# 关于人脸识别技术的调研报告

**姓　名**　　**孙彬彬**

**日　期**　　**2017 年 4 月 23 日**

# 目录


前言 . . . . . ii

第1章 概 述 . . . . . 1

1.1 人脸识别研究背景和意义 . . . . . 1

1.2 人脸识别流程 . . . . . 2

1.3 人脸识别研究状况 . . . . . 3

1.3.1 人脸识别的发展历程 . . . . . 3

1.3.2 人脸识别技术难点分析 . . . . . 5

1.4 人脸识别常用数据库 . . . . . 6

1.5 本报告的章节安排 . . . . . 7

第2章 基于深度学习的人脸识别 . . . . . 8

2.1 深度学习算法和训练人脸库大小 . . . . . 9

2.2 研究人脸识别的企业及主要算法 . . . . . 9

2.2.1 Facebook . . . . . 9

2.2.2 香港中文大学 . . . . . 10

2.2.3 Google . . . . . 11

2.2.4 Face++ . . . . . 11

2.2.5 百度 . . . . . 11

2.2.6 腾讯优图 . . . . . 11

2.3 深度学习网络模型 . . . . . 11

2.3.1 DeepFace . . . . . 12

2.3.2 VGG . . . . . 13

2.3.3 DeepID系列 . . . . . 13

2.3.4 FaceNet . . . . . 15

第3章 总结与展望 . . . . . 16

# 1 概　　述

## 1.1 人脸识别研究背景和意义

随着信息化社会的迅速发展，以生物特征识别技术为核心的身份认证技术得到全球范围内的普遍关注，在诸多场合都有着重要的应用，如公共安全，监控系统，网络安全，门禁系统，人机交互等。传统的身份认证技术主要通过用户名，密码，证件等信息来进行身份鉴别，而这些信息比较容易遗失，泄露，盗取甚至是伪造，可靠性不高。伴随着功能强大且低耗的掌上设备的广泛应用，以图像处理或语音处理为基础的生物特征识别技术有了更多的用武之地。生物特征识别技术主要是通过生物传感器来获取人体固有的生理特征和行为特征数据，利用计算机技术，声学，光学，统计学等学科的紧密结合有效的对特征数据进行分析，最终获得个人身份的准确鉴别。通过生物特征识别技术能够更容易将计算机与监控，安全和管理系统有效结合，实现设备的自动化管理。正是由于其广阔的市场，伴随着巨大的经济效益和社会效益，生物特征技术获得了国内外研究机构的普遍关注和高度重视。

常见的生物特征包括人脸，指纹，手掌，语音，瞳孔，签名等。在众多的生物识别技术中，目前备受关注且发展迅速的是人脸识别技术。相较于其他生物特征识别技术如指纹和虹膜，人脸识别的优势在于其自然性和非侵犯，其最重要的优势在于能够在一定距离获取人脸信息，同时，其获取的方式也具有隐蔽性。人脸识别信息作为一个重要的生物特征识别技术，它具有良好的独立性和不可复制的特性，这就为个人身份的有效鉴定提供了必要的前提条件。

人脸识别的研究与众多科学领域都有着紧密的联系，包括图像处理，机器学习，模式识别，人工智能，计算机视觉，人工神经网络等。对其的研究已经成为人工智能，模式识别与计算机视觉相关领域中一个极其活跃的课题。对于人脸识别的深入研究能够极大的促进相关领域的发展和进步，具有十分重要的科研价值。由于其应用场景广泛，使其其市场前景较为广阔。人脸识别相关的产业化产品广泛应用于我们的衣食住行各个方面。表1.1.1 汇总了人脸识别的主要应用场景。

表 1.1.1　人脸识别的主要应用场景

| 应用领域 | 具体场合 |
| --- | --- |
| 娱乐业 | 视频游戏、虚拟现实（Virtual Reality，VR）、人机交互等 |
| 公共安全 | 电子驾照、电子身份证、电子护照等使用人脸识别方便检查个人身份的真实合法性 |
| 信息安全 | 桌面设备的登录、掌上设备登录、数据库安全、文件使用等场合进行身份认证和授权，避免由于密码被盗从而造成的信息被窃取所造成人身安全和经济损失 |
| 国家安全 | 犯罪嫌疑人人脸比照、重要场所视频监控、网上追捕逃犯等 |
| 出入管理 | 门禁考勤管理、监狱宿舍管理、车辆管理、刷脸防盗门等 |

## 1.2　人脸识别流程

人脸识别是一个视觉模式匹配问题，其中的人脸取自现实中三维人体对象，受到不同的照明，姿态，表情和其他因素的影响，其识别依赖于所获图像。通常大多数应用中进行识别的图像都是在二维人脸图像中，一些特定场景下需要依据人脸图像获取更多的信息以用于人脸识别。传统的人脸识别系统通常包括四个模块：人脸检测，人脸归一化，特征提取和分类识别。模块流程如下图1.2.1所示

