基于深度卷积神经网络的人脸识别技术综述

景晨凯[1] 宋 涛[1] 庄 雷[1] 刘 刚[2] 王 乐[2] 刘凯伦[1]

[1](郑州大学信息工程学院 河南 郑州 450001)

[2](河南省招生办公室 河南 郑州 450046)

摘 要 人脸识别是计算机视觉的重要应用之一,广义的人脸识别包含图像采集、人脸检测、人脸对齐、特征表示等过程。人脸识别的发展史主要是人脸特征表示方法的变迁史。针对特征的表示方法,从人脸识别技术的发展历史、研究现状和未来发展三个方面进行综述:分阶段对传统的几类经典的人脸识别算法进行回顾和总结;以深度学习算法的诞生过程为切入点,重点分析了在人脸识别中取得突破性进展的深度卷积神经网络 DCNN(deep convolutional neural networks)的技术思想和关键问题;针对人脸识别和深度学习算法的重大挑战,展望了未来可能存在的发展方向。

关键词 人脸识别 特征表示 深度学习 深度卷积神经网络

中图分类号 TP391.41 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2018.01.039

A SURVEY OF FACE RECOGNITION TECHNOLOGY BASED ON DEEP CONVOLUTIONAL NEURAL NETWORKS

Jing Chenkai[1] Song Tao[1] Zhuang Lei[1] Liu Gang[2] Wang Le[2] Liu Kailun[1]

[1](*School of Information Engineering, Zhengzhou University, Zhengzhou 450001, Henan, China*)

[2](*Higher Education Admission Office of Henan, Zhengzhou 450046, Henan, China*)

Abstract Face recognition is one of the important applications of computer vision. Generalized face recognition includes image acquisition, face detection, face alignment, and feature representation and so on. However, the development history of face recognition is mainly the history of the change of face feature representation method and summarizes the three aspects of the development history, research status and future development of face recognition technology. Firstly, reviews and summarizes several kinds of classical phases of traditional face recognition algorithm. Secondly, based on the process of the depth learning algorithm, the technical ideas and key problems of deep convolutional neural networks (DCNN), which are the breakthrough progress in face recognition, are analyzed emphatically. Based on this, the paper finally talks about the prospect of face recognition in the direction of development may exist in the future under the challenge of face recognition and deep learning algorithms.

Keywords Face recognition Feature representation Deep learning DCNN

0 引 言

人脸识别属于计算机视觉的范畴,特指计算机利用分析比较人脸视觉特征信息自动进行身份鉴别的“智能”技术。相比于指纹、虹膜等传统生物识别手段,人脸识别具有无接触、符合人类识别习惯、交互性强、不易盗取等优势,因此在保障公共安全,信息安全,金融安全,公司和个人财产安全上等有强烈的需求。尤其是近些年来随着深度卷积神经网络 DCNN 的引入,人脸识别的准确率得以跨越式提升,各类相关应用如人脸识别考勤、考生身份验证、刷脸支付、人脸归类查询等已开始逐步投入使用,效果显著。

对于一张静态 2D 人脸图片,影响识别的因素主

收稿日期:2017-03-01。国家自然科学基金项目(61379079);河南省国际合作项目(152102410021)。景晨凯,硕士生,主研领域:深度学习,人脸识别。宋涛,博士生。庄雷,教授。刘刚,博士。王乐,硕士生。刘凯伦,硕士生。

要可分为基础因素、外在因素和内在因素。基础因素是指人脸本身具有的全局相似性,即人的五官、轮廓大致相同;外在因素是指成像质量,主要包括人与摄像设备的位置关系(距离、角度、尺度),摄影器材的性能,光照的强弱,外物(眼镜、围巾)遮挡等;内在因素是指个人的内部属性,如性别、年龄变化、精神健康状态、面部毛发、化妆整容、意外损伤等。但人类似乎天生具有面部识别能力,可以很轻松地剔除掉这些因素的影响[1],并且可以通过人的其他部位、神情、习惯等辅助手段快速确定一个人。而对于计算机,这些辅助手段由于其特征不稳定性反而容易被不法分子利用。目前大多数情况下计算机读取的是一张静态二维图片,这些图片本质上又是由繁多的多维数字矩阵组成,如一张 256 × 256 的 RGB 彩色图片就有 196 608 个数字。可想而知,其识别任务难度巨大。幸运的是计算机可以使用人类设计的算法从图像中提取特征或者学习到特征。计算机自动人脸识别 AFR(automatic face recognition)的主要任务就是如何为减少个人内部的变化,同时扩大人外部差异制定低维有效的特征表示。

1 回顾传统的人脸识别算法

人脸识别的发展史主要上还是人脸特征表示方法的变迁史,从最初的几何特征,到经验驱动的“人造特征”,最后到数据驱动的“表示学习”,人脸识别已历经了近 60 年的发展历程。

英国心理学家 Galton 于 1888 年和 1920 年便在《Nature》上发表了两篇关于人脸识别的论文,他将不同人脸的侧面特征用一组数字代表,但并未涉及 AFR 问题。1965 年,Bledsoe 等在 Panoramic Research Inc 上发表了第一篇 AFR 的报告[2],他们用脸部器官间的间距(如两眼之间)、比率等参数作为特征,构建了一个半自动人脸识别系统,开始了真正意义上的人脸识别研究。1965 年至 1990 年的人脸识别研究主要基于几何结构特征的方法以及基于模板匹配的方法。基于几何结构特征的方法一般通过提取人眼、口、鼻等重要特征点的位置,以及眼睛等重要器官的直观几何形状作为分类特征,计算量小。但当受光照变化、外物遮挡、面部表情变化等内外在因素影响时,所需特征点将无法精确定位,进而造成特征急剧变化。而基于模板匹配的方法则通过计算模板和图像灰度的自相关性来实现识别功能,但忽略了局部特征,造成部分信息丢失。这一阶段可以称为人脸识别的初级阶段,该阶段的研究只适用于人脸图像的粗略识别,无法在实际中应用。1992 年 Brunelli 等通过实验得出基于模板匹配的方法优于基于几何结构特征的方法[3]的结论。

