

# 基于 MTCNN 和 Facenet 的临时人脸识别系统

## 摘要

本系统是针对人群的临时身份识别，在没有前期训练的情况下，实现基于面部的身份区分。本次实验所构建的系统所使用的技术框架为 MTCNN 以及 Facenet。

其中 MTCNN(Multi-task convolutional neural network)，多任务卷积神经网络，主要负责将一张图片中的人脸用一个矩形框划分出来。而 Facenet 则用于提取人脸的特征，本系统将 InceptionResNetV1 网络作为 Facenet 的 Backbone 网络，将提取的特征映射到欧式空间的特征向量上，得到一个 128 维的关于输入人脸的特征向量。

在得到向量后，会将特征向量和人脸的图片存入后方的数据库，便于以后进行比对，即临时身份识别。除此之外，设计一套 UI 界面，便于更方便的使用和操作这一系统。

关键字：人脸识别；MTCNN；Facenet；神经网络；

## 一、项目概述

本项目是针对人群的临时身份识别，在没有前期训练的情况下，实现基于面部的身份区分。本次实验所构建的系统所使用的技术框架为 MTCNN 以及 Facenet。

其中 MTCNN(Multi-task convolutional neural network)，多任务卷积神经网络，主要负责将一张图片中的人脸用一个矩形框划分出来。而 Facenet 则用于提取人脸的特征，本系统将 InceptionResNetV1 网络作为 Facenet 的 Backbone 网络，将提取的特征映射到欧式空间的特征向量上，得到一个 128 维的关于输入人脸的特征向量。

在得到向量后，会将特征向量和人脸的图片存入后方的数据库，便于以后进行比对，即临时身份识别。除此之外，设计一套 UI 界面，便于更方便的使用和操作这一系统

## 二、MFCNN 简介

多任务卷积神经网络(MTCNN)实现人脸检测与对齐是在一个网络里实现了人脸检测与五点标定的模型，主要是通过 CNN 模型级联实现了多任务学习网络。整个模型分为三个阶段，第一阶段通过一个浅层的 CNN 网络快速产生一系列的候选窗口；第二阶段通过一个能力更强的 CNN 网络过滤掉绝大部分非人脸候选窗口；第三阶段通过一个能力更加强的网络找到人脸上面的五个标记点；