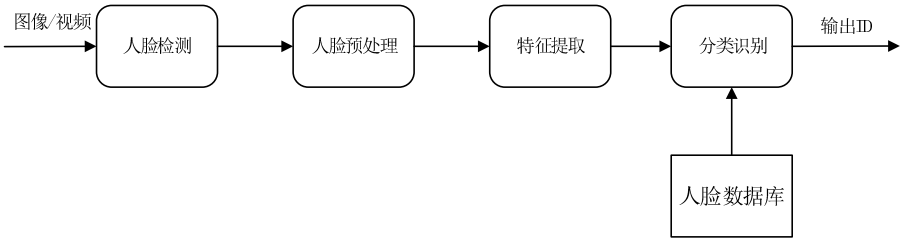


图 1.2.1　人脸识别系统流程框图

人脸识别所包含的四个模块中，包括了正常情况下的人脸识别步骤。人脸检测是指从含有背景中的图像中将人脸分离出来，区分人脸和非人脸区域。以视频为例，人脸的检测可能会涉及到目标追踪和多帧的人脸识别。对于检测得到的人脸，使用人脸预处理尽可能降低非可控因素（姿态，光照，表情等）对整体识别效果的影响，通过裁减，缩放等以使得获得人脸图像在同一尺度和分辨率下。然后对于获取的人脸图像通过特征提取处理进而获得不同人脸的能够用于区分的信息。特征提取人脸特征提取的好坏是影响人脸识别系统性能的关键因素。好的提取特征即使使用简单的分类器也能取得较高的正确率。识别一个目标就是，获得的特征对于同类人脸的差异信息不大，对于不同类人脸的差异较大，只有这样，才能够便于接下来的分类识别。分类识别操作是将输入的人脸和人脸库中进行匹配，选择最优的匹配结果并输出，继而完成人脸

识别。

## 1.3 人脸识别研究状况

### 1.3.1 人脸识别的发展历程

人脸识别技术的研究历史悠久，早在19世纪末，英国心理学家Galton就在Nature杂志上发表过相关的文章，到上个世纪六七十年代，研究人员开始关注人脸识别问题，并开始着手开发自动人脸识别技术。但是由于早期的相关技术和理论达不到要求，研究并未取得较大进展。直至20 世纪九十年代，随着计算机技术和信息技术的发展，同时相关理论的研究也有一定的基础，如模式识别，图像处理，认知学等，人脸识别技术取得了重大进展。Turk 等人提出了基于子空间分析的方法，该方法主要从高维的人脸图像空间学习低维的人脸特征，而后在低维的特征空间中完成人脸识别。这种方法的提出，带来了人脸识别领域的一次变革，为更加鲁棒的自动人脸识别系统提供了基础和保障。现在，一些重要的国际期刊如模式分析与机器智能汇刊（Transactions on Pattern Analysis and Machine Intelligence,PAMI)，图像处理汇刊(IEEE Transcations on Image Processing) 和模式识别（Pattern Recognition）等，对于人脸识别也是十分关注，国际计算机视觉方面的三大顶级会议IEEE International Conference on Computer Vision （ICCV)，International Conference on Computer Vision Pattern Recognition(CVPR) 和Europeon Conference on Computer Vision(ECCV) 也有不少关于人脸识别相关的论文。众多的美国高校，如麻省理工大学的MIT 媒体实验室和人工智能实验室，卡内基梅隆大学（CMU）的机器人研究所等也对人脸识别做出了大量的理论研究工作。随着人工智能的兴起，计算机CPU和GPU运算能力的突飞猛进，人脸识别的新方法和技术也是层出不穷。近来，国际知名公司微软，谷歌，Facebook 等也宣布其基于深度学习的人脸识别软件能够获得不错的识别率。于此同时，国内一些公司如百度，腾讯，旷视和汉王等在人脸识别领域获得的结果也能达到世界顶尖水平。

经过近三十年的快速发展，人脸识别方法也随着理论的进步不断推陈出新，大体上的人脸识别方法大体上可以分为如下五类。

**基于几何特征的人脸识别方法**

基于几何特征的特征提取方法是最早的人脸识别算法。其基本思想时通过获取人脸的面部拓扑结构信息，利用基于结构的方法提取人脸主要特征，并用一组几何特征

向量表示，通过对特征矢量之间的相互匹配，从而完成人脸识别。常用的几何特征包括：距离信息，如眉毛相对于眼睛的垂直距离，嘴巴与鼻子间的距离等；形状信息，如眉毛和眼睛的弧度等。基于几何特征的人脸特征提取方法比较直观，且计算量较少。但是当环境发生变化，或是加入遮挡，表情，姿态等变化信息时，其往往准确度不高。实际环境的变化对这种特征的影响较大，故而单纯依赖于几何特征的人脸识别已经不多，更多的是将该特征与其他特征结合，进而提高整体的识别率。