1991 年到 1997 年是人脸识别研究的第二阶段,尽管时间短暂,却是非常重要的时期。大量的人力物力投入其中,如美国国防部发起的 FERET(Face Recognition Technology Test)资助了多项人脸识别研究,并创建了著名的 FERET 人脸图像数据库,该项目极大地促进了人脸识别算法的改进及实用化,许多经典的人脸识别算法也都在这个阶段产生。具有里程碑意义的研究是麻省理工学院的 Turk 等提出特征脸 Eigenface[4],该方法是后来其他大多数算法的基准。还有基于子空间分析的人脸识别算法 Fisherface[5],它首先通过主成分分析方法 PCA(Principal Component Analysis)[4]将人脸降维,之后采用线性判别分析 LDA(Linear Discriminant Analysis)[5]期望获得类间差异大且类内差异小的线性子空间,但正因如此,它无法对复杂的非线性模型进行建模。基于弹性图匹配的方法[6]是一种将几何特征与对灰度分布信息的小波纹理分析相结合的识别算法,它利用人脸的基准特征点构造拓扑图,使其符合人脸的几何特征,然后获取人脸关键点的特征值进行匹配。该算法能够在局部结构的基础上保留全局结构,而且能自动定位面部特征点,因此对角度变化具有一定的鲁棒性。其缺点是时间复杂度高,实现复杂。基于模型的方法如主动表观模型 AAMs(Active Appearance Models)[7]是人脸建模方面的一个重要贡献。AAMs 将人脸图像的形状和纹理分别用统计的方法进行描述,然后通过 PCA 将二者融合来对人脸进行统计建模,该算法常用在人脸对齐上。另外比较经典的还有 SVD 分解[8]、人脸等密度线分析匹配[9]、隐马尔可夫模型(Hidden Markov Model)[10]以及神经网络等方法。总的来说,这一阶段的人脸识别研究发展迅速,所提出的算法直接采用人脸图像中所有像素的颜色或灰度值作为初始特征,然后通过在训练数据上学习得到更具区分力的人脸表示。从技术方案上看,2D 人脸图像线性子空间判别分析、统计模式识别方法是这一阶段的主流技术。这一阶段的人脸识别系统在较理想图像采集条件、用户配合、中小规模数据库上的情况下较为适用。

第三阶段(1998 年—2013 年)重点研究真实条件下,以及基于其他的数据源(如视频、近红外和素描)的人脸识别问题,并深入分析和研究不同影响下的人脸识别,如光照不变人脸识别、姿态不变人脸识别和表情不变人脸识别等。为了克服直接使用像素灰度值对光照敏感等问题的限制,这一时期涌现出了很多对局部邻域像素亮度或颜色值进行手工特征提取的方法,比如对人

脸较为有效 Gabor Face、LBP Face[11] 以及基于无监督学习的特征 learning Descriptors[12] 等。分类识别上主要采用以线性判别分析为代表的线性建模方法[13-14]，以核方法为代表的非线性建模方法[15-16] 和基于 3D 人脸重建的人脸识别方法[17-18]。LBP 特征是这一时期的典型特征描述子，其将图像分成若干区域，在每个区域用中心值对邻域作阈值化，将结果表示成二进制数，然后基于区域的频率直方图做统计。LBP 特征对单调灰度变化保持不变，并对图像中的噪声和姿态具有一定的鲁棒性。在子空间分析改进上，如针对 Eigenface 算法的缺点，中科院计算所提出的特定人脸子空间(FSS)算法[13]，FSS 为每个对象建立一个私有的人脸子空间，更好地描述了不同个体人脸之间的差异性。香港中文大学的王晓刚等提出的统一子空间分析[14] 方法将 PCA、LDA 和贝叶斯子空间[19] 三种子空间方法进行比较，并将三者有机结合提高了识别性能。基于 3D 人脸重建的人脸识别一般基于形变模型(morphable model)[18]，其主要思想是首先将 2D 人脸图像映射到 3D 模型表面，之后将这个 3D 模型转到正脸提取特征。虽然对姿态变化具有鲁棒性，但需要定位大量基准点，并且 3D 数据难以收集。值得一提的是 2007 年以后，LFW[20] 数据库成为真实条件下最权威的人脸识别测试基准。它的样本来自互联网的 5 749 人的 13 233 张名人人脸照片，采用十折平均精度作为性能评价指标。2012 年 Huang 等首次采用深度学习的无监督的特征学习方法[21] 在 LFW 取得了 87% 的识别率，与当时最好的传统人脸识别算法相比还有一定差距。总之，这一阶段提取的面部特征是人为设计或基于无监督学习的局部描述子。此后以 DCNN 为代表的深度学习算法的有监督学习在 AFR 的应用彻底颠覆了这种经验驱动的“人造特征”范式，开启了数据驱动的“表示学习”的革命。

