

· 综 述 ·

基于深度学习的人脸分析研究进展

严 严, 陈日伟, 王菡子*

(厦门大学 信息科学与技术学院, 福建省智慧城市感知与计算重点实验室, 福建 厦门 361005)

摘要: 近年来, 基于深度学习的人脸分析取得了巨大的进步, 成为计算机视觉领域最为活跃的研究方向之一. 为了进一步推动深度学习和人脸分析的研究, 结合近年已发表的相关文献, 对基于深度学习的人脸分析技术进行综述. 首先, 简要概述深度学习及其发展历史, 并分析深度学习有效性原因. 然后, 按照任务目的的不同, 将人脸分析分成了人脸检测、人脸关键点检测、人脸识别、人脸属性识别等任务进行详细的介绍和讨论, 重点分析各种任务现阶段存在的主要问题. 接着, 介绍人脸分析中常用的人脸数据库. 最后, 讨论深度学习和人脸分析面临的主要挑战, 并给出结论.

关键词: 深度学习; 卷积神经网络; 人脸数据库; 人脸识别; 人脸分析;

中图分类号: TP 391

文献标志码: A

文章编号: 0438-0479(2017)01-0013-12

计算机视觉(computer vision, CV)是一门研究如何利用计算机模拟人类视觉的科学, 其主要任务是通过采集到的图像(或视频)进行分析和理解, 从而做出判断或决策. 在过去几十年间, CV 取得了巨大的进步和发展. 其中, 人脸分析由于其重要的理论意义和巨大的实际应用价值, 一直以来都是 CV 领域的热点. 其通过自动检测(或跟踪)图像(或视频)中的人脸, 进而对检测和跟踪到的人脸提取相关信息, 比如身份、表情、年龄、性别等.

如何从图像(或视频)中获取有效的人脸表征一直以来都是人脸分析的一个核心问题. 各种有效的特征, 如局部二值模式特征^[1]、尺度不变特征变换特征^[2]、梯度方向直方图特征^[3]等, 被人为构建出来进行人脸的表征, 并结合各种降维和分类方法^[4-6]进行人脸分析, 取得了较好的性能. 然而, 这些人为设计的特征通常无法捕获不同任务的高层语义信息, 成为制约人脸分析性能进一步提高的瓶颈. 此外, 传统的人脸分析方法通常将任务分解为两个独立的步骤(即特征提取和分类器设计). 而在第 1 个步骤中所挑选的特征可能无法跟第 2 个步骤中的分类算法很好匹配, 从而造成人脸分析方法性能的下降.

作为近年来一种流行的机器学习方法, 深度学

习^[7-12]已经在 CV、语音识别、自然语言处理等领域中得到了广泛的研究和应用. 目前, 针对人脸分析的研究, 基于深度学习的方法已经成为主流研究方向. 相比于传统的人脸分析方法, 基于深度学习的方法能够自动地从海量数据中学习人脸表征. 同时, 该类方法把特征提取和分类器统一在一个框架下, 能够快速适应不同的人脸分析任务, 有效地提高了方法的性能.

国内外众多的大学和研究机构, 都对基于深度学习的人脸分析进行了广泛而深入的研究. 早期的深度学习算法(即早期的神经网络算法)主要应用于人脸检测^[13-14]、人脸识别^[15]等任务. 而目前的深度学习算法, 不仅仅应用于人脸检测和人脸识别, 在各种人脸分析任务中, 如人脸关键点检测^[16-17]、人脸属性识别^[18-19]、人脸重构^[20]等也得到更为广泛的应用.

目前现有的综述文献主要针对深度学习技术^[7-12]或者针对人脸分析某一类具体任务(如人脸检测^[4]、人脸识别^[5-6]等), 因此有必要对基于深度学习的人脸分析研究现状进行分析和总结, 期望能够更好地指导未来的研究工作. 本文将简要介绍深度学习的发展历史, 并分析其有效性原因, 重点评述基于深度学习的人脸分析研究新进展、现阶段常用的人脸数据库和面临的主要挑战, 最后给出结论.

收稿日期: 2016-09-11 录用日期: 2016-10-18

基金项目: 国家自然科学基金(61571379, 61472334)

* 通信作者: hanzhi.wang@xmu.edu.cn

引文格式: 严严, 陈日伟, 王菡子. 基于深度学习的人脸分析研究进展[J]. 厦门大学学报(自然科学版), 2017, 56(1): 13-24.

Citation: YAN Y, CHEN R W, WANG H Z. Recent advances on deep-learning-based face analysis[J]. J Xiamen Univ Nat Sci, 2017, 56(1): 13-24. (in Chinese)



1 深度学习

本节主要介绍深度学习的概述、发展历史以及分析深度学习有效性原因。

1.1 概述

通常认为,深度学习是指具有多层非线性转换函数的网络结构^[9]。而深度神经网络(deep neural network, DNN)是一种特殊的深度学习模型,也是目前大多数深度学习方法的主要实现手段,它通过学习深层的非线性神经网络结构,实现任意复杂函数的逼近。为了从大规模数据中建立有效模型,DNN自动学习多层信息表示,高层的特征利用低层的特征进行构建,并且以一种级联的方式逐层构成深度结构。因此,DNN能够自动学习从底层特征到高层特征的组合,从而有效减少对人工提取特征的依赖。

1.2 发展历史

深度学习的历史最早可以追溯到20世纪40年代。概括来讲,深度学习经历了3次发展浪潮^[9]:20世纪40年代到60年代,称为控制论主义;20世纪80年代到90年代,称为连接主义;2006年之后,称之为深度学习。

第1次发展:代表性的算法为人工神经网络,由于其借鉴了生物大脑的方式,因此称之为神经网络。感知机模型^[21]是第一个被提出的神经网络模型,用于区分两种类别。原始的感知机的连接权重需要手工设定,后来发展为可以自动学习。同时期的模型,还包括自适应线性单元(adaptive linear neuron, ADALINE)用于预测实数值。该时期的学习算法深刻地影响了现在的机器学习领域。比如ADALINE算法采用的权重修正训练方法实际上就是现在普遍采用的随机梯度下降法的一个特例。然而,由于感知机模型是一种线性模型,当它被发现不能解决著名的异或(XOR)函数问题^[22]时,神经网络陷入了第1次衰落期。

第2次发展:神经网络的第2次发展,是随着连接主义的浪潮和分布式处理技术的流行而发展起来的。连接主义是基于简单的计算单元,通过网络连接的形式实现复杂的功能,其等价于在神经元中,引入隐藏层单元。该时期提出的很多概念至今依然有着重要的作用。比如,分布式表示的思想,其通过组合不同的特征表示进行学习并利用反向传播算法(backpropagation, BP)训练神经网络。第2次深度学习发展浪潮持续到了20世纪90年代中期。但是由于当时对神经网络

