

深度学习算法研究进展

田启川¹, 王满丽²

1. 北京建筑大学 电气与信息工程学院, 北京 100044

2. 北京建筑大学 建筑大数据智能处理方法研究北京市重点实验室, 北京 100044

摘要:深度学习是人工智能领域的研究热点,利用深度学习支持人工智能的研究工作已经是必然趋势,在图像、语音、文本等领域已经展现出性能优势。对深度学习相关文献进行了分析研究,介绍了深度学习的概念和方法框架,综述了当前深度学习模型及其应用情况,分析了深度学习需要突破的瓶颈,指出了深度学习未来的研究方向。

关键词:深度学习;卷积神经网络(CNN);建模;人脸识别

文献标志码:A **中图分类号:**TP391.41 **doi:**10.3778/j.issn.1002-8331.1908-0359

田启川,王满丽.深度学习算法研究进展.计算机工程与应用,2019,55(22):25-33.

TIAN Qichuan, WANG Manli. Research progress on deep learning algorithms. Computer Engineering and Applications, 2019, 55(22):25-33.

Research Progress on Deep Learning Algorithms

TIAN Qichuan¹, WANG Manli²

1. School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

2. Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

Abstract: Deep learning is a hot research topic in the field of artificial intelligence, and it is an inevitable trend to use deep learning to support the research of artificial intelligence. Its performance advantages have been demonstrated in the fields of image, voice and text. Based on the analysis of deep learning related literