

深度人脸识别方法综述

王海勇, 潘海涛*

南京邮电大学计算机学院, 江苏 南京

收稿日期: 2023年6月3日; 录用日期: 2023年8月9日; 发布日期: 2023年8月17日

摘要

深度人脸识别通过大规模数据集训练卷积神经网络获取更鲁棒的人脸表示,极大的提升了人脸识别性能。文中总结了深度人脸识别方法的发展脉络,首先根据卷积神经网络的不同发展阶段回顾了现有的深度人脸识别方法,其次对基于欧几里得距离以及基于角余弦裕度的损失函数进行了回顾,同时总结了一些针对特定任务的人脸识别方法。然后总结了现有的人脸识别数据集以及人脸识别性能的评价指标,并对主流深度人脸识别方法进行了比较。最后总结了人脸识别当前面临的挑战和未来的发展趋势。

关键词

人脸识别, 卷积神经网络, 人脸表示, 损失函数

A Survey of Deep Face Recognition

Haiyong Wang, Haitao Pan*

College of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing Jiangsu

Received: Jun. 3rd, 2023; accepted: Aug. 9th, 2023; published: Aug. 17th, 2023

Abstract

Deep face recognition greatly improves the performance of face recognition by training convolutional neural networks on large-scale data sets to obtain more robust face representation. This paper summarizes the development of depth face recognition methods. First, the existing depth face recognition methods are reviewed according to the different development stages of convolutional neural networks. Secondly, the loss functions based on Euclidean distance and angular cosine margin are reviewed, and some task-specific face recognition methods are summarized. Then, the existing data sets and the evaluation indicators of face recognition performance are summarized, and the mainstream depth face recognition methods are compared. Finally, the current challenges and future trends of face recognition are summarized.

*通讯作者。

Keywords

Face Recognition, Convolutional Neural Networks, Face Representation, Loss Function

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人脸识别在计算机视觉和模式识别领域广受关注, 具有很多现实应用, 包括身份认证、访问控制、人机交互等。由于人工智能技术的兴起, 人脸识别技术取得了重大进展, 早期研究兴趣主要集中在受控条件下的人脸识别。现如今的研究重点在于无约束条件下的人脸识别, 传统的人脸识别方法无法提取鲁棒的人脸表示, 面对光照、姿态变化等无约束场景下的识别性能较差。基于深度学习的人脸识别方法克服了传统手工提取人脸特征的局限性, 成为了目前的研究热点。人脸识别也可以看成是一个多分类任务, 其将人脸图像投影到特征空间, 要求不同身份的人脸特征分布尽可能分散, 而相同身份的人脸特征分布尽可能紧凑, 所提取的人脸特征必须具有较高的可分辨性, 卷积神经网络强大的特征提取能力使其在图像分类、目标检测等计算机视觉领域广受关注, 同时也推动了深度人脸识别技术的发展。深度人脸识别过程如图 1 所示, 主要分为三个部分, 包括人脸预处理、人脸表征和人脸匹配。人脸预处理主要包括人脸检测和人脸对齐, 人脸检测是在一张图像中检测并截取人脸区域, 之后将人脸区域进行人脸对齐, 即获取人脸关键点并进行相似性变换, 将人脸区域映射回标准人脸。人脸表征是将预处理完成后的人脸图像输入深度神经网络提取人脸特征, 将每幅人脸图像映射为一维向量, 最后进行人脸匹配, 将代表两幅图像的人脸特征向量进行比对, 使用欧几里得距离或余弦相似度进行度量, 看是否属于同一身份。人脸识别可进一步分为一对一的人脸验证任务以及一对多的人脸辨认任务, 人脸验证任务是将待测人脸与给定的人脸进行比对, 看是否为同一身份, 主要用于刷脸支付、手机解锁等场景。人脸辨认任务则是将待测人脸与给定的人脸图像库中的人脸进行比对, 看是否属于图像库中的人脸身份, 主要用于门禁系统、视频监控等场景。

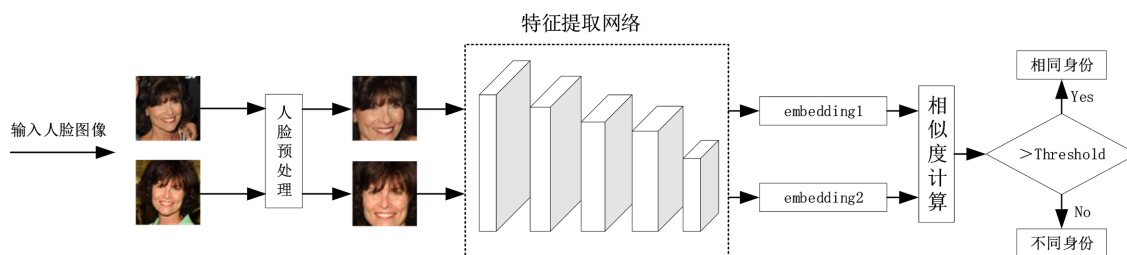


Figure 1. Process of deep face recognition

图 1. 深度人脸识别流程

近年来深度人脸识别的研究主要针对网络架构以及损失函数进行优化改进。网络架构主要基于现有的卷积神经网络进行改进, 目的是使提取出的人脸特征对光照、姿态变化等因素具有鲁棒性, 同时具备可区分性。为了更好的划分决策边界, 使类内更聚, 类间更开, 提出了一系列用于人脸识别的损失函数, 包括对比损失、三元组损失、中心损失以及基于角余弦裕度的损失。本文主要归纳总结了深度人脸识别

技术,总结了基于不同网络架构的人脸识别方法,回顾了深度人脸识别中用于监督训练的损失函数,包括基于欧几里得距离的损失和基于角余弦量的损失。接着总结了人脸识别常用的数据集和评价指标,并在数据集上对主流的深度人脸识别方法进行比较。最后总结了人脸识别目前面临的挑战和未来的发展趋势。

2. 网络模型

深度人脸识别主要采用卷积神经网络作为特征提取网络提取人脸特征,卷积神经网络的关键思想就是局部感受野、权值共享以及池化。目前深度人脸识别的网络模型可分为两类,一类是深度卷积神经网络,通过设计更复杂的网络模块提升网络的特征提取能力,但急剧增加的模型参数和计算量使其无法应用于实时人脸识别任务。另一类则是轻量级网络,设计轻量级架构降低模型参数数量和计算量,同时保持良好的识别性能,提升现实场景的应用能力。本文将重点关注基于卷积神经网络的深度人脸识别方法,并依据不同的网络模型对深度人脸识别方法进行归纳总结。

2.1. 基于 AlexNet 的人脸识别方法

自 AlexNet 被提出并在 ImageNet 分类竞赛中获得第一,卷积神经网络受到计算机视觉领域的关注。AlexNet 分别在两个 GPU 上进行同等操作,每个卷积层都包含卷积、池化和激活,使用 ReLu 激活函数替代 Sigmoid 并引入 dropout 策略使神经元随机失活,保持网络的稀疏性,防止过拟合。受 AlexNet 模型的启发,Deepface 被提出,首次使用卷积神经网络提取人脸特征,使用 9 层 CNN 结构提升了人脸验证性能。DeepID2 通过 4 层卷积加池化操作提取人脸特征。这些深度人脸识别方法借助卷积神经网络极大的提升了人脸识别性能,使得基于深度学习的人脸识别方逐渐取代了传统基于图像处理的人脸识别方法。

2.2. 基于 VGGNet 的人脸识别方法

AlexNet 网络深度有限,一个自然的想法就是增大网络深度,进一步提取图像的高维特征,于是 VGGNet 被提出,其将卷积核大小减小为 3×3 以减少参数量,虽然卷积核大小减少会导致小的感受野,但通过堆叠多个 3×3 卷积可以获得相当于 5×5 或 7×7 的感受野,更多的非线性映射可以增加网络的拟合能力。为进一步提取高层的人脸特征,为了使网络更好的学习中间层特征,使深层网络更容易优化,Yi Sun 等人将 VGGNet 最后几层的卷积改为局部卷积,将全局权值共享改为局部区域权值共享,同时在多个中间卷积层分支出的全连接层中添加监督信号。虽然通过增加网络深度可以一定程度的提升网络的特征提取能力,但也会使参数量和计算量的大幅增加,同时也会增大过拟合的风险。

2.3. 基于 GoogleNet 的人脸识别方法

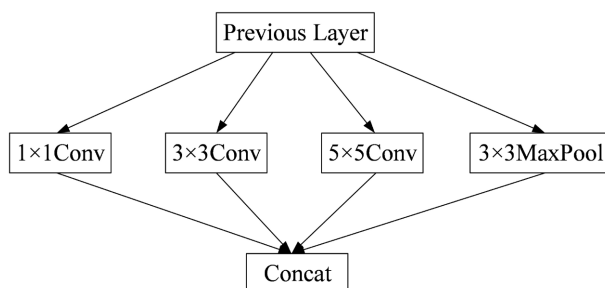


Figure 2. Example of inception structure
图 2. Inception 结构示例

GoogleNet 通过设计一种 Inception 模块化结构来保持神经网络的稀疏性, 使用稀疏连接代替全连接和卷积, 缓解了过拟合和计算复杂度过高的问题, Inception 结构如图 2 所示, 在同一层中同时部署多个卷积和池化操作, 将每个经过卷积和池化后的特征图进行通道拼接。每个 Inception 结构中包含不同大小的卷积尺寸, 图 2 仅给出一种常用的卷积尺寸。受 GoogleNet 启发, Hana Ben Fredj 等人分别在第 2、3、4 个池化层之后加入 2、5、2 个 Inception 结构, 同时还使用了一些数据增强技术, 在 LFW 和 YTF 数据集上提升了人脸识别性能。Inception 的结构相比于堆叠卷积的方式减少了参数和计算量, 同时多尺度卷积的方式有利于提取人脸图像的多尺度特征, 进而提升人脸识别性能, 但随着 Inception 模块的增多, 梯度消失或梯度爆炸的问题仍然存在。

