基于VGG-face在线人脸识别系统概述

张军琛

（华中科技大学计算机科学与技术学院 湖北 武汉 430074）

**摘要：**人脸识别是计算机视觉领域的重要分支，近年来随着深度学习研究取的不断突破，由于深度网络对复杂问题的强大表述能力，深度学习方法在人脸识别领域体现出的性能已超越传统的人脸识别方法，最新的代表性的深度学习人脸识别算法模型在LFW和YFC等数据库上测试的识别准确度已达99%，超过了人眼鉴别的极限。VGG-face模型是用于解决大规模图形识别的深度卷积网络，本文基于VGG-face实现在线实时人脸识别系统，有效解决输入的待测人脸图多姿态时识别准确性变差等问题，系统识别结果获的可观的准确率，满足工程应用实时性。

**关键词：**人脸识别，深度学习，VGG-face，图像处理

**Abstract:** Face recognition is an important branch of computer vision, in recent years, with the breakthrough of deep learning, because of the strong ability of deep network to express complex problems, performance of deep learning method in face recognition field has been beyond the traditional methods. The recognition accuracy of new representative deep learning algorithm for face recognition test on the LFW and YFC database has reached 99%, more than the limit of human’s. The VGG-face model is convolutional neural networks used to solve visual recognition in large scale, this paper realizes the real-time online face recognition system based on VGG-face, effectively solves the recognition accuracy problems of input test image with multi pose variation, identification results obtains considerable accuracy, meets the general real-time application.

**Key words:** face recognition/identification, deep learning, VGG-face, Image Processing

# 1 引言

随着现今社会互联网等的快速发展，信息安全的重要性日益突出，身份鉴别的技术在社会中的应用更加广泛，目前身份鉴别技术主要用于金融、安保、网络传输、司法等领域，主要通过证件、口令、密码等手段实现，虽然该类方法在技术上比较成熟，并且可以结合一系列先进的加密保密认证策略，但由于其本质是对个体附加区分性信息，因而容易发生鉴别证件丢失、盗窃、伪造等问题。

人脸识别是基于人的脸部特征信息进行身份识别的一种生物识别技术，人脸是指人的面部五官以及轮廓的分布，这些分布特征因人而异，与生俱来，具有良好的防伪、防欺诈、准确、直观、方便的特点。用摄像机或摄像头采集含有人脸的图像或视频流，在图像中自动检测和跟踪人脸，进而对检测到的人脸进行一系列的相关处理。

深度学习起源于1943年由神经科学家W.S.McCulloch和数学家W.Pitts建立了神经网络和数学模型（M-P模型，是按照生物神经元的结构和工作原理构造出来的一个抽象和简化了的模型）。到1957年Frank Rosenblatt首次提出感知机，感知机是具有学习功能的M-P模型，这种学习功能表现为神经元连接权值可以被调节，引起科学界对神经网络的研究热潮。但在1969年，Marvin Minsky证明了感知器的致命弱点，指出感知器并不能够解决异或问题，并且两层神经网络的计算相当复杂，几乎不可能解决，由此神经网络的研究进入冰河期。Rumelhar和Hinton 在1986年提出了反向传播算法（Backpropagation algorithm），BP算法是一种监督学习算法，解决了两层神经网络计算的复杂性，再次引起人工神经领域的研究热潮。在20世纪90年代中期，机器学习领域的支撑向量机（SVM）算法诞生，并显示出不凡的潜力，基于统计学的SVM算法再次让神经网络学习从人们视线中淡出。直到2006年，一直专注于神经网络领域研究的Geoffery Hinton提出了深度学习（多层神经网络）的概念，由此开启了各大科研机构和公司企业对深度学习如火如荼的研究。深度学习也很快应用到图像处理、语言处理、数字识别等各大领域，并且在很多领域取的了state-of-the-art的成果。

传统的人脸识别（浅层学习方法）准确率最多也就能到70%，比如多数用于安保的人脸监控系统，即便部署后仍需安保人员时刻注意；而基于深度学习的人脸识别系统准确率全天候平均能达到90%之多，部署后可大幅节省安保人员精力。主要差异在于基于深度学习的人脸识别系统采用全局人脸特征检索，对人脸全局特征进行建模分析，而不局限于人眼等局部特征，因此人脸识别算法准确率已超过人眼极限，比如香港中文大学Yi Sun等人研究的最新Deep ID3算法，在LFW数据集上测试精度达到了99.53%。