抱有过的预期,所以当实际效果没有达到预期的时候,引起了人们的质疑。与此同时,其他浅层的机器学习算法,包括核机器学习和图模型等快速发展,在许多计算机视觉任务中都表现出良好的性能。上述两个原因导致了这一阶段神经网络的衰落。但是需要指出的是,在这个时期,神经网络依然在一些领域取得了不错的成绩,比如卷积神经网络(convolutional neural network, CNN)^[23]被提出并用于手写体识别等问题中。

第3次发展:2006年,Hinton等^[24]提出了逐层贪婪预训练的方法来高效训练神经网络,有效地解决了一直以来多层深度网络难以训练的问题。在这一阶段,人们开始使用深度学习来表示多层神经网络。这一阶段发展初期,人们集中研究无监督深度模型。而今天,深度学习的研究者更多研究基于海量的数据并利用有监督的深度学习方法进一步提高机器学习的性能。目前,常见的深度学习模型包括:基于限制玻尔兹曼机的深度模型^[25],基于自编码器(autoencoder, AE)的深度模型^[26],基于卷积的深度模型^[23]以及基于递归的深度模型^[27-29]等。

1.3 有效性原因

从深度学习发展历史可以看出,深度学习并不是一种全新的模型。常见的DNN的组成结构在20世纪90年代就已经被提出。这些年,DNN又重新引起了学术界和工业界的关注。总的来说,深度学习在各种CV任务中的有效性原因可以大致归结为以下3点:

1) 大规模数据的使用。例如,在AlexNet^[30]中使用的ImageNet^[31]图像分类数据库包含了1 000类不同类别图像,共计超过100万张训练图像。

2) 计算水平的提高。现在计算机的计算能力不断增强。特别是图形处理器(graphics processing unit, GPU)计算能力的提高,使得如今能够较为快速地训练一个DNN。此外,大规模计算机集群的出现和发展也为深度学习的推进提供了巨大动力。

3) 网络训练方式的发展。比如,上述1.2节提到的逐层贪婪预训练方法^[24]。还有,如今常用的丢包方法^[32]可以减少深度模型中存在的过拟合风险。另外,相比较于传统的Sigmoid和Tanh激活函数,新的激活函数,如修正线性单元^[33]以及其改进方法参数化修正线性单元^[34],能够极大地提高网络训练收敛的速度。

2 基于深度学习的人脸分析研究进展

本节主要介绍基于深度学习的人脸分析研究进

展.按照任务目的的不同,本文中人脸分析研究分成了人脸检测、人脸关键点检测、人脸识别、人脸属性识别以及其他人脸分析任务展开综述.表 1 列出了各种人脸分析任务及其代表性算法.

表 1 各种人脸分析任务及其代表性算法

Tab. 1 Different types of face analysis tasks and their representative algorithms

人脸分析任务	代表性算法
人脸检测	1998 年,基于 BP 神经网络 ^[13] 2001 年,基于 CGM 的新型神经网络 ^[14] 2004 年,基于 CNN ^[35] 2014 年,基于 CNN 的多任务人脸检测 ^[36] 2015 年,基于级联 CNN ^[37] 2015 年,基于 CNN 的人脸部件检测 ^[38] 2015 年,结合 Adaboost 和 CNN ^[39]
人脸关键点检测	2013 年,基于级联 CNN ^[16] 2013 年,基于粗到细级联 CNN ^[40] 2014 年,基于 AE 的 CFAN ^[17] 2014 年,基于多任务 CNN ^[41] 2016 年,结合级联 CNN 回归器和 3DMM ^[42] 2016 年,基于 CNN 的 3DDFA ^[43] 2016 年,基于 DRN 和 DA 的 DRDA ^[44]
人脸识别	1997 年,基于 SOM 和 CNN ^[15] 2014 年,基于深度 CNN 的 DeepFace ^[45] 2014 年,基于深度 CNN 的 DeepID ^[46] 和 Deep ID2 ^[47] 2015 年,基于深度 CNN 的 DeepID2+ ^[48] 2015 年,基于深度 CNN 的 FaceNet ^[49]
人脸属性识别	2014 年,基于 Boost 的 DBN 人脸表情识别 ^[19] 2015 年,基于 CNN 的性别识别和年龄分类 ^[18] 2015 年,基于 CNN 的年龄估计 ^[50] 2015 年,基于 DEX 的年龄估计 ^[51] 2016 年,基于多模态深度网络的视频表情识别 ^[52]
其他	2012 年,基于 DBN 的人脸解析 ^[53] 2013 年,基于 CNN 的人脸重构 ^[20] 2014 年,基于 CNN 的人脸反欺诈 ^[54] 2015 年,基于 CNN 的亲属验证识别 ^[55] 2016 年,基于对冲的目标跟踪 ^[56]

注:CGM 表示限制生成模型;Adaboost 表示自适应增强算法;CFAN 表示由粗到细自编码网络;3DMM 表示三维形变模型;3DDFA 表示三维稠密人脸对齐;DRN 表示深度回归网络;DA 表示去遮挡自编码器;SOM 表示自组织映射;DBN 表示深度置信网络;DEX 表示深度期望.

2.1 人脸检测

人脸检测是指给定图像(或视频),寻找出图像或视频中所有人脸的位置、大小和姿态.人脸检测是各类人脸分析任务的基础,其检测精度直接影响着后继任务的性能.但是由于人脸存在的剧烈内在变化,如表

情、姿态等变化,以及外在环境影响,如光照、遮挡等,造成人脸检测一直以来在人脸分析中都是一个非常具有挑战性的问题.

早在 1998 年,Rowley 等^[13]首先提出了利用 BP 神经网络进行人脸检测,并基于重抽样自举技术选择非人脸图像,有效地解决了手工选择非人脸图像的难度.Feraund 等^[14]提出了利用基于 CGM 的新型神经网络用于人脸检测中,具有快速检测人脸的优点.Garcia 等^[35]提出了基于 CNN 的人脸检测方法,该方法直接将整幅图像作为输入,让 CNN 自动从数据中学习有效的人脸表征,并且能够检测不同姿态的人脸图像.然而,受限于当时的计算能力和有限的数量,这些方法很快就被后来 Viola 等^[57]提出的基于 Adaboost 的人脸检测方法取代.