如下图 1 所示，该 MTCNN 由 3 个网络结构组成（P-Net，R-Net，O-Net）

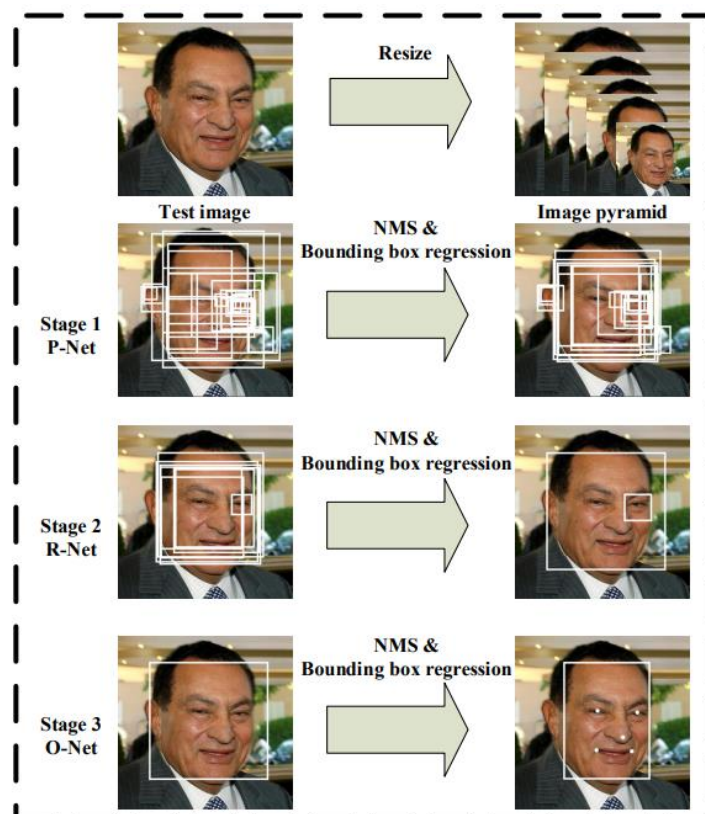


图 1. MTCNN 示意图

关于三个网络的简介如下：

- 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）。

同时，三个网络的结构示意图如图 2 所示：

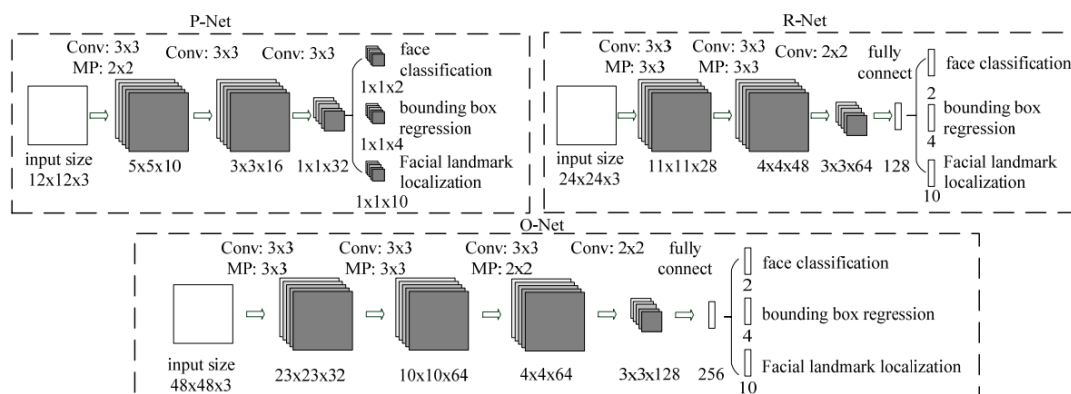


图 2. P-Net, R-Net, O-Net 结构示意图

该项目经过 MTCNN 网络的处理过后，会将一张图像中的人脸部分所识别和提取出来，方便后续的特征提取等操作

### 三、Facenet 简介

Facenet 是谷歌提出的一种人脸检测算法，发表于 CVPR 2015。与其他的深度学习方法在人脸上的应用不同，FaceNet 并没有用传统的 softmax 的方式去进行分类学习，然后抽取其中某一层作为特征，而是直接进行端对端学习一个从图像到欧式空间的编码方法，然后基于这个编码再做人脸识别、人脸验证和人脸聚类等。

FaceNet 算法有如下要点：

- 去掉了最后的 softmax 层，而是用元组计算距离的方式来进行模型的训练。使用这种方式学到的图像表示非常紧致，使用 128 位足矣。
- 元组的选择非常重要，选则适合的可以很快的收敛。

Facenet 的原理如下图所示：

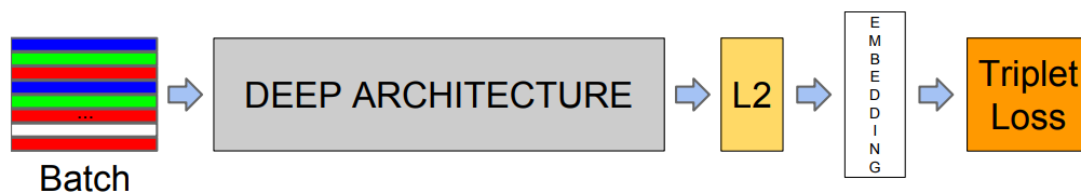


图 3. Facenet 原理图

本项目在使用 Facenet 时，主要进行的是如下步骤：

- 输入一张经过 MTCNN 识别截取后的人脸图片
- 通过深度卷积网络（**InceptionResNetV1**）提取特征
- L2 标准化
- 得到一个长度为 128 特征向量。

该项目经过 Facenet 处理后，会得到关于一个特定用户面部的 128 维的特征向量。此后若是录入新用户的信息则将该特征存入后方数据库，若为识别身份则将图片的特征向量与数据库中的向量相对比，在不超过阈值的情况下选择数据库中最近的向量对应的用户作为识别出的身份。

