第38卷 第4期 工 程 数 学 学 报 Vol. 38 No. 4
2021年08月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Aug. 2021

doi: 10.3969/j.issn.1005-3085.2021.04.001 文章编号: 1005-3085(2021)04-0451-19

基于深度学习的人脸识别方法综述

余璀璨, 李慧斌†

(西安交通大学数学与统计学院 大数据算法与分析技术国家工程实验室，西安 710049)

摘 要: 人脸识别与虹膜识别、指纹识别、步态识别等其它生物特征识别技术相比，具有自然、便捷、用户体验友好等独特优势，因而受到了学术界和工业界的广泛关注. 近年来，在深度学习技术的驱动下，人脸识别技术取得了突破性进展，在面对表情、姿态、光照、遮挡等外在干扰因素时，仍表现出较好的鲁棒性. 特别地，基于深度学习的人脸识别技术已广泛应用于安防、金融、教育、交通、新零售等应用领域. 我们认识到，在人脸识别技术不断走向大众化的过程中，急需一些综述性的和普及性的文献来总结人脸识别技术的基本原理和基本方法. 基于此，本文首先简要回顾了人脸识别的发展脉络，之后从人脸预处理、深度特征学习、特征比对、人脸数据集、评价标准五个方面重点介绍了基于深度学习的人脸识别技术. 最后指出了人脸识别技术未来的发展趋势.

关键词: 人脸识别；深度学习；卷积神经网络；特征学习

分类号: **AMS**(2010) 68T10 **中图分类号**: TP389.1 **文献标识码**: A

1 引言

人脸识别[1]是一种依据人脸图像进行身份识别的生物特征识别技术. 人脸识别的研究始于20世纪60年代，与虹膜识别、指纹识别、步态识别等生物特征识别技术相比，人脸识别因其便捷、高效、易普及的优点成为最受关注的研究问题之一. 通常，其难点在于人脸结构相似性导致不同个体之间差异不显著，而同一个体在不同表情、姿态、年龄、光照、遮挡、妆饰等干扰因素下变化显著. 因而人脸识别技术需要在类内变化干扰的情况下尽可能增大类间差距以区分不同个体，其关键在于从人脸图像中提取有利于识别的特征. 早期基于人脸几何特征的识别方法[2–4]使用眼睛、鼻子、嘴巴等关键部位之间的关系(如角度、距离)构建人脸描述子，此类方法忽略了人脸纹理、外观包含的有用信息，因此，识别效果一般. 基于子空间学习的识别方法如Eigenfaces[5]、Fisherfaces[6]，将原始数据整体映射到低维人脸子空间，这类方法很大程度上推动了人脸识别技术的发展. 基于局部特征分析的识别方法使用合适的滤波器提取

收稿日期: 2019-03-12. **作者简介**: 余璀璨(1995年10月生)，女，硕士. 研究方向：人脸识别、深度学习.

基金项目: 国家自然科学基金(61976173)；国家重点研发计划(2018AAA0102201)；教育部-中国移动人工智能建设资助项目(MCM20190701)；中央高校基本科研业务费(xzy012019041)；陕西省自然科学基础研究计划(2019JQ-628).

†**通讯作者**: 李慧斌 E-mail: huibinli@xjtu.edu.cn

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

人脸局部特征，Gabor[7]、LBP[8]、HOG[9]等常用于此类方法．在光照、姿态和表情变化较小时这类人脸识别方法的效果一定程度上比较稳定．2014年以来，深度卷积神经网络为人脸识别技术带来了巨大突破．无需人工设计特征，深度卷积神经网络能够针对训练数据学习如何提取特征．在特定数据集上，这类方法的识别能力已超过人类识别水平[10]．

深度学习是一类使用多层线性及非线性处理单元通过组合底层特征而形成更加抽象的高层特征表示的机器学习算法，基于深度学习的人脸识别方法使用端到端的方式学习提取特征的能力，并使用提取到的特征进行分类，在损失函数的指导下利用一些优化方法如随机梯度下降、自适应学习率算法优化神经网络中的参数．

近年来，基于深度学习的人脸识别方法受到了广泛研究．据了解，现有综述文献[4,11–14]主要针对传统识别方法，基于此，本文综述了2014年以来基于深度学习的二维人脸识别方法．特别地，本文将从人脸预处理、深度特征学习、特征比对、人脸数据集和评价标准五个方面进行介绍．最后对人脸识别的未来发展进行展望．

2 基于深度学习的人脸识别方法

人脸识别技术通过采集人脸图片或视频等数据进行身份识别和认证．身份识别指给出一张人脸图像和已注册的人脸数据库，判断该图像在数据库中的身份，本质是$1:N$的多分类问题，常见的应用场景有门禁系统和会场签到系统等．身份认证指判断两幅人脸图像是否属于同一身份，并不需要知道图像的身份所属，是1:1的二分类问题，通常应用于人证比对和身份核验等场景．

如图1所示，基于深度学习的人脸识别流程主要包括人脸预处理(检测、对齐、标准化、数据增强等)、特征学习、特征比对等步骤，其中特征学习是人脸识别的关键，如何提取强判别性、强鲁棒性的特征是人脸识别的研究重点．本节首先对人脸预处理的各个环节进行简要介绍，然后重点介绍基于深度卷积神经网络的人脸特征学习方法．

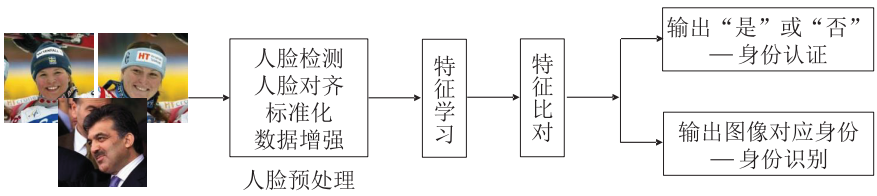


图1 基于深度学习的人脸识别模型训练流程图

2.1 人脸预处理

如图2所示，基于深度学习的人脸识别方法预处理流程通常包括人脸检测、关键点定位、人脸姿态及灰度标准化、人脸数据裁剪及增强．

1) 人脸检测

人脸检测指检测出人脸图像中人脸的具体位置，通常用矩形框框出人脸．人脸检测技术是人脸识别不可或缺的重要环节，随着深度学习的发展该技术也不断得到提升．基于深

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

度学习的人脸检测方法主要分为Fast R-CNN系列[15]、级联CNN系列[16,17]以及SSD系列[18]．其中，Fast R-CNN系列方法用于人脸检测时通常能够获得较低的误检率，但检测速度难以达到实时．级联CNN系列方法如MTCNN[16]速度非常快，即便基于CPU也能对单张人脸进行实时检测．SSD系列方法不仅能达到Fast R-CNN系列方法的误检率，而且能保证检测速度，代表性方法FaceBoxes[19]可以在CPU上实现实时检测．人脸检测方面的综述有[20]，除了检测性能和速度之外，目前比较受关注的问题还有低质量图像中的人脸检测[21]．