2012 年,Krizhevsky 等^[30]在图像分类任务中首次验证了 CNN 的有效性后,基于 CNN 的人脸检测方法研究又开始得到重视.Li 等^[37]提出了基于级联 CNN 的人脸检测方法,其通过在多个尺度的图像上快速地过滤掉非人脸图像区域,并对剩下的区域进行精细分类来提高人脸检测的性能.Zhang 等^[36]提出了一种基于 CNN 的多任务人脸检测方法,其通过构造人脸姿态估计和人脸关键点检测两个辅助任务的共同学习来提高人脸检测的性能.Yang 等^[38]提出了一种基于 CNN 的人脸部件检测,并基于检测的各个部件及其对应的关系建立人脸模型.该方法对严重遮挡和有姿态变化的人脸图像,均取得了较好的检测结果.然而,该方法却无法处理低分辨率的人脸检测问题.

目前,传统的基于 Adaboost 的人脸检测方法依然具有明显的速度优势.而基于深度学习的方法可以取得非常好的性能表现(比如在人脸检测数据评测集 FDDB 上,传统方法只能达到 85% 的准确率,而深度学习方法^[58-60]已超过 95% 的准确率).因此,许多方法试图结合传统方法和深度学习.例如,Zhan 等^[39]提出了结合 Adaboost 和 CNN 的人脸检测方法.

2.2 人脸关键点检测

人脸关键点检测也称为人脸对齐,是指给定人脸图像,定位出人脸面部的关键区域位置,包括眉毛、眼睛、鼻子、嘴巴、脸部轮廓等.和人脸检测类似,由于受到姿态和遮挡等因素的影响,人脸关键点检测也是一个富有挑战性的任务.

传统的人脸关键点检测方法包括主动形状模型^[61]和主动外观模型^[62],以及后继发展出来的一些方法(比如,基于限制局部模型的方法^[63],基于主动外观模型的方法^[64]和基于回归的方法^[65]等).

与人脸检测、人脸识别等任务相比,基于神经网络的方法直到最近才被应用到人脸关键点检测任务中。比如,Sun 等^[16]首先提出了利用 CNN 对人脸的 5 个关键点(包括左眼、右眼、鼻尖、左嘴角、右嘴角)进行回归,并用不同级别的 CNN 微调关键点组合,取得了较好的性能。但是该方法无法有效处理遮挡问题。Zhang 等^[17]提出了 CFAN 方法,采用级联的方式逐步进行人脸关键点检测。由于采用多个 CNN 结构,因此上述方法检测的人脸关键数目均不多(≤ 6 个)。Zhou 等^[40]采用级联 CNN 进行人脸关键点检测(共 68 个关键点),将人脸关键点检测分为 4 个级别的网络,每个网络分别回归不同位置的人脸关键点。Zhang 等^[41]使用基于多任务的 CNN(包括人脸关键点检测、人脸姿态估计和人脸属性分类等任务),利用多个任务共同学习来提高人脸关键点检测的准确率。

目前,人脸关键点检测的关键是解决大规模姿态变化以及遮挡等因素所带来的检测性能的下降。比如,为了解决大规模人脸姿态变化,Jourabloo 等^[42]结合级联 CNN 回归器和 3DMM 进行人脸对齐,该方法把人脸对齐建模为 3DMM 拟合问题,利用级联 CNN 估计摄像机投影矩阵和三维形状参数,但由于采用了 3DMM 模型使得算法整体复杂度过高。Zhu 等^[43]提出了一个 3DDFA 框架,利用 CNN 把三维人脸模型拟合到人脸图像上,有效处理大规模人脸姿态变化(如侧脸)时的人脸对齐问题。Zhang 等^[44]基于级联的结构,将多个 DRN 和多个 DA 耦合起来进行人脸关键点检测,该方法能有效处理遮挡问题,并且可以对遮挡的人脸进行重构。

2.3 人脸识别

人脸识别的研究历史悠久,一直以来都是 CV 领域一个研究热点方向。人脸识别问题可以描述为^[6]:输入(查询)场景中的图像或视频,使用人脸数据库辨识或验证场景中的一个人或者多个人。

早在 1997 年,Lawrence 等^[15]就提出了基于 SOM 和 CNN 的人脸识别方法,该方法通过 SOM 保持输入输出空间的邻域结构,利用 CNN 逐层自动学习特征,并在 40 个人的人脸图像库中,取得了较好的效果。但是由于训练数据有限,限制了该方法的推广。在深度学习方法流行之前,人脸识别的研究^[5-6]主要集中在特征提取(如 Gabor 小波特征、局部二值模式特征等)、降维(如主成分分析、鉴别成分分析等子空间学习方法)和分类器设计(如最近邻、 k 最近邻等)等方面。

2014 年,Facebook 公司的 Taigman 等^[45]提出了

基于深度 CNN 的人脸验证方法,即 DeepFace,该方法利用 4 000 个人,每个人超过 1 000 张人脸图像的大数据进行训练,并利用 3D 模型对齐人脸图像,用于训练 9 层的神经网络。在户外标记人脸数据库(labeled faces in the wild, LFW)人脸验证测试集中,取得了 97.25% 的准确率,首次接近人类水平。从 2014 年开始,香港中文大学多媒体实验室。Sun 等^[46-48]先后提出了一系列深度学习人脸识别算法。他们首先提出了深度隐藏身份特征(deep hidden identity features, DeepID)^[46]算法,利用深度 CNN 训练超过 1 万人的脸分类器,再将训练完后的最后一层全连接层输出视为提取的人脸特征,并结合联合贝叶斯方法进行人脸验证。在 LFW 测试集上,DeepID 算法取得了 97.45% 的准确率。之后他们又提出了 DeepID2 算法^[47]和 DeepID2+ 算法^[48]。相比于 DeepID 算法,DeepID2 和 DeepID2+ 在训练网络的过程中,使用了人脸验证和人脸分类 2 种信号进行网络训练。在 LFW 测试集中,这 2 种算法分别取得了 99.15% 和 99.47% 的精度。2015 年,谷歌公司的 Schroff 等^[49]提出了 FaceNet 算法。该算法利用三元组损失函数进行网络训练以直接学习人脸特征,并将这些特征用于人脸验证和人脸聚类中,在 LFW 测试集上,该算法取得了 99.63% 的精度。

目前,基于深度学习的人脸识别算法在 LFW 测试集上取得了非常优异的成绩,甚至超过了人类的水平。但是最新研究表明^[66],当数据库中存在大量的干扰人脸(即非查询人脸)时,人脸辨识率和验证率仍然比较低,尤其是在姿态和年龄变化情况下识别性能下降严重。

2.4 人脸属性识别

与人脸识别不同,人脸属性识别试图分析人脸特有的属性,包括性别、年龄、表情、种族、肤色、发型等属性,是一类重要的人脸分析任务。与早期的人脸识别研究类似,早期的人脸属性识别研究也主要集中在特征提取、降维和分类器设计等方面。

近年来,Levi 等^[18]提出了利用 CNN 分别进行人脸性别识别和人脸年龄分类,显著地提高了人脸性别和年龄的识别率。Liu 等^[19]提出了基于 boost 的 DBN,并用于人脸表情识别中,有效提高了人脸表情识别的性能。Kahou 等^[52]提出了多模态深度网络用于视频表情识别,其利用深度 CNN 提取视频中的人脸特征信息,并利用 DBN 提取视频中的声音信息。Wang 等^[50]同样利用 CNN 提取多级特征进行融合并用于人脸年龄估计。Rothe 等^[51]提出了一种 DEX 算法,其利用深度 CNN 进行人脸年龄估计。

最近,有些工作是利用人脸属性辅助提高其他人脸分析任务的性能。比如,Devries 等^[67]提出了一种基于多任务学习的人脸表情识别和人脸关键点检测算法,利用表情识别辅助提高人脸关键点检测的性能。Zhang 等^[68]提出了利用人脸是否带眼镜、人脸微笑、性别、姿态等人脸属性来改进人脸关键点检测的性能。Kumar 等^[69]利用人脸属性有效地提高人脸识别的性能。合理利用人脸属性提高相关人脸分析任务的性能是值得研究的问题。比如,对人脸识别任务而言,肤色是一个对识别有利的属性,而表情可能无法辅助提高人脸识别的性能。

2.5 其他

此外,深度学习还广泛应用在人脸分析的其他领域。比如,Tang 等^[53]利用 DBN 进行人脸解析,把人脸图像分割成不同的区域,针对不同输入姿态和表情的人脸图像,Zhu 等^[20]提出了一种基于 CNN 的人脸特征提取算法,并用获取到的人脸特征重构出正面的中立人脸图像。Kan 等^[70]根据同样的思路,提出了运用堆叠自编码器(stacked autoencoder, SAE)算法进行逐步人脸正面重构。Yim 等^[71]提出了利用多任务学习进行人脸正面重构。Zhang 等^[55]提出了利用深度 CNN 进行亲属验证。Yang 等^[54]提出了利用深度 CNN 进行人脸反欺诈研究。Qi 等^[56]提出了一种新型的基于 CNN 的目标跟踪算法,并应用在人脸跟踪上。

3 常用人脸数据库

为了验证基于深度学习的人脸分析技术有效性,各种人脸数据库被采集和发布供算法训练和测试使用。近年来出现的人脸数据库呈现出训练样本和测试样本大量增加、复杂环境(如户外)下采集等特点。本节主要介绍现阶段常用人脸数据库。按照任务目的的不同,将人脸数据库分成了人脸检测数据库、人脸关键点检测数据库、人脸识别数据库、人脸属性识别数据库和其他人脸数据库分别介绍。

3.1 人脸检测数据库

CMU+MIT 人脸数据库^[13,72]包含 3 个正面人脸测试子集和一个旋转人脸测试子集,其中正面人脸测试子集有 130 幅图像,共 511 个人脸;旋转人脸测试子集有 50 幅图像,共 223 个人脸。FDDB 人脸数据库^[73]提供 2 845 幅图像,共 5 171 个人脸。AFW 人脸数据库^[74]由从 Flickr 采集的 205 幅图像组成,共 468 个人脸,其包含复杂的背景变化和人脸姿态变化等。MALF

人脸数据库^[75]是一个大规模人脸数据库,其包含 5 250 幅图像,共 11 931 个人脸,可用于人脸的细粒度评估。IJB-A 人脸数据库^[76]由美国 NIST 发布,包含 24 327 幅图像,共 49 759 个人脸,可用于人脸检测和人脸识别。最近发布的 WIDER 人脸数据库^[77]含有 32 203 幅图像,共 393 703 个人脸,并指定专门的训练集、验证集和测试集。表 2 总结了常用的人脸检测数据库,并给出相应的下载链接。

表 2 常用的人脸检测数据库
Tab. 2 Commonly-used face detection databases

数据库名	发布年份	数据库描述和下载链接
CMU+MIT	1999	180 幅图像,共 734 个人脸 http://vasc.ri.cmu.edu/idb/html/face/frontal_images/
FDDB	2010	2 845 幅图像,共 5 171 个人脸 http://vis-www.cs.umass.edu/fddb/index.html
AFW	2012	205 幅图像,共 468 个人脸 http://www.ics.uci.edu/~xzhu/face/
MALF	2015	5 250 幅图像,共 11 931 个人脸 http://www.cbsr.ia.ac.cn/faceevaluation/
IJB-A	2015	24 327 幅图像,共 49 759 个人脸 www.nist.gov/itl/iad/ig/ijba_request.cfm
WIDER	2016	32 203 幅图像,共 393 703 个人脸 http://mmlab.ie.cuhk.edu.hk/projects/WIDER-Face/

3.2 人脸关键点检测数据库

传统人脸关键点检测数据库为室内环境下采集的数据库,比如 Multi-PIE、FERET、FRGC、AR、BioID 等人脸数据库^[4-6]。而现阶段人脸关键点检测数据库通常为复杂环境下采集的数据库。LFPW 人脸数据库^[78]有 1 132 幅训练人脸图像和 300 幅测试人脸图像,大部分为正面人脸图像,每个人脸标定 29 个关键点。AFLW 人脸数据库^[79]包含了 25 993 幅从 Flickr 采集的人脸图像,每个人脸提供 21 个关键点标定。COFW 人脸数据库^[80]包含 LFPW 人脸数据库训练集中的 845 幅人脸图像以及其他 500 幅遮挡人脸图像,而测试集为 507 幅严重遮挡(同时包含姿态和表情的变化)的人脸图像,每个人脸标定 29 个关键点。MVFWD 人脸数据库^[81]为多视角人脸数据集,包括 2 050 幅训练人脸图像和 450 幅测试人脸图像,每个人脸标定 68 个关键点。OCFW 人脸数据库^[81]包含 2 591 幅训练人脸图像(均为未遮挡人脸)和 1 246 幅测试人脸图像(均为遮挡人脸),每个人脸标定 68 个关键点。

300-W 人脸数据库^[82]包含了 300 幅室内测试人脸图像(从 222 幅图像中采集)和 300 幅室外测试人脸图像(从 177 幅图像中采集),每个人脸标定 68 个关键点.表 3 总结了常用人脸关键点检测数据库,并给出相应的下载链接.