**基于子空间分析的人脸识别方法**

基于子空间分析的人脸识别方法，主要是考虑到高维的人脸图像矢量，在现实情况下无法准确分类。于是考虑将高维数据投影到低维的子空间之中，保留其中的有效信息部分以达到降维的目的，同时便于之后的分类识别。这类方法包括Turk等人提出的主成分分析(Principal Component Analysis, PCA)人脸识别方法，这个方法首先是对由人脸图像组成的矢量矩阵做K-L变换得到其协方差矩阵正交的特征向量，待识别的人脸图像投影到以特征向量张成的空间中，用投影系数作为人脸特征，用于人脸识别。由于PCA在进行提取时考虑的是人脸在降维过程中尽量保持能量不丢失，并未考虑到每个类别的差异信息。故此，Lu 等人基于此提出一种线性判别分析（Linear Discriminant Analysis,LDA）的人脸识别方法，依据fisher准则，在人脸图像组成矢量矩阵进行投影时，寻找能够使样本的类内误差最小，类间误差较大的投影方向。线性判别分析是一种有监督的判别方法，需要一定数目的有标签的人脸图像。

**基于局部特征描述子的人脸识别方法**

局部特征描述子主要是将高维的人脸图像矢量通过某种模式转换为低维的特征描述子，再使用描述子训练分类器，进行分类识别，进而降低计算复杂度。其中在人脸识别领域应用较广的是Ahonen等人提出的使用LBP (Local Binary Patterns)算子进行人脸识别。LBP算子是一种描述图像局部空间结构的非参数算子，并用该算子来分析图像的纹理特征，其描述了它在纹理分类中的强区分能力。LBP 的中心思想是，用中心像素的灰度值作为阈值，而与其邻域相比较得到的二进制码来表述局部纹理特征。除此之外，Gabor 特征，HOG特征，SITF 特征作为一些局部特征描述子，同样获得了比较广泛的应用。

**基于人工神经网络的人脸识别方法**

人工神经网络是一种模拟人脑神经突触结构的网络模型，它由大量的网络节点通过某种拓扑结构连接而成。它将输入信息通过一定方式进行加权求和，并对求和结果进行激活作为下一层网络节点的值，这种推导方式可以很好地模拟信号在生物神经元中的传递。人工神经网络已经在人脸识别领域得到了广泛的应用。人工神经网络具有一定的非线性，能够很好地映射输入与输出之间的关系，并且它是一种并行的系统，易于并行计算，能够实现网络的快速计算。早期的人工神经网络系统，通常只取1至2层隐藏层，与近年发展的深层网络系统相比，可称为浅层人工神经网络。浅层人工神经网络对未经训练过的图像推理能力(或称泛化能力)有限。

**基于深度学习的人脸识别方法**

近年来提出的深度学习(Deep Leaming)方法为人脸识别技术提供了新的理论依据，翻开了人脸识别领域的新篇章。深度学习采用多层非线性神经网络的方式来模拟人脑的工作机理，分析和解释数据。深层网络结构具有更强的非线性拟合能力，经过大样本训练的深层网络往往具有较强的泛化能力。深度学习具有分层特征表达的能力，每一层提取的特征都是对样本数据不同尺寸或不同层面的表达，高层的特征具有更强的抽象性和判别力，能够更好地描述数据的内涵特征，对不同类别的数据具有很强的分辨能力。深度学习的基本概念起源于人工神经网络，多隐层多感知就是一种深度学习结构，但是这个深度学习结构的训练存在容易收敛到局部最小等诸多问题，因此需要对深度学习结构分层训练，在每一层结构中采用无监督的学习方法就可以避免传统神经网络在训练过程中遇到的问题。近年来随着理论的进步，各种新型的深度学习网络也是层出不穷，这极大的推动了人脸识别效果的提升。

### 1.3.2 人脸识别技术难点分析

虽然人脸识别有着高效、简单和易于实现的优点，但在实际非可控环境中应用中进行人脸识别仍然存在着许多问题。在实际环境中是实现人脸的高精度识别仍然是一个富有挑战性的课题，在实现人脸识别过程中，仍存在如下技术问题亟待解决。

(1) 人脸数据样本少。在实际应用中，由于诸多的条件限制，导致常用场景下，人脸的样本数较少。当面临人脸发生各种因环境而造成的变化，在使用这种样本进行训练或是建立模型时，很容易导致算法不收敛或是分类算法的过拟合问题，使得整体识别率较低。

(2)算法的有效性和时效性。人脸识别的有效性指的是人脸识别的正确率，时效性

指的是人脸识别的速率。这两个是一个非常重要的因素。一般而言，想要获得较高的识别率，需要耗时更多。如何在这两者进行权衡，将是人脸识别技术能否广泛应用的前提。

(3)抵抗非法攻击的人脸识别。人脸识别技术中经常需要面临的一个情况是，非法情况的攻击。正常情况下确定人脸的所属类别，非法情况下，需要面临一些视频或是照片攻击，在这种情况下的技术能否准确抵御非法攻击，是人脸识别技术发展的挑战。

(4)不可控环境下的人脸识别。人脸识别最大的困难在于其环境的不可控性，现今的算法还不能够解决所有的在不可控环境下的干扰。有时，这些干扰因素常人都难以分辨。常见的不可控因素如下所示。

