

DOI:10.13364/j.issn.1672-6510.20160050

## 基于深度学习的人脸识别方法研究综述

杨巨成<sup>1</sup>, 刘 娜<sup>1,2</sup>, 房珊珊<sup>1</sup>, 谢 迎<sup>1</sup>

(1. 天津科技大学计算机科学与信息工程学院, 天津 300222; 2. 天津农学院工程技术学院, 天津 300384)

**摘 要:** 针对基于深度学习的人脸识别这一生物特征识别领域的研究热点进行了综述. 阐释了人脸识别及深度学习模型的基本结构; 总结了该技术在国内外外的研究现状及其应用, 如基于卷积神经网络(CNN)的人脸识别方法、深度非线性人脸形状提取方法、基于深度学习的人脸姿态鲁棒性建模、有约束环境中的全自动人脸识别、基于深度学习的视频监控下的人脸识别、基于深度学习的低分辨率人脸识别以及其他基于深度学习的人脸信息的识别等; 分析了当前人脸识别技术在深度学习应用中存在的问题及发展趋势.

**关键词:** 深度学习; 人脸识别; 卷积神经网络; 生物特征识别

**中图分类号:** TP301.6      **文献标志码:** A      **文章编号:** 1672-6510(2016)06-0001-10

## Review of Face Recognition Methods Based on Deep Learning

YANG Jucheng<sup>1</sup>, LIU Na<sup>1,2</sup>, FANG Shanshan<sup>1</sup>, XIE Ying<sup>1</sup>

(1. College of Computer Science and Information Engineering, Tianjin University of Science & Technology, Tianjin 300222, China; 2. College of Engineering and Technology, Tianjin Agricultural University, Tianjin 300384, China)

**Abstract:** Face recognition based on deep learning, which has already become a hot research topic in the field of biometric recognition at present, was reviewed. Firstly, face recognition and the basic structure of deep learning were introduced. Then, the current international and domestic research status quo and application of the technology were summarized, such as face recognition method based on convolutional neural network(CNN), deep nonlinear face shape extraction method, robust modeling of face pose based on deep learning, fully automatic face recognition in the constrained environment, face recognition based on deep learning under video surveillance, low resolution face recognition based on deep learning, and other face information recognition based on deep learning. Finally, a general analysis was made on the existing problems and future development trend of face recognition in the application of deep learning.

**Key words:** deep learning; face recognition; convolutional neural network; biometric recognition

人脸具有高度非刚性的特点, 存在着大量体现个体差异的细节. 人脸识别是通过从静态图像或者动态视频中检测出的人脸图像与数据库中的人脸图像进行比对, 找到与之匹配的人脸的过程, 通常用于身份识别和鉴定的目的<sup>[1]</sup>, 是属于生物特征识别领域的课题.

人脸识别研究是从 20 世纪 50 年代开始的, 作为一种重要的生物特征识别技术, 其具有直接、友好、方便、交互性好等优点, 一直被研究学者所关注<sup>[2]</sup>.

人脸识别技术在安防和金融支付等方面有许多实际的应用, 如视频监控、智能支付、访问控制等, 是目前

### 机器学习和模式识别中最热门的研究方向.

目前, 基于机器视觉的人脸识别方法已经取得了丰硕成果. 研究中需要考虑受到人面部表情、姿态、年龄、位置和遮盖物等因素引起的类内变化, 及来自外界光照、背景等身份不同引起的类间变化, 这两种变化的分布是高度复杂且非线性的. 传统的基于浅层学习的人脸识别方法, 对于类内和类间这两种变化的复杂分布和非线性的人脸数据进行识别, 往往达不到预期效果. 深度学习是模拟人类视觉感知神经系统的认知学习, 能够获得更具表征力的高层特征, 可以用来解决人脸识别中的类内和类间的变化分布这

收稿日期: 2016-02-23; 修回日期: 2016-06-10

基金项目: 国家自然科学基金资助项目(61502338); 天津市科技支撑计划重点资助项目(15ZCZDZX00200)

作者简介: 杨巨成(1980—), 男, 湖北天门人, 教授, jcyang@tust.edu.cn.

一难题。

本文对基于深度学习的人脸识别技术进行了综述,列举了深度学习基本的模型结构,总结了该技术的国内外研究现状,介绍了深度学习进行人脸识别的典型技术及常用的大规模人脸识别数据库,最后对基于深度学习的人脸识别研究存在的问题和发展趋势进行了分析与展望。

## 1 深度学习

近年来,随着机器学习的不断发展,深度学习作为一个崭新的研究方向引起人工智能领域的广泛关注。2012年12月29日《纽约时报》的头版报道称“深度学习让机器执行人类的活动,如看、听和思考,为模式识别提供了可能性,促进了人工智能技术的进步”。2013年,《麻省理工科技评论》(MIT Technology Review)将深度学习列为世界十大技术突破之首。深度学习的输入数据分布式表示,并具有强大的集中学习数据集本质特征的能力,从而可以提高学习效率。

算法速度、系统性能以及易用性等是目前制约机器学习普及应用的瓶颈问题。2015年5月21日,由百度牵头的分布式深度机器学习开源平台“深盟”正式面向公众开放,降低了分布式机器学习的门槛。腾讯深度学习优图团队也于2014年国际机器学习顶级会议ICML2014亮相。国内的阿里巴巴、科大讯飞、搜狗、云知声等公司也分别开始在其产品中使用深度学习技术。典型的深度学习模型的结构大致可以分为3类:

(1)生成性深度结构。深度信念网络(deep belief networks, DBN)<sup>[3]</sup>是目前研究比较广泛的深度学习结构。如图1所示,DBN由一系列受限波尔兹曼机(restricted Boltzmann machine, RBM)单元组成,RBM是一个随机无向图模型,其可见层和隐层单元彼此互连,层内无连接,隐层单元可获取输入可视单元的高阶相关性。为了获取生成性权值,预训练采用无监督贪心逐层方式实现。DBN可获取观测数据和标签的联合概率分布,这方便了对先验概率和后验概率的估计,而区分型模型仅能估计后验概率<sup>[4]</sup>。