2.4. 基于 ResNet 的人脸识别方法

在之前的深层网络优化过程中, 若每一层的误差梯度小于 1, 反向传播时网络越深梯度越趋近于 0, 导致梯度消失, 若每一层的误差梯度大于 1, 反向传播时网络越深梯度越大, 导致梯度爆炸。ResNet 提出一种残差连接方法, 缓解了网络过深导致的性能退化问题, 其中提出的残差模块结构如图 3 所示, 其在多个权重层之后添加快捷连接方式, 将经过多层特征提取之后的特征 $F(x)$ 与原始特征 x 对应相加, 保证在经过多层特征提取之后原始特征信息不会丢失。因此近年来主流的深度人脸识别方法多采用残差连接的网络结构, 实现了最先进的人脸识别性能。该残差连接方式不受模块内具体结构的影响, 所以可以很好的兼容其他网络模型。

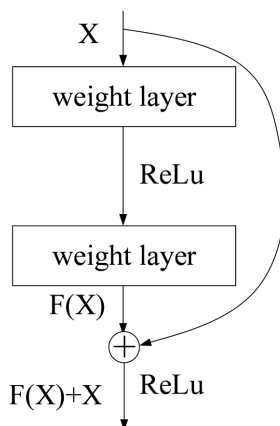


Figure 3. Structure of residual module
图 3. 残差模块结构

2.5. 基于轻量级网络的人脸识别方法

深度卷积神经网络通过堆叠卷积和池化操作提取图像的深层特征, 一方面可以提升网络的特征提取能力, 但另一方面也带来了较大的资源消耗, 大量的参数和复杂的计算量使其在移动端的应用受到限制。轻量级网络模型的提出显著降低了卷积神经网络的参数量和计算量。受这些轻量级模型启发, 一些基于轻量级网络的人脸识别方法被提出, 使得实际应用可以在移动端部署人脸识别系统。为了降低计算复杂度, MobileFaceNet [1] 使用深度可分离卷积替代普通卷积, 同时使用全局深度卷积替代全局平均池化, 突出人脸图像中心单元的特征。为进一步提升 MobileFaceNet 的速度, Mobiface [2] 使用快速下采样, 在提取特征的最开始阶段连续地应用下采样步骤, 以避免特征图的大空间维度。然后在后期添加更多的特征图, 以支撑整个网络的信息流。ShuffleFaceNet 通过使用全局深度卷积层和参数校正线性单元(PReLU)来

扩展 ShuffleNetV2, 用于实时人脸识别应用。Jintao Zhang 等[3]使用 Seesaw 模块, 用非均匀的组卷积替代逐点卷积并添加通道混洗操作实现不同卷积组之间的信息流动。Fadi Boutros 等[4]使用混合卷积, 对输入通道进行分组并将不同的内核大小应用于每个组, 并结合通道混洗操作。这些人脸识别方法通过设计轻量级的网络架构或高效的卷积方式极大的减少了模型的复杂度, 提升了人脸特征提取效率, 促进了深度人脸识别在移动端的应用。

3. 损失函数

人脸识别损失函数的设计对于网络模型的监督训练起到关键作用, 其将人脸特征映射到特征空间进行相似性比较, 根据特征空间的不同可分为基于欧几里得距离的损失函数与基于角余弦裕度的损失函数。基于欧几里得距离的损失函数将人脸特征映射到欧几里得空间, 通过比较人脸特征之间的欧几里得距离判断人脸身份。基于角余弦裕度的损失函数则将人脸特征映射到超球空间, 再使用人脸特征之间的余弦相似度判断人脸身份。