- 光照因素：同样的人脸在不同的光照强弱，方向条件下，会有不同人脸表现形式，有时，这种差异甚至会超过同类人脸的相似程度。
- 噪声大小：人脸图像在传输或是压缩过程中会受到自然界或是电子相关的干扰，进而影响图片质量，关键信息的丢失对于识别来说也是一个困难。
- 表情变化：人脸识别的一个技术难点在于其表情和姿态的变化，由于人脸在识别中的不对齐，甚至会发生检测不出人脸的情况。人类表情的多样性和易变性增加了识别的困难程度。
- 遮挡因素：人脸的图像在遮挡情况下的识别亦是一个难点，由于季节或是时间的变化，人脸可能会被围巾，眼睛，口罩等遮挡。
- 年龄变化：随着年龄的增长，人脸会有不同程度的改变，这就对识别算法的鲁棒性提出了一个要求。同时，人脸表现出的年龄易受角度和化妆的影响，如何挖掘图像表面内更加深层的内容将是人脸识别面临的挑战之一。

## 1.4 人脸识别常用数据库

在人脸识别中算法的研究，建立模型和算法的对比中，一个统一的人脸数据库是必不可少的。人脸数据库由于采集时，尽量满足了在不同环境和条件下的采集，考虑了算法在实际应用中可能出现的非可控因素的影响。人脸识别的结果通常与每类人脸的数目和人脸图片的大小有密切的关系。同时在算法的比较中由于人脸库的采集条件不同，算法在不同人脸库下的识别效果仍会有所差异，而这帮助我们找出算法的适用

性环境，从而进一步优化算法性能。如下是一些在人脸识别算法对比中常用的人脸数据库。

ORL（Oliver Research Laboratory）人脸数据库：ORL人脸数据库是由剑桥大学建立。其中包含有从1992年到1994年在实验室中采集的一系列人脸。有40个人脸类别，每类人包含10个不同的人脸，是以92 × 112大小的灰度图片。每类人脸在不同的时间采集，具有不同光照和表情变化。所有的图像以黑色背景拍摄，脸部允许轻微的移动。它是当前应用较为广泛的人脸库之一。

AR人脸数据库：该数据库由阿拉巴马大学（UAB）的计算机视觉中心的Alex 等人创建。其中包含有126人（70个男人和56个女人），每人有26 张人脸，人脸图像的分辨率为165 × 120。每类人脸分成两组，其中一组包含只有光照和表情变化(微笑，愤怒和冷酷)，另外一组包含有遮挡变化(不戴/戴眼镜，不戴/戴围巾)。

LFW（Labeled Faces in the wild）人脸数据库：该人脸库由美国马萨诸塞大学阿姆斯特分校计算机视觉实验室创建，用于研究在非受限条件下的人脸识别问题。该人脸库包含了超过13000张从网上收集的人脸图像。人脸库中包括5749个人，其中1680个人脸有二张及以上人脸图像。其人脸采集是由Viola-Jones人脸检测器获得。由于其更接近人脸识别中真实的识别环境，对于在LFW识别库取得较为优异结果的算法通常在实际中也有不错结果。LFW人脸识别库是如今主流人脸识别识别进行对比的最为重要的人脸库，其识别结果已成为学术界和工业界评价人脸识别功能的基准。

## 1.5 本报告的章节安排

本文的研究重点是基于压缩感知理论的人脸识别的对比研究，总共分为三个章节，论文的主要内容安排如下：

第一章：概述。在本章中，首先主要阐述人脸识别的研究背景和研究意义，然后介绍人脸识别的发展历程及研究现状，归纳整理了常用的人脸识别方法。

第二章：基于深度学习的人脸识别。本章主要了各个公司和机构人脸识别算法和基于深度学习的神经网络的模型。

第五章：总结与展望：对本报告的内容进行了总结。

# 2　基于深度学习的人脸识别

自从2006年Hiton提出深度学习的概念以来，对于深度学习的研究便广泛进行，在理论和运用方面都有巨大进展。深度学习的主要框架如下。无监督+有监督的有受限波尔兹曼机和自动编码机两种框架。自动编码机又拓展为稀疏自动编码机（降低隐层维度）和降噪自动编码机（加入随机噪声）。纯有监督的主要是卷积神经网络。在实际的运用中主要还是采用深度卷积神经网络的模式。在人脸识别算法比较中经常作为基准的是LFW人脸库，由于其图片比较接近真实情况，通常在LFW 库中取得不错结果的在实际中也会有优异表现，其中最新的LFW的人脸识别结果如下所示：