DBN应用广泛、灵活性强、易于拓展,但其输入是简单的图像矩阵一维向量化,没有考虑到图像的二维结构。卷积DBN网络(convolutional deep belief networks, CDBN)<sup>[5-7]</sup>则利用了邻域像素的空域关系,通过变换不变性变换到高维图像。2011年,Liu等<sup>[8]</sup>

提出一种新的半监督学习算法,即判别深度置信网络(discriminative deep belief networks, DDBN),使用一个新的深层架构,结合DBN的抽象能力和指数损失函数的区分能力,应用于可视化数据分析中。

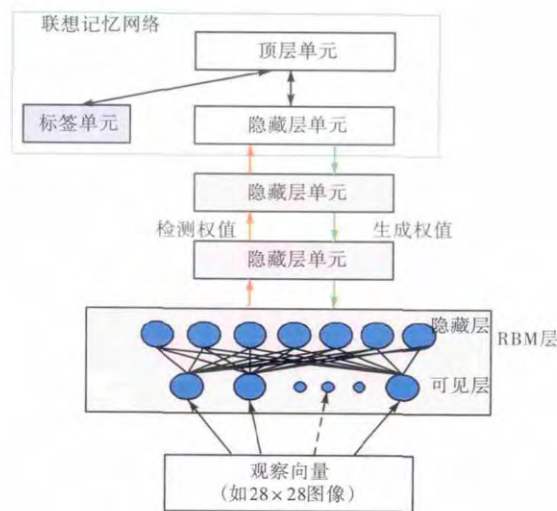


图1 深度信念网络结构

Fig. 1 DBN framework

(2)区分性深度结构。卷积神经网络(convolutional neural networks, CNN)<sup>[9-11]</sup>是第一个真正成功训练多层网络结构的学习算法,利用BP算法设计并训练。CNN是一种适应二维人脸图像识别场景的有效学习方式,被大量文献用于解决人脸识别问题,主要用来识别位移、缩放及其他形式扭曲不变性的二维图像。由于CNN的特征检测层通过训练数据进行学习,避免了显式的特征提取,隐式地从训练数据中进行学习;而且由于同一特征映射面上的神经元权值相同,所以网络可以并行学习,这也是CNN相对于神经元彼此相连网络的优势。其布局更接近实际的生物神经网络,权值共享降低了网络的复杂性,尤其是多维输入向量的图像可以直接输入网络,避免了特征提取和分类过程中数据重建的复杂度。图2是图像分类的两层卷积神经网络的示例。

(3)混合型结构。混合型结构<sup>[12-13]</sup>的学习过程包含生成性部分和区分性部分,一般使用最优化和区分性深度网络模型来解决。这种区分性网络寻优过程通常是附加一个顶层变量来估计任何深度生成模型或者非监督深度网络中的参数。BP算法用于优化DBN权值,初始权值通过在RBM和DBN预训练中得到,而不是随机产生的。这样的网络一般会比仅通过BP算法单独训练的网络性能优越,与前馈神经网络相比加速了训练和收敛时间。

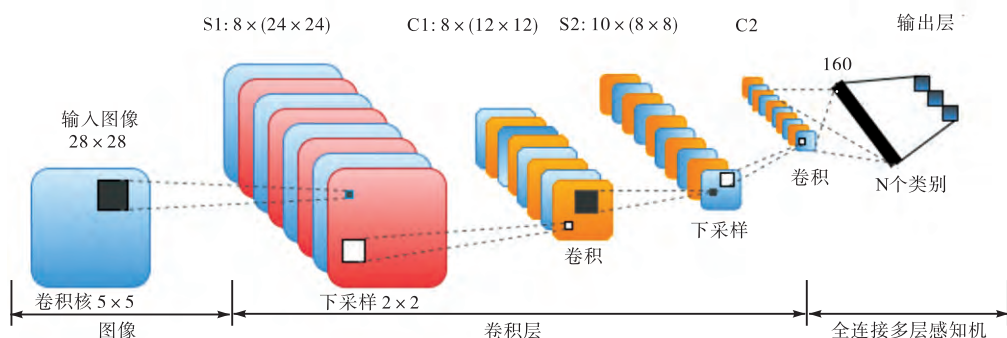


图 2 图像分类的两层卷积神经网络示例

Fig. 2 Two-layers CNN architecture of image classification

## 2 基于深度学习的人脸识别技术

传统的人脸识别方法有多种,如主动形状模型(active shape model, ASM)<sup>[14]</sup>和主动外观模型(active appearance models, AAM)<sup>[15]</sup>;基于局部的方法,如利用局部描述子 Gabor、局部二值模式(local binary pattern, LBP)等进行识别;还有基于全局的方法,包括经典的人脸识别算法,如特征脸方法(Eigenface)<sup>[16]</sup>、线性判别分析法(linear discriminant analysis, LDA)等子空间学习算法以及局部保持投影算法(locality preserving projection, LPP)等流行学习算法;3D 人脸识别也是一个新的方向。但是,由于受到光照、姿态及表情变化、遮挡、海量数据等因素的影响,传统的人脸识别方法由于自身的局限性,其识别精度受到制约。

在深度学习框架下,学习算法直接从原始图像学习判别性的人脸特征。在海量人脸数据支撑下,基于深度学习的人脸识别在速度和精度方面已经远远超过人类。深度学习借助于图形处理器(GPU)组成的运算系统作大数据分析,人脸识别是图像处理及人工智能的一个重要指标,证明深度学习模型有助于推动人工智能发展,将来甚至可能超越人类的智能水平。