(a) 原图　(b) 人脸检测　(c) 关键点检测　(d) 人脸矫正　(e) 裁剪

图2　基于深度学习的人脸识别方法预处理流程

2)　人脸对齐

检测出人脸在图像中的位置后需要进行人脸对齐操作，人脸对齐指检测人脸特征点，如眉眼、鼻子、嘴角以及其它轮廓点．人脸对齐方法可分为判别式方法和生成式方法：生成式方法根据形状和外观构建人脸生成模型，以AAM (Active Appereance Model)[22]和ASM (Active Shape Model)[23]为代表；判别式方法通常学习独立的局部检测器或回归器来定位每个面部关键点，具体实现方法包括CLMs[24]、级联形状回归[25]以及深度学习方法[26,27]．特别地，DeepFace[28]为了使卷积神经网络发挥最大作用，得到二维对齐图像后，进行了三维人脸对齐．随着网络性能不断提升以及数据集的种类和数量不断扩大，大多数人脸识别方法[29–31]只需进行二维人脸对齐甚至弱对齐[32]就能达到很高的识别精度．人脸对齐的难点在于人脸尺度、光照、遮挡、姿态、复杂表情等带来的影响，人脸对齐的更多内容可参考综述[33,34]．

3)　人脸标准化

为了算法的稳定性，一般会对图像进行一些数值标准化的处理，对不同光强、不同光源方向下得到的人脸图像进行补偿，以减弱由于光照变化造成的图像信号变化．例如人脸识别方法SphereFace[30]将所有像素值减去127.5再除以128，使图片像素值范围从$[0,255]$变成$[-1,1]$，VGGFace[35]则是将图片中所有图像减去平均脸，而文献[36,37]对图像进行了灰度处理．

4)　人脸数据增强

数据增强是基于深度学习的人脸识别方法常用的预处理步骤，目的是为了增加数据量．需要说明的是，基于深度学习的人脸识别模型在训练阶段使用数据增强，而测试阶段则不使用．数据增强的方式多种多样，常见的方法是随机裁剪和镜像翻转[38,39]．随机裁

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

剪将图片随机裁剪成不同的图像块，镜像翻转指水平镜像翻转图片，全部翻转或以一定的概率翻转．在使用深度卷积神经网络的人脸识别方法中，数据增强被大量使用[38,40–42]．

2.2 人脸图像深度特征学习

深度卷积神经网络的网络结构和损失函数是影响人脸深度特征学习及识别性能的两个关键因素．2012年，Hinton和其学生Krizhevsky首次将深度卷积神经网络成功应用于解决计算机视觉领域的关键问题[38]．之后，VGGNet[43]，GoogLeNet[44]以及ResNet[45]这三类网络相继被提出并成功被应用于物体识别和人脸识别．在经典的多分类损失函数Softmax loss基础上，损失函数的设计问题受到广泛关注，通过引入分类间隔及度量学习等机制使得人脸深度特征学习具有强的判别性[29–31,46–48]，人脸识别的性能不断得到提高．下面主要从人脸识别常用的网络结构和损失函数两方面展开讨论．

2.2.1 人脸识别网络结构

使用深度学习进行人脸识别的早期，研究人员倾向于使用多个深度卷积神经网络提取人脸特征，再将特征融合．在文献[28,49]中，作者提出首先将多个深度卷积神经网络提取的特征拼接并使用PCA降维得到更有效的特征．文献[50]中使用60个深度卷积神经网络(DCNN)从不同的面部图像块提取出19,200维融合特征，然后通过PCA将特征压缩至150维．多达60个DCNN使DeepID在Labeled Faces in the Wild (LFW)数据集[10,51]上取得97.45%的人脸认证准确率．类似的，文献[42,48,52]均使用了25个DCNN用于提取人脸深度特征并融合．而基于深度学习的人脸识别方法的趋势是使用单个网络，多网络融合特征逐渐被VGGNet[43]、GoogLeNet[44]和ResNet[45]这三种深度人脸识别的代表性网络架构所取代．

1) VGGNet

牛津大学视觉几何组在2014年提出的VGGNet系列深度卷积神经网络一共有5种结构，层数在11层至19层之间，其中应用最广的是VGG16和VGG19．VGGNet的突出表现在于使用多个3×3的卷积核替代AlexNet中7×7的卷积核，小的卷积核一方面可以减少参数，另一方面增加了非线性映射，有助于提升网络的拟合能力．并且，VGGNet增加了网络的深度，使用多种结构验证了增加网络深度可以提升分类准确性．以VGG16为例，该网络由13个卷积层和3个全连接层组成，每个卷积层后连接一个ReLU激活函数层，池化方式与AlexNet相同，前两个全连接层都有4096个通道，最后一个全连接层的通道数与分类的类别数一致．文献[35]使用VGGNet在LFW数据集上获得了99.13%的人脸认证准确率．VGGNet系列网络结构的参数量仍然很庞大，五种结构的参数量均在1.3亿以上．

2) GoogLeNet

同是2014年，由谷歌团队提出的网络结构GoogLeNet通过增加网络结构的稀疏性来解决网络参数过多的问题．不同于VGGNet和AlexNet，GoogLeNet使用Inception模块构建模块化结构，在模块中使用不同大小的卷积核实现多尺度特征的融合．图3是一个Inception模块，为了方便对齐选用了1×1、3×3和5×5的卷积核．由于较大的卷积核会带来巨大计算量，分别在3×3和5×5的卷积层之前增加了一层1×1的卷积层

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

用于降维，并且在模块中加入池化层．最后，将四个通道的输出合并．FaceNet[32] 使用 Inception 模块实现了轻量级的深度人脸识别模型，可以在手机上实时运行．

3)　ResNet

网络加深理应有利于提升网络的性能，但深度增加也给训练带来难度．He 等人针对这类问题提出了 ResNet[45]，允许网络结构尽可能加深．ResNet 的核心策略是增加跨层连接，直接学习层与层之间的残差．图 4 是一个残差模块，该模块的输入为 x，输出是 $F(x)+x$，$F(x)$ 即残差，中间的参数层只需要学习残差部分，可以有效减小训练误差，并且这个恒等映射的跨层连接避免了反向传播过程中的梯度消失，有利于训练更深的网络．ResNet 收敛速度快，目前最新的基于深度学习的二维人脸识别方法[30,31,37,47] 大部分都采用残差模块．

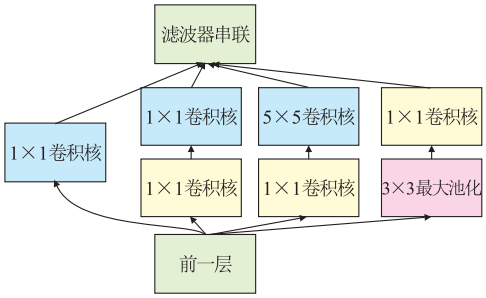


图 3　Inception 模块[44]

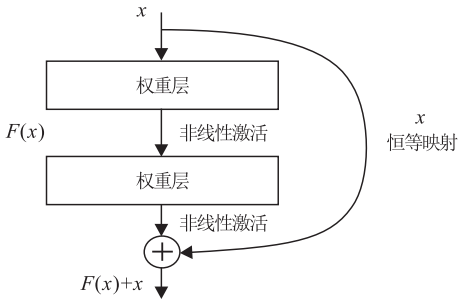


图 4　残差模块[45]

2.2.2　人脸识别损失函数

除了网络结构之外，用于衡量模型识别能力的损失函数同样对基于深度学习的人脸识别方法有重要作用．损失函数可以指导神经网络将人脸图像映射到不同的特征空间，选择合适的损失函数有利于在特征空间将不同类别的人脸图像区分开，提升人脸识别的精度．