VGG-face是牛津大学工程科学系视觉图形组（Visual Geometry Group）开发的基于深度学习的人脸识别模型。VGG组开发的深度卷积神经网络VGG-net继承了LeNet网络以及AlexNet网络的一些框架，其在2014年ImageNet竞赛取的很好的成绩。VGG-face提供已经训练好了的Caffe版本模型，除了直接使用外，用户还可使用自己的数据集在该模型上进行微调（fine-tuning），本文选择已训练好的Caffe版本模型作为人脸特征提取器实现在线人脸识别系统。

本文将先简要陈述神经网络理论以及笔者重点研究内容，其后将基于笔者实现的在线人脸识别系统，详细阐述人脸识别过程的问题及解决方案，最后对已解决和待解决的问题进行总结，并展望后续研究内容。

# 2 研究内容

## 2.1 基础理论

神经网络现今存在多种定义，使用的最广泛的一种是：神经网络是由具有适应性的简单单元组成的广泛并行互连的网络，它的组织能够模拟生物神经系统对真实世界物体所作出的交互反应。神经网络中最基本的成分是神经元模型。



图1. M-P模型

图1. M-P模型是最简单的M-P神经元模型，由McCulloch和Pitts在1943年提出，信号的输入和输出关系由式1. M-P模型计算公式表示。

式1. M-P模型计算公式

感知机由两层神经元组成，其学习能力非常有限，甚至不能解决异或这样简单的非线性可分问题，要解决非线性可分问题，需要考虑使用多层功能神经元。输入层和输出层之间的神经元层被称为隐层或隐含层。将多个神经元组合起来并具有分层结构时，就形成了神经网络模型，如图2所示为包含一个隐层的神经网络。



图2. 包含一个隐层的神经网络

图2所示神经网络的计算公式由式2表示。

式2. 图2神经网络计算公式

常见的多层级神经网络，每层神经元与下一层神经元全互连，神经元之间不存在同层互连，也不存在跨层互连，通常称作多层前馈神经网络。对于这样的神经网络，当模型的复杂度变高网络层数增加时，模型参数将迅速增多，模型的训练效率变低、易陷入过拟合。

20世纪60年代，Hubel和Wiesel在研究猫脑皮层中用于局部敏感和方向选择的神经元时发现其独特的网络结构可以有效地降低反馈神经网络的复杂性，继而提出了卷积神经网络（Convolutional Neural Networks）。

卷积神经网络具有局部感知、参数共享、多卷积核、降采样和多层卷积特性，相比全连接网络，有效约简模型参数，提高模型训练效率。具有代表性的卷积神经网络先后有LeNet-5、AlexNet、Googlenet、VGGNet和DeepFace等。

## 2.2 VGG-face

VGG-face模型网络基于16层深度的VGGNet网络模型，如图3所示。

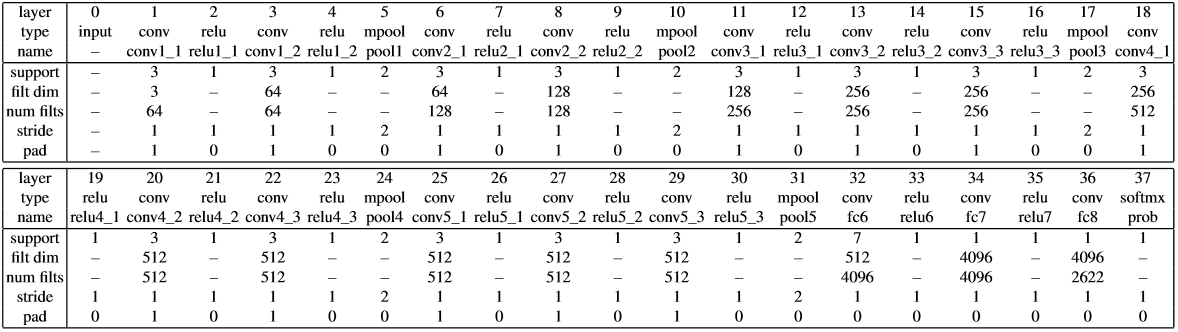


图3. VGG-face神经网络结构图

VGG-face官网提供Matlab、Torch、Caffe版本的预训练模型下载，本文选取Caffe版本预训练模型。

Caffe的全称是Convolution Architecture For Feature Extraction，即用于提取特征的卷积网络框架，由伯克利大学的贾扬清开发，是纯粹的C++/cuda架构，支持命令行、Python和Matlab接口，支持CPU/GPU之间的无缝切换。