### 2.1 国外研究现状

2012 年,勒尼德·米勒研究小组<sup>[17]</sup>率先将深度学习用于 LFW(labeled face in the wild)数据库的人脸识别。他们采用无监督的特征学习方法,取得了 87% 的识别率,与当时最好的人脸识别算法相比还有一定差距。经典的人脸识别算法 Eigenface<sup>[17]</sup>在 LFW 中的识别率只有 60%,而最新深度学习算法的识别率是 99.47%,甚至超过了人眼的识别率(99.25%)。近几年,国际上许多项目将深度学习成功地运用到人脸

识别中,例如:DeepFace<sup>[18]</sup>、DeepID<sup>[19]</sup>、FaceNet<sup>[20]</sup>等。以上的算法都基于海量的训练数据,让深度学习算法自己从海量的数据中学习对于光照、表情、角度等不变的特性。在以上算法中,准确率最高的是 FaceNet,在 LFW 数据库上已经达到 99.63% 的准确率,超过了人眼的识别结果。

LFW<sup>[21]</sup>数据库是由美国马萨诸塞大学阿姆斯特分校计算机视觉实验室整理完成,用于研究非受限情形下的人脸识别问题,已成为学术界评价识别性能的标准参照。其他数据库如 LFPW(labeled face parts in the wild)<sup>[22]</sup>数据库,包含 3 000 张从互联网采集到的人脸图片,得到的图像经人脸检测得到人脸在图像中的位置,并给出 3 组人工标注的 35 个人脸特征点的结果。WDFace<sup>[23]</sup>数据库包含了 2 995 个人的 99 773 张图片,其中有超过 2 000 个人有 15 张以上的图片。

LFW 项目是人脸识别技术从初始阶段提升到原型系统阶段的动力,参加 LFW 测试的数据已经成为现在论文验证的参照标准,该数据库极大地推动了人脸识别技术在深度学习上的进步。表 1 和表 2 分别列出了深度学习及其他各种方法在 LFW 人脸数据库中的识别率。

表 1 深度学习在 LFW 人脸数据库中的识别率

Tab. 1 Recognition rate of deep learning in the LFW face database

| 方法                        | 配准的关键点数 | 训练样本数      | 识别率/% |
|---------------------------|---------|------------|-------|
| 文献 <sup>[24]</sup>        | 3       | 非监督        | 87    |
| 文献 <sup>[18]</sup>        | 5       | 8.762 8 万  | 92.52 |
| DeepFace <sup>[25]</sup>  | 73      | 700 万      | 97.35 |
| DeepFace <sup>[26]</sup>  | 73      | 5 亿        | 98.40 |
| DeepID <sup>[27]</sup>    | 5       | 20.259 9 万 | 97.45 |
| DeepID2 <sup>[28]</sup>   | 18      | 20.259 9 万 | 99.15 |
| DeepID2 + <sup>[27]</sup> | 18      | 45 万       | 99.47 |



表2 各方法在 LFW 人脸数据库中的识别率

Tab. 2 Recognition rate of each method in the LFW face database

| 方法                           | Rank-1/% |
|------------------------------|----------|
| COST-S1 <sup>[29]</sup>      | 3.0      |
| COST-S1 + S2 <sup>[29]</sup> | 56.7     |
| DeepFace <sup>[18]</sup>     | 66.5     |
| DeepFace <sup>[25]</sup>     | 64.9     |
| DeepID2 <sup>[28]</sup>      | 91.1     |
| DeepID2 + <sup>[27]</sup>    | 95.0     |

注: Rank-1 为第一次找到正确人脸的识别率。

国际学术界的研究团队还有加拿大蒙特利尔大学的 Yoshua Bengio 研究组、英国谢菲尔德大学的 Neil Lawrence 研究组。此外,企业团队有 Microsoft 亚洲研究院,其在深度学习的研究上处于领先地位,不仅研究算法,还研究硬件结构;Facebook 的 Yann LeCun 成立了人工智能实验室;Google 公司有 Geoffrey Hinton 团队。

深度学习的主流开源项目同样适用于人脸识别的检测,对于在应用中整合深度学习功能的开发者,以下开源项目在 GitHub 均可以免费下载。Caffe 是一个清晰而且高效的深度学习框架,应用于 Google 公司于 2015 年开发的 DeepDream 项目。Tensorflow 是 Google 公司开发,于 2016 年开源,并公布了系统参数,此举在深度学习领域影响巨大。另外,还有 Torch7、Theano、Deeplearn Tool-Box、Deeplearning 4j、Brainstorm、Chainer、Marvin、ConvNetJs、MXNet、Neon 等开源深度学习框架,为广大研究人员研究深度学习算法提供了多种途径。

## 2.2 国内研究现状

近年来,国内研究机构也在人脸识别领域进行了大量深入的研究。在 2014 年的国际计算机视觉与模式识别会议上,来自 Facebook 人工智能实验室的 Taigman 等<sup>[18]</sup>和中国香港中文大学的团队,在允许利用有标签外部数据且非限定的测试条件下,均采用卷积神经网络,分别得到 97.35% 和 97.45% 的平均分类精度。香港中文大学汤晓鸥和王晓刚领导的研究团队,研发了一套准确度极高的自动人脸识别系统,该系统在 LFW 数据库中的识别率高达 99.15%。

此外,在深度学习和人脸识别领域,中国科学院陈熙霖和山世光等<sup>[30-35]</sup>利用深度学习在人脸特征识别和其他特征识别方面进行了深入的探索。其研究不仅包括人脸识别,还包括人脸检测与跟踪、面部对准(面部标签本地化)、性别识别、老龄化建模与评估、皮肤颜色分类等。李子青等<sup>[36-40]</sup>在人脸识别和智