2　深度学习革命下的人脸识别研究

2006 年，Hinton 等在《Science》上首次提出了深度学习的概念[22]。深度学习本质上也是一种特征学习方法，传统方法需要有相关专业背景的专家设计特征表示方式，而深度学习各层的特征是使用一种通用的学习过程从数据中学到的。其也可以看作是使用像素灰度值特征，它把原始数据通过一些简单的非线性的模型转变成为更高层次的、更加抽象的表达，经过足够多转换的组合来学习非常复杂的函数。2012 年，Hinton 又带领学生在目前最大的图像数据库 ImageNet[23] 上，将 Top5 的分类错误率 26% 降低至 15%，在学术界一鸣惊人，并引起了工业界的强烈关注，特别是以谷歌、百度、微软、脸谱等为首的拥有大量数据和高性能计算的科技巨头企业。深度学习俨然已成为当今人工智能界具有统治地位的算法，而深度学习前身就是 NN。由此，本节先从人工智能和 NN 的起源开始逐步深入分析这一算法。

2.1　深度学习的前世今生

1956 年，John McCarthy 与 Marvin Minsky，Herbert Simon 等在达特茅斯学院正式创立了人工智能的概念，并形成以 Herbert Simon 为代表的理性学派和以 Marvin Minsky 为代表的感性学派。NN 正是感性学派的代表。1957 年康奈尔大学心理学教授 Rosenblatt 利用神经网络原理首次成功制作了能够读入并识别简单的字母和图像电子感知机。1959 年，霍普金斯大学的 Hubel 和 Wiesel 通过观察猫脑部视觉中枢对视网膜进入图像的处理方式发现，提出了简单细胞和复杂细胞的概念。这一工作对后来从事 NN 研究的计算机专家提供了重要的建模思路，比如神经元是分工分层对信息进行处理，不同神经元关注的对象特征不同。CNN 中的卷积和池化层灵感也直接来源于视觉神经科学中的简单细胞和复杂细胞。对于人脸图像来说，前几层的神经元抽象出脸部的部分特征如边角或线条，然后经过逐层激发逐渐形成不同的形状，如眼睛和鼻子，最后在中枢的最高层激发对整个对象产生认知的“祖母神经元”，也就是整张人脸的特征。

但好景不长，1969 年 Minsky 在《感知机》的书中证明两层神经网络不能解决 XOR(异或)这一个基本逻辑问题直接导致了 NN 研究经历了第一次长达十几年的寒冬。这一时期理性学派的专家系统得以盛行，感性流派虽没有专家系统那样成功，但也取得了一些进步。如 1974 年，哈佛 Werbos 的博士论文证明在输入层和输出层之间添加一个隐层，可以解决 XOR 问题，但并未引起重视，另外层数的增加为各个层的神经节点连接的权重选取带来新的困难。1986 年 Rumelhart 等在《nature》提出的反向传播 BP(back propagation)算法[24] 一定程度上解决了权重选取问题。多层感知机和 BP 算法为 NN 研究点燃了新的希望，在此基础上分支联结主义开始流行，其核心领导者是两位心理学家 Rumelhart 和 McLelland 和未来的“深度学习之父” Hinton。但是很快由于多层网络训练困难：如梯度不稳定，训练数据和计算能力不足等问题，NN 在 20 世纪末再次进入寒冬。值得关注的是，在此期间专家系统及 NN 维度的深化推动了超级计算技术的发展。这一领域衍生出的计算机集群技术成为 20 世纪 90 年代信息领域的互联网公司的计算平台，业务量和数据量

的增加使这些网络平台不断扩张，存储和计算能力相应越来越强大，由此也产生了大量的数据，为 NN 的第三次复苏埋下伏笔。

2006 年，NN 脱胎换骨成为深度学习，Hinton 等所提出的深度信念网络 DBN（deep belief networks）[22] 指出具有大量隐层的网络具有优异的特征学习能力，而网络的训练可以采用非监督的逐层初始化与反向传播实现。2012 年机器学习界的泰斗 Andrew Ng 等发起的 Google Brain 项目在包含 16 000 个 CPU 的分布式并行计算平台上构建一种被称为“深度神经网络”的类脑学习模型 [25]，并成功地“认识”了猫。而近些年 GPU 强大的并行计算能力更是加快了训练速度，深度学习势如破竹。深度学习能取得如今的成就，离不开三个长期专注 NN 领域的计算机科学家，分别是以上提到的深度学习开创者 Geoffrey Hinton、CNN 的重要研究与发扬者 Yann LeCun 以及加拿大蒙特利尔大学教授 Yoshua Bengio。而 DCNN 是深度学习算法的一种，目前主要在计算机视觉领域取得突破进展。

2.2 DCNN 算法及其在人脸识别中的应用

1979 年日本京都大学的 Fukushima 基于感受野概念提出了神经认知机来进行手写字母的图像识别，这可以看作是 CNN 的第一个实现网络，也是感受野概念在神经网络领域的首次应用。1989 年 LeCun 选择将 BP 算法用于训练多层卷积神经网络来识别手写数字 [26]，这是 CNN 概念提出的最早文献。但是建立起现代卷积网络学科的开创性论文是 1998 年 LeCun 提出的 LeNet-5 [27]，并且 LeCun 认为 CNN 不应看作是生物学上的神经系统原型，因此他更倾向于称其为卷积网络，并把网络中的节点称为单元。尽管如此，卷积网络由于使用了与许多神经网络相同的思想。因此，本文遵循惯例，把它看作是神经网络的一种类型。

2014 年，脸谱的团队 [30] 和香港中文大学的团队 [31] 在 LFW 上分别报告了 97.35% 和 97.45% 的平均分类精度，人脸识别的主要技术路线由人工设计特征与分类识别转变为基于 DCNN 的端到端的自主学习特征。2015 年 Google 的 FaceNet [34] 在 LFW 数据集上平均分类精度达到 99.63%，基本上宣告了在 LFW 上 8 年性能竞赛的结束。DCNN 同样使用 BP 算法进行有监督的学习，因此在卷积核中的权值都能得到训练。BP 算法是训练深度网络的核心算法，其利用链式求导法则求解目标函数关于多层神经网络权值梯度。巧妙之处在于目标函数对于某层输入的梯度可以通过向后传播对该层输出的导数求得，它首先从最高层的输出一直到最底层的输入计算目标函数对每层输入的导数（残差）然后一次性地求解每一层残差对权值 w 和偏置 b 的梯度。

