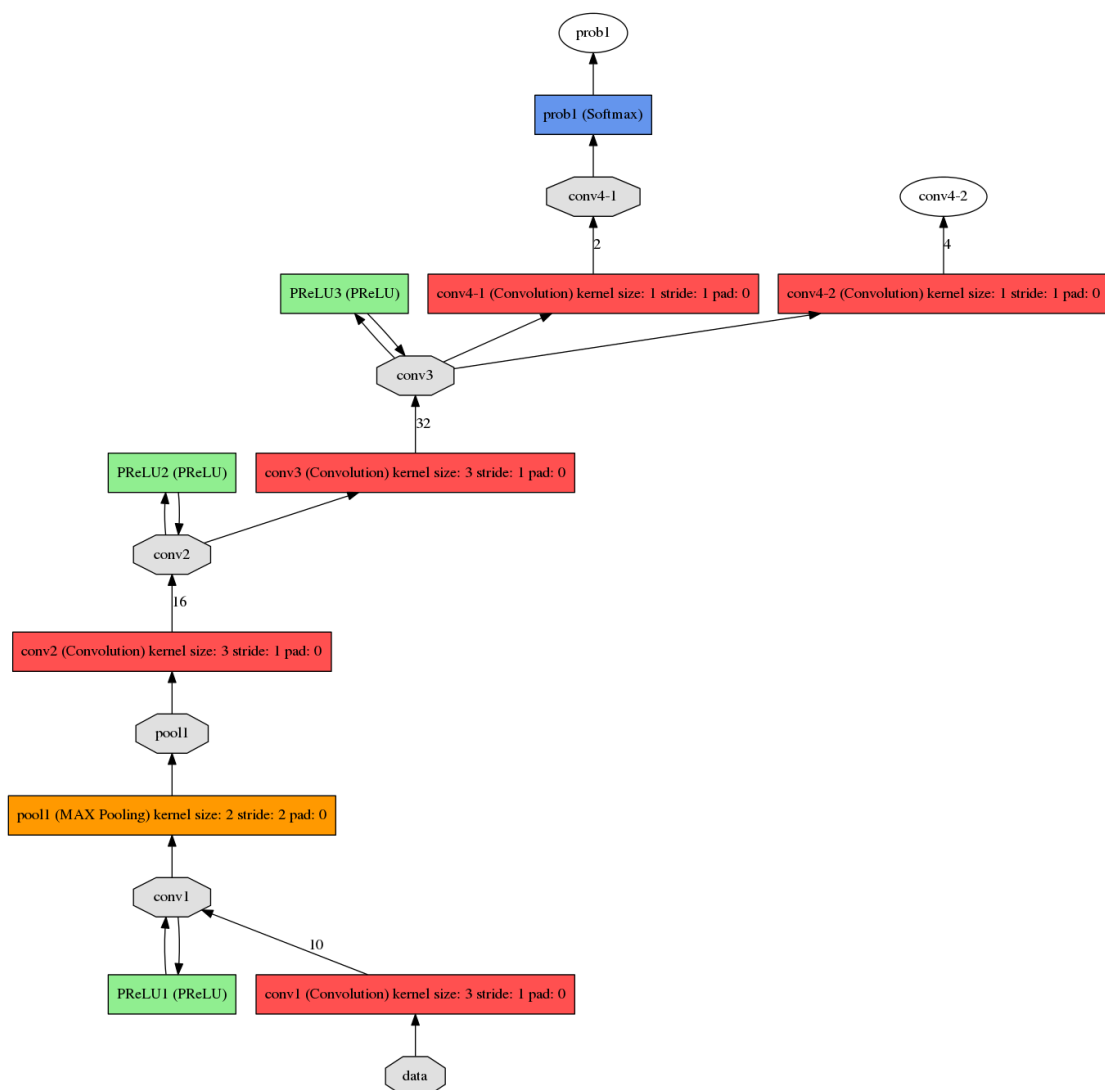
MTCNN 在测试第一阶段的 PNet 是全卷积网络（FCN），全卷积网络的优点在于可以输入任意尺寸的图像，同时使用卷积运算代替了滑动窗口运算，大幅提高了效率。

除了增加人脸 5 个关键点的回归任务，另外在 calibration 阶段采用了直接回归真实位置坐标的偏移量的思路替代了 Cascade CNN 中的固定模式分类方式，整个思路更为合理。