表 3 常用人脸关键点检测数据库

Tab. 3 Commonly-used facial landmark detection databases

数据库名	发布年份	数据库描述和下载链接
BioID	2001	约 1 000 幅图像,每个人脸标定 20 个关键点 https://www.bioid.com/About/BioID-Face-Database
LFPW	2011	1 132 幅图像,每个人脸标定 29 个关键点 http://neerajkumar.org/databases/lfpw/
AFLW	2011	25 993 幅图像,每个人脸标定 21 个关键点 https://lrs.icg.tugraz.at/research/aflw/
COFW	2013	1 852 幅图像,每个人脸标定 29 个关键点 http://www.vision.caltech.edu/xpburgos/ICCV13/
MVFW	2014	2 500 幅图像,每个人脸标定 68 个关键点 https://sites.google.com/site/junliangxing/codes
OCFW	2014	3 837 幅图像,每个人脸标定 68 个关键点 https://sites.google.com/site/junliangxing/codes
300-W	2016	600 幅图像,每个人脸标定 68 个关键点 http://ibug.doc.ic.ac.uk/resources/300-W_IMAVIS/

3.3 人脸识别数据库

传统人脸识别数据库为室内环境下采集的数据库,常见的包括 AR、Multi-PIE、FERET、FRGC、CAS-PEAL 等人脸数据库^[4-6].而人脸识别的真正挑战在于实际复杂场景中的识别问题,因此各种复杂环境下采集的大规模人脸数据库被广泛使用.LFW 人脸数据库^[83]包含 5 749 个人的共 13 233 幅人脸图像(主要来源于新闻图片),并且提供相应的训练集和测试集.PubFig 人脸数据库^[84]包括 200 个人的 58 797 幅人脸图像.CASIA-WebFace 人脸数据库^[85]包含 10 575 个人的共 49 414 幅人脸图像,主要用来训练深度网络模型.FaceScrub 人脸数据库^[86]由 530 个名人的共 106 863 幅人脸图像(每个人大约有 200 幅人脸图像)组成.MegaFace 人脸数据库^[59]是一个 100 万级人脸识别数据集.表 4 总结了常用人脸识别数据库,并给出相应的下载链接.

3.4 人脸属性识别数据库

常见的人脸属性包括表情、年龄和性别等.JAFFE

表 4 常用人脸识别数据库

Tab. 4 Commonly-used face recognition databases

数据库名	发布年份	数据库描述和下载链接
CAS-PEAL	2004	约 1 000 个人,共约 3 万幅人脸图像 http://www.jdl.ac.cn/peal/index.html
Multi-PIE	2008	337 个人,共约 75 万图像 http://www.flintbox.com/public/project/4742/
LFW	2007	5 749 个人,共 13 233 幅人脸图像 http://vis-www.cs.umass.edu/lfw/
PubFig	2009	200 个人,共 58 797 幅人脸图像 http://www.cs.columbia.edu/CAVE/databases/pubfig/
CASIA-WebFace	2014	10 575 个人,共 49 414 幅人脸图像 http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html
FaceScrub	2014	530 个人,共 106 863 幅人脸图像 http://vintage.winklerbros.net/facescrub.html
MegaFace	2016	约 69 万个人,共约 100 万幅人脸图像 http://megaface.cs.washington.edu/

人脸数据库^[87]由 10 个日本女性的 213 幅人脸图像组成.每个人采集 7 种不同表情.CK+人脸数据库^[88]包含 123 个人的 593 段视频,其中每段视频对应一种人脸表情.MMI 人脸数据库^[89]包含 75 个人的 2 900 段人脸表情视频.FG-NET 人脸数据库包括 82 个人的 1 002 幅人脸图像,年龄分布在 0~69 岁.MORPH 人脸数据库^[90]包括 13 673 个人的 55 608 幅人脸图像,年龄分布在 16~99 岁.FG-NET 和 MORPH 人脸数据库主要用来进行年龄识别.Adience 人脸数据库^[91]包含 2 284 个人的 26 580 幅人脸图像,每幅图像提供年龄和性别信息.IMDB-WIKI 人脸数据库^[51]由 IMDB 中 20 284 个名人的 460 723 幅人脸图像与维基百科的 62 328 幅人脸图像构成(共 523 051 幅人脸图像),每幅图像提供年龄和性别信息.CACD2000 人脸数据库^[92]由 2 000 个名人 163 446 幅人脸图像组成,其中每幅图像提供年龄标签.CelebA 人脸数据库^[93]包括 10 177 个人的 202 599 幅人脸图像.每幅人脸图像提供 5 个关键点标定和 40 种人脸属性信息.表 5 总结了常用人脸属性识别数据库,并给出相应的下载链接.

3.5 其他人脸数据库

本小节介绍其他人脸分析任务中常见人脸数据库.YouTube 名人数据库^[94]由 47 个人的 1 910 段视频序列组成.YouTube 人脸数据库^[95]包括 1 595 个人的 3 425 段视频.上述两个数据库可用于人脸跟踪和基于视频的人脸识别算法验证.KFW 人脸数据库^[96]主要用

表 5 常用人脸属性识别数据库

Tab. 5 Commonly-used face attribute recognition databases

数据库名	发布年份	数据库描述和下载链接
JAFFE	1999	10 个人,共 213 幅人脸图像(表情识别) http://www.kasrl.org/jaffe.html
CK+	2010	123 个人,共 593 段视频(表情识别) http://www.pitt.edu/~emotion/ck-spread.htm
MMI	2010	75 个人,共 2 900 段视频(表情识别) http://mmifacedb.eu/
FG-NET	2003	82 个人,共 1 002 幅人脸图像(年龄识别) http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html
MORPH	2006	13 673 个人,共 55 608 幅图像(年龄识别) http://www.faceaginggroup.com/morph/
Adience	2014	2 284 个人,共 26 580 幅人脸图像(年龄、性别识别) http://www.openul.ac.il/home/hassner/Adience/data.html
IMDB-WIKI	2015	20 284 个人,共 523 051 幅人脸图像(年龄、性别识别) https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/
CACD2000	2015	2 000 个人,共 163 446 幅人脸图像(年龄识别) http://bcsiriuschen.github.io/CARC/
CelebA	2015	10 177 个人,共 202 599 幅人脸图像(属性识别) http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

于亲属验证,包括 KFW-I 和 KFW-II 两个数据库。CASIA 人脸数据库^[97]由 50 个人,每个人 12 段视频(3 段真实和 9 段假冒)组成。Reply-Attack 人脸数据库^[98]由 50 个人,每个人 24 段视频(4 段真实和 20 段假冒)组成。CASIA 和 Reply-Attack 数据库可用于人脸反欺诈算法验证。表 6 对上述人脸数据库进行了总结,并给出相应的下载链接。