## 四、项目运行结果展示

本次实验的 GUI 部分，主要使用了 python 的 tkinter 库和 PIL. ImageTk 库。其中，tkinter 用于提供一些基础的 GUI 设置, 包括窗口、对话框、输入框和按钮等，ImageTk 用于将电脑环境的照片导入程序，并转换成相应的格式。加入了自适应照片的模块，使导入的任意大小的照片都能按等比例缩放出现在 500\*500 的界面上。在编写代码过程中，有出现参数难以传递的情况，后来通过全局变量以及改变程序设计思路的方法解决。

本次程序界面如下：

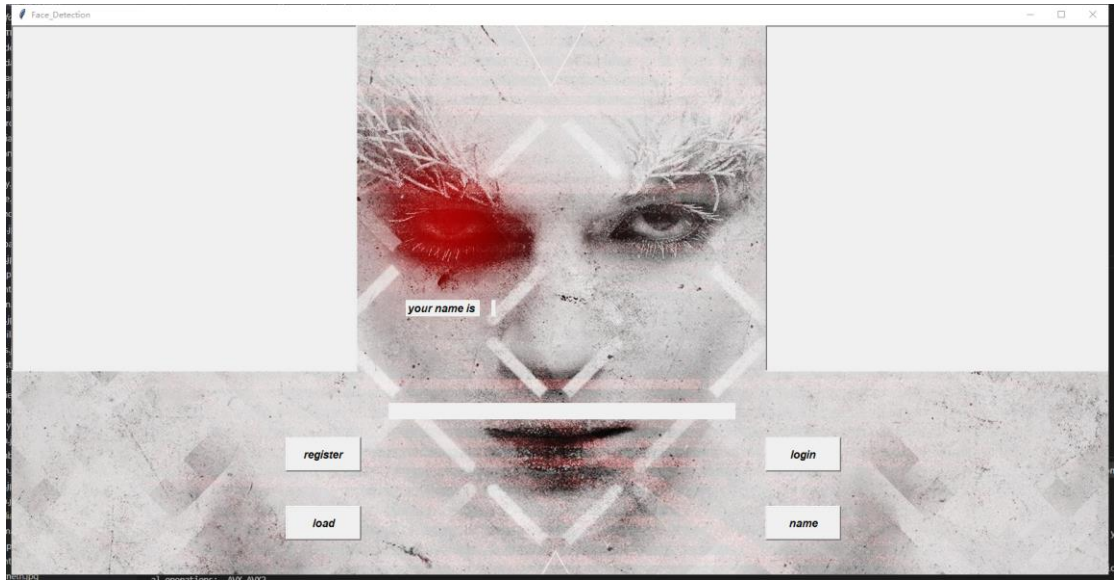


图 3. GUI 界面示意图

识别的结果展示如下图所示，图 3, 4 展示的是本小组成员的人脸识别示例，图 5 展示的是其他数据库中人脸的测试（可以看到都正确的识别出了身份），图 6 展示的是没有录入数据库人脸的测试（显示未找到人脸信息）。