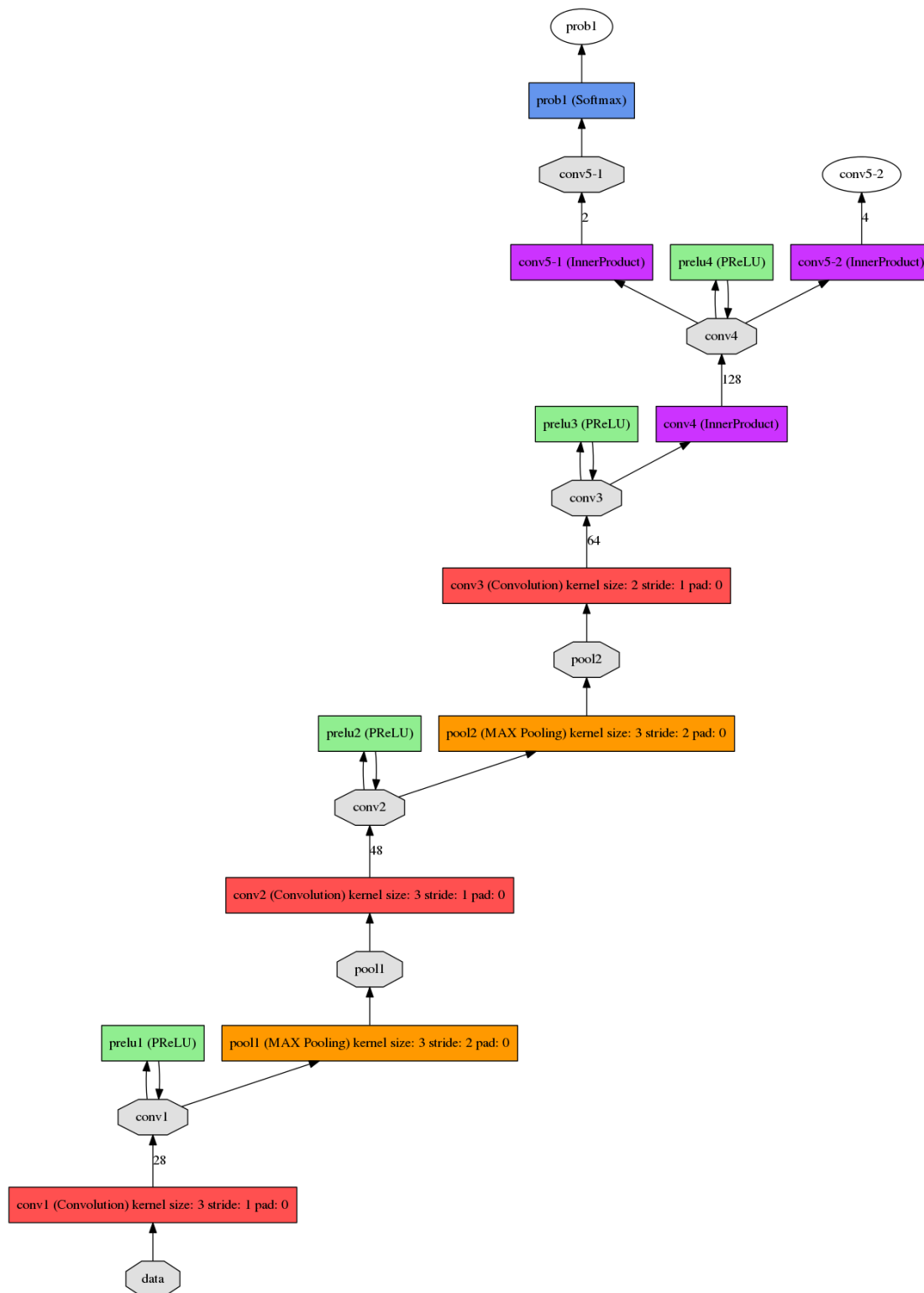
MTCNN 的整体设计思路很好，将人脸检测和人脸对齐集成到了一个框架中实现，另外整体的复杂度得到了很好的控制，可以在中端手机上跑 20~30FPS。该方法目前在很多工业级场景中得到了应用。

prototxt 的更加详细的网络结构如下：分别为 det1,det2,det3。

Det1.prototxt 结构：

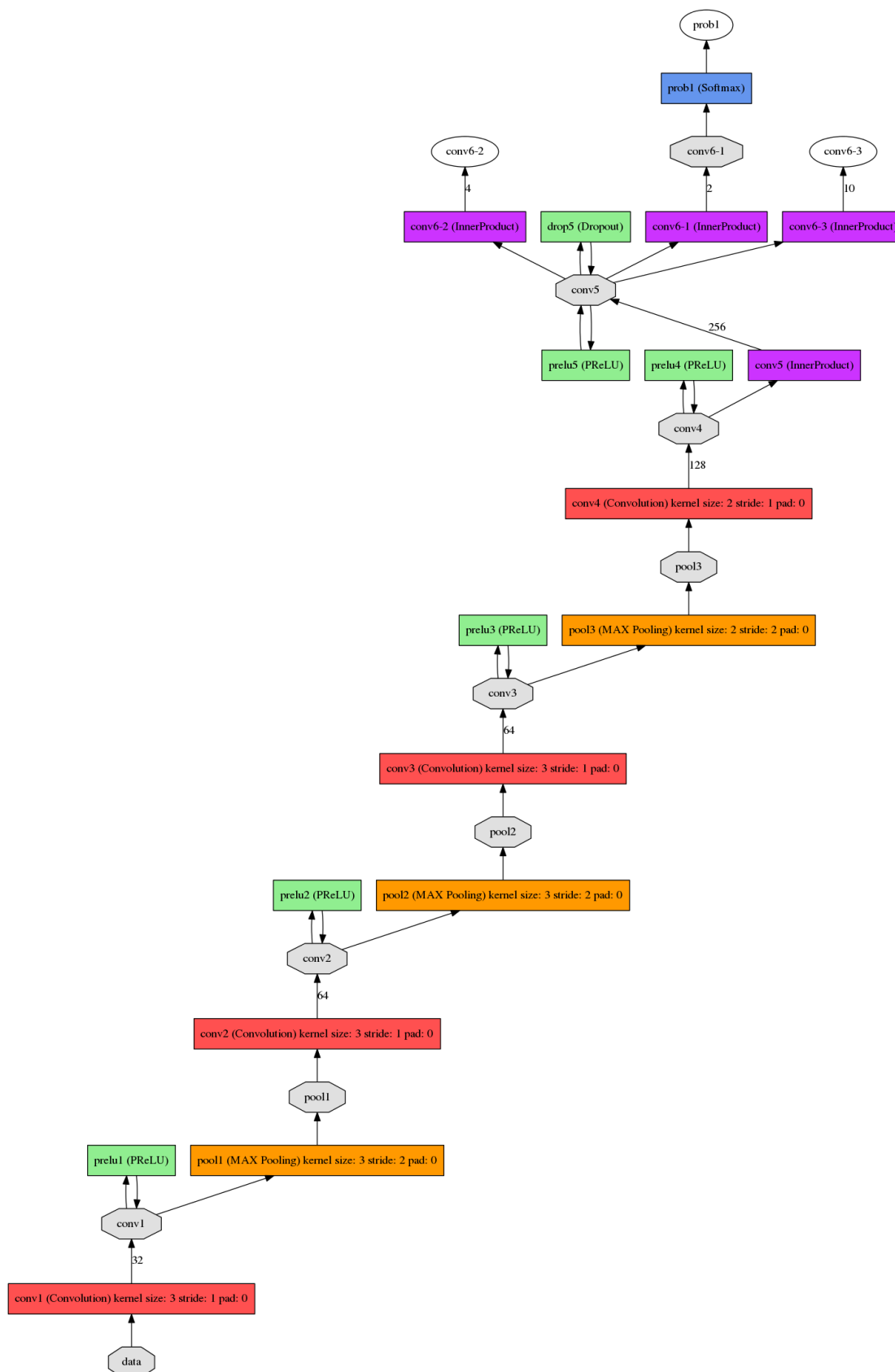


Det2.prototxt 结构:



Det3.prototxt 结构:

人工智能应用实践开题报告——人脸检测任务



4 开源代码

[Face recognition using Tensorflow](#)

[Reproduce MTCNN using Tensorflow](#)

[OpenFace](#)

5 参考文献

Deep Face recognition

FaceNet: A Unified Embedding for Face Recognition and Clustering

Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks

Selective Refinement Network for High Performance Face Detection

Improved Selective Refinement Network for Face Detection

6 预期成果

首先对参考论文加以研究，然后选择一套合适的开源代码配置相应的环境训练相应的模型，针对不同数据集进行测试，并且在不同的数据集上进行训练，挑选出最佳模型，在人脸验证、人脸识别、人脸聚类方面都能实现相应的功能，争取在此代码的基础上做一些添加，能够实现变脸的技术。