4 面临的挑战

尽管目前基于深度学习的人脸分析取得了巨大的进步,但是,现阶段依然存在许多问题亟待解决。本节从深度学习和人脸分析两个方面分析面临的主要挑战。

4.1 深度学习面临的挑战

1) 大数据训练问题。现今的深度学习算法的训练数据量已经达到了数百万和数千万,甚至是更大的规

表 6 其他常用人脸数据库

Tab. 6 Other commonly-used face databases

数据库名	发布年份	数据库描述和下载链接
YouTube 名人	2008	47 个人,共 1 910 段视频 http://seqam.rutgers.edu/site/media/data_files/ytcelebrity.tar
YouTube	2011	1 595 个人,共 3 425 段视频 http://www.cs.tau.ac.il/~wolf/yt-faces/
KFW	2012	533 对亲属关系(KFW-I)和 1 000 对亲属关系(KFW-II) http://www.kinfacew.com/download.html
CASIA	2012	50 个人,每个人 12 段视频 http://www.cbsr.ia.ac.cn/english/FaceAntiSpoofDatabases.asp
Replay-Attack	2012	50 个人,每个人 24 段视频 https://www.idiap.ch/dataset/replayattack

模量。现有的一些技术在大数据上已经不太适用。因此,如何高效地构建大规模集群设备处理这些大数据,是深度学习面临的一个巨大挑战。

2) 小数据训练问题。深度学习常常依赖于大数据。然而在现实世界中,有很多问题没有为数众多有标签的数据。研究如何将深度学习应用于小数据的训练上,构建能够动态自主学习的深度学习方法,也是深度学习面临的主要问题和挑战之一。

3) 理论问题。相比于传统的浅层学习,如支持向量机、随机森林等方法,深度学习存在着很多理论问题没有解决。比如,如何通过理论进行最优的超参数选择;如何采用理论指导的快速训练方式;如何针对不同实际问题设计出最佳的网络结构等。

4.2 人脸分析面临的挑战

1) 人脸的相似性和多变性。人脸图像有相似的五官结构和基本特征。这些内在的相似结构特征有利于特定人脸分析任务(如人脸检测、人脸关键点检测),但是对于人脸识别任务却是不利的。人脸是可塑性强的三维柔性曲面,而人脸图像是三维人脸模型在二维平面上的投影。由于采集角度的不同,使得获取到的人脸姿态是多样的,导致不同的角度获取到的人脸图像往往差别很大。此外,人脸具有非常丰富的表情变化,这些变化丰富的表情给人脸分析(如人脸识别任务等)造成困难。同时,由于人脸随着年龄的增加,同一个人的人脸图像,发生巨大的变化。最后,世界人口众多(约 70 多亿人),目前在大规模人脸数据库(至少百万量级)场景中的特定人脸分析任务(如人脸识别)的性

能仍然离真正的实际应用有很大的差距.因此,针对大规模人脸数据库,构建能够有效克服姿态、表情、年龄变化影响的基于深度学习的人脸分析方法,是人脸分析面临的一个巨大挑战.

2) 应用环境的复杂性.人脸分析技术在常规的应用环境中,能够取得较好的性能.然而,当受到外界光照变化、遮挡物等环境影响时,算法的性能通常会发生显著下降.另外,现有的通用摄像头(比如视频监控摄像头)获取到的人脸图像质量普遍不高,如何利用这些低质量的人脸图像进行分析和理解需要深入研究.同时,一些基于深度学习的人脸分析方法(如三维模型)常常需要耗费比较大的计算资源,如何将人脸分析方法应用于手机等移动设备,以扩大应用领域,也是现阶段人脸分析面临的困难.因此,如何有效解决应用环境的复杂性问题是人脸分析面临的另一个巨大挑战.

5 结 论

基于深度学习的人脸分析由于广泛的市场应用前景和巨大的研究价值受到重视.本研究对基于深度学习的人脸分析进行了深入的介绍,全面综述了常用的人脸分析任务(包括人脸检测、人脸关键点检测、人脸识别和人脸属性识别等)研究进展,同时介绍了常用人脸数据库,并着重讨论了深度学习和人脸分析各自面临的挑战.

参考文献:

- [1] AHONEN T, HADID A, PIETIKAINEN M. Face description with local binary patterns: application to face recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 2037-2041.
- [2] LOWE D J. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [3] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington D C: IEEE, 2005: 886-893.
- [4] 梁路宏,艾海舟,徐光祐,等.人脸检测研究综述[J]. *计算机学报*, 2002, 25(5): 449-458.
- [5] ZHAO W, CHELLAPPA R, PHILLIPS P J, et al. Face recognition: a literature survey[J]. *ACM Computing Surveys*, 2003, 35(4): 399-458.
- [6] 严严,章毓晋.基于视频的人脸识别研究进展[J]. *计算机学报*, 2009, 32(5): 878-886.
- [7] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [8] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 5(8): 1798-1828.
- [9] IAN G, YOSHUA B, ARON C. Deep learning[M]. Massachusetts: MIT Press, 2016.
- [10] 余凯,贾磊,陈雨强,等.深度学习的昨天、今天和明天[J]. *计算机研究与发展*, 2013, 50(9): 1799-1804.
- [11] 山世光,阚美娜,刘昕,等.深度学习:多层神经网络的复兴与变革[J]. *科技导报*, 2016, 34(13): 60-70.
- [12] 尹宝才,王文通,王立春.深度学习研究综述[J]. *北京工业大学学报*, 2015, 41(1): 48-59.
- [13] ROWLEY H, BALUJA S, KANADE T. Neural network-based face detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(1): 23-38.
- [14] FERAUND R, BERNIER O J, VIALLET J, et al. A fast and accurate face detector based on neural networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(1): 42-53.
- [15] LAWRENCE S, GILES C L, TSOI A C, et al. Face recognition: a convolutional neural-network approach[J]. *IEEE Transactions on Neural Networks*, 1997, 8(1): 98-113.
- [16] SUN Y, WANG X, TANG X. Deep convolutional network cascade for facial point detection[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington D C: IEEE, 2013: 3476-3483.
- [17] ZHANG J, SHAN S, KAN M, et al. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment[C]// *Proceedings of the European Conference on Computer Vision*. Berlin: Springer, 2014: 1-16.
- [18] LEVI G, HASSNER T. Age and gender classification using convolutional neural networks[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Washington D C: IEEE, 2015: 34-42.
- [19] LIU P, HAN S, MENG Z, et al. Facial expression recognition via a boosted deep belief network [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington D C: IEEE, 2014: 1805-1812.
- [20] ZHU Z, LUO P, WANG X, et al. Deep learning identity-preserving face space[C]// *Proceedings of the IEEE In-*