能视频两个技术领域,结合现在流行的深度学习理论,进行了一定的研究。

在最权威的大规模人脸识别评测数据库 FRVT 2006 (the face recognition vendor test, FRVT) 上,清华大学丁晓青等所在的课题组完成的测试达到了国际领先水平;浙江大学何晓飞教授课题组<sup>[41-42]</sup>针对大数据时代机器学习开展了一系列研究;南京大学周志华教授课题组在机器学习和数据挖掘等方面进行了长期研究,并提出了许多新的算法;西安电子科技大学陈渤教授<sup>[43-44]</sup>对基于分层超完备字典稀疏表示的深度学习算法研究及应用开展深入的研究。这些研究表明国内学者对深度学习以及人脸识别相关领域已经有很大的关注。

## 3 深度学习进行人脸识别的典型应用

深度学习的出现使人脸识别技术取得了突破性进展。人脸识别的最新研究成果<sup>[21]</sup>表明,深度学习得到的人脸特征表达具有手工特征表达所不具备的重要特性,例如它是中度稀疏的、对人脸身份和人脸属性有很强的选择性、对局部遮挡具有良好的鲁棒性。这些特性是通过大数据训练自然得到的,并未对模型加入显式约束或后期处理,这也是深度学习能成功应用在人脸识别中的主要原因。

深度学习在人脸识别上有 7 个方面的典型应用:基于卷积神经网络(CNN)的人脸识别方法,深度非线性人脸形状提取方法,基于深度学习的人脸姿态鲁棒性建模,有约束环境中的全自动人脸识别,基于深度学习的视频监控下的人脸识别,基于深度学习的低分辨率人脸识别及其他基于深度学习的人脸相关信息的识别,具体见图 3。

### 3.1 基于卷积神经网络(CNN)的人脸识别方法

卷积神经网络(CNN)是一种深度的监督学习下的机器学习模型,能挖掘数据局部特征,提取全局训练特征和分类,其权值共享结构网络使之更类似于生物神经网络,在模式识别各个领域都得到成功应用。CNN 通过结合人脸图像空间的局部感知区域、共享权重、在空间或时间上的降采样来充分利用数据本身包含的局部性等特征,优化模型结构,保证一定的位移不变性<sup>[45]</sup>。

利用 CNN 模型,香港中文大学的 DeepID 项目<sup>[27]</sup>以及 Facebook 的 DeepFace 项目<sup>[18]</sup>在 LFW 数据库上的人脸识别正确率分别达 97.45% 和 97.35%, 只比人类视觉识别 97.5%<sup>[46]</sup>的正确率略低。在取得突破性

成果之后, 香港中文大学的 DeepID2 项目<sup>[28]</sup>将识别率提高到了 99.15%。DeepID2 通过学习非线性特征变换使类内变化达到最小, 而同时使不同身份的人脸图像间的距离保持恒定, 超过了目前所有领先的深度学习<sup>[18-19]</sup>和非深度学习算法<sup>[47]</sup>在 LFW 数据库上的识别率以及人类在该数据库的识别率<sup>[46]</sup>。深度学习已经成为计算机视觉中的研究热点, 关于深度学习的新算法和新方向不断涌现, 并且深度学习算法的性能逐渐在一些国际重大评测比赛中超过了浅层学习算法。

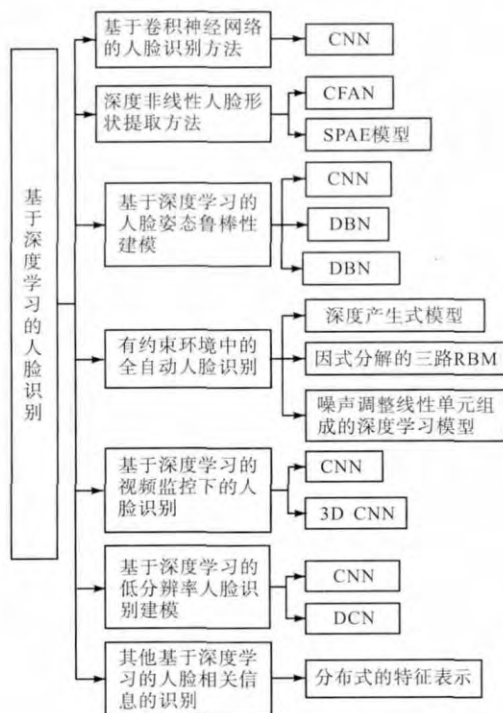


图3 基于深度学习的人脸识别算法

Fig. 3 Face recognition methods based on deep learning

### 3.2 深度非线性人脸形状提取方法

面部特征点定位(人脸形状提取或人脸对齐)在人脸识别、表情识别、人脸动画合成等诸多任务中具有非常重要的作用。由于姿态、表情、光照和遮挡等因素的影响, 在真实场景下的人脸对齐任务是非常困难的。

主动形状模型(active shape model, ASM)<sup>[14]</sup>和主动表观模型(active appearance models, AAM)<sup>[15-16]</sup>是经典的人脸对齐方法。它们使用线性的主成分分析技术对人脸形状和纹理变化建模, 并通过优化模型参数使之适配测试人脸图像。由于线性模型难以刻画复杂的人脸形状和纹理变化, 在大姿态、夸张表情、剧烈光照变化和部分遮挡下的效果欠佳。该问题的

最新进展是通过级联多个线性回归模型直接从人脸纹理特征预测人脸形状。

人脸识别中的主要难点在于人脸形态和纹理高度复杂性的特点, 为进一步提高算法的非线性回归能力以获得对形状等变化的鲁棒性, zhang 等<sup>[32]</sup>提出了一种由粗到细的深度非线性人脸形状提取方法(coarse-to-fine auto-encoders networks, CFAN)。CFAN 级联多个由栈式自编码网络实现的非线性回归模型, 每一级刻画从人脸表观到人脸形状的部分非线性映射。在对人脸图像识别过程中, 深度学习方法不仅能够提取有用的人脸纹理特征, 而且能够得到精确的人脸形状和几何结构信息。