1)　Softmax loss

Softmax loss 是一种常用于人脸图像多分类问题的损失函数．Softmax 激活函数

$$f_j(\mathbf{x}_i)=\exp(\mathbf{W}_{y_i}^T\mathbf{x}_i+b_{y_i})\Big/\sum_j^C\exp(\mathbf{W}_j^T\mathbf{x}_i+b_j)$$

的作用是将模型预测结果进行归一化操作，使输出结果为 $[0,1]$ 区间内的概率值．而交叉熵损失函数用于计算模型判别的分类结果与人脸图像真实标签之间的误差．将 Softmax 函数取负对数得到交叉熵损失

$$L_i=-\log\frac{\exp(\mathbf{W}_{y_i}^T\mathbf{x}_i+b_{y_i})}{\sum\limits_j^C\exp(\mathbf{W}_j^T\mathbf{x}_i+b_j)},\tag{1}$$

其中 $\mathbf{x}_i$ 表示第 i 个人脸图像的特征向量，y_i 是 $\mathbf{x}_i$ 真实的类别标签，L_i 表示损失，b 为偏置，$\mathbf{W}_{y_i}^T$ 与 $\mathbf{W}_j^T$ 分别表示将 $\mathbf{x}_i$ 判别为 y_i 类和 j 类的权向量，C 表示总类别数，$\mathbf{W}_{y_i}^T\mathbf{x}_i+$

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

b_{y_i} 表示人脸图像在类别 y_i 上的得分，上式对应的人脸图像类标签形式为独热编码，在真实类别上的得分越高则损失越低．总体 N 个训练样本的 Softmax loss 如下

$$L_s = -\frac{1}{N}\sum_i^N \log \frac{\exp(\mathbf{W}_{y_i}^T \mathbf{x}_i + b_{y_i})}{\sum\limits_j^C \exp(\mathbf{W}_j^T \mathbf{x}_i + b_j)}.$$

2) Large Margin Softmax (L-Softmax)[46]

Softmax loss 分类的原理是当人脸图像特征 $\mathbf{x}_i$ 来自类别 y_i 时满足 $\mathbf{W}_{y_i}^T \mathbf{x}_i > \mathbf{W}_{j\neq y_i}^T \mathbf{x}_i$，可写成

$$\|\mathbf{W}_{y_i}^T\|\|\mathbf{x}_i\| \cos\theta_{y_i,i} > \|\mathbf{W}_{j\neq y_i}^T\|\|\mathbf{x}_i\| \cos\theta_{j\neq y_i,i},$$

其中 $\theta_{y_i,i}$ 表示向量 $\mathbf{W}_{y_i}^T$ 与 $\mathbf{x}_i$ 之间的夹角．L-Softmax[46] 通过增加一个正整数变量 m 将原式变成

$$\|\mathbf{W}_{y_i}^T\|\|\mathbf{x}_i\| \cos(m\theta_{y_i,i}) > \|\mathbf{W}_{j\neq y_i}^T\|\|\mathbf{x}_i\| \cos\theta_{j\neq y_i,i}, \tag{2}$$

使原约束条件变得更加严格从而保证不同类别人脸图像特征之间有分类间隔．于是改进的损失函数 L-Softmax 形式如下

$$L_{lm} = -\frac{1}{N}\sum_i^N \log \frac{\exp(\|\mathbf{W}_{y_i}\|\|\mathbf{x}_i\|\psi(\theta_{y_i,i}))}{\exp(\|\mathbf{W}_{y_i}\|\|\mathbf{x}_i\|\psi(\theta_{y_i,i})) + \sum\limits_{j\neq y_i}^C \exp(\|\mathbf{W}_j\|\|\mathbf{x}_i\| \cos\theta_{j,i})}, \tag{3}$$

其中 $\mathbf{W}_i$ 为权向量，C 表示总类别数，$\theta_{j,i}$ 为 $\mathbf{W}_j$ 和 $\mathbf{x}_i$ 之间的夹角，m 用于控制类间距离，$\psi(\cdot)$ 是为便于梯度反向传播而设计的单调递减函数

$$\psi(\theta) = (-1)^k \cos(m\theta) - 2k, \quad \theta \in \Big[\frac{k\pi}{m}, \frac{(k+1)\pi}{m}\Big].$$

3) Angular Softmax (A-Softmax)[30]

A-Softmax[30] 是对 L-Softmax 的进一步改进，将权向量进行 L_2 归一化，只利用角度进行分类，并用余弦角度间隔代替欧几里得距离间隔，具体形式如下

$$L_{as} = -\frac{1}{N}\sum_i^N \log\left(\frac{\exp(\|\mathbf{x}_i\|\psi(\theta_{y_i,i}))}{\exp(\|\mathbf{x}_i\|\psi(\theta_{y_i,i})) + \sum\limits_{j\neq y_i}^C \exp(\|\mathbf{x}_i\| \cos\theta_{j,i})}\right). \tag{4}$$

使用 A-Softmax loss 进行人脸识别的 SphereFace[30] 虽然有效，但是优化目标与测试方式不一致．针对 SphereFace 存在的问题，NormFace[29] 中的损失函数将权向量和特征向量都进行了归一化，弥补了 A-Aoftmax 的不足，使其更具解释性．

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

4)　Additive Margin Softmax[48] 及 CosFace[31]

Additive Margin Softmax (AM-Softmax)[48] 与 CosFace[31] 将 (4) 中形式复杂的 $\psi(\cdot)$ 替换成形式更简单的 $\psi(\theta)=\cos\theta-m$，增加余弦距离间隔．并且与 NormFace[29] 一致，将权向量和特征向量都进行了 L_2 归一化得到如下形式

$$L_{am}=-\frac{1}{N}\sum_{i}^{N}\log\frac{\exp(s\cos\theta_{y_i,i}-m)}{\exp(s\cos\theta_{y_i,i}-m)+\sum\limits_{j\neq y_i}^{C}\exp(s\cos\theta_{j,i})},\tag{5}$$

其中 $\cos\theta_{j,i}$ 表示归一化后的权向量 $\mathbf{W}_j$ 和特征向量 $\mathbf{x}_i$ 夹角的余弦值，C 表示总类别数．m 表示余弦距离间隔，用于控制不同类别人脸图像之间的距离．s 为尺度参数，用于控制人脸图像特征所在超球面的半径大小．

5)　ArcFace[47]

ArcFace 使用了与式 (5) 不同的间隔控制方式，将控制人脸图像类间距离的超参数 m 放置于余弦函数内，得到如下损失函数

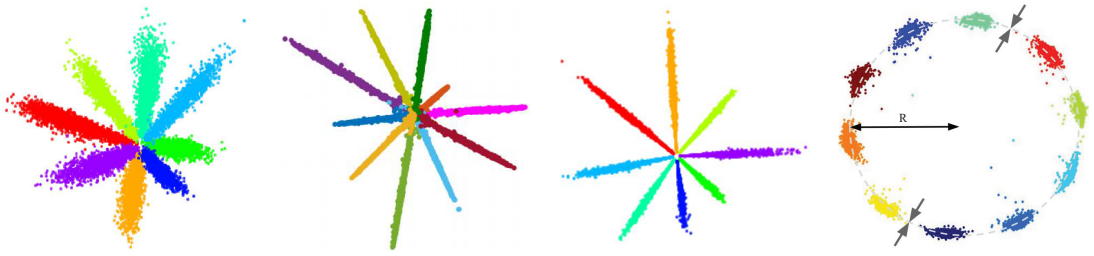
$$L_{arc}=-\frac{1}{N}\sum_{i}^{N}\log\frac{\exp(s\cos(\theta_{y_i,i}+m))}{\exp(s\cos(\theta_{y_i,i}+m))+\sum\limits_{j\neq y_i}^{C}\exp(s\cos\theta_{j,i})}.\tag{6}$$