- ternational Conference on Computer Vision, Washington D C; IEEE, 2013; 113-120.
- [21] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological Review, 1958, 65(6): 386-408.
- [22] MINSKY M L, PAPER T S A. Perceptrons: expanded edition[M]. Massachusetts: MIT Press, 1988; 1-308.
- [23] 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.
- [24] HINTON G, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [25] SMOLENSKY P. Information processing in dynamical systems: foundations of harmony theory[J]. Parallel Distributed Processing, 1986, 1: 194-281.
- [26] RIFAI S, VINCENT P, MULLER X, et al. Contractive auto-encoders: explicit invariance during feature extraction[C]// Proceedings of the International Conference on Machine Learning, Madison: Omnipress, 2011: 833-840.
- [27] WILLIAMS D, HINTON G. Learning representations by back-propagating errors [J]. Nature, 1986, 323: 533-536.
- [28] GRAVES A, JAITLY N. Towards end-to-end speech recognition with recurrent neural networks[C]// Proceedings of the International Conference on Machine Learning, Madison: Omnipress, 2014; 1764-1772.
- [29] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM[J]. Neural Computation, 2000, 12(10): 2451-2471.
- [30] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems, Lake Tahoe: NIPS, 2012; 1097-1110.
- [31] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C; IEEE, 2009; 248-255.
- [32] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [33] NAIR V, HINTON G. Rectified linear units improve restricted Boltzmann machines[C]// Proceedings of the International Conference on Machine Learning, Madison: Omnipress, 2010; 807-814.
- [34] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human level performance on ImageNet classification[C]// Proceedings of the IEEE International Conference on Computer Vision, Washington D C; IEEE, 2015; 1026-1034.
- [35] GARCIA C, DELAKIS M. Convolutional face finder: a neural architecture for fast and robust face detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(11): 1408-1423.
- [36] ZHANG C, ZHANG Z. Improving multiview face detection with multi-task deep convolutional neural networks [C]// Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Washington D C; IEEE, 2014; 1036-1041.
- [37] LI H, LIN Z, SHEN X, et al. A convolutional neural network cascade for face detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C; IEEE, 2015; 5325-5334.
- [38] YANG S, LUO P, LOY C C, et al. From facial parts responses to face detection: a deep learning approach[C]// Proceedings of the IEEE International Conference on Computer Vision, Washington D C; IEEE, 2015; 3676-3684.
- [39] ZHAN S, TAO Q, LI X. Face detection using representation learning[J]. Neurocomputing, 2015, 187: 19-26.
- [40] ZHOU E, FAN H, CAO Z, et al. Extensive facial landmark localization with coarse-to-fine convolutional network cascade[C]// Proceedings of the IEEE International Conference on Computer Vision Workshops, Washington D C; IEEE, 2013; 386-391.
- [41] ZHANG Z, LUO P, LOY C C, et al. Facial landmark detection by deep multi-task learning[C]// Proceedings of the European Conference on Computer Vision, Berlin: Springer, 2014; 94-108.
- [42] JOURABLOO A, LIU X. Large-pose face alignment via CNN-based dense 3D model fitting[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C; IEEE, 2016; 4188-4196.
- [43] ZHU X, LEI Z, LIU X, et al. Face alignment across large poses: a 3D solution[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C; IEEE, 2016; 146-155.
- [44] ZHANG J, KAN M, SHAN S, et al. Occlusion-free face alignment: deep regression networks coupled with de-corrupt autoencoders [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C; IEEE, 2016; 3428-3437.
- [45] TAIGMAN Y, YANG M, RANZATO M, 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, Washington D C: IEEE, 2014: 1701-1708.
- [46] SUN Y, WANG X, TANG X. Deep learning face representation from predicting 10, 000 classes [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C: IEEE, 2014: 1891-1898.
- [47] SUN Y, CHEN Y, WANG X, et al. Deep learning face representation by joint identification-verification [C] // Advances in Neural Information Processing Systems, Montreal: NIPS, 2014: 1988-1996.
- [48] SUN Y, WANG X, TANG X. Deeply learned face representations are sparse, selective, and robust [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C: IEEE, 2015: 2892-2900.
- [49] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C: IEEE, 2015: 815-823.
- [50] WANG X, GUO R, KAMBHAMETTU C. Deeply learned feature for age estimation [C] // Proceedings of the IEEE Winter Conference on Applications of Computer Vision 2014, Washington D C: IEEE, 2015: 534-541.
- [51] ROTHE R, TIMOFTE R, GOOL L. DEX: deep expectation of apparent age from a single image [C] // Proceedings of the IEEE International Conference on Computer Vision Workshops, Washington D C: IEEE, 2015: 10-15.
- [52] KAHOU S, BOUTHILLIER X, LAMBLIN P, et al. EmoNets: multimodal deep learning approaches for emotion recognition in video [J]. Journal on Multimodal User Interfaces, 2016, 10(2): 99-111.
- [53] TANG X, WANG X, LUO P. Hierarchical face parsing via deep learning [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C: IEEE, 2012: 2480-2487.
- [54] YANG J, LEI S, LI S Z. Learn convolutional neural network for face anti-spoofing [J]. Computer Science, 2014, 9218: 373-384.
- [55] ZHANG K, Y HUANG, SONG C, et al. Kinship verification with deep convolutional neural networks [C] // British Machine Vision Conference, British: BMVA, 2015: 148. 1-148. 12.
- [56] QI Y, ZHANG S, QIN L, et al. Hedged deep tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C: IEEE, 2016: 4303-4311.
- [57] VIOLA P, JONES M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [58] LI Y, SUN B, WU T, et al. Face detection with end-to-end integration of a convnet and a 3D model [C] // Proceedings of the European Conference on Computer Vision, Berlin: Springer, 2016: 420-436.
- [59] 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, Washington D C: IEEE, 2016: 3456-3465.
- [60] SUBBURAMAN V B, MARCEL S. Fast bounding box estimation based face detection [C] // Proceedings of EC-CV Workshop on Face Detection: Where We Are, and What Next? Berlin: Springer, 2010: 5325-5334.
- [61] COOTES T F, COOPER D, TAYLOR C J, et al. Active shape models-their training and application [J]. Computer Vision and Image Understanding, 1995, 61(1): 38-59.
- [62] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models [C] // Proceedings of the European Conference on Computer Vision, Berlin: Springer, 1998: 484-498.
- [63] SARAGIH J M, LUCEY S, COHN J. Face alignment through subspace constrained mean-shifts [C] // Proceedings of the IEEE International Conference on Computer Vision, Washington D C: IEEE, 2009: 1034-1041.
- [64] LIU X. Discriminative face alignment [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(11): 1941-1954.
- [65] CAO X, WEI Y, WEN F, et al. Face alignment by explicit shape regression [J]. International Journal of Computer Vision, 2014, 107(2): 177-190.
- [66] SHLIZERMAN I, SEITZ S, MILLER D, et al. The MegaFace benchmark: 1 million faces for recognition at scale [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington D C: IEEE, 2016: 4873-4882.
- [67] DEVRIES T, BISWARANJAN K, TAYLOR G W. Multi-task learning of facial landmarks and expression [C] // Proceedings of the Canadian Conference on Computer and Robot Vision, Washington D C: IEEE, 2014: 98-103.
- [68] ZHANG Z, LUO P, CHEN C L, et al. Learning deep representation for face alignment with auxiliary attributes [J]. IEEE Transactions on Pattern Analysis