### 3.3 基于深度学习的人脸姿态鲁棒性建模

人脸识别技术在可控条件及半可控条件下已经基本趋于成熟, 然而在非可控条件下, 由于人脸容易受到姿态、表情、年龄和遮挡等因素的影响, 识别率不高。其中, 姿态变化会导致极大的面部表观变化, 是对人脸识别影响最大的因素之一。

姿态变化导致的人脸表观变化是一种复杂的非线性变化, 利用 3D 模型生成虚拟图像的方式固然可以较好地解决不同姿态间的非线性变化问题, 但从 2D 图像恢复准确的 3D 模型非常困难。文献[33]提出了一种栈式渐进自编码(SPAE)神经网络模型, 以实现较小规模数据下对姿态变化的非线性建模。侧面图像到正面图像变化虽然非常复杂, 但却是缓慢平滑的。根据这一特点, 将侧面图像到正面图像的建模划分为若干子任务, 每个子任务仅负责将变化较大的姿态变换到变化较小的姿态而非直接变换到正面姿态, 由此控制了每个子问题的难度, 使用一个浅层的神经网络即可有效建模, 进而将多个浅层的神经网络叠联到一起即可得到一个深层的神经网络, 实现侧面图像到正面图像的平滑变换。这种渐进学习的思想将深度神经网络划分为若干浅层的网络, 使其模型能力与有限的的数据相匹配, 避免数据规模小带来的过学习问题。文献[2]是利用深度学习解决人脸识别姿态和分辨率上的问题, 具体研究了 DBN 在人脸姿态处理中的应用, 如姿态映射和姿态分类。人脸由于姿态的变化造成的旋转在人脸成像中带来的变化更不是一个简单的线性方法。而深度学习是一种多层的复杂神经网络, 它是一个良好的非线性的生成模型, DBN 可以学习侧面人脸图像到正面人脸图像的一个全局映射, 但个体细节差异易丢失。文献[48]采用局部二值模式提取人脸纹理信息, 将其纹理信息作为 DBN 的

输入实现分类,用以提供识别度更好的人脸特征,获得非限制条件下人脸图像的特征.文献[49]在RBM堆栈的顶层增加了一个回归层,用来在一个统一的深度学习框架下完成特征提取以及分类两种任务.深度神经网络有很强的非线性建模能力,可以采用深度学习来对人脸姿态鲁棒性建模,这些文献提供了一系列的解决思路.然而,深度学习需要大规模的有监督、多姿态人脸图像进行训练,而这类数据在实际中很难收集.

### 3.4 有约束环境中的全自动人脸识别

文献[50]用噪声调整线性单元组成的深度结构神经网络将深度学习应用于目标识别和人脸验证.文献[51]中的DBN是非常善于在人脸图像预测表达类别中应对遮挡,并且可以执行SIFT描述子用于区分不同类型的场景特征.基于SIFT分类的方法能得到较好的识别效果,缺点是特征计算过程复杂,点匹配过程时间长,有一定的限制.文献[52]将因式分解的三路RBM模型应用到多人脸图像匹配,匹配性均优于以前同类生成模型.文献[53]研究如何从局部遮挡的人脸图像解析面部成分,人脸被划分为若干互相重叠的块,每个块只与部分隐节点发生关联,在块等级上检测人脸,通过DBN执行训练过程,再用逻辑回归进行判别式调整,然后计算对像素敏感的标记映射.从LFW、BioID和CUFSF三个数据集中挑选2239个图像进行实验,说明了该方法的有效性,不仅对局部遮挡的人脸图像具有鲁棒性,提供了更丰富的人脸表情分析的人脸合成,人脸关键点检测.研究表明,在有约束环境中利用深度学习的方法可以自动学习到人脸特征,与浅层方法相比,可使复杂的特征提取工作更加简单,并且可以学习到人脸图像中的一些隐性规律和规则.

### 3.5 基于深度学习的视频监控下的人脸识别

在智能监控环境下,对可疑人物进行识别是人脸识别的一个重要用途.准确、快速地识别视频中人的身份对于视频搜索和视频监控具有非常重要的意义.虽然已开展了许多研究工作,但结果并不理想.近年出现了一些模拟相似场景的数据集,如美国NIST发布的(Point and Shoot Face Recognition Challenge, PaSC)视频人脸数据库<sup>[54]</sup>,在2014年生物特征识别国际联合会议(International Joint Conference on Biometrics 2014, IJCB'14)上进行首次测试,算法为特征概率弹性匹配模型,其中FAR为1%时正确识别率仅为26%.

2012年,文献[55]把视频数据里连续的帧作为卷积神经网络的输入数据,加上时间维度等信息,用来识别人的动作.文献[56]把视频数据里连续的帧进行数据整合,构造数据立方体,调节RBM,然后对深度模型的反馈微调,达到识别人脸的目的.在第11届国际人脸与姿态自动识别会议(the 11th IEEE Conference on Automatic Face and Gesture Recognition, FG2015)组织的对PaSC的再次评测中,山世光带领团队所采用的基于深度学习的技术方案有两个核心步骤:针对视频中每一帧人脸的卷积神经网络特征提取方法和集成视频片段中所有视频帧中人脸卷积神经网络特征的集合建模方法.该方法在FAR为1%时验证率分别达到了58%和59%,明显高于第二名的48%和38%.

深度学习的本质是学习到多层的非线性的函数关系,这种多层的非线性的函数关系使得人们能更好地对视觉信息进行建模,从而更好地理解图像和视频,更好地处理好诸如视频人脸目标和行为识别这类复杂的问题.上述研究成果表明了深度学习在准确、快速地视频人脸识别中起到一定的作用.