6)　Ring loss[53]

大量研究工作[30,31,46] 表明人脸图像深度特征向量归一化有利于提升基于深度学习的人脸识别方法性能．Ring loss 提供了一种软归一化方式，从数据中学习归一化尺度而不是直接将人脸图像的深度特征归一化至人工设定的尺度，具体形式如下

$$L_{ring}=L_S+\frac{\lambda}{2}L_r=L_S+\frac{\lambda}{2N}\sum_{i}^{N}\left(\|\mathbf{x}_i\|_2-r\right)^2,\tag{7}$$

其中 L_S 代表主损失函数，例如 Softmax loss、L-Softmax loss 等．L_r 代表 Ring loss．$\mathbf{x}_i$ 代表第 i 个样本的深度特征，r 代表归一化尺度．图 5 为几种不同损失函数对应的深度特征可视化．



(a) Softmax[53]　(b) L-Softmax[46]　(c) AM-Softmax[31]　(d) Ring loss[53]

图 5　使用不同损失函数训练深度卷积神经网络得到的特征可视化

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

7) Center loss[54]

Center loss[54]的主要思想是通过增加惩罚让同类人脸图像特征向类中心靠拢. 实验验证了单独使用Center loss不如与Softmax loss结合效果好，因此在实际应用中将Center loss与Softmax loss结合，并使用超参数λ平衡这两种损失函数的作用，即

$$L_{center}=L_S+\lambda L_c=\frac{\lambda}{2}\sum_{i}^{m}\|\mathbf{x}_i-c_{y_i}\|_2^2, \tag{8}$$

其中L_S表示Softmax loss，L_c表示center loss, c_{y_i}是人脸图像特征$\mathbf{x}_i$对应类别y_i的中心，m表示人脸图像样本数.

8) Contrastive loss[36,49,55,56]

Contrastive loss原本由Yann LeCun[55]提出用于数据降维，其目标是让原本相似(不相似)的样本在低维特征空间仍然相似(不相似)，形式如下

$$L_{contra}=\frac{1}{2N}\sum_{i}^{N}\{y\|\mathbf{x}_i^a-\mathbf{x}_i^b\|_2^2+(1-y)\left[\max\{0,m-\|\mathbf{x}_i^a-\mathbf{x}_i^b\|_2\}\right]^2\}, \tag{9}$$

其中$\mathbf{x}_i^a$和$\mathbf{x}_i^b$代表第i对样本特征，$y=1$代表两个样本特征相似，$y=0$代表两个样本特征不相似，m代表阈值. DeepID2[42]将Contrastive loss和Softmax loss分别作为身份验证和身份识别的监督信息，用于训练人脸识别模型.

9) Triplet loss

FaceNet[32]中使用的Triplet loss是度量学习[36,57,58]的方法之一，使用Contrastive loss训练人脸识别模型时每次比较两个人脸图像特征之间的距离，Triplet loss则需要比较三个特征向量之间的距离，包括两个同类人脸图像特征和一个与之不同类的人脸图像特征，也被称为三元组. 如图6，通过训练使得在特征空间中，同一个人的不同人脸图像的特征距离较小，而不同人的人脸图像特征距离较大. 损失函数的具体形式为

$$L_{tri}=\sum_{i}^{N}\left[\|\mathbf{x}_i^a-\mathbf{x}_i^p\|_2^2-\|\mathbf{x}_i^a-\mathbf{x}_i^n\|_2^2+\alpha\right]_+, \tag{10}$$

其中$\mathbf{x}_i^p$与$\mathbf{x}_i^a$为同一个人的人脸图像特征，而$\mathbf{x}_i^n$则代表与$\mathbf{x}_i^a$不同类的人脸图像特征. $[\,]_+$表示取非负数，当$[\,]$内数值为负时取0. α为阈值.

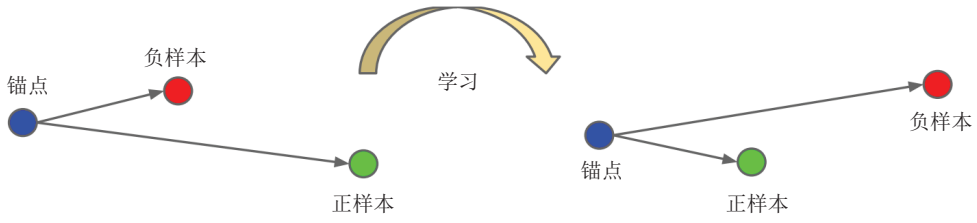


图6 Triplet loss作用示意图[32]

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

从另一个角度考虑，如果使用海量数据，如FaceNet使用800万人的2亿张图片训练网络，若选Softmax loss为损失函数则输出层的节点多达800万，而使用Triplet loss可避免该问题．Triplet loss的原理符合认知规律，在应用中也取得了不错的表现，但其难点在于采样，如果采样选择得当则损失函数很快收敛，否则会需要很长时间用于训练模型．因此VGGFace[35]为了加速损失函数的收敛速度，用Softmax loss训练好模型再使用Triplet loss对特征提取层进行微调．

10)　Multi-class N-pair loss (N-pair-mc loss)[59]

Contrastive loss与Triplet loss每次更新只选一个负样本，而N-pair-mc loss的不同之处在于每次更新时与更多不同人的人脸图像进行交互，并且控制特征比对次数，有效地减少了计算量．为了分析N-pair-mc loss的作用，将其与$(N+1)$-tuple loss进行了比较，$(N+1)$-tuple loss是将Triplet loss中用于比较的负样本数量由1提升到$N-1$，$N=2$时与Triplet等价

$$L_{(N+1)}\left(\{x,x^+,\{x_i\}_{i=1}^{N-1}\};\mathbf{x}\right)=\log\left(1+\sum_{i=1}^{N-1}\exp\left(\mathbf{x}^T\mathbf{x}_i-\mathbf{x}^T\mathbf{x}^+\right)\right),\tag{11}$$

其中x^+是x的正样本，即属于同一人的人脸图像，而$\{x_i\}_{i=1}^{N-1}$是$N-1$个与x不同类的负样本，即与x属于不同人的人脸图像，$\mathbf{x}$、$\mathbf{x}_i$和$\mathbf{x}^+$分别是人脸图像x、x_i和x^+的特征．

N-pair-mc loss采用一种技巧性的方式选择负样本，假设$\{(x_i,x_i^+)\}_{i=1}^N$是N对分别属于N个人的人脸图像，如果$i\neq j$，则$y_i\neq y_j$．

$$L_{Npair}\left(\{x_i,x_i^+\}_{i=1}^N;\mathbf{x}\right)=\frac{1}{N}\sum_{i=1}^{N}\log\left(1+\sum_{j\neq i}\exp\left(\mathbf{x}_i^T\mathbf{x}_j^+-\mathbf{x}_i^T\mathbf{x}_i^+\right)\right),\tag{12}$$

对于N次采样，使用Triplet loss需要进行$3N$次特征提取，$(N+1)$-tuple loss需要进行$(N+1)N$次特征提取，而N-pair-mc loss只需要进行$2N$次特征提取．文献[59]实验验证了使用N-pair loss能达到比使用Triplet loss更快的收敛速度和人脸识别精度．