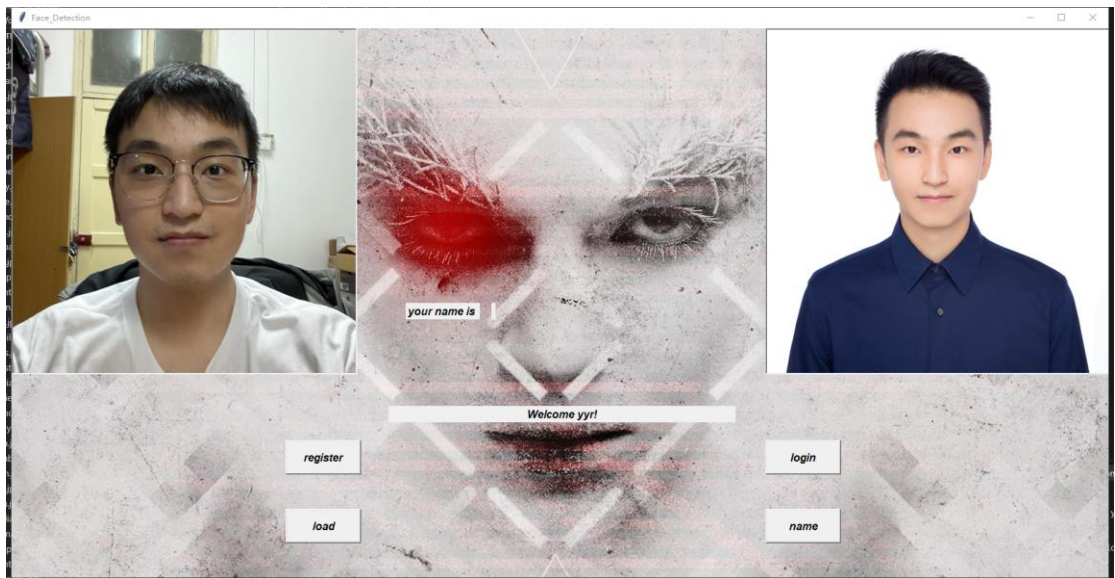


图 3. 小组成员的人脸识别示例 1



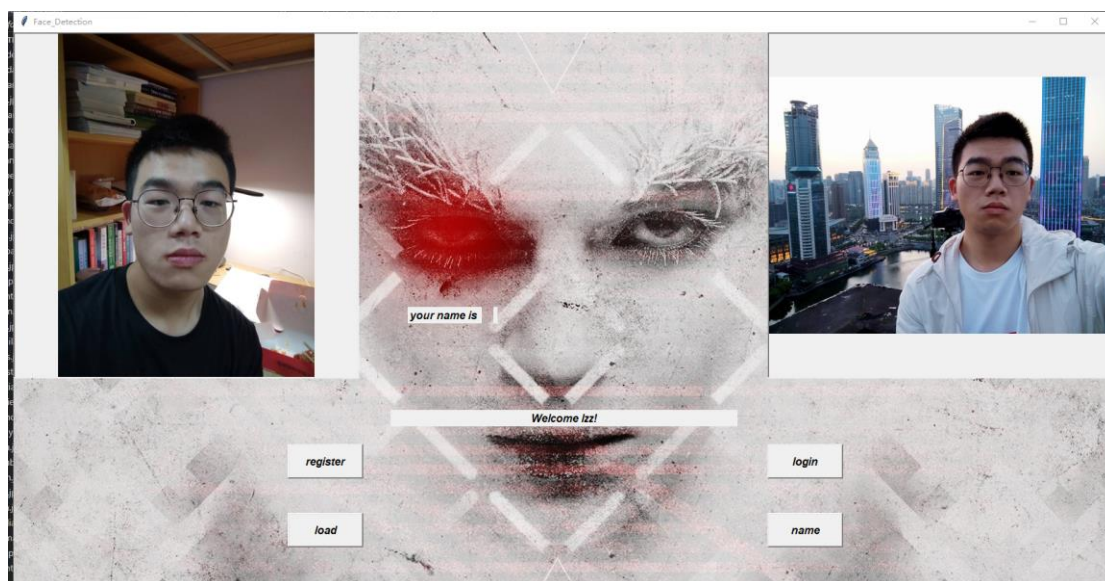


图 4. 小组成员的人脸识别示例 2

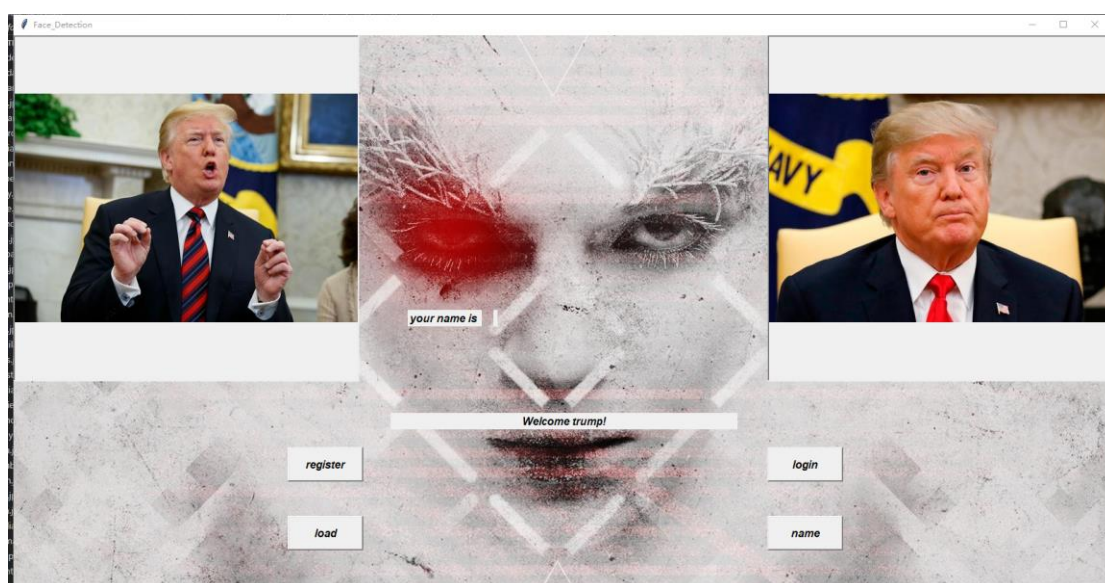


图 5. 其他数据库中人脸的测试

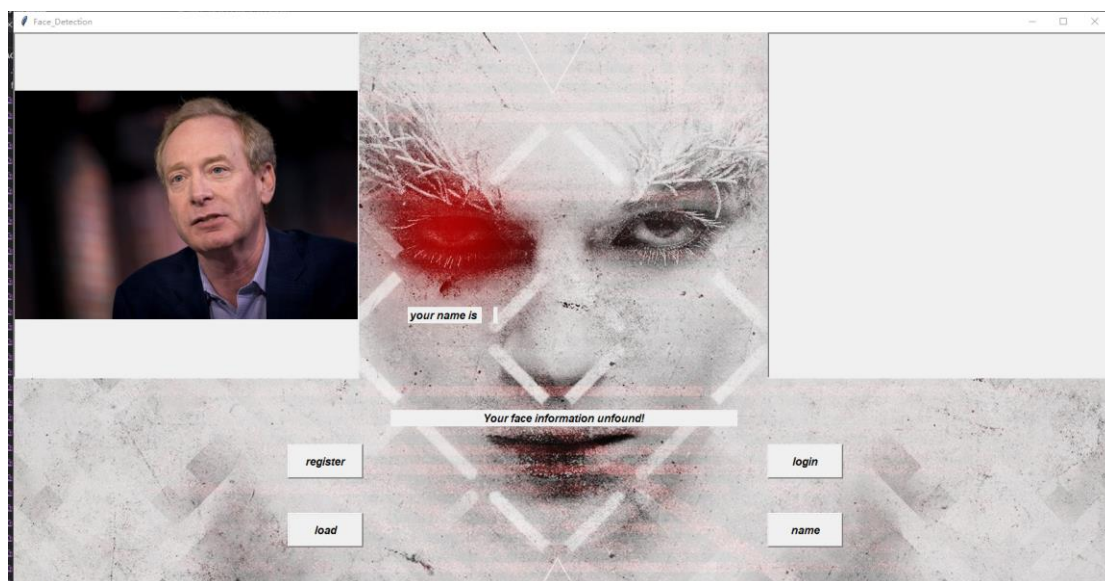


图 6. 没有录入数据库人脸的测试

综合来看，我们共录入了超过 100 位用户的人脸信息，并且识别的准确率在 95%以上。



## 五、参考文献

- [1] Florian Schroff, Dmitry Kalenichenko, James Philbin: “*FaceNet: A Unified Embedding for Face Recognition and Clustering*”, 2015;
- [2] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Yu Qiao: “*Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks*”, 2016;