- and Machine Intelligence, 2016, 38(5):918-930.
- [69] KUMAR N, BERG A C, BELHUMEUR P N, et al. Describable visual attributes for face verification and image search[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(10):1962-1977.
- [70] KAN M, SHAN S, CHANG H, et al. Stacked progressive auto-encoders (SPAEC) for face recognition across poses[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2014:1883-1890.
- [71] YIM J, JUNG H, YOO B, et al. Rotating your face using multi-task deep neural network[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2015:676-684.
- [72] SUNG K, POGGIO T. Example-based learning for view-based human face detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 20(1):39-51.
- [73] JAIN V, LEARNED-MILLER E. FDDB: a benchmark for face detection in unconstrained settings, Technical Report UM-CS-2010-009 [R]. Amherst: University of Massachusetts, 2010.
- [74] ZHU X, RAMANAN D. Face detection, pose estimation, and landmark localization in the wild[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2012:2879-2886.
- [75] YANG B, YAN J, LEI Z, et al. Fine-grained evaluation on face detection in the wild[C] // Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition Conference and Workshops. Washington D C: IEEE, 2015:111-117.
- [76] KLARE B F, KLEIN B, TABORSKY E, et al. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2015:1931-1939.
- [77] YANG S, LUO P, CHEN C, et al. WIDER FACE: a face detection benchmark [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2016:5525-5533.
- [78] BELHUMEUR P, JACOBS D, KRIEGMAN D, et al. Localization parts of faces using a consensus of exemplars [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2011:545-552.
- [79] KOSTINGER M, WOHLHART P, ROTH P M, et al. Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization[C] // Proceedings of IEEE International Conference on Computer Vision Workshop on Benchmarking Facial Image Analysis Technologies. Washington D C: IEEE, 2011:2144-2151.
- [80] BURGOS-ARTIZU X P, PERONA P, DOLLAR P. Robust face landmark estimation under occlusion [C] // Proceedings of the IEEE International Conference on Computer Vision. Washington D C: IEEE, 2013:1513-1520.
- [81] XING J, NIU Z, HUANG J, et al. Towards multi-view and partially-occluded face alignment [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2014:1829-1836.
- [82] SAGONAS C, ANTONAKOS E, TZIMIROPOULOS G, et al. 300 faces in-the-wild challenge: database and results[J]. Image and Vision Computing, 2016, 47:3-18.
- [83] HUANG G B, RAMESH M, BERG T, et al. Labeled faces in the wild: a database for studying face recognition in unconstrained environments [R]. Amherst: University of Massachusetts, 2007.
- [84] KUMAR N, BERG A, BELHUMEUR P N, et al. Attribute and simile classifiers for face verification [C] // Proceedings of the International Conference on Computer Vision. Washington D C: IEEE, 2009:365-372.
- [85] YI D, LEI Z, LIAO S, et al. Learning face representation from scratch [DB/OL]. [2016-09-01]. <https://arxiv.org/pdf/1411.7923.pdf>.
- [86] NG H W, WINKLER S. A data-driven approach to cleaning large face datasets [C] // Proceedings of the IEEE International Conference on Image Processing. Washington D C: IEEE, 2014:343-347.
- [87] LYONS M J, BUDYNEK J, AKAMATSU S. Automatic classification of single facial images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(12):1357-1362.
- [88] LUCEY P, COHN J F, KANADE T, et al. The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington D C: IEEE, 2010:94-101.
- [89] VALSTAR M F, PANTIC M. Induced disgust, happiness and surprise: an addition to the MMI facial expression database [C] // Proceedings of International Conference on Language Resources and Evaluation. Washington D C: IEEE, 2010:65-70.
- [90] RICANEK K, TESAFAYE T. MORPH: a longitudinal image database of normal adult age-progression [C] //

- Proceedings the IEEE International Conference on Automatic Face and Gesture Recognition. Washington D C: IEEE, 2006: 341-345.
- [91] EIDINGER E, ENBAR R, HASSNER T. Age and gender estimation of unfiltered faces[J]. IEEE Transactions on Information Forensics and Security, 2014, 9 (12): 2170-2179.
- [92] CHEN B, CHEN C, HSU W. Face recognition using cross-age reference coding with cross-age celebrity dataset[J]. IEEE Transactions on Multimedia, 2015, 17 (6): 804-815.
- [93] LIU Z, LUO P, WANG X, et al. Deep learning face attributes in the wild [C] // Proceedings of the International Conference on Computer Vision. Washington D C: IEEE, 2015: 3730-3738.
- [94] KIM M, KUMAR S, PAVLOVIC V, et al. Face tracking and recognition with visual constraints in real-world videos[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2008: 1-8.
- [95] WOLF L, HASSNER T, MAOZ I. Face recognition in unconstrained videos with matched background similarity[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2011: 529-534.
- [96] LU J, HU J, ZHOU X, et al. Neighborhood repulsed metric learning for kinship verification[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Washington D C: IEEE, 2012: 2594-2601.
- [97] ZHANG Z, YAN J, LIU S, et al. A face antispoofing database with diverse attacks[C] // Proceedings of the 5th IAPR International Conference on Biometrics. Washington D C: IEEE, 2012: 26-31.
- [98] CHINGOVSKA I, ANJOS A, MARCEL S. On the effectiveness of local binary patterns in face anti-spoofing[C] // Proceedings of the IEEE International Conference of the Biometrics Special Interest Group. Washington D C: IEEE, 2012: 1-7.

Recent Advances on Deep-Learning-Based Face Analysis

YAN Yan, CHEN Riwei, WANG Hanzi *

(Fujian Key Laboratory of Sensing and Computing for Smart City,

School of Information Science and Engineering, Xiamen University, Xiamen 361005, China)

Abstract: In recent years, the face analysis based on deep learning has made great progress, and has become one of the most active research areas in the field of computer vision. In order to further promote the study of deep learning and face analysis, this paper overviews recent advances on the deep-learning-based face analysis techniques in the literature. First, a brief overview of deep learning and its history are given and reasons for the effectiveness of deep learning are also analyzed. Then, according to different objectives, four face analysis tasks, i.e., face detection, facial key-point detection, face recognition, face attribute recognition, are introduced and discussed in detail, and the key problems existing in these tasks at present are analyzed. After that, the commonly used face databases in the face analysis are described. At last, main challenges of face analysis based deep learning are shown and the conclusion is presented.

Key words: deep learning; convolutional neural network; face database; face recognition; face analysis