3.1. 基于欧几里得距离的损失函数

对比损失向网络输入两种类型的样本对, 即正样本对(来自同类的两张脸)和负样本对(两张来自不同类的人脸图像), 最小化正样本对的欧几里得距离, 并惩罚距离小于阈值的负样本对。假设有一对人脸特征, D_w 表示两个人脸特征的欧氏距离, Y 为两个特征是否匹配的标签, 当两个特征匹配时取 1, 否则取 0, m 为设定的距离阈值, N 为样本数量, 则对比损失如式 1 所示。对比损失引入两种监督信号, 其中识别信号使用交叉熵损失, 使不同人脸身份的特征之间有足够的时间距, 而验证信号采用对比损失, 最小化相同身份人脸特征对的欧氏距离。

$$L = \frac{1}{N} \sum_{n=1}^N Y D_w^2 + (1-Y) \max(m - D_w, 0)^2 \quad (1)$$

FaceNet 尝试使用三元组损失代替 SoftMax 损失来监督训练网络模型, 与对比损失不同的是, 三元组损失将每张人脸图像与其他人脸图像构成的每个图像对之间设置边距, 而对比损失试图将一个身份的所有人脸投射到嵌入空间的单个点上。三元组损失使用三元组 (x^a, x^p, x^n) 作为输入, 其中 x^a 表示锚点, 即能代表类中心的样本, x^p 表示正类样本, 与 x^a 属于同一人脸身份, x^n 表示负类样本, 与 x^a 属于不同人脸身份。在训练时我们选择三元组使其满足式 3 表示的不等式, 函数 f 表示特征提取过程, α 为距离阈值, 若违背式 2 则更新网络参数, 只有不满足式 2 的三元组才会对网络参数更新起到积极作用。该损失函数的目的是在训练过程中将正样本对距离拉近, 负样本对推开。三元组损失的提出有效的提升了人脸识别性能, 但在每个训练批次中选择硬三元组的方式, 增加了训练难度。

$$\|f(x^a) - f(x^p)\|_2^2 + \alpha < \|f(x^a) - f(x^n)\|_2^2 \quad (2)$$

由于对比损失和三元组损失训练难度较大, 会出现训练不稳定的情况, 中心损失被提出, 该损失通过限制类内样本与其类中心的距离使同类样本之间的距离更近, 其损失公式如式 3 所示, c_{y_i} 代表属于第 y_i 类人脸的类中心, 在每轮迭代的小批次中选择与类中心同属一个类别的人脸特征更新 c_{y_i} 。由式 3 可以看出该损失只能将相同身份的人脸特征距离拉近, 无法保证不同身份的人脸特征之间的距离, 所以该损失需要配合 Softmax 交叉熵损失一起训练。中心损失避免了挖掘硬样本, 简化训练步骤, 但该损失要求各个身份的人脸图像足够多且平衡, 保证每个类中心都被充分优化。

$$L_c = \frac{1}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \quad (3)$$

3.2. 基于角余弦裕度的损失函数

基于欧几里得距离的损失函数的本质是使同身份人脸特征更紧凑而不同身份人脸特征分布更分散, 但基于欧式距离的度量方式使得网络模型训练过程复杂且不容易收敛。为了不增加额外的损失, SphereFace [5] 基于原始 Softmax 损失进行改进, 提出了 A-Softmax Loss, 将式 1 中的 $W_j \cdot x_j$ 转换为 $\|W_j\| \cdot \|x_j\| \cdot \cos \theta_{j,j}$, $\theta_{j,j}$ 表示权值矩阵第 j 列的向量 W_j 与特征向量 x_j 之间的夹角, 同时为了便于训练将 $\|W_j\|$ 归一化为 1, 在样本与其所属类中心的夹角添加乘法裕度, 进一步扩大类间距, 损失公式如式 4 所示, 为保证余弦函数单调递减, 规定 $\theta \in [k\pi/m, (k+1)\pi/m]$, k 为 0 到 $m-1$ 之间的整数, m 即为乘法角裕度。该损失函数通过缩小特征向量与类中心的夹角使同类人脸特征分布紧凑, 同时乘法角裕度的引入使不同类的人脸分布更分散。

$$L = \frac{1}{N} \sum_i -\log \left(\frac{e^{\|x_i\| \left[(-1)^k \cos(m\theta_{y_i,i}) - 2k \right]}}{e^{\|x_i\| \left[(-1)^k \cos(m\theta_{y_i,i}) - 2k \right]} + \sum_{j \neq y_i} e^{\|x_i\| \cos \theta_{j,i}}} \right) \quad (4)$$