总结 BP 算法的一般形式如下：

（1）输入 x，为输入层设置对应的激活值 h^1；

（2）前向传播：对每层 $l = 1,2,\cdots,l$，计算相应的 $z^l = w^l z^{l-1} + b, h^l = f(z^l)$；

（3）计算输出层误差：计算向量 $\delta^l = \nabla_h L \odot f(z^l)$；

（4）反向误差传播：对每层 $l = l-1, l-2, l-3, \cdots, 2$，计算 $\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot f(z^l)$；

（5）输出：代价函数对 w 和 b 的梯度，$\frac{\partial L}{\partial w_{jk}^l} = h_k^{l-1}\delta_j^l$，$\frac{\partial L}{\partial w_j^l} = \delta_j^l$。其中 w_{jk}^l 表示 l 层第 j 个神经元与 $l-1$ 层第 k 个神经元之间的连接权重。

DCNN 被设计用来处理图像等多维数据，其用了 4 个关键思想来利用自然信号的属性：局部连接、权值共享、池化以及多网络层，与人工设计的特征（LBP 等）不同，其能够端到端地自主学习到具有高层次、抽象的特征表达向量。一般情况下卷积层后面都紧随有一个非线性激活层，如图 1、图 2 所示。图 1 最左侧是 $l-1$ 层的输出，同样也是 l 层的输入，是一个单通道的 5×5 的特征映射图，l 层有一个 3×3 的卷积核 w 和一个偏置 b，卷积核从 l 层特征映射图的左上方以步长为 1 滑动，依次与对应局部位置求加权和，并与偏置 b 相加后得到线性输出 z，继续传入非线性激活函数 $f(x)$。图 2 中例子为 ReLU [28-29]，最终得到 l 层的 3×3 的非线性输出。一个特征图的各个局部共享一个卷积核，使用不同的卷积核形成新的不同的特征映射图。使用这种局部连接、权值共享的结构基于两方面的原因：一方面是对于人脸等图像，一个像素与周围的像素经常是高度相关的，能够形成有区分性的局部特征；另一方面是自然图像有其固有特性，一部分的统计特性与其它部分是相关的，在一个位置出现的特征也可能出现在别的位置。

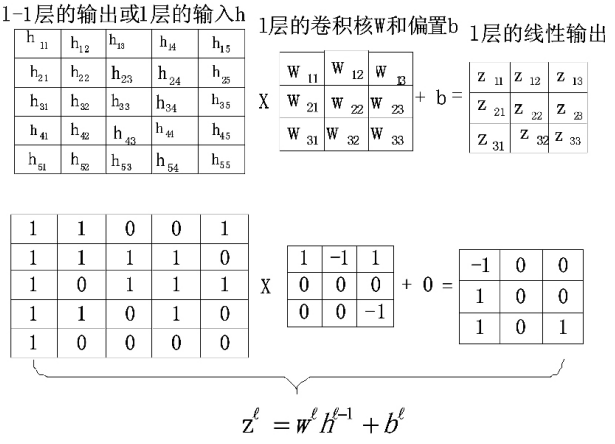

图 1 卷积层运算实例

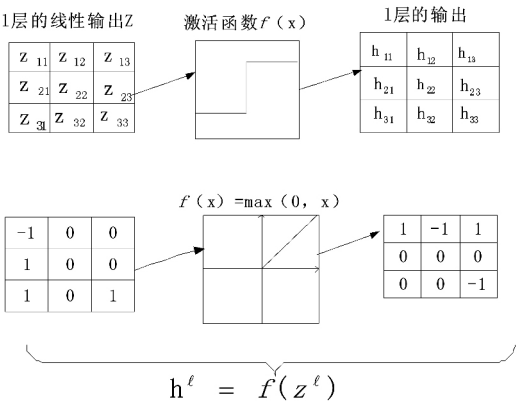

图 2　激活函数层运算实例

卷积层用来探测特征图的局部连接,池化层则在语义上把相似的特征进行融合,池化也具有平移不变性,大量经验验证,加入池化层能够提升识别率。常见的池化方式有:平均池化(取局部平均值),最小池化(取局部最小值),最大池化(取局部最大值)等。如图 3 所示是最大池化操作,池化单元计算特征图中的一个局部块(图 3 中的尺寸大小是 2 ×2)的最大值,池化单元通过移动一行或者一列(图 3 步长为 1)最终提取出一个 2 ×2 的特征图(图 3 右侧)。卷积层和池化层除了以上所述的优点外,还有一个直接原因就是它们大大降低了可训练参数的同时也降低了特征图的维度。对于图 1,如果是全连接层,则需要学习 5 ×5 + 1 =26 个参数,而对于一个卷积核来说,则只需要学习 3 ×3 +1 =10 个参数,并最终使一个 5 ×5 的特征图转化成一个 2 ×2 的特征图。对于输入的多维人脸,随着深度的增加,卷积与池化的层层叠加,神经元的数目也相应的减少,并最终形成一个特定的、紧凑的、低维度、全局性的人脸特征表达向量(一般是倒数第二层的隐藏层)用于人脸识别(通过 knn 分类器等),人脸验证(计算距离)等任务。