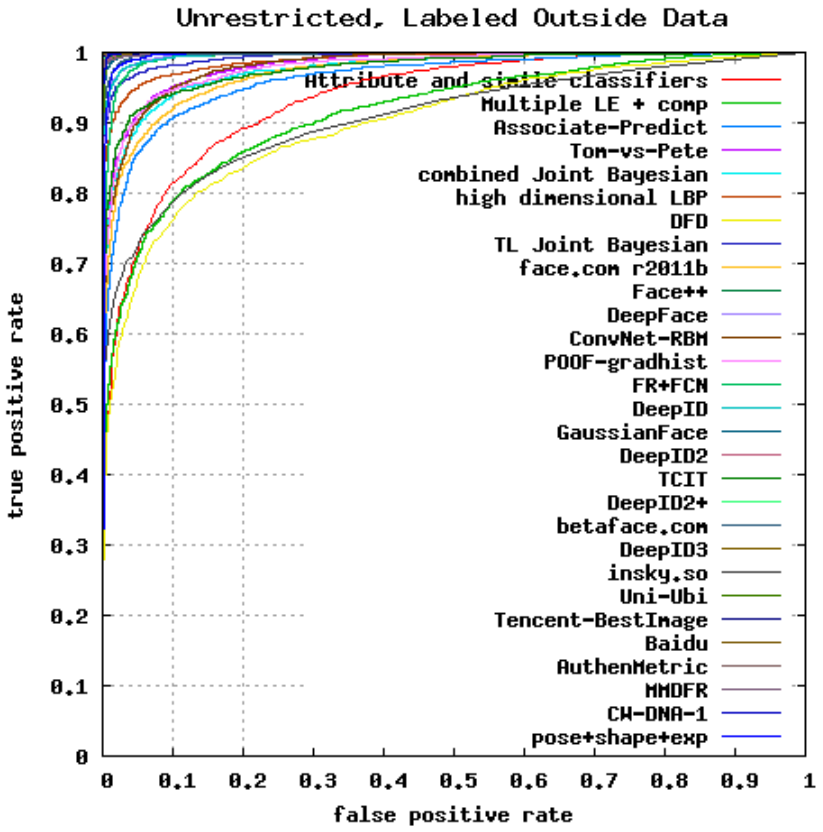


图 2.0.1　LFW识别率对比

上图可以看出，基于深度学习模型的人脸识别技术在一定的错误率下，部分算法的准确率几乎达到该库的最高识别效果。

## 2.1 深度学习算法和训练人脸库大小

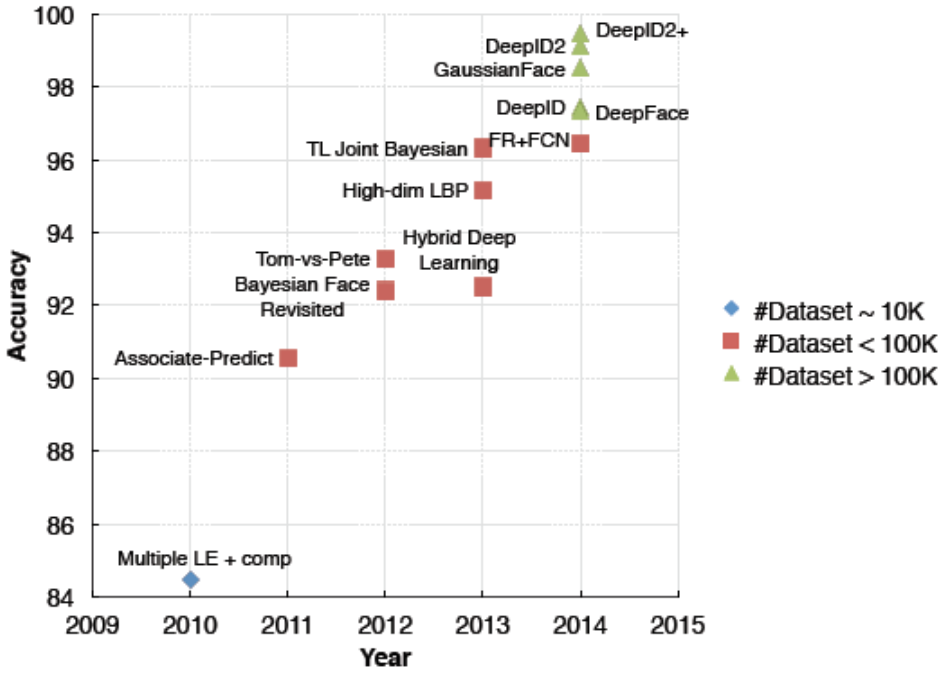


图 2.1.1 识别率和训练库大小

通过上图2.3.3能够看出，一方面随着实验中所获得人脸训练库数目量级的增加，另一方面随着深度学习神经网络的优化，在LFW库上的识别结果一直有稳步上升的趋势。而这也几乎达到在这个库中所能获得的最好识别结果。

## 2.2 研究人脸识别的企业及主要算法

由于深度学习在人脸识别领域中比较优异的性能，LFW库中获得较高识别率的都是依此为基础上获得的。下表中列举了部分公司和机构的在该库下获得的人脸识别结果的对比。

### 2.2.1 Facebook

Deepface：Deepface是Facebook公司研发的人脸识别算法。DeepFace利用面部多点的稠密对齐，借助于3D模型对有姿态的人脸进行校正，同时利用一个9层深度卷积神