虽然 A-Softmax Loss 取得了先进的性能, 但特征范数的存在使得训练难度较大, 难以收敛。为进一步降低训练难度, Rajeev Ranjan 等向人脸特征添加 L2 约束, 然后通过常数因子 α 进行缩放, 限制特征位于固定半径的超球面上。CosFace [6] 提出 Large Margin Cosine Loss (LMCL) 将特征范数归一化, 人脸特征被映射到半径为 s 的超球面上, 同时使用余弦裕度扩大类间距, 相比乘法角裕度更容易优化, 损失公式如式 5 所示。ArcFace [7] 提出一种附加角裕度的损失函数, 其损失公式如式 6 所示, 可以看出与式 5 的 LMCL 具有类似的结构, 不同的只有样本特征与其正类中心的相似度的裕度表示形式, 使用附加角裕度的方式与使用乘法角裕度或加法余弦裕度相比可以获得更具辨别性的深度人脸特征, 同时具有更好的可解释性。AirFace 将 ArcFace 损失中目标 logit 的余弦函数改为角度, 对于嵌入特征尺寸较小的模型具有较好的收敛性。为了避免手动调节超参数, Xiao Zhang 等人 [8] 提出一种无需调节超参数的自适应损失函数 AdaCos, 可以在训练过程中自动加强监督。

$$L = \frac{1}{N} \sum_i -\log \left(\frac{e^{s(\cos \theta_{y_i,i} - m)}}{e^{s(\cos \theta_{y_i,i} - m)} + \sum_{j \neq y_i} e^{s \cdot \cos \theta_{j,i}}} \right) \quad (5)$$

$$L = -\frac{1}{m} \sum_{i=1}^m \log \frac{e^{s(\theta_{y_i,i} + m)}}{e^{s(\cos(\theta_{y_i,i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_{j,i}}} \quad (6)$$

最近样本挖掘的思想被融入到损失函数中, 式 7 为损失函数的基本形式, $p(x_i)$ 表示第 i 个样本 x_i 被预测正确的概率, G 为指示函数, 通常固定值为 1。 $T(\cos \theta_{y_i})$ 为正余弦相似度, 代表样本特征与其正类中心的相似度, $N(t, \cos \theta_j)$ 为负余弦相似度, 代表样本特征与其负类中心的相似度。MV-Arc-Softmax [9] 通过额外的超参数调节, 强调错误分类的样本特征, 正余弦相似度使用附加角裕度, 负余弦相似度如式 8 所示, 每轮迭代的负余弦相似度可能不同, 当 $\cos(\theta_{y_i} + m) \geq \cos \theta_j$ 时, 被认为是简单样本, 负余弦相似度保持 $\cos \theta_j$ 不变, 否则被认为是困难样本, 取参数 t 大于 1, 则 $t \cos \theta_j + t - 1$ 始终大于 $\cos \theta_j$, 困难样本的损失占比加大, 在训练阶段始终被强调, 提高了训练效率。但超参数需要手动设置, 较大的超参数可能会导致收敛问题。为了更有效的利用训练样本, CurricularFace [10] 将课程式学习融入损失函数中, 在训练前期关注简单样本, 后期关注困难样本, 其负余弦相似度的定义如式 9, 对于每轮训练所使用的参数 t 使用上一轮迭代的 t 以及当前批次所有正余弦相似度的平均值自适应调节, 避免了手动调整超参数。AdaFace [11] 提供了一种新的数据挖掘思路, 基于图像质量来强调不同难度的样本而不是仅强调错误分类

的样本, 忽略训练样本中模糊难以分辨的人脸图像。

$$L = -\frac{1}{N} \sum_{i=1}^N G(p(x_i)) \log \frac{e^{sT(\cos \theta_{y_i})}}{e^{sT(\cos \theta_{y_i})} + \sum_{j=1, j \neq y_i}^n e^{sN(t, \cos \theta_j)}} \quad (7)$$

$$N(t, \cos \theta_j) = \begin{cases} \cos \theta_j & \cos(\theta_{y_i} + m) \geq \cos \theta_j \\ t \cos \theta_j + t - 1 & \cos(\theta_{y_i} + m) < \cos \theta_j \end{cases} \quad (8)$$

$$N(t, \cos \theta_j) = \begin{cases} \cos \theta_j & \cos(\theta_{y_i} + m) \geq \cos \theta_j \\ \cos \theta_j (t + \cos \theta_j) & \cos(\theta_{y_i} + m) < \cos \theta_j \end{cases} \quad (9)$$

此外一些损失函数进一步规范样本的特征分布, UniformFace 为了使人脸类中心均匀分布在特征空间中, 施加一个等分布约束, 通过利用特征空间来最大化类中心之间的最小距离。MagFace [12] 引入了一种自适应机制, 通过将简单样本拉到类中心, 并将困难样本推开, 来学习结构良好的类内特征分布。针对数据不平衡问题, Fair Loss 提出一种公平损失, 使用自适应裕度代替固定的余弦裕度或角裕度, 结合强化学习优化公平损失。类似的, AdaptiveFace [13] 引入自适应裕度, 使模型学习每个类的特殊裕度, 以自适应地压缩类内变化。

4. 针对特定任务的人脸识别方法

除了设计网络架构以及损失函数来提升人脸识别性能, 针对如大姿态变化、遮挡、数据不平衡以及大的数据噪声等的特定任务场景, 需要设计特殊的适用性方法满足实际场景的需要。为了减少除人脸身份以外的其他信息干扰, Yu Liu 等人使用自动编码器提取身份信号并通过对抗训练减少其他因素干扰。针对人脸遮挡问题, Lingxue Song 等人提出了一种掩模学习策略, 以从识别中发现并丢弃损坏的特征元素。人脸识别常用的训练数据中各类的图像数量不平衡, 某些身份图像数量较少导致数据的长尾分布, Yandong Guo 等人在交叉熵损失中引入正则化器, 同时为样本缺少的类引入一种新的损失以平衡训练, 将样本稀少的类的权重向量的范数与正常类的权重向量对齐。Yue Wu 等人提出了一种中心不变损失, 该损失惩罚每个类中心之间的距离, 将每个人的类中心均匀分布, 以强制学习的特征对所有人具有通用表示。

通常深度人脸识别将人脸图像映射为特征空间中的确定性表示, 这对于一些低质量的模糊图像并不适用, 近年来不确定性学习被引入深度人脸识别来解决此类问题。Yichun Shi 等[14]提出了概率人脸嵌入, 它将每个人脸图像表示为潜在空间中的高斯分布, 分布的平均值估计最可能的特征值, 而方差显示特征值的不确定性。Jie Chang 等[15]将数据不确定性学习应用于人脸识别, 从而同时学习特征和不确定性, 同时提供了关于合并不确定性估计如何帮助减少噪声样本的不利影响并影响特征学习的深入分析。Shen Li 等人[16]提出了一种新的球面空间人脸特征不确定性学习框架。

5. 数据集

深度人脸识别需要大规模的数据训练网络模型, 数据集的发展也很大程度上促进了深度人脸识别技术的发展。对于训练集, CASIA 从 IMDb 网站上搜索 1940 年到 2014 年的名人信息, 通过聚类保证每个身份的图像数量不少于 15 张, 并且删除了与 LFW 数据集中同名的人脸图像, 使训练数据独立于测试数据。VGGFace 从 IMDb 和 Google 上搜集了更多的公众人物图像, 去除了与 LFW 和 YTF 身份相同的图像。UMDFaces 提供了静态图片以及视频, 同时还提供了人脸框、21 个关键点和性别注释。MS-Celeb-1M 从搜索引擎获取图片, 每个人约 20 张图片并经过微软标注。VGGFace2 包含了不同年龄、姿态、光照和种

族的人脸图像。IMDb-Face 是一个经过人工清理标签的干净人脸识别数据集, 图片质量较高。WebFace260M 是全球目前最大的人脸识别数据集, 其从互联网公开渠道收集, WebFace42M 为过滤后的干净版本。对于测试集, LFW 主要用来研究非受限情况下的人脸识别场景, 图片主要来自互联网, YTF 是从 YouTube 下载的人脸视频数据集, 主要用于视频人脸识别研究。IJB-A 包含了来自不同国家、地区和种族的人脸图像和视频, 具有广泛的地域性, IJB-B 和 IJB-C 为其扩展版本。MegaFace 是第一个百万规模级别的人脸识别算法测试基准, CFP 中每个人除了正面人脸图片还有 4 张侧面人脸图片, 测试识别对于大姿态变化的鲁棒性。AgeDB 为跨年龄人脸识别数据集, 最低年龄 1 岁, 最高年龄 101 岁, 平均年龄大概为 50 岁。CPLFW 和 CALFW 是基于 LFW 的跨姿态人脸数据集和跨年龄人脸数据集。为有效助力疫情防控中的人脸识别, 武汉大学提出了佩戴口罩的人脸数据集, RMFRD 包含真实世界佩戴口罩的人脸图像, SMFRD 包含模拟佩戴口罩的人脸图像。表 1 详细列举了人脸识别常用的公开数据集信息, 包括人脸身份数量、图片数量、视频数量以及发布时间。

Table 1. Dataset of face recognition

表 1. 人脸识别数据集

数据集	身份数量	图片数量	视频数量	时间
CASIA	10575	494414	0	2014
VGGFace	2622	2.6M	0	2015
UMDFaces	8277	367888	22075	2016
MS-Celeb-1M	100000	1980681	0	2016
VGGFace2	9131	3.31M	0	2017
IMDb-Face	590000	1.7M	0	2018
WebFace260	4M	260M	0	2021
WebFace42M	2M	42M	0	2021
LFW	5749	13233	0	2007
YTF	1595	0	3425	2011
IJB-A	500	5712	2085	2015
MegaFace	690572	4.7M	0	2016
CFP	500	7000	0	2016
CALFW	5749	13233	0	2017
AgeDB	440	12240	0	2017
IJB-B	1845	11754	7011	2017
IJB-C	28936	138000	11000	2018
CPLFW	5749	13233	0	2018
RMFRD	525	95000	0	2020
SMFRD	10000	500000	0	2020