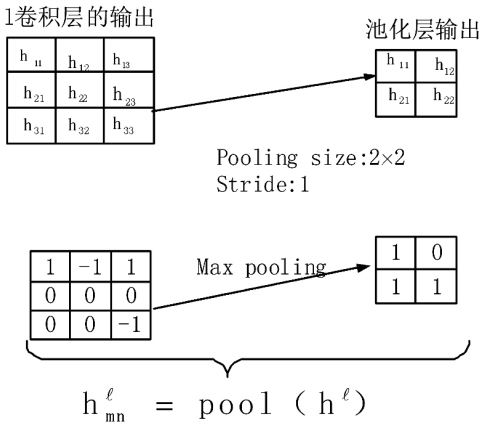

图 3　池化层运算实例

如表 1 中列举了近些年几种比较成功的基于 DCNN 的人脸识别模型及在 LFW 上的测试情况,DCNN 作为一种特征提取器,在人脸识别中的主要目的还是通过 DCNN 自动学习到更具区分力的人脸特征表达进而具有更强的泛化能力。这主要通过两方面来提升:1) 通过表达能力更强的网络结构;2) 通过更有效的损失估计。在网络结构上,DeepFace [30] ,DeepID [31] 和 VGGFace 均使用了直线型网络结构, Deepface 后面三层采用了参数不共享的卷积核,但导致了参数的膨胀; DeepId 系列 [31-33] 则将卷积层的输出与上一层的池化层的输出进行融合来增强特征表达;FaceNet 则采用了 inception [35] 局部多分支型网络结构同时融合了多尺度的特征,并采用 1 ×1 的卷积核减少训练参数数量。从表 1 中也可以看出 DCNN 的网络结构正在变大变深: VGGFace16 层、FaceNet22 层。2015 年的 ResNet [36] 已经达到 152 层;更深的网络意味着更加强大的特征抽象能力,但同时也意味着训练难度的加大,训练参数的增多和计算效率的下降,文献 [37] 通过对卷积核的有效分解等操作在不明显增加参数和降低计算效率的情况下提升了网络的特征表达能力。

表 1　几种经典的 DCNN 模型在 LFW 数据集上的测试结果

名称	年份	网络数量	训练集数据量/万	LFW 十则平均精度/%
DeepFace	2014	3	700	97.35
DeepID	2014	25	20	97.45
DeepID2	2014	25	20	99.15
DeepID2 +	2015	25	45	99.47
VGGFace	2015	1	2 600	98.95
FaceNet	2015	1	20 000	99.63

在损失估计上,有效的损失计算不但能加快网络的训练,而且有利于学习到更强有力的人脸特征表达,在早期 DeepFace 及 DeepId 中直接采用了 softmax 分类器(人脸分类信号)作为损失计算。这种情况下一般需要人脸的类别数达到一定数量(万人),并且每一个人的样本数也应足够的多(数十甚至上百)的情况下较为适用。分类的数目越多,对应的样本数越多,学习到人脸特征的区分性和泛化性就越好。但是当人脸类别数少且类内样本不足的情况下,采用度量学习的判别式学习方法是必要的。在 DeepId2 则同时采用人脸确认损失和人脸分类损失作为监督信号进行联合深度学习,使用联合信号使类内变化达到最小,并使不同类别的人脸图像间距保持恒定 [38] ,其验证信号仅考虑了一对样本的误差;在 FaceNet 中则直接放弃了 softmax 分类损失,采用 Triplet Loss 作为损失函数,通过构建三元组,将最近负样本距离的大于最远正样本的距离作为目标函数,使最终的特征表示不需要额外训练模型进行分类,人脸验证只需直接计算倒数第二层隐层

输出的 128 维向量的距离即可,简单有效。从最初的单一的多分类器到度量学习到引入,再到仅需要度量学习便可学习优异的特征,这种转变也直接反映出了度量学习对于人脸特征学习所起到的关键作用。但是值得注意的是,在度量学习中样本对的选择是一个不可回避的重要问题,不恰当的选择策略将很有可能引起过拟合问题。

另外还有其他的提升人脸特征表达能力的方法。如在 deepid 系列中也通过将人脸分割多个区域、尺度,对人脸做镜像和反转等作为输入形成互补和数据增强。deepface 则在三维人脸图像对齐后再输入到网络便于提取更有效的特征。还有不得不重视的是 DCNN 作为一种非常适合大数据的算法,更多的数据依然能够带来更鲁棒,更具抽象能力的特征。从表 1 也看出了这一趋势: DeepID 系列从 20 万到 45 万,DeepFace 的 700 万, VGGFace 有 2 600 万,FaceNet 则达到 2 个亿。

为了解释分析 DCNN 内部神经元的特性,在 deepid2 + [33] 中研究发现通过 DCNN 学习得到的高层次的人脸特征是中度稀疏的、对人脸身份和人脸属性有很强的选择性(特定的神经元对特定的属性会有持续的响应和抑制)、对局部遮挡具有良好的鲁棒性,不过本文对此目前仍抱有怀疑态度,有待今后更深入的研究成果去证明。

2.3 应用 DCNN 算法的障碍

虽然 DCNN 目前已经在人脸识别以及其他的计算机视觉任务中得以成功应用并有成为一种通用的 AI 算法之势,但应用 DCNN 算法本身依然是一个不小的挑战,也可以说是一个主要问题,本文主要划分为以下四点:

1) 有监督的学习,需要大量的标记样本 从目前的发展状况来看,有监督的学习已经远远盖过了无监督学习的风头。而训练深度网络需要大量的数据,尽管网络上有大量的数据,但都杂乱无章,需要人工标注,并且近些年的标注成本也是水涨船高。