2.3　特征比对

基于深度学习的人脸识别方法的基本思路：训练阶段，在损失函数的指导下利用海量有标记的人脸图像样本对网络参数进行有监督训练．测试阶段，将待测试的人脸图像输入训练好的神经网络提取人脸深度特征，使用最近邻分类器通过比较深度特征之间的距离进行身份识别或认证．通常使用欧几里得距离或余弦相似度作为特征距离的度量，假设人脸图像x_i和x_j的特征分别为$f(x_i)$和$f(x_j)$，当特征之间的距离在预先设定的阈值τ范围内时，即

$$d\left(f(x_i)-f(x_j)\right)\leq\tau,\tag{13}$$

则认为这两幅图像来自同一个人．

2.4　人脸数据集

作为数据驱动的方法，基于深度学习的人脸识别方法需要大量训练数据，数据集的发展也反映了人脸识别技术的发展．与早期实验室环境下采集获得的人脸数据

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

不同，2007年公开的LFW数据集开启了无约束场景下人脸识别研究的新阶段，有力地推动了无约束人脸识别的发展．随后不断有更大、更多样化的人脸数据集被发布，例如CASIA WebFace、MS-Celeb-1M、MegaFace为训练人脸识别算法提供了海量样本数据．IARPA Janus Benchmark-A (IJB-A)、IARPA Janus Benchmark-B (IJB-B)、IARPA Janus Benchmark-C (IJB-C)在不断扩充人脸图片数据量的同时，增加了被拍摄者姿态、拍摄环境的变化．VGGFace2则侧重跨年龄条件下的人脸识别场景．YouTube Faces (YTF)的任务是基于人脸视频进行动态人脸识别．目前常用的二维人脸数据集见表1，表中列出了数据集名称、所含身份个数，人脸图像总数以及发布时间．各数据集的详细介绍如下．

表1 常用于二维人脸识别的公开数据集

数据集	身份数量	图像数量	发布时间
LFW[10,51]	5,749	13,233	2007
YTF[60]	1,595	3,425	2011
CASIA WebFace[36]	10,575	494,414	2014
IJB-A[61]	500	5,712幅图片，2,085段视频	2015
VGGFace[35]	2,622	2.6M	2015
MegaFace[62]	690,572	4.7M	2016
UMDFaces[63,64]	8,277	367,888幅图片，22,075段视频	2016
IJB-B[65]	1,845	11,754幅图片，7,011段视频	2017
VGGFace2[66]	9,131	3.31M	2018
IJB-C[67]	28,936	138,000幅图片，11,000段视频	2018

1) LFW[10,51]

由美国马萨诸塞大学阿姆斯特朗分校计算机视觉实验室发布的LFW数据集包含5,749人的13,233幅人脸图片，图片来自于雅虎网，属于无约束场景．由于LFW数据集中有4,069人仅有一张人脸图片，通常该数据库不用于训练深度神经网络，而是作为测试集使用，常用的任务是分别判断LFW提供的6,000对人脸图片是否属于同一人．

2) YTF[60]

除了静态图片，也可用视频图像进行人脸识别[68–70]．YTF数据集[60]包含1,595人的3,425段视频，每个人平均有2.15段视频，视频长度介于48至6070帧之间，平均长度为181.3帧，视频均来自视频网站YouTube．该数据集的任务是判断每两段视频中的人是否属于同一身份，对于一段视频，一般通过离散采样转换成多帧图片，再基于图片进行特征提取及比对．

3) CASIA WebFace[36]

该数据集包含10,575人的494,414幅人脸图片，图像来自于IMDb网站，已被广泛用于训练基于深度卷积神经网络的人脸识别模型[29–31]．CASIA WebFace的作者指出使

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

用CASIA WebFace数据集训练，在LFW进行测试是一个较好的评价人脸识别模型性能的方案.

4) IJB-A[61]

该数据集包含来自500人的5,712幅图片和2,085段视频数据．与LFW和YTF相比，IJB-A的特点是图片和视频取自完全无约束环境，光照条件与被拍摄者面部姿态的变化比较大，且具有不同的分辨率，除了静态图片，还包含被拍摄者的动态视频，该数据集非常符合实际应用场景．而IJB-B[65]和IJB-C[67]数据集是该研究院随后发布的更大的数据集，IJB-B包含1,845个对象的11,754幅图片和7,011段视频，内容囊括了IJB-A数据集．而IJB-C囊括了IJB-A与IJB-B的内容，包含138,000幅人脸图像，11,000段视频.

5) MegaFace[62]

由华盛顿大学举办的MegaFace挑战赛有两种挑战，其一是将训练好的模型在一百万干扰项条件下进行识别和验证测试，其二是使用MegaFace提供的67万人的470万张人脸图片训练模型，在百万规模的测试集上进行测试．MegaFace的目的是挑战从百万人的干扰选项中寻找同一个人的不同图片之间的匹配．在此之前，用于测试的身份一般在一万左右，MegeFace超大规模的测试对于评估和提升人脸识别算法很有意义.

6) VGGFace2[66]

该数据集包含9,131人的大约3百万人脸图片，平均每人有362.6幅图片，该数据集中的数据来源于谷歌．数据集分为训练集和测试集，其中训练集包含8,631人的图片，测试集包含500人的图片．图片涵盖了不同的年龄、姿势、光照、种族和职业，除了身份信息之外，数据集中还提供每幅人脸图像的人脸框、5个关键点、以及估计的年龄和姿态.

7) MS-Celeb-1M[71]

该数据集包含1百万名人的1千万幅图片．这些图像均来自于互联网，其中测试集包含1,000人．经过微软标注，每人大约有20幅人脸图片，并且用于测试的图片并未公开，以保证公平性.

8) UMDFaces[63, 64]

该数据集包含8,277人的367,888幅静态图片和3,100人的22,075段视频．数据集提供的人脸信息包括人脸框、姿势估计、21个关键点以及性别，并且该数据集提供了容易、中等、困难三个等级的人脸验证测试，每个等级的测试集包含100,000对人脸图像.

9) Face Recognition Vendor Test (FRVT)

由美国国家标准技术局NIST (National Institute of Standards and Technology)设定的人脸识别测试集FRVT的测试权威性是全球工业界黄金标准，使用来自美国国土安全局的百万量级真实业务场景图片进行评估，并且为了保障公平性，FRVT不公开用于测试的数据．与学术上常用的LFW、YTF甚至MegaFace相比，FRVT更贴近真实场景，也更公平.