## 2.3人脸识别

人脸识别的过程主要分人脸检测、人脸对齐和人脸识别，如图4所示。

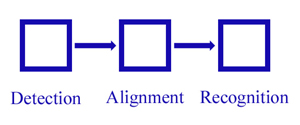


图4. 人脸识别过程

对于输入原始待测图片，图中人脸可能呈现多种姿态，需要进行人脸检测和人脸对齐预处理，以提高识别的准确度。

人脸检测用以确定输入图片中人脸所在的位置，检测准确度容易受姿态、光线、遮挡物等因素影响。人脸对齐时需要先定位出面部关键特征点（基准点），一般选取眼睛、鼻尖、嘴角点等（为使结果准确可再定位人脸其他轮廓点），对齐的目的是使待提取特征的人脸尽量维持正脸（即两个眼睛在同一水平线，嘴角两个点在同一水平线）以提高识别准确度。然后将两个眼睛连成一条直线，计算其与水平线的角度，估算需要旋转的角度，可以根据三维模型或二维仿射变换将待测人脸变换为正脸。

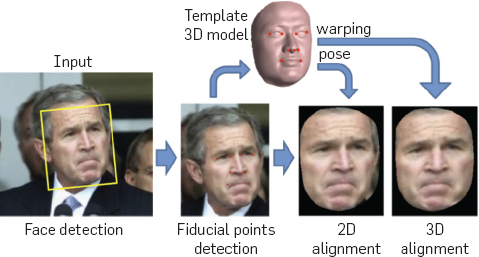


图5. 人脸对齐示意图

完成图片预处理之后，采用人脸模型进行特征提取，深度模型的一大特性在于其网络的任一中间层结果可以被提取出来做处理，表示执行某种变换操作。对于提取出的特征，和事先建立的人脸特征库中数据逐一比对，计算两者相似度，相似度越高则约有可能是预期识别结果，比对过程如图所示。选取Top-k特征人脸作为最后的识别结果，即对相似度进行降序排列，取前k人脸进行投票获取最终识别结果。

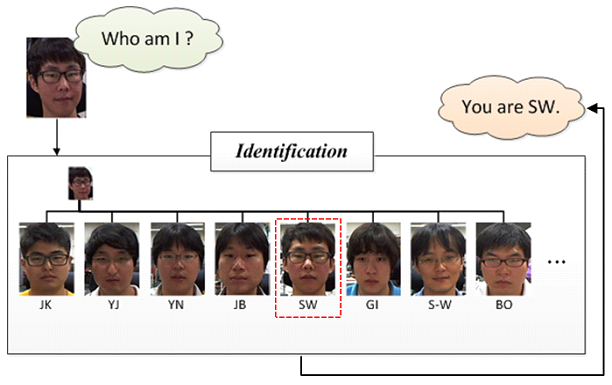


图6. 人脸识别示意图

# 3 系统框架

在线人脸识别系统面向深度学习班，以深度学习班同学人脸像为人脸数据库，同时支持人脸数据库扩充，提供在线人脸识别功能，由C++、python、QT语言协同开发实现。系统框架图如图7所示。



图7. 人脸识别系统框架图

系统主要由如下模块构成：用户界面，图像预处理，特征提取，特征对比、特征数据库。以下详细阐述各模块的功能。

## 3.1 用户界面

处理用户的输入数据，显示运算结果。用户可以通过摄像头抓取或者打开本地图片的方式加载输入文件，输入后最终的识别结果显示到用户界面的结果显示框中。

## 3.2图像预处理

系统采用opencv提供的基于haar的面部特征分类器来进行人脸检测，具有不错的效果，对于多姿态人脸和光线明暗有较强的适应性，满足基本的规范使用需求；但对于处理光线亮度很强或很弱的、人脸姿态只有侧脸以及人脸存在遮挡物等情形缺乏鲁棒性。本系统同时提供另一种人脸检测的工具以增强适应性，即采用开源的flandmark\_model来检测人脸。人脸检测完成后，进行人脸对齐操作，使需要进行特征提取的人脸尽量保持正脸，以提高识别精度。

将检测到且已对齐的人脸从输入图片中裁剪出来，对裁剪出的图片提取特征。

## 3.3特征提取

特征提取是人脸识别中非常关键的一步，特征表示的准确程度直接影响识别的准确度，提取出的特征必须充分代表图片的具体属性，才能在特征向量接近时更准确的识别出真实结果。

对于已经过对齐的人脸图，使用Caffe网络模型进行特征提取，Caffe支持三通道彩色图像，在输入之前先进行像素均值处理，此操作可以提高运行速度和精度。使用预训练的Caffe网络模型运行处理过的输入图片，提取其在VGG-face网络fc7层的4096维特征向量作为输入图片的特征。

## 3.4特征对比

在本系统中，待比较的特征都是4096维的特征向量，通过余弦定理计算向量夹角来计算特征向量的相似度，特征向量夹角越小则越相似。式3即余弦定理的数学公式表示。

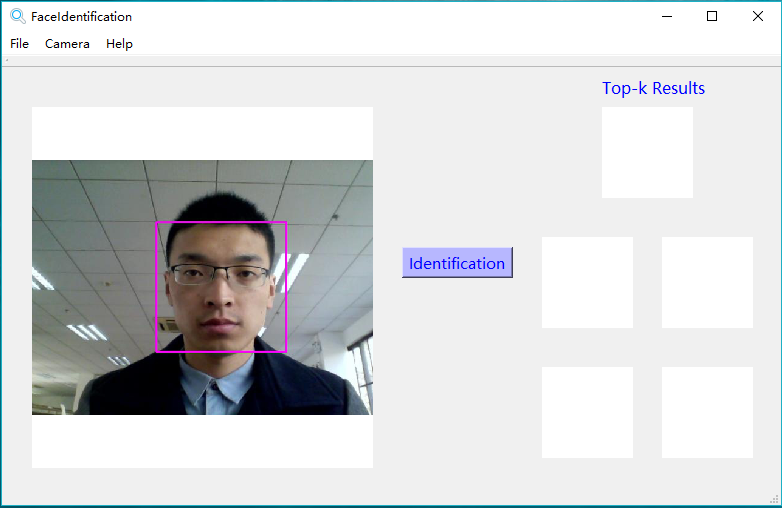
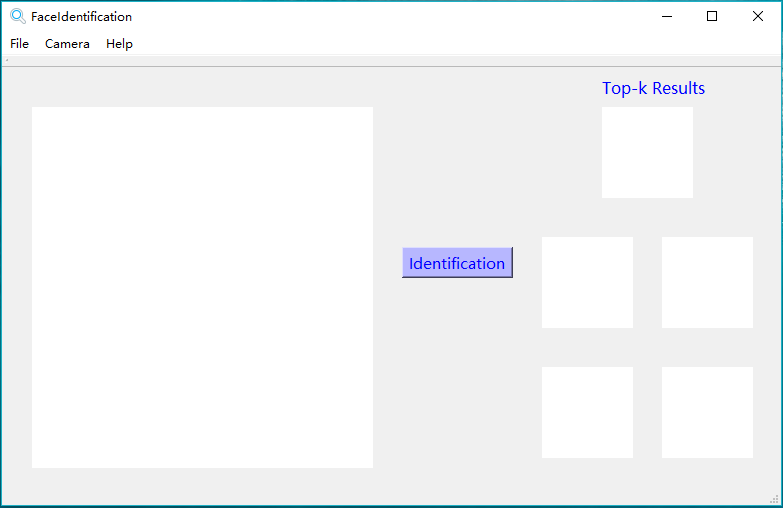
式3. 余弦定理

## 3.5特征数据库

为了增加人脸识别的速度，需要事先根据人脸数据库图像建立人脸特征数据库，过程类同特征提取的过程，最后对每张图像提取其在VGG-face网络的fc7层4096维特征向量，将该特征和图像ID建立映射关系，并存储为二进制文件，以在特征对比阶段快速加载。