### 3.6 基于深度学习的低分辨率人脸识别建模

在实际应用中,采集到的人脸图像有多种姿态变化,其图像分辨率偏低,造成人脸图像识别性能迅速下降.姿态变化将非线性因素引入人脸识别中,目标对象具有丰富含义.

由于被监控的人群距离摄像头一般较远,导致被检测到的人脸区域较小,因此小尺寸和低质量的人脸图像识别性能下降,此种情况称为低分辨率人脸识别(low-resolution face recognition, LR FR).因为绝大多数的人脸识别算法在低分辨率人脸识别场合中的识别率不高,并且可供识别的人脸特征信息很少.应用卷积神经网络对视频中的低分辨率人脸进行处理,可以得到较好的实验效果<sup>[35]</sup>.文献[57]图像超分辨率的实验表明,该DNC(deep network cascade)方法可以逐渐迎合随之增加的网络层的低分辨率图像,并实现更好的视觉质量和数量性能.

现阶段,基于深度学习的低分辨率人脸识别模型一般是将人脸识别的问题归结为感兴趣区域的划分和如何对感兴趣区域进行分类两个子问题,因此低分辨率人脸检测问题比分类问题难度更大,更加复杂,对构建模型的性能要求也更高.在这一领域的发展过程中,深度学习本身的结构得到了改进,更多的模型侧重于优化训练方法与流程.在低分辨率人脸识

别的准确率不断提升的同时, 运行时间也相应减少, 从而使其能够更好地投入到实际应用中。

### 3.7 其他基于深度学习的人脸信息识别

人脸具有高度非刚性的特点, 在成像中形成了一个复杂的非线性的人脸数据。而深度学习是一个有效的非线性的神经网络模型, 在人脸的其他特征中也得到了广泛的应用。文献[58]将卷积神经网络特征用于人脸表情识别。表情识别过去采用与人脸识别类似的局部特征(如 HOG、SIFT、LBP、LGBP 等), 并在此基础上用支持向量机等分类器进行分类<sup>[30]</sup>。由于某些数据库缺少表情类别标签, 有些模型并不是针对表情识别精调的, 而是针对人脸识别精调的。文献[59]将可变形部分学习到 3D CNN 框架, 它可以检测到特定的面部动作部分下的结构性空间限制, 并同时获得判别部分的代表性。文献[60]引入深度信念网络(DBN)来模拟输入图像的数据分布, 完成对疲劳特征的分层自动抽取, 最后基于时间窗实现视频流图像的疲劳状态识别。文献[61]的基本思想是通过深度信念网络(DBN)的非监督贪心逐层训练来提取人脸图像的高层特征, 结合传统的图像预处理和相似性度量技术, 达到人脸核实的目的。文献[62]提出了融合深度信念网络和多层感知机的人脸表情识别的方法。

深度学习提供的是分布式的特征表示。在最高的隐含层, 每个神经元代表一个属性分类器, 也可以用来进行如男女性别<sup>[63]</sup>、人种和头发颜色<sup>[64]</sup>等的识别和分类。

## 4 分析与展望

相对于利用其他机器学习的人脸识别技术而言, 深度学习具有关键性的优势: 首先, 能从特征中检测复杂的相互作用; 其次, 可以从几乎没有处理的原始数据中学习低层次的特征; 而且, 它既能处理高基数类数据, 还可以处理未标记的数据。因此, 深度学习可以学习得到更有用的数据, 并且能构建更精确的模型。但是, 深度学习也有一些缺点, 例如训练模型的时间比较长, 需要不断地迭代来进行模型优化, 不能保证得到全局最优解等, 这些也需要在未来不断地探索。

目前深度学习的理论方面需要解决的主要问题有四个方面: 首先, 深度学习的理论极限在哪里, 是否存在某个固定的层数, 达到这个层数之后计算机可以真正实现人工智能; 其次, 如何决定某类问题深度

学习的层数和隐层节点的个数; 再次, 如何评价深度学习得到的特征的优劣; 最后, 对于梯度下降法如何进行改进以达到更好的局部极值点甚至是全局最优优点。

在实际的应用中, 深度学习虽然能够自动学习模式的特征, 并可以达到很好的识别精度, 但这种学习算法工作的前提是用户能够提供“相当大”量级的数据库。所以在小数据库的应用场景下, 深度学习算法不能够对数据的规律进行无偏差的估计, 因此在人脸识别效果上可能还不如一些已有的传统人脸识别算法。

由于不同的识别系统适用于不同的实际环境, 同时不同的深度学习结构有不同的实际要求。利用深度学习的人脸识别在各个方面取得了一定的进展, 后续研究方向如下:

(1) 在无标记数据的特征学习方面, 大规模人脸搜索相关应用受到学术界和高科技企业越来越多的关注, 除了利用人脸生物特征的特殊性进行传统的人脸识别, 应更多地关注对无标记数据进行自动添加标签技术的研究, 进而快速准确的在大规模人脸数据库中搜索到相似的人脸。

(2) 应进一步丰富数据库的资源。在引入外部训练数据且无约束训练协议的情况下, 鉴于在大规模人脸数据库 LFW 人脸识别率已经达到 95% 以上, 如果继续扩充此人脸数据库, 并设计一个更合理的全面评测协议, 这将会是一个非常具有挑战性的工作。

(3) 利用深度学习的人脸识别要与其他方法结合。在允许使用特殊设备的情况下, 可以考虑采用 3D 模型和深度信息以提高系统的稳定性; 对姿态变化、遮挡及表情变化引起的表观变化, 也许需要重新设计目前的人脸设计方式, 采用多个局部模型而非一个整体模型来进行表示<sup>[65]</sup>; 深度学习还可以结合一些人脸姿态或表情矫正方法, 进行相关的探索。

就目前的发展趋势来看, 深度学习将是未来人脸识别的发展方向。新的方法如强化学习与深度学习相结合, 已经取得了很大成功<sup>[66]</sup>。由于目前深度学习的计算复杂度较高, 在训练过程中, 需要 GPU 等并行设备支持, 而且训练时间比较长。在测试过程中, 由于模型参数过多, 并不方便在嵌入式设备上直接运行。使用快速的学习方法如极限学习<sup>[67-68]</sup>等方法构建深度学习单元将会是一个研究方向。现在基于深度学习的算法一般是通过云服务提供, 如 Face++、Microsoft Project Oxford 等, 这些也是未来的趋势。

致谢: 本文受天津食品安全低碳制造协同创新中心资助, 特此感谢!