目前已有近30种人脸识别方法在LFW数据集达到了99%以上的识别精度，最高达到99.83%．类似于YTF的人脸视频数据集增大了识别难度，由于视频中的人脸是动态的，比静态图片多了一些姿态的变化，在静态图片上效果好的算法在处理视频时未必仍

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

然能保持很好的效果，因此YTF数据集对于评测人脸识别方法的性能很有意义．IJB-A、IJB-B以及IJB-C也是人脸识别方法常用的测试数据集，与LFW、YTF的图像相比，IJB的系列数据集中的图像更贴近实际应用场景．以上的测试数据集包含的对象在几千人至一万人，而MegaFace开启了超大规模的人脸识别任务，使用大规模的人脸识别测试有助于发现人脸识别方法的优点和缺陷．CASIA WebFace人脸数据集常用于训练深度卷积神经网络，在很多机构不公开数据库的情况下，CASIA WebFace人脸数据集为推动基于深度学习的人脸识别技术的发展起到了很大的作用．而VGGFace2人脸数据集的优点在于覆盖了很大范围的姿态、年龄以及种族，除了进行身份识别外，还可以进行姿态、年龄识别等．MS-Celeb-1M人脸数据集中每个对象有多个属性，并且数据量非常大，但缺点在于这个数据集有很大噪声，即存在大量标注错误的图片，因此，在使用前需要针对标注问题对数据集进行清洗处理．

2.5 评价标准

1) 身份认证

一般使用ROC曲线作为人脸识别方法的评价指标，ROC曲线由两项指标确定，分别是接受率(Ture Alarm Rate, TAR)，误识率(False Alarm Rate, FAR)．将所有正样本(i,j)、负样本(i,j)的集合分别记为P_{same}和P_{diff}，用$D(x_i,x_j)$表示特征之间的距离，距离根据测试数据集要求选择欧氏距离、余弦距离等．由此可计算接受率TAR和误识率FAR，如

$$\text{TAR}(\tau)=\frac{|(i,j)\in P_{\text{same}},D(x_i,x_j)\leq\tau|}{|P_{\text{same}}|},\tag{14}$$

$$\text{FAR}(\tau)=\frac{|(i,j)\in P_{\text{diff}},D(x_i,x_j)\leq\tau|}{|P_{\text{diff}}|}.\tag{15}$$

易知两个比率都在0到1之间．通过改变阈值τ可以调节接受率和误识率的值，分别以接受率和误识率为横纵坐标轴绘制ROC曲线．ROC曲线与误识率轴之间的面积被定义为AUC(Area Under Curve)，AUC始终不会超过1．ROC曲线下方的面积越大说明该方法的准确度越高．有时也会直接用认证精度作为评价人脸识别方法的指标．随着深度学习技术的发展，大家对FAR很低的条件下TAR值的关注程度越来越高，即对安全度的要求越来越高．例如IJB-A要求在$\text{FAR}=10^{-3}$时评估TAR；而Megaface关注$\text{FAR}=10^{-6}$时对应的TAR；在MS-Celeb-1M挑战中，需要考量$\text{FAR}=10^{-9}$时对应的TAR．

2) 身份识别

一般使用身份识别精度作为识别方法的评价指标，计算方式简单明了，与认证准确度类似，计算识别正确的比例即可．比较特别的是，在大规模分类问题中常使用K次命中率作为评价的标准，即真实标签出现在预测结果前K名之内，则认为预测正确．早期的论文经常使用5次命中率进行比较，但是随着身份识别准确度不断提升，现在一般使用首位命中率，即模型以最高概率将样本分类到真实标签的比例．

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

3)　基于深度学习人脸识别方法比较

表2从训练样本数量、使用网络个数以及在LFW和YTF数据集的表现等方面比较了一些具有代表性的基于深度学习的人脸识别方法．从表中可知，早期研究人员倾向于使用多个深度卷积神经网络学习人脸图像的多尺度融合特征，如DeepFace[28]和DeepID系列[42,50,52]．随着深度卷积神经网络的发展，目前人脸识别方法一般只使用单个网络，并且采用的网络结构以ResNet为主，例如DeepVisage[37]、SphereFace[30]、CosFace[31]等．研究热点也从网络结构设计转移至损失函数的设计，例如L-Softmax[46]、NormFace[29]、ArcFace[47]等方法将度量学习的思想引入Softmax loss并提升了人脸识别模型的性能．

表2　人脸识别方法在LFW、YTF数据集验证精度的比较

方法	训练样本	网络个数	LFW准确度	YTF准确度
DeepFace[28]	4M	3	97.35	91.4
DeepID[50]	0.2M	60	97.45	—
DeepID2[42]	0.29M	25	99.15	—
DeepID2+[48]	0.29M	25	99.47	91.9
DeepID3[52]	0.29M	25	99.53	—
Sparse ConvNet[54]	0.29M	25	99.55	92.7
CASIA-WebFace[36]	0.49M	1	97.73	92.2
VGGFace[35]	2.6M	1	98.95	97.3
FaceNet[32]	200M	1	99.65	95.1
Baidu[72]	1.2M	10	99.77	—
CenterFace[54]	0.7M	1	99.28	94.9
DeepVisage[37]	4.48M	1	99.62	99.2
L-Softmax[46]	0.49M	1	98.71	—
SphereFace[30]	0.49M	1	99.42	95.0
NormFace[29]	0.49M	1	99.19	94.7
CosFace[31]	5M	1	99.73	97.6
ArcFace[47]	10M	1	99.83	—

3　总结与展望

本文首先介绍了人脸识别的发展脉络，然后着重从人脸预处理、特征学习、特征比对、人脸数据集、评价标准五个方面综述了近几年基于深度学习的二维人脸识别方法．本文从网络结构和损失函数两部分总结了特征学习方法：对于网络结构，深度学习方法从早期使用多个网络发展为使用单个网络，并且多采用VGGNet、GoogLeNet以及ResNet这三类常用网络结构；对于损失函数，本文总结了基于欧式距离的损失函数如Contrastive

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

loss、Triplet loss、N-pair loss以及Softmax loss及其变种，度量学习的引入使得深度人脸识别模型更易区分同类和不同类人脸图像的特征．对于人脸数据集，本文总结和归纳了常用于深度人脸识别的大规模人脸图像数据集，包括常用的训练数据集以及测试集．本文还介绍了人脸预处理流程、特征比对方式以及两种人脸识别任务分别对应的评价标准．总体而言，随着深度学习技术的不断发展以及真实环境下大尺度人脸数据库的不断公开，人脸识别技术受到了广泛研究，获得了长足进步．近年来，随着人脸识别方法精度的不断提升，人脸识别已广泛应用于手机解锁、安防、金融、教育、交通等各个方面，出现了“刷脸”吃饭、“刷脸”购物，甚至“刷脸”登机等现象．可以说，人脸识别技术从学术研究和产业化应用均取得了丰硕成果，但在以下方面仍然面临严峻挑战．

1) 低质量图像人脸识别

通常，人脸图像质量受采集环境、采集设备和采集距离等因素影响．人脸图像的分辨率、模糊程度、姿态变化、光照变化、遮挡物等是影响人脸图像质量的关键因素．基于深度学习的人脸识别方法，特别是动态视频监控下人脸识别方法受人脸质量影响较大．如何提升低质量人脸图像识别精度是一个值得关注的问题．大姿态人脸识别的解决方案通常利用三维人脸模型将人脸姿态矫正之后再进行识别．对于光照变化，三维人脸识别技术和近红外人脸识别技术为解决该问题提供了一定的可能性．对于非配合场景下因佩戴墨镜、口罩等造成采集到的人脸图像严重遮挡问题，目前没有较为有效的解决方法．

2) 跨年龄人脸识别

随着年龄增长，人的相貌会发生显著变化，特别是少年、成年到老年各个阶段的相貌会有明显差异，这使得跨年龄人脸识别成为一大难点．跨年龄人脸识别的实际应用场景包括人证比对、失踪人群追踪等．目前主要的解决方案是使用生成模型生成目标年龄段的人脸图像辅助跨年龄人脸识别．

3) 跨模态及多模态识别

跨模态及多模态人脸识别能够利用多重传感器的优势，通常能够克服单一模态人脸识别的诸多问题，因此也是一个重要的研究问题．目前，人脸主要模态包括素描、图像、红外图像、三维人脸等四种．跨模态人脸识别的难点在于如何挖掘异构信息中共同的、本质的身份判别信息．多模态人脸识别的难点在于如何有效融合多模态之间的互补信息．