6. 评价指标

人脸识别任务可分为验证任务以及识别任务, 验证任务即一对一匹配, 判断待检测图像与给定的图像是否属于同一个身份, 其评价指标通常为 ROC 曲线, 横坐标为接受率 TAR, 纵坐标为误识率 FAR, 其计算公式如式 10、11 所示。TP 表示被正确匹配的正图像对数量, FN 表示未被匹配的正图像对数量, TN 表示未被匹配的负图像对数量, FP 表示被错误匹配的负图像对数量。不同的相似度阈值可得到曲线上不同的 TAR/FAR 点, ROC 曲线与坐标轴构成的图形面积被称为曲线下面积 AUC, 面积越大表示识别

效果越好。实际考察要求在较低的误识率下保持较高接受率。

$$TAR = \frac{TP}{TP + FN} \quad (10)$$

$$FAR = \frac{FP}{FP + TN} \quad (11)$$

识别任务即一对多匹配, 判断待测图像身份是否属于图像库中的身份, 根据图像库中是否包含待检测图像可分为开集识别和闭集识别。开集识别中待测图像不在图像库中出现, 其也有对应的 ROC 曲线, 横坐标为真阳性率 TPIR, 表示在图像库中成功搜索出待测图像所占的比例, 纵坐标为假阳性率 FPIR, 表示待测图像不在图像库中, 但被误认为图像库中人脸身份的待测图像所占比例。rank-k 表示真实标签出现在预测结果前 k 个即算预测正确, 通过固定 rank 值, 改变相似度阈值可以得到曲线上不同的 TPIR/FPIR 点, 类似的实际考察要求在较低的假阳性率下保持较高的真阳性率。闭集识别中待测图像在图像库中, 此时评价指标为 CMC 曲线, 其横轴为 rank 值, 纵轴为正确率即正确识别的待测图像所占的比例, 要求对于 rank 值较小的情况下仍然能保持较高的正确率。

随着深度人脸识别的发展, 其研究热点从网络架构设计转移到损失函数的设计, 现阶段深度人脸识别已经达到了可靠的精度, 如何降低资源消耗、减少计算成本逐渐成为另一个研究热点。同时针对特定任务的深度人脸识别方法也在探究设计一个对遮挡、大姿态变化、光照等更鲁棒的识别框架。表 2 比较了近五年来主流的深度人脸识别方法, 主要从模型大小、训练集大小以及验证精度五个方面进行比较, 选用 6 个测试集上的验证精度进行比较, 包括 LFW、YTF、AgeDB、CFP-FP、CPLFW 和 CALFW。

Table 2. Comparison of mainstream deep facial recognition methods

表 2. 主流深度人脸识别方法比较

方法	模型大小	训练集数量	LFW	YTF	AgeDB	CFP-FP	CALFW	CPLFW
SphereFace [5]	160 M	0.49 M	99.42	95	—	—	—	—
CosFace [6]	160 M	5 M	99.73	97.6	—	—	—	—
ArcFace [7]	250 M	3.8 M	99.83	—	98.08	94.04	—	—
AdaptiveFace [13]	125 M	5 M	99.62	—	—	—	—	—
AdaCos [8]	125 M	0.45 M	99.71	—	—	—	—	—
MobileFaceNets [1]	4 M	3.8 M	99.55	—	—	—	—	—
Mobiface [2]	9.3 M	3.8 M	99.72	—	—	—	—	—
SeesawFaceNets [3]	—	5.8 M	99.80	—	97.55	97.24	95.98	91.98
MixFaceNets [4]	—	5.8 M	99.60	—	96.63	—	—	—
PFE [14]	170 M	4.4 M	99.82	97.36	—	93.34	—	—
DUL [15]	50 M	3.6 M	99.76	96.38	—	97.11	—	—
SCF [16]	260 M	5.8 M	99.82	—	98.3	98.59	96.18	93.26
MV-Arc-Softmax [9]	—	3.9 M	99.79	—	98	95.3	95.36	89.19
CurricularFace [10]	250 M	5.8 M	99.80	—	98.32	98.37	96.2	93.13
MagFace [12]	250 M	5.8 M	99.83	—	98.17	98.46	96.15	92.87
AdaFace [11]	250 M	4.2 M	99.80	—	97.9	99.17	96.05	94.63

7. 结束语

本文对深度人脸识别方法进行了总结和梳理。目前人脸识别技术已经成熟, 但同样面临着许多挑战:

- 1) 面对各种面部变化如大姿态、运动模糊、低照度、低分辨率、遮挡等情况下的识别率较低。
- 2) 大规模

训练数据不可避免的存在噪声干扰。3) 随着人脸身份的增加, 计算成本增大, 训练周期变长。4) 训练数据不平衡, 长尾数据的人脸表征效果较差。5) 手动标记的成本较大, 未标记的人脸图像未被充分利用。