# 4 结果展示

在线人脸识别系统实际运行效果图如图8所示。



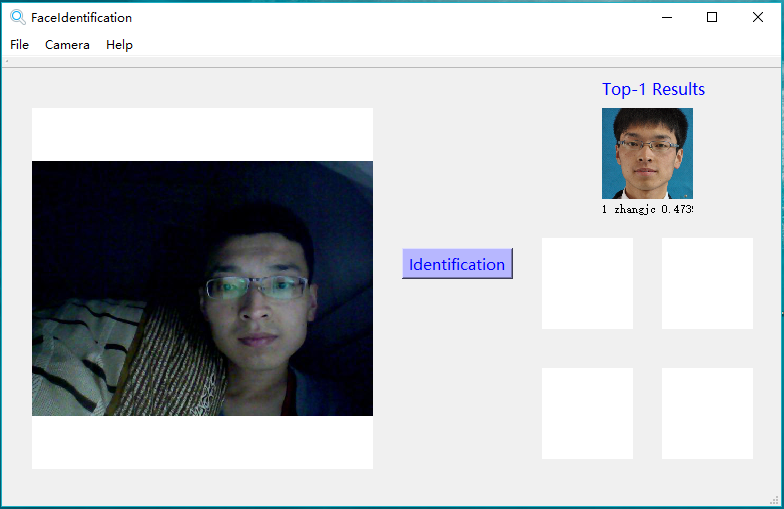
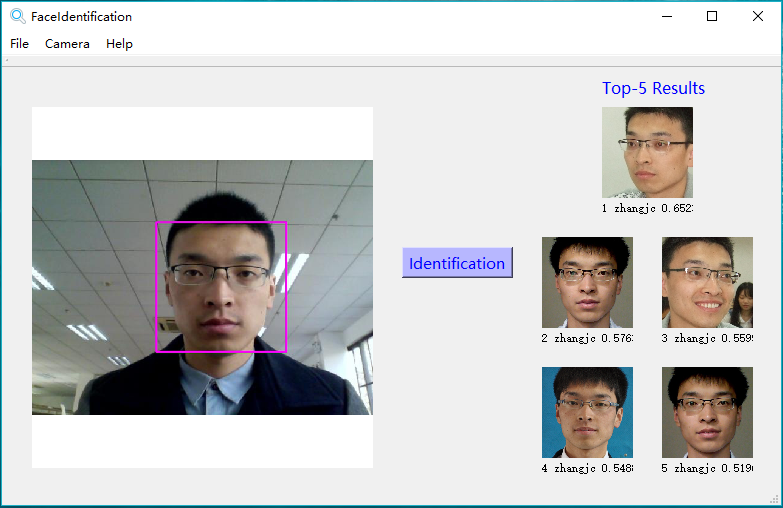


图8. 在线人脸识别系统运行结果图

# 5 结论与展望

由上文叙述可知，人脸识别系统的关键在于特征提取和图像预处理阶段，这两个阶段处理效果的好坏直接影响最终的识别结果，同时也是识别系统主要的时间开销所在。

通过对人脸识别系统的测试，其识别准确率基本满足预期，依靠人脸对齐操作，人脸识别系统对于输入的多姿态人脸图片的识别也具有较好的适应性。系统在普通PC设备上对于一般像素的人脸输入图的识别可在1秒内完成，满足普通应用的实时性要求。系统基于组件化构建，具有良好的扩展性，如对于新增人脸图像库，只需运行相关例程即可增量生成人脸特征库。

同时，人脸识别系统还有很多需要优化的地方，比如人脸检测和对齐模块的时间开销与业内相关state-of-the-art算法有一定差距，需要从图像处理和深度学习两方面入手寻求更优的方法。此外，系统对光照反差较大、侧脸姿态、人脸有遮挡等情形的输入图像处理的鲁棒性不足，后续将以上述几点为切入点继续展开研究，提升人脸识别系统运行速度和鲁棒性。

在此衷心感谢深度学习课程邹复好老师的精彩讲授和悉心指导，同时也感谢人脸识别组的各位同学，让我在这较短的时间内获得巨大的收获。

**References**

1. http://blog.csdn.NET/stdcoutzyx/article/details/41596663
2. Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[C]//British Machine Vision Conference. 2015, 1(3): 6.
3. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.
4. Kemelmacher-Shlizerman I, Shechtman E, Garg R, et al. Moving portraits[J]. Communications of the ACM, 2014, 57(9): 93-99.
5. Felzenszwalb P F, Girshick R B, McAllester D. Cascade object detection with deformable part models[C]//Computer vision and pattern recognition (CVPR), 2010 IEEE conference on. IEEE, 2010: 2241-2248.
6. Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1701-1708.