4) 人脸防伪

随着人脸识别技术逐渐被应用到日常生活中，人脸识别防伪技术也迫切需要被重视．常见的欺骗手段包括使用合法用户的人脸照片、视频或者三维人脸面具等攻击人脸识别系统．人脸防伪方面的研究目前集中在活体检测、基于图像纹理区分以及基于三维人脸重建防伪等．

5) 隐私保护

随着人脸识别技术的普及，该技术也可能会被不法分子利用，造成隐私安全隐患．门禁系统采集的人脸数据的保护、第三方通过移动终端恶意收集人脸数据等问题亟需解决．对此，部分学者提出了保护生物特征隐私的方式以及反人脸识别技术．

我们相信，随着基于深度学习的人脸识别技术的不断发展，数据样本的不断积累以及

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

国家相关法律法规的不断完善，上述问题均能够得到较好的解决．人脸识别技术也能够被合理利用，服务大众生活．

参考文献：

[1] LI S Z, JAIN A K. Handbook of face recognition[M]. 2nd ed. New York: Springer, 2011.

[2] BRUNELLI R, POGGIO T A. Face recognition: features versus templates[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(10): 1042-1052.

[3] LADES M, VORBRÜGGEN J C, BUHMANN J M, et al. Distortion invariant object recognition in the dynamic link architecture[J]. IEEE Transactions on Computers, 1993, 42(3): 300-311.

[4] SAMAL A, IYENGAR P A. Automatic recognition and analysis of human faces and facial expressions: a survey[J]. Pattern Recognition, 1992, 25(1): 65-77.

[5] MATTHEW T, ALEX P. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71-86.

[6] BELHUMEUR P N, HESPANHA J P, KRIEGMAN D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 19(7): 711-720.

[7] LIU C J, WECHSLER H. A Gabor feature classifier for face recognition[C]// Proceedings 8th IEEE International Conference on Computer Vision, Vancouver, BC, Canada. IEEE, 2001: 270-275.

[8] AHONEN T, HADID A, PIETIKÄINEN M. Face recognition with local binary patterns[C]// The 8th European Conference on Computer Vision, Prague, Czech Republic. Berlin, Heidelberg: Springer-Verlag, 2004: 469-481.

[9] SINGH G, CHHABRA I. Integrating global Zernike and local discriminative HOG features for face recognition[J]. International Journal of Image and Graphics, 2016, 16(4): 1650021.

[10] HUANG G B, RAMESH M, BERG T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments[R]. University of Massachusetts, Amherst, 2007: 7-49.

[11] ZHAO W Y, CHELLAPPA R, JONATHON PHILLIPS P, et al. Face recognition: a literature survey[J]. ACM Computing Surveys, 2003, 35(4): 399-458.

[12] ABATE A F, NAPPI M, RICCIO D, et al. 2D and 3D face recognition: a survey[J]. Pattern Recognition Letters, 2007, 28(14): 1885-1906.

[13] BOWYER K W, CHANG K, FLYNN P. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition[M]. Amsterdam: Elsevier Science Incorporated, 2006: 1-15.

[14] JAFRI R, ARABNIA H R. A survey of face recognition techniques[J]. Journal of Information Processing Systems, 2009, 5(2): 41-68.

[15] GIRSHICK R. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision, Santiago, Chile. IEEE, 2015: 1440-1448.

[16] ZHANG K P, ZHANG Z P, LI Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

[17] LI H X, LIN Z, SHEN X H, et al. A convolutional neural network cascade for face detection[C]// IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA. IEEE, 2015: 5325-5334.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[18] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 21-37.

[19] ZHANG S F, ZHU X Y, LEI Z, et al. FaceBoxes: a cpu real-time face detector with high accuracy[C]// 2017 IEEE International Joint Conference on Biometrics, Denver, CO, USA. IEEE, 2017: 1-9.

[20] ZAFEIRIOU S, ZHANG C, ZHANG Z Y. A survey on face detection in the wild: past, present and future[J]. Computer Vision and Image Understanding, 2015, 138: 1-24.

[21] ZHOU Y Q, LIU D, HUANG T S. Survey of face detection on low-quality images[C]// 13th IEEE International Conference on Automatic Face and Gesture Recognition, Xi'an, China. IEEE, 2018: 769-773.

[22] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 681-685.

[23] BEHAINE C A R, SCHARCANSKI J. Enhancing the performance of active shape models in face recognition applications[J]. IEEE Transactions on Instrumentation and Measurement, 2012, 61(8): 2330-2333.

[24] ZADEH A, BALTRUSAITIS T, MORENCY L P. Convolutional experts constrained local model for facial landmark detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA. IEEE, 2017: 2051-2059.

[25] SANCHEZ-LOZANO E, MARTINEZ B, VALSTAR M F. Cascaded regression with sparsified feature covariance matrix for facial landmark detection[J]. Pattern Recognition Letters, 2016, 73(C): 19-25.

[26] WU W Y, QIAN C, YANG S, et al. Look at boundary: a boundary-aware face alignment algorithm[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA. IEEE, 2018: 2129-2138.

[27] WU Y, HASSNER T, KIM K, et al. Facial landmark detection with tweaked convolutional neural networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(12): 3067-3074.

[28] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: closing the gap to human-level performance in face verification[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA. IEEE, 2014: 1701-1708.

[29] WANG F, XIANG X, CHENG J, et al. NormFace: L_2 hypersphere embedding for face verification[C]// Proceedings of the 2017 ACM on Multimedia Conference, Mountain View, CA, USA, 2017: 1041-1049.

[30] LIU W Y, WEN Y D, YU Z D, et al. SphereFace: deep hypersphere embedding for face recognition[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA. IEEE, 2017: 6738-6746.

[31] WANG H, WANG Y T, ZHOU Z, et al. CosFace: large margin cosine loss for deep face recognition[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 5265-5274.

[32] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]// IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA. IEEE, 2015: 815-823.

[33] WANG N N, GAO X B, TAO D C, et al. Facial feature point detection: a comprehensive survey[J]. Neurocomputing, 2018, 275: 50-65.

[34] JIN X, TAN X Y. Face alignment in-the-wild: a survey[J]. Computer Vision and Image Understanding, 2017, 162: 1-22.

[35] PARKHI O M, VEDALDI A, ZISSERMAN A. Deep face recognition[C]// Proceedings of the British Machine Vision Conference, Swansea, UK, 2015: 41.1-41.12.

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[36] YI D, LEI Z, LIAO S C, et al. Learning face representation from scratch[EB/OL]. (2014-11-28)[2017-06-07]. https://arxiv.org/abs/1411.7923.

[37] HASNAT A, BOHNÉ J, MILGRAM J, et al. DeepVisage: making face recognition simple yet with powerful generalization skills[C]// 2017 IEEE International Conference on Computer Vision Workshops, Venice, Italy. IEEE, 2017: 1682-1691.

[38] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// 26th Annual Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, United States, 2012: 1106-1114.

[39] YANG H, PATRAS I. Mirror, mirror on the wall, tell me, is the error small?[C]// IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA. IEEE, 2015: 4685-4693.

[40] DING C X, TAO D C. Robust face recognition via multimodal deep face representation[J]. IEEE Transactions on Multimedia, 2015, 17(11): 2049-2058.