### 参考文献:

- [1] 吴巾一, 周德龙. 人脸识别方法综述[J]. 计算机应用研究, 2009, 26(9): 3205–3209.
- [2] 林妙真. 基于深度学习的人脸识别研究[D]. 大连: 大连理工大学, 2013.
- [3] Hinton G, Osindero S, Teh Y. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527–1554.
- [4] 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述[J]. 计算机应用研究, 2012, 29(8): 2806–2810.
- [5] Lee H, Grosse R, Ranganath R, et al. Unsupervised learning of hierarchical representations with convolutional deep belief networks[J]. Communications of the ACM, 2011, 54(10): 95–103.
- [6] Huang G B, Lee H, Miller E L. Learning hierarchical representations for face verification with convolutional deep belief networks[C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 223–226.
- [7] Lee H, Grosse R, Ranganath R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations[C]//Proceedings of International Conference on Machine Learning. 2009: 21–30.
- [8] Liu Y, Zhou S S, Chen Q C. Discriminative deep belief for visual data classification[J]. Pattern Recognition, 2011, 44(10/11): 2287–2296.
- [9] Lecun Y, Boser L, Denker J S, et al. Backpropagation applied to handwritten zip code Recognition[J]. Neural Computation, 1989, 1(4): 541–551.
- [10] Lecun Y, Botou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [11] Zhou S S, Chen Q C, Wang X L. Convolutional deep networks for visual data classification[J]. Neural Process Letters, 2013, 38(1): 17–27.
- [12] Ouyang W, Wang X. Joint deep learning for pedestrian detection[C]//Proceedings of International Conference on Computer Vision. Piscataway: IEEE, 2013: 2056–2063.
- [13] Ouyang W, Luo P, Zeng X, et al. DeepID-Net: Multi-stage and deformable deep convolutional neural networks for object detection[EB/OL]. (2014-09-11) [2016-02-23]. <http://arxiv.org/abs/1409.3505v1>.
- [14] Cootes T F, Taylor C J, Cooper D H, et al. Active shape models: Their training and application[J]. Computer Vision and Image Understanding, 1995, 61(1): 38–59.
- [15] Cootes T F, Taylor C J. Statistical models of appearance for medical image analysis and computer vision[J]. Proceedings of SPIE, 2004, 4322(1): 236–248.
- [16] Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71–86.
- [17] Huang G B, Lee H, Learned-Miller E. Learning hierarchical representations for face verification with convolutional deep belief networks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 2518–2525.
- [18] Taigman Y, Yang M, Ranzato M, et al. Deepface: Closing the gap to human-level performance in face verification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 1701–1708.
- [19] Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10000 classes[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 1891–1898.
- [20] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 815–823.
- [21] Huang G B, Ramesh M, Berg T, et al. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments[R]. Workshop on Faces in ‘Real-Life’ Images: Detection, Alignment, and Recognition, 2007: 1–14.
- [22] Belhumeur P N, Jacobs D W, Kriegman D J, et al. Localizing parts of faces using a consensus of exemplars[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2930–2940.
- [23] Chen D, Cao X D, Wang L, et al. Bayesian face revisited: A joint formulation[J]. European Conference on Computer Vision, 2012, 7574(1): 566–579.
- [24] Sun Y, Wang X, Tang X. Hybrid deep learning for computing face Verification[C]//Proceedings of IEEE International Conference on Computer Vision. Piscataway: IEEE, 2013: 1489–1496.
- [25] Taigman Y, Yang M, Ranzato M, et al. Web-scale train-