表 2.2.1　各公司或机构在LFW库上识别率比较

| 企业或机构名称 | 算法名称 | 识别率 |
|---|---|---|
| FaceBook | DeepFace-ensemble | 0.9735 ± 0.0025 |
| 香港中文大学 | DeepID | 0.9745 ± 0.0026 |
| Google | FaceNet | 0.9963 ± 0.0009 |
| Face++ | Face++ | 0.9950 ± 0.0036 |
| 云从科技 | CW-DNA-1 | 0.9950 ± 0.0022 |
| 百度 | IDL Ensemble model | 0.9977 ± 0.0006 |
| 腾讯优图 | YC-T | 0.9980 ± 0.0023 |
| 宇范智能 | Uni-Ubi | 0.9900 ± 0.0032 |

经网络在400万规模的人脸库上训练了一个深度人脸表示模型，在LFW数据集上取得了97.25%的平均精度。

### 2.2.2　香港中文大学

DeepID是香港中文大学汤晓鸥课题组发明的一种人脸识别算法。孙祎团队研发的技术和产品已经在当今备受关注的互联网金融领域被应用于“人脸在线身份认证”，让用户无需面签即可通过人脸识别在手机端认证身份，继而开户、转账、借贷，这项身份认证服务已达到了百万级/天的调用量。其经过了三代的发展。第一代DeepID结构与普通的卷积神经网络的结构相似，但是在隐含层，也就是倒数第二层，与Convolutional layer 4和Max-pooling layer3相连，鉴于卷积神经网络层数越高视野域越大的特性，这样的连接方式可以既考虑局部的特征，又考虑全局的特征。DeepID2在DeepID的基础上添加了验证信号。具体来说，原本的卷积神经网络最后一层softmax使用的是Logistic Regression作为最终的目标函数，也就是识别信号；但在DeepID2中，目标函数上添加了验证信号，两个信号使用加权的方式进行了组合。第三代DeepID2+有如下贡献，第一点是继续更改了网络结构；第二点是对卷积神经网络进行了大量的分析，发现了几大特征，包括：1:神经单元的适度稀疏性，该性质甚至可以保证即便经过二值化后，仍然可以达到较好的识别效果；2:高层的神经单元对人比较敏感，即对同一个人的头像来说，总有一些单元处于一直激活或者一直抑制的状态；3:DeepID2+的输出对遮挡非常鲁棒。

### 2.2.3 Google

FacenNet是Google开发的深度学习算法。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，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，选的好可以很快的收敛。三元组比softmax的优势在于：softmax不直接，（三元组直接优化距离），因而性能也不好。softmax产生的特征表示向量都很大，一般超过1000维。

### 2.2.4 Face++

Face++使用的是依照自行从网上获取的5百万人脸库直接训练而得的一个深度卷积网络。其构建的Face++TM是北京旷视科技有限公司旗下的新型视觉服务平台，Face++TM 平台通过提供云端API、离线SDK、以及面向用户的自主研发产品形式，将人脸识别技术广泛应用到互联网及移动应用场景中，人脸识别云计算平台市场前景广阔。和蚂蚁金服共同开发的Smile to Pay，可以进行扫脸支付。

### 2.2.5 百度

百度的深度学习研究院（IDL）开发的人脸识别技术同样是基于深度学习来进行人脸识别。提出的人脸识别算法是结合了深度CNN和深度度量学习（Deep metric learning）,能够在使用获得的较低维度的特征用于人脸识别。

### 2.2.6 腾讯优图

腾讯优图在训练阶段使用了自行建立的优图数据库，包含了20,000个人的共计2,000,000张脸。使用多机和多GPU的tensorflow集群，训练了三个极深的深度残差网络nception-ResNet-v2）,三个网络的深度分别为360, 540, 720，借此来完成人脸识别任务。使用最后的fc 层作为输出作为特征，获得了均值为0.9977的识别效果。

## 2.3 深度学习网络模型

下面将简要介绍部分深度学习的网络模型：

### 2.3.1 DeepFace

Deepface是Facebook公司研发的人脸识别算法。其实现的基本流程为：分为如下几步：首先是人脸对齐，通过以下流程实现：

第一步：a. 人脸检测，使用6个基点。b. 二维剪切，将人脸部分裁剪出来。c. 67个基点，然后Delaunay三角化，在轮廓处添加三角形来避免不连续。d. 将三角化后的人脸转换成3D形状。e. 三角化后的人脸变为有深度的3D三角网。f. 将三角网做偏转，使人脸的正面朝前。g. 最后放正的人脸。h. 获取一个新角度的人脸。总体上说，第一步的作用就是使用3D模型来将人脸对齐，从而使CNN发挥最大的效果。