针对深度人脸识别当前面临的挑战提出未来的发展方向: 1) 设计对面部变化鲁棒的人脸识别方法, 更好的应用于实际场景。2) 设计轻量级模型和更高效的模型训练方法, 使现有的深度人脸识别方法能够应用到实际场景。3) 设计合理利用不平衡数据的训练方法, 减少长尾数据的干扰。4) 设计无监督或半监督的人脸识别方法, 充分利用未标记的人脸图像。

基金项目

国家自然科学基金(61872190); 江苏省博士后科研资助计划项目(2020Z058)。

参考文献

- [1] Chen, S., Liu, Y., Gao, X. and Han, Z. (2018) MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices. In: Zhou, J., *et al.* Eds., *CCBR 2018: Biometric Recognition*, Springer, Cham, 428-438. https://doi.org/10.1007/978-3-319-97909-0_46
- [2] Duong, C.N., Quach, K.G., Jalata, I., Le, N. and Luu, K. (2019) MobiFace: A Lightweight Deep Learning Face recognition on Mobile Devices. 2019 *IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Tampa, 23-26 September 2019, 1-6. <https://doi.org/10.1109/BTAS46853.2019.9185981>
- [3] Zhang, J. (2019) Seesawfacenets: Sparse and Robust Face Verification Model for Mobile Platform. <https://arxiv.org/abs/1908.09124>
- [4] Boutros, F., Damer, N., Fang, M., Kirchbuchner, F. and Kuijper, A. (2021) MixFaceNets: Extremely Efficient Face Recognition Networks. 2021 *IEEE International Joint Conference on Biometrics (IJCB)*, Shenzhen, 4-7 August 2021, 1-8. <https://doi.org/10.1109/IJCB52358.2021.9484374>
- [5] Liu, W., Wen, Y., Yu, Z., *et al.* (2017) SphereFace: Deep Hypersphere Embedding for Face Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6738-6746. <https://doi.org/10.1109/CVPR.2017.713>
- [6] Wang, H., Wang, Y., Zhou, Z., *et al.* (2018) CosFace: Large Margin Cosine Loss for Deep Face Recognition. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 5265-5274. <https://doi.org/10.1109/CVPR.2018.00552>
- [7] Deng, J., Guo, J., Xue, N. and Zafeiriou, S. (2019) ArcFace: Additive Angular Margin Loss for Deep Face recognition. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 4 4685-4694. <https://doi.org/10.1109/CVPR.2019.00482>
- [8] Zhang, X., Zhao, R., Qiao, Y., Wang, X.G. and Li, H.S. (2019) AdaCos: Adaptively Scaling Cosine Logits for Effectively Learning Deep Face Representations. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 10815-10824. <https://doi.org/10.1109/CVPR.2019.01108>
- [9] Wang, X., Zhang, S., Wang, S., *et al.* (2020) Mis-Classified Vector Guided Softmax Loss for Face Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 12241-12248. <https://doi.org/10.1609/aaai.v34i07.6906>
- [10] Huang, Y., Wang, Y., Tai, Y., *et al.* (2020) Curricularface: Adaptive Curriculum Learning Loss for Deep Face recognition. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 5900-5909. <https://doi.org/10.1109/CVPR42600.2020.00594>
- [11] Kim, M., Jain, A.K. and Liu, X. (2022) AdaFace: Quality Adaptive Margin for Face Recognition. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 18729-18738. <https://doi.org/10.1109/CVPR52688.2022.01819>
- [12] Meng, Q., Zhao, S., Huang, Z. and Zhou, F. (2021) MagFace: A Universal Representation for Face Recognition and Quality Assessment. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 14220-14229. <https://doi.org/10.1109/CVPR46437.2021.01400>
- [13] Liu, H., Zhu, X., Lei, Z. and Li, S.Z. (2019) AdaptiveFace: Adaptive Margin and Sampling for Face Recognition. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 11939-11948. <https://doi.org/10.1109/CVPR.2019.01222>
- [14] Shi, Y. and Jain, A.K. (2019) Probabilistic Face Embeddings. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October 2019-2 November 2019, 6901-6910. <https://doi.org/10.1109/ICCV.2019.00700>
- [15] Chang, J., Lan, Z., Cheng, C. and Wei, Y.C. (2020) Data Uncertainty Learning in Face Recognition. 2020 *IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 13-19 June 2020, 5709-5718.
<https://doi.org/10.1109/CVPR42600.2020.00575>

- [16] Li, S., Xu, J., Xu, X., *et al.* (2021) Spherical Confidence Learning for Face Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 15629-15637.