[41] WANG D Y, OTTO C, JAIN A K. Face search at scale: 80 million gallery[EB/OL]. (2015-07-26)[2017-10-27]. http://arxiv.org/abs/1507.07242.

[42] SUN Y, CHEN Y H, WANG X G, et al. Deep learning face representation by joint identification-verification[C]// Annual Conference on Neural Information Processing Systems, Montreal, Quebec, Canada, 2014: 1988-1996.

[43] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2019-07-17]. http://arxiv.org/abs/1409.1556.

[44] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]// IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA. IEEE, 2015: 1-9.

[45] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA. IEEE, 2016: 770-778.

[46] LIU W Y, WEN Y D, YU Z D, et al. Large-margin softmax loss for convolutional neural networks[C]// Proceedings of the 33rd International Conference on Machine Learning, New York City, NY, USA, 2016: 507-516.

[47] DENG J K, GUO J, XUE N N, et al. ArcFace: additive angular margin loss for deep face recognition[C]// 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA. IEEE, 2019: 4690-4699.

[48] WANG F, CHENG J, LIU W Y, et al. Additive margin softmax for face verification[J]. IEEE Signal Processing Letters, 2018, 25(7): 926-930.

[49] SUN Y, WANG X G, TANG X O. Deeply learned face representations are sparse, selective, and robust[C]// IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA. IEEE, 2015: 2892-2900.

[50] SUN Y, WANG X G, TANG X O. Deep learning face representation from predicting 10000 classes[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA. IEEE, 2014: 1891-1898.

[51] HUANG G B, LEARNED-MILLER E. Labeled faces in the wild: updates and new reporting procedures[R]. University of Massachusetts, Amherst, 2014.

[52] SUN Y, LIANG D, WANG X G, et al. DeepID3: face recognition with very deep neural networks[EB/OL]. (2015-02-03)[2017-06-07]. http://arxiv.org/abs/1502.00873.

[53] ZHENG Y T, PAL D K, SAVVIDES M. Ring loss: convex feature normalization for face recognition[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA. IEEE,

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2018: 5089-5097.

[54] WEN Y D, ZHANG K P, LI Z F, et al. A discriminative feature learning approach for deep face recognition[C]// European Conference on Computer Vision, Cham: Springer, 2016: 499-515.

[55] CHOPRA S, HADSELL R, LECUN Y. Learning a similarity metric discriminatively, with application to face verification[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA. IEEE, 2005: 539-546.

[56] SUN Y, WANG X G, TANG X O. Sparsifying neural network connections for face recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA. IEEE, 2016: 4856-4864.

[57] XING E P, NG A Y, JORDAN M I, et al. Distance metric learning with application to clustering with side-information[C]// Advances in Neural Information Processing Systems 15, Vancouver, British Columbia, Canada, 2002: 505-512.

[58] WEINBERGER K Q, SAUL L K. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research, 2009, 10: 207-244.

[59] SOHN K. Improved deep metric learning with multi-class N-pair loss objective[C]// Annual Conference on Neural Information Processing Systems, Barcelona, Spain, 2016: 1849-1857.

[60] WOLF L, HASSNER T, MAOZ I. Face recognition in unconstrained videos with matched background similarity[C]// 24th IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA. IEEE, 2011: 529-534.

[61] KLARE B F, KLEIN B, TABORSKY E, et al. Pushing the frontiers of unconstrained face detection and recognition: IARPA janus benchmark-A[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA. IEEE, 2015: 1931-1939.

[62] KEMELMACHER-SHLIZERMAN I, SEITZ S M, MILLER D, et al. The MegaFace benchmark: 1 million faces for recognition at scale[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA. IEEE, 2016: 4873-4882.

[63] BANSAL A, NANDURI A, CASTILLO C D, et al. UMDFaces: an annotated face dataset for training deep networks[C]// 2017 IEEE International Joint Conference on Biometrics, Denver, CO, USA. IEEE, 2017: 464-473.

[64] BANSAL A, CASTILLO C D, RANJAN R, et al. The do's and don'ts for CNN-based face verification[C]// 2017 IEEE International Conference on Computer Vision Workshops, Venice, Italy. IEEE, 2017: 2545-2554.

[65] WHITELAM C, TABORSKY E, BLANTON A, et al. IARPA janus benchmark-B face dataset[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA. IEEE, 2017: 90-98.

[66] CAO Q, SHEN L, XIE W D, et al. VGGFace2: a dataset for recognising faces across pose and age[C]// 13th IEEE International Conference on Automatic Face and Gesture Recognition, Xi'an, China. IEEE, 2018: 67-74.

[67] MAZE B, ADAMS J, DUNCAN J A, et al. IARPA janus benchmark-C: face dataset and protocol[C]// 2018 International Conference on Biometrics, Gold Coast, QLD, Australia. IEEE, 2018: 158-165.

[68] WOLF L, LEVY N. The SVM-minus similarity score for video face recognition[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA. IEEE, 2013: 3523-3530.

[69] CUI Z, LI W, XU D, et al. Fusing robust face region descriptors via multiple metric learning for face recognition in the wild[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland,

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

OR, USA. IEEE, 2013: 3554-3561.

[70] LI H X, HUA G, LIN Z, et al. Probabilistic elastic matching for pose variant face verification[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA. IEEE, 2013: 3499-3506.

[71] GUO Y D, ZHANG L, HU Y X, et al. MS-Celeb-1M: a dataset and benchmark for large-scale face recognition[C]// 14th European Conference, Amsterdam, Netherlands, 2016: 87-102.

[72] LIU J T, DENG Y F, BAI T, et al. Targeting ultimate accuracy: face recognition via deep embedding[EB/OL]. (2015-07-23)[2017-06-07]. http://arxiv.org/abs/1506.07310.

Deep Learning Based 2D Face Recognition: a Survey

YU Cui-can, LI Hui-bin†

(National Engineering Laboratory for Big Data Analytics,
School of Mathematics and Statistical, Xi'an Jiaotong University, Xi'an 710049)

Abstract: Compared with iris, fingerprint, gait, and other biometric recognition technologies, face recognition has attracted wide attention from academia to industry due to its unique advantages such as natural, convenient, and user-friendly experience. In recent years, driven by deep learning technology, face recognition has made a breakthrough, which shows strong robustness even when suffering from obstacles like facial expression, head pose, illumination, and external occlusions. In particular, deep face recognition technologies have been widely used in security, finance, education, transportation, new retail, and other applications. We realise that in the process of deep face recognition technology becoming widespread, there is an urgent need for some review articles to summarise the basic principles and methods of deep face recognition. This paper first briefly reviews the development of face recognition and then introduces the deep learning based face recognition methods from five aspects: face preprocessing, deep feature learning, feature comparison, face datasets, and evaluation. Finally, the development trend of deep face recognition is discussed.

Keywords: face recognition; deep learning; convolutional neural network; feature learning

Received: 12 Mar 2019. **Accepted:** 15 July 2019.

Foundation item: The National Natural Science Foundation of China (61976173); the National Key Research and Development Program of China (2018AAA0102201); the Ministry of Education-CMCC Artificial Intelligence Construction Project (MCM20190701); the Fundamental Research Funds for the Central Universities (xzy012019041); the Natural Science Basic Research Plan in Shaanxi Province (2019JQ-628).

†**Corresponding author:** H. Li. E-mail address: huibinli@xjtu.edu.cn

(C)1994-2022 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net