第二步：人脸表示经过3D对齐以后，形成的图像都是152×152的图像，输入到下述网络结构中，进行卷积神经网络处理，该结构的参数如下：

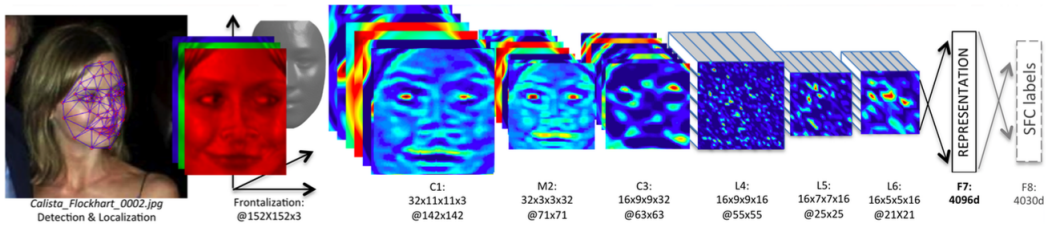


图 2.3.1　Deepface网络模型

Conv：32个11×11×3的卷积核

max-pooling: 3×3，stride=2

Conv: 16个9×9的卷积核

Local-Conv: 16个9×9的卷积核，Local的意思是卷积核的参数不共享

Local-Conv: 16个7×7的卷积核，参数不共享

Local-Conv: 16个5×5的卷积核，参数不共享

Fully-connected: 4096维

Softmax: 4030维

第三步：人脸表示归一化对于输出的4096-d向量：先每一维进行归一化，即对于结果向量中的每一维，都要除以该维度在整个训练集上的最大值。每个向量进行L2归一化最后进行分类：得到表示后，使用了多种方法进行分类：直接算内积、加权的卡方距离、使用Siamese网络结构。

### 2.3.2 VGG

模型框架为：输入为尺寸224X224的RGB图，由A-E5个卷积网络构成，深度由浅到深。所有的卷积核都是用很小的3x3,步长为1，只有一个是1x1的卷积核，可视为输入通道的线性变换。一共有5个池化层，不是每个卷积层后都有池化层。池化用2x2像素的窗口，步长为2.。最后三层是全连接层，前两层有4096个通道，最后一层有1000个通道分别对应1000个类别，最后一层用softmax分类。每个隐层用ReLU做输出函数，整个网络都不用LRN，因为没效果，还占内存消耗和计算时间。

训练的目的是为了最优化多项逻辑回归，通过基于BP算法的mini-batch gradient descent来实现。训练通过权重衰减和对前两层全连接层的dropout regularisation来调整参数。学习速率初始设为0.01，并以10倍减少当正确率不再提高时。一共衰减了三次，学习到370K次迭代后停止。

初始化网络的权重：首先用随机初始化训练结构A，因为网络较浅；训练其他深的结构，就用A来初始化他们的前四层和最后三个全连接层，中间层随机初始化，不改变预初始化的学习速率，让他们随着学习改变。随机初始化用，使用0平均和$10^{-2}$的方差的标准正态分布。有两种确定训练图像大小的方法，出于速度的考虑，采取单一大小的方式，并定为384.

### 2.3.3 DeepID系列

DeepID结构：由四层卷积神经网络构成，前三层后跟池化层，第三层池化层和第四层卷积层一起全连接够成最后的DeepID层，提取图片特征。最后用Soft-max来分类，结构如下：

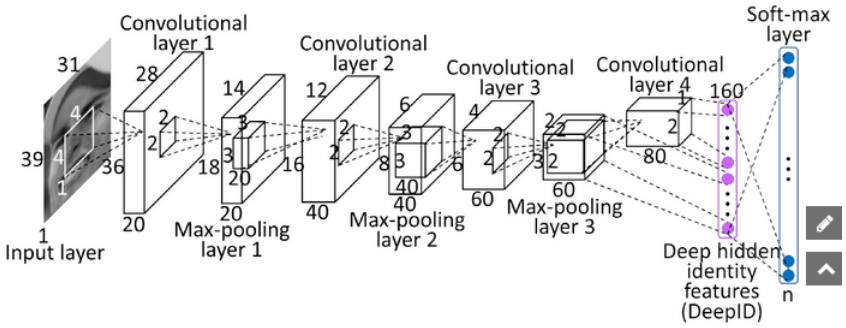


图 2.3.2 DeepID网络模型

在DeepID的实验过程中，使用的外部数据集为CelebFaces+，有10177人，202599张图片；8700人训练DeepID，1477人训练Joint Bayesian分类器。切分的patch数目为100，使用了五种不同的scale。每张图片最后形成的向量长度为32000，使用PCA降维到150。如此，达到97.20%的效果。使用某种Transfer Learning的算法后，达到97.45% 的最终效果。

DeepID2:结构框架：输入为55x47的RGB图，网络包含4个卷积层，前三个后面跟着池化层。为了学到不同的高层特征，高层卷积网络权值不共享。尤其，第三层卷积网络权值只在2x2的局部区域共享，第四层本地连接层完全不共享。最后的一层是第三和第四两层卷积网络的全连接层，提取了160维的向量。使用ReLU作为激活函数。另外，加上了face identification signal和face verification signal两个监督信号。face identification signal通过在DeepID2层之后加上n路的softmax层，通过训练最小化交叉熵损失，来确保正确分类。那么DeepID2层就要找到最具类间区分度的特征，这样就最大化了类间差距。face verification signal用来认准两个人是否是同一个人。通过L1/L2标准化或余弦相似性来衡量，通过训练，使DeepID2层对于同一个人尽量显示一致的特征，以此来减小类内差距。结构如下图：

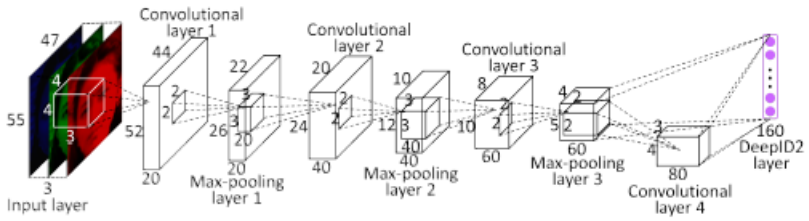


Figure 1: The ConvNet structure for DeepID2 extraction.