2) 理论研究不足 深度学习包括 DCNN 是一个端到端的学习。神经网络,反向传播算法,卷积神经网络等基本的方法原理早已存在,近些年的发展也主要得益于大数据,高性能计算以及各种网络结构和训练方法的改进,而实际上却并无深层次的本质理解,大量的研究思路简单粗糙。因此在设计 DCNN 结构以及在训练当中经常碰到的过拟合问题、梯度不稳定的问题,除了遵循一些基本原则,更多需要通过经验和直觉来进行,这种试验性的研究思路增加了运用难度。如表 2 所示。

表 2 训练 DCNN 模型的建议

方法	具体操作
数据增强	图像的反转,随机剪切,多尺度,颜色渲染等
预处理	归一化数据等
初始化	使用 gaussion/Xavier [39] 等小的随机数初始化卷积核,卷积核或者池化层的大小或步长,是否做 fine-ture
超参数	学习率(一般为 0.1),动量(0.9),批处理大小等
激活函数	如经典的 Sigmoid,tanh,目前常用的 ReLu 及其变形,leaky ReLu,Parametric Relu,Randomized ReLU
正则化	L2 regularization, L1 regularization,Dropout,Batch normalization [40] 等
看图修正	训练过程中,从误差下降的曲线,以及在验证集上的表现,及时地调整如学习率等参数,进而加快训练速度,并且也能尽量避免过拟合的现象

3) 局部最优解 由于深度学习算法需要学习的目标函数是非凸的,存在着大量的局部最小值。而训练中用的梯度下降算法,理论情况下会很容易停留在一个局部最小值上面。并且如果初始值的不同,即使是同样的训练集也会朝着不同的方向优化,这就给最终的结果带来了很大的不确定性。不过大量实践证明,对于非小网络,这个问题并不会引起太大的麻烦。

4) 训练时间长且计算资源代价不菲 深度学习由于参数较多,相比其他机器学习算法训练周期要长很多,近些年来其能够流行的一大因素离不开 GPU 的发展。GPU 成倍加快了训练速度,但是这些 GPU 售价昂贵,建立大规模的 GPU 集群并非一般院校能够负担。使用浮点计算的深度网络要求大存储空间和大计算量,使其在手机、移动机器人等设备上的应用大大受阻。

3 人脸识别的未来之路

3.1 更具挑战的人脸数据集

LFW 作为前些年来最流行的人脸测试数据集,识别率频频被刷新,如香港中文大学的 DeepID2 +, Google 的 FaceNet 在 2015 年均取得了 99% 以上的识别率,这基本宣告了 LFW 竞争之战的结束。在 LFW 上的刷分已然没有太大意义,但现有脸部识别系统仍难以准确识别超过百万的数据量。因此,未来急需更多更具挑战的公开人脸数据集。这些数据集首先应当满足大规模,标签准确等基本条件,可以是针对特定任务(如特定的年龄层或特定的场景等)的数据集,也可

以是综合(如包括各个年龄层或者各类复杂场景等)的数据集。2015 年华盛顿大学为了研究当数据集规模提升数个量级时,现有的脸部识别系统能否继续维持可靠的识别率,发起了一个名为 "MegaFace Challenge"的公开竞赛,MegaFace 数据集有 690 572 个体 1 027 060张公开人脸图像[41],难度颇大,对大规模数据的人脸识别起到了促进作用。

3.2　特定问题的深入研究

影响人脸识别的诸如光照、姿态、年龄、遮挡等问题并没有得到根本解决。对特定问题的研究有助于整体人脸识别研究的进步。在 CVPR2016 上,就有许多关于人脸识别特定问题的研究工作,例如南加州大学的 Masi 关注了人脸识别中的大姿态变化问题。与当前大部分利用大量数据训练单一模型或者矫正人脸到正脸来学习姿态不变性的方法不同,该作者通过使用五个指定角度模型和渲染人脸图片的方法处理姿态变化[42]。中科院计算所 Kan 等通过尝试移除人脸数据之间的跨模态差异性,并寻找跨模态之间的非线性的差异性和模态不变性表达解决人脸识别中的跨视图或跨姿态问题[43]。还有意大利特伦托大学做了人脸老龄化预测的有关工作[44],这对跨越年龄的人脸识别具有很大的参考意义。

3.3　新型有效的网络结构和训练方法

生物神经系统的连接极为复杂,既有自下而上的前馈和同层递归,又有自上而下的反馈和来自其他神经子系统的外部连接,目前的深度模型尚未对这些建模。去年 MSRA 的 ResNet 达到了惊人的 152 层,解决了极深网络在增加层数的同时也能保持准确率的问题,也证明了极深网络在其他任务中也有很好的泛化性能。而芝加哥大学的 Gustav 提出了一个不依赖于残差的极深架构 FractalNet[45],作者称该分形结构可以自动容纳过去已有的强大结构。但是需要明白,这些网络结构本身也是人为设计,哪个网络结构最佳,卷积层的数量多少才合适,我们不得而知。近期的网络剪枝,网络简化等工作对此进行了探讨[46-47],并认为稀疏性对于卷积神经网络应用于人脸识别效果有提升,但该研究还处于起步阶段。

另外,DCNN 早在 20 世纪 80 年代就已经基本成型,当时未能普及的原因之一,就是缺少高效地优化多层网络的方法,如对多层神经网络进行初始化的有效方法。尽管有 Mini-Batch SGD、ResNet 中的 shortcut、ReLU 激活函数、Batch Normalization 等促进表达能力和加快收敛的方法。但对此仍然缺乏一个完善的理论作指导。对于人脸识别,深度度量学习(deep metric learning)是一个最常用的方法,更好的目标函数能够学到更具有区分力的特征。如上文提到的 DeepFace 和 DeepID 的 contrastive loss 度量,Facenet 的 triplet loss 度量等都有用到 deep metric learning 的方法。最近的如在 CVPR2016 斯坦福大学提出利用训练批处理中所有相同标签的人脸对和不同标签的人脸对的信息进行语义特征映射,来减少同类间距离同时增加异类间距离[48]。