- ing for face identification[EB/OL]. (2014-06-20) [2016-02-23]. <http://arxiv.org/abs/1406.5266>.
- [26] Sun Y, Wang X D, Tang X. Deep learning face representation from predicting 10,000 classes[C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 1891-1898.
- [27] Sun Y, Wang X G, Tang X. Deeply learned face representations are sparse, selective, and robust[EB/OL]. (2014-12-03) [2016-02-23]. <http://arxiv.org/abs/1412.1265>.
- [28] Sun Y, Wang X D, Tang X. Deep learning face representation by joint identification-verification[J]. Advances in Neural Information Processing Systems, 2014, 27: 1988-1996.
- [29] Best-Rowden L, Han H, Otto C, et al. Unconstrained face recognition: Identifying a person of interest from a media collection[J]. IEEE Transactions on Information Forensics and Security, 2014, 9(12): 2144-2157.
- [30] Liu M Y, Li S X, Shan S G. AU-inspired deep networks for facial expression feature learning[J]. Neurocomputing, 2015, 159(1): 126-136.
- [31] Chen J, Chen X L, Yang J, et al. Optimization of a training set for more robust face detection[J]. Pattern Recognition, 2009, 42(11): 2828-2840.
- [32] Zhang J, Shan S G, Kan M N, et al. Coarse-to-fine auto-encoder networks(CFAN) for real-time face alignment [M]//Computer Vision-ECCV 2014. Springer International Publishing, 2014: 1-16.
- [33] Kan M N, Shan S G, Chang H, et al. Stacked progressive auto-encoder(SPAE) for face recognition across poses [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014, 1883-1890.
- [34] Liu M Y, Shan S G, Wang R P, et al. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 1749-1756.
- [35] 李绍欣, 山世光, 阚美娜, 等. 基于多视统一子空间的跨姿态人脸识别应用技术研究[J]. 警察技术, 2014(5): 8-11.
- [36] Liao S C, Lei Z, Yi D, et al. A benchmark study of large-scale unconstrained face recognition[C]//Proceedings of IEEE International Joint Conference on Biometrics. Piscataway: IEEE, 2014: 1749-1756.
- [37] Yi D, Lei Z, Liao S C, et al. Deep metric learning for person re-identification[C]//Proceedings of the 22nd International Conference on Pattern Recognition. Piscataway: IEEE, 2014: 24-28.
- [38] Yan J J, Lei Z, Yi D, et al. Learn to combine multiple hypotheses for accurate face alignment[C]//Proceedings of IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2013: 392-396.
- [39] Hu Y, Liao S C, Lei Z, et al. Exploring structural information and fusing multiple features for person re-identification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2013: 794-799.
- [40] Yi D, Lei Z, Li S Z. Towards Pose Robust Face Recognition[C]//Proceedings of the 26th IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3539-3545.
- [41] Zhao J D, Lu K, He X F. Locality sensitive semi-supervised feature selection[J]. Neurocomputing, 2008, 71(10): 1842-1849.
- [42] He X F, Sun C Y. Special issue on advances in intelligence science and intelligent data engineering[J]. Neurocomputing, 2014, 139(4860): 1-2.
- [43] Chen B, Polatkan G, Sapiro G, et al. Deep Learning with hierarchical convolutional factor analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1887-1901.
- [44] Chen B, Polatkan G, Sapiro G, et al. The hierarchical beta process for convolutional factor analysis and deep learning[C]//Proceedings of the 28th International Conference on Machine Learning. Piscataway: IEEE, 2011: 361-368.
- [45] 许可. 卷积神经网络在图像识别上的应用的研究[D]. 杭州: 浙江大学, 2012.
- [46] Kumar N, Berga A C, Belhumeur P N, et al. Attribute and simile classifiers for face verification[C]//Proceedings of the 12th International Conference on Computer Vision, Piscataway: IEEE, 2009: 365-372.
- [47] Lu C C, Tang X. Surpassing human-level face verification performance on LFW with GaussianFace [EB/OL]. (2014-04-15) [2016-02-23]. <http://arxiv.org/abs/1404.3840>.
- [48] 梁淑芬, 刘银华, 李立琛. 基于 LBP 和深度学习的非限制条件下人脸识别算法[J]. 通信学报, 2014, 35(6):

- 154-160.
- [49] 王莹,樊鑫,李豪杰,等. 基于深度网络的多形态人脸识别[J]. 计算机科学,2015,42(9):61-65.
- [50] Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines[C]//Proceedings of the 27th International Conference on Machine Learning. Piscataway: IEEE, 2010: 807-814.
- [51] Ranzato M, Susskind J, Mnih V, et al. On deep generative models with applications to recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2011: 2857-2864.
- [52] Susskind J, Hinton GE, Memisevic R, et al. Modeling the joint density of two images under a variety of transformations[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2011: 2793-2800.
- [53] Luo P, Wang X G, Tang X. Hierarchical face parsing via deep learning[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 2480-2487.
- [54] Beveridge J R, Phillips P J, Bolme D S, et al. The challenge of face recognition from digital point-and-shoot cameras[C]//Proceedings of the 6th International Conference on Biometrics: Theory, Applications and Systems. Piscataway: IEEE, 2013: 1-8.
- [55] Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for human action recognition[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231.
- [56] 由清圳. 基于深度学习的视频人脸识别方法[D]. 哈尔滨: 哈尔滨工业大学, 2012.
- [57] Cui Z, Chang H, Shan S, et al. Deep network cascade for image super-resolution[M]//Computer Vision-ECCV 2014. Springer International Publishing, 2014: 49-64.
- [58] 王剑云, 李小霞. 一种基于深度学习的表情识别方法[J]. 计算机与现代化, 2015(1): 84-87.
- [59] Liu M Y, Li S X, Shan S G, et al. Deeply learning deformable facial action parts model for dynamic expression analysis[M]//Computer Vision-ACCV 2014. Springer International Publishing, 2014: 143-157.
- [60] 周慧, 周良, 丁秋林. 基于深度学习的疲劳状态识别算法[J]. 计算机科学, 2015, 42(3): 191-200.
- [61] 张媛媛, 霍静, 杨婉琪, 等. 深度信念网络的二代身份证异构人脸核实算法[J]. 智能系统学报, 2015, 10(2): 193-200.
- [62] 施徐敢, 张石清, 赵小明. 融合深度信念网络和多层感知器的人脸表情识别[J]. 小型微型计算机系统, 2015, 36(7): 1629-1632.
- [63] Li S X, Xing J L, Niu Z H, et al. Shape driven kernel adaptation in convolutional neural network for robust facial traits recognition[C]//Proceedings of the 28th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 222-230.
- [64] 王丹, 山世光, 张洪明, 等. 一种由粗到细的头发表分割方法[J]. 软件学报, 2013, 24(10): 2391-2404.
- [65] Chen D, Cao X, Wen F, et al. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3025-3032.
- [66] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [67] Wang Y Q, Xie Z G, Xu K, et al. An efficient and effective convolutional auto-encoder extreme learning machine network for 3d feature learning[J]. Neurocomputing, 2016, 174: 988-998.
- [68] Tissera M D, McDonnell M D. Deep extreme learning machines: Supervised autoencoding architecture for classification[J]. Neurocomputing, 2016, 174: 42-49.

责任编辑: 常涛