图 2.3.3　DeepID2网络模型

训练过程：首先初始化权重参数，从训练集中随机取两个样例输入网络，计算各自的输出，采用随机梯度下降法来更新各参数，进行迭代，知道模型收敛，输出各权重参数。

人脸验证：

首先使用SDM算法对每张人脸检测出21个landmarks，然后根据这些landmarks，再加上位置、尺度、通道、水平翻转等因素，每张人脸形成了400张patch，使用200个CNN对其进行训练，水平翻转形成的patch跟原始图片放在一起进行训练。这样，就形成了400×160维的向量。这样形成的特征维数太高，所以要进行特征选择，

不同于之前的DeepID直接采用PCA的方式，DeepID2先对patch进行选取，使用前向-后向贪心算法选取了25个最有效的patch，这样就只有25×160维向量，然后使用PCA进行降维，降维后为180维，然后再输入到联合贝叶斯模型中进行分类。

DeepID2使用的外部数据集仍然是CelebFaces+，但先把CelebFaces+进行了切分，切分成了CelebFaces+A(8192个人)和CelebFaces+B(1985个人)。首先，训练DeepID2，CelebFaces+A做训练集，此时CelebFaces+B做验证集；其次，CelebFaces+B切分为1485人和500人两个部分，进行特征选择，选择25个patch。最后在CelebFaces+B整个数据集上训练联合贝叶斯模型，然后在LFW上进行测试。在上一段描述的基础上，进行了组合模型的加强，即在选取特征时进行了七次。第一次选效果最好的25个patch，第二次从剩余的patch中再选25个，以此类推。然后将七个联合贝叶斯模型使用SVM进行融合。最终达到了99.15%的结果。

### 2.3.4 FaceNet

FaceNet是谷歌提出的人脸识别算法。结构：从图像中获取输入快，经过卷积神经网络来获取特征，但并没有使用softmax来分类，而是经过L2归一化后，用训练好的三元组计算三元损失，直接判断图片是否是一个人.

实施过程：三元损失由三部分组成，需要被判断的图像（anchor），属于同一个人的另一张图像（positive），和不是同一个人的一张图像(negative).训练目的就是为了让所有anchor和positive之间的距离都要比anchor和negative之间的距离都要短，那么就能正判断出图中的人是否是同一个。而三元组的选择很重要，选的好就能迅速收敛。文章提出两种方法：1.每N步线下在数据的子集上生成一些triplet 2.在线生成triplet，在每一个mini-batch中选择hard pos/neg 样例。使用线上生成时，为了使mini-batch中生成的triplet合理，生成mini-batch的时候，保证每个mini-batch中每个人平均有40张图片。然后随机加一些反例进去。在生成triplet的时候，找出所有的anchor-pos对，然后对每个anchor-pos对找出其hard neg样本。这里，并不是严格的去找hard的anchor-pos对，找出所有的anchor-pos对训练的收敛速度也很快。

# 3　总结与展望

随着信息化社会的迅速发展，公共安全问题的日益突出，以生物特征识别技术在诸多场合都有着重要的应用，如公共安全，监控系统，网络安全，门禁系统，人机交互等。在这其中的人脸识别由于其非侵犯性成本低等特点，获得了较为广泛的关注与研究。本报告主要是按照以人脸识别技术为基础，简要介绍了人脸识别的流程和人脸识别的技术研究和应用的现状。重点介绍了以深度学习为基础的人脸识别算法。

传统的人脸识别算法主要都是基于统计学习分类理论，寻找人脸图像数据中的内在联系，利用不同类人脸的差异信息从而提高人脸识别率。近些年来，伴随着技术的进步和硬件水平的提高，以深度学习为基础的算法，在人脸识别领域的测试和应用中获得了不错的识别效果。然而同样需要注意的是，虽然在科研领域能够获得较为不错的识别结果，当到实际环境中，由于其他的因素的影响，对于最终的人脸识别结果仍然有所影响。具体场景需要具体研究。

人脸识别作为一个经典的问题正逐渐成为IT产业下一轮浪潮，不少知名科技公司的加入推动了整体技术的发展。经过近些年的深入研究，人脸识别技术获得较为快速的进步。不少商业组织和机构已经将人脸识别技术应用现实生活中，并取得了获得不错的效果，提高了人机交互性。正是由于人脸识别的非侵犯性和隐蔽性，人脸识别在金融，国防，安全和教育领域前景巨大。随着技术的逐渐成熟，以人脸识别为基础的衍生品将会逐渐走向各家各户。