3.4　其他的学习算法

在使用 DCNN 训练出的模型时可以发现,在某个数据集上表现好的模型在另外一个数据集结果可能并不如意,比如使用东方人训练出的模型去识别西方人的人脸,或者反之。这种训练数据和应用数据之间的偏差便可通过迁移学习进行消除,简而言之,如果这两个领域之间有某种联系、某种相似性,就只需小部分数据在新的领域中重新学习即可。中科院 Kan 等提出的对于人脸识别的领域自适应学习[49]做了相关的工作。

强化学习相对深度学习更古老,但由于计算瓶颈使它长时间处于静默状态,不能处理大数据。但 2015 年 Google 的 DeepMind 把深度学习和强化学习相结合,隐藏了很多强化学习的状态个数,这种隐藏使得强化学习能够应付大数据,强化学习比 DCNN 在图像上面的应用更加复杂,更加契合人的行为。

大量有标签数据是 DCNN 的局限性之一,无监督学习在人类和动物的学习中却占据主导地位,但目前几乎所有由人工智能创造的经济价值都来自监督学习。CNN 虽然与神经认知架构有点相似,但是在神经认知中并不需要类似 BP 算法这种端到端的监督学习算法。并且获取大量无监督数据的成本相比有标签数据微乎其微。各方面讲,无监督学习都是未来的趋势,代表了人工智能的一种关键技能。但直接从大量的无监督数据中学习确实非常困难,也许少量有监督数据与大量无监督数据结合的半监督学习是现阶段需要重点研究的方向。

另外还有如增量学习、终生学习、对抗学习、注意力模型等都是未来可能应用在人脸识别甚至影响整个人工智能领域。

4　结　语

AFR 经过几十年的研究发展,已经逐渐成为一个

成熟的研究领域。DCNN 的到来，为这个领域注入了新的活力，并取得了显著的效果，甚至说在某些数据集上已经超越人类，但是否真的超越，还言之过早。对于实际应用中的光照、抖动、模糊、遮挡、分辨率、姿态等的外在因素或性别、年龄变化、精神健康状态、面部毛发、化妆整容、意外损伤等内在因素依然没有得到完全解决。对于深度学习算法的内在原理，甚至还无从知晓，本质上仍然是弱人工智能。两者的结合是历史的必然，但未来的发展还需要计算机视觉研究者们的共同努力。

参 考 文 献

[1] 山世光. 人脸识别中若干关键问题的研究 [D]. 中国科学院研究生院(计算技术研究所),2004.

[2] Bledsoe W W. Man-machine facial recognition [J]. Rep. PRi,1966,22.

[3] Brunelli R,Poggio T. Face recognition:Features versus templates [J]. IEEE transactions on pattern analysis and machine intelligence,1993,15(10):1042-1052.

[4] Turk M,Pentland A. Eigenfaces for recognition [J]. Journal of cognitive neuroscience,1991,3(1):71-86.

[5] 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,1997,19(7):711-720.

[6] Lades M,Vorbruggen J C,Buhmann J,et al. Distortion invariant object recognition in the dynamic link architecture [J]. IEEE Transactions on computers,1993,42(3):300-311.

[7] Qin H,Yan J,Li X,et al. Joint training of cascaded CNN for face detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:3456-3465.

[8] Hong Z Q. Algebraic feature extraction of image for recognition [J]. Pattern recognition,1991,24(3):211-219.

[9] Nakamura O,Mathur S,Minami T. Identification of human faces based on isodensity maps [J]. Pattern Recognition,1991,24(3):263-272.

[10] Samaria F,Young S. HMM-based architecture for face identification [J]. Image and vision computing,1994,12(8):537-543.

[11] Chen D,Cao X,Wen F,et al. Blessing of dimensionality:High-dimensional feature and its efficient compression for face verification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2013:3025-3032.

[12] Winder S A J,Brown M. Learning local image descriptors [C]//Computer Vision and Pattern Recognition,2007. CVPR07. IEEE Conference on. IEEE,2007:1-8.

[13] Shan S,Gao W,Zhao D. Face identification from a single example image based on face-specific subspace (FSS) [C]//Acoustics,Speech,and Signal Processing (ICASSP),2002 IEEE International Conference on. IEEE,2002,2:II-2125-II-2128.

[14] Wang X,Tang X. A unified framework for subspace face recognition [J]. IEEE Transactions on pattern analysis and machine intelligence,2004,26(9):1222-1228.

[15] Yang M H. Kernel Eigenfaces vs. Kernel Fisherfaces:Face Recognition Using Kernel Methods [C]//IEEE International Conference on Automatic Face and Gesture Recognition,2002. Proceedings. IEEE,2002:215-220.

[16] Zhou S K,Chellappa R. Multiple-exemplar discriminant analysis for face recognition [C]//Pattern Recognition,2004. ICPR 2004. Proceedings of the 17th International Conference on. IEEE,2004,4:191-194.

[17] Blanz V,Vetter T. A morphable model for the synthesis of 3D faces [C]//Proceedings of the 26th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co.,1999:187-194.

[18] Blanz V,Vetter T. Face recognition based on fitting a 3D morphable model [J]. IEEE Transactions on pattern analysis and machine intelligence,2003,25(9):1063-1074.

[19] Moghaddam B,Jebara T,Pentland A. Bayesian face recognition [J]. Pattern Recognition,2000,33(11):1771-1782.

[20] Huang G B,Ramesh M,Berg T,et al. Labeled faces in the wild:A database for studying face recognition in unconstrained environments [R]. Technical Report 07-49,University of Massachusetts,Amherst,2007.

[21] Huang G B,Lee H,Learned-Miller E. Learning hierarchical representations for face verification with convolutional deep belief networks [C]//Computer Vision and Pattern Recognition (CVPR),2012 IEEE Conference on. IEEE,2012:2518-2525.

[22] Hinton G E,Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science,2006,313(5786):504-507.

[23] Deng Jia,Dong Wei,Socher R,et al. Imagenet:A large-scale hierarchical image database [C]//Computer Vision and Pattern Recognition,2009. CVPR 2009. IEEE Conference on. IEEE,2009:248-255.

[24] Rumelhart D E,Hinton G E,Williams R J. Learning representations by back-propagating errors [J]. Cognitive model-

ing,1988,5(3):1.

[25] Markoff J. How many computers to identify a cat? 16 000 [N]. New York Times,2012-06-25.

[26] LeCun Y,Boser B,Denker J S,et al. Backpropagation applied to handwritten zip code recognition [J]. Neural computation, 1989,1(4):541-551.

[27] LeCun Y,Bottou L,Bengio Y,et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE,1998,86(11):2278-2324.

[28] Nair V,Hinton G E. Rectified linear units improve restricted boltzmann machines [C] //Proceedings of the 27th international conference on machine learning (ICML - 10). 2010: 807-814.

[29] Glorot X,Bordes A,Bengio Y. Deep Sparse Rectifier Neural Networks [C] //International Conference on Artificial Intelligence and Statistics,2012.

[30] 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.

[31] Sun Yi,Wang Xiaogang,Tang Xiaoou. Deep learning face representation from predicting 10 000 classes [C] //Proc of the IEEE Conference on Computer Vision and Pattern Recognition,2014:1891-1898.

[32] Sun Yi,Chen Yuheng,Wang Xiaogang,et al. Deep learning face representation by joint identification-verification [C] // Advances in Neural Information Proc Systems, 2014: 1988-1996.

[33] Sun Yi,Wang Xiaogang,Tang Xiaoou. Deeply learned face representations are sparse,selective,and robust [C] //Proc of the IEEE Conference on Computer Vision and Pattern Recognition,2015:2892-2900.

[34] Schroff F,Kalenichenko D,Philbin J. Facenet:A unified embedding for face recognition and clustering [C] //Proc of the IEEE Conference on Computer Vision and Pattern Recognition,2015:815-823.

[35] Szegedy C,Liu W,Jia Y,et al. Going deeper with convolutions [C] //Proc of the IEEE Conference on Computer Vision and Pattern Recognition,2015:1-9.

[36] He Kaiming,Zhang Xiangyu,Ren Shaoqing,et al. Deep residual learning for image recognition [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:770-778.

[37] Szegedy C,Vanhoucke V,Ioffe S,et al. Rethinking the inception architecture for computer vision [C] //2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016:2818-2826.

[38] 王晓刚,孙祎,汤晓鸥. 从统一子空间分析到联合深度学习:人脸识别的十年历程 [J]. 中国计算机学会通讯, 2015,11(4):8-15.

[39] Glorot X,Bengio Y. Understanding the difficulty of training deep feedforward neural networks [J]. Journal of Machine Learning Research,2010,9:249-256.

[40] Ioffe S,Szegedy C. Batch normalization:Accelerating deep network training by reducing internal covariate shift [C] // Proceedings of the 32nd International Conference on Machine Learning,Lille,France,2015.

[41] Kemelmachershlizerman I,Seitz S M,Miller D,et al. The MegaFace Benchmark:1 Million Faces for Recognition at Scale [C] //Computer Vision and Pattern Recognition. IEEE, 2016:4873-4882.

[42] Masi I,Rawls S,Medioni G,et al. Pose-Aware Face Recognition in the Wild [C] //Proc of the IEEE Conference on Computer Vision and Pattern Recognition,2016:4838-4846.

[43] Kan M,Shan S,Chen X. Multi-view Deep Network for Cross-View Classification [C] //IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2016:4847-4855.

[44] Wang W,Cui Z,Yan Y,et al. Recurrent Face Aging [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society,2016:2378-2386.

[45] Larsson G,Maire M,Shakhnarovich G. FractalNet:Ultra-Deep Neural Networks without Residuals [J]. arXiv preprint arXiv: 1605.07648,2016.

[46] Sun Y,Wang X,Tang X. Sparsifying Neural Network Connections for Face Recognition [J]. Computer Science,2015: 4856-4864.

[47] Han S,Pool J,Tran J,et al. Learning both weights and connections for efficient neural network [C] //Advances in Neural Information Proc Systems,2015:1135-1143.

[48] Song H O,Xiang Y,Jegelka S,et al. Deep metric learning via lifted structured feature embedding [J]. arXiv preprint arXiv: 1511.06452,2015.

[49] Kan Meina,Wu Junting,Shan Shiguang,et al. Domain Adaptation for Face Recognition:Targetize Source Domain Bridged by Common Subspace [J]. International Journal of Computer Vision,2014,109(1-2):94-109.

(上接第 216 页)

[27] You Q,Jin H,Wang Z,et al. Image captioning with semantic attention [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:4651-4659.

[28] Tang S,Han S. Generate Image Descriptions based on Deep RNN and Memory Cells for Images Features [DB]. arXiv preprint arXiv:1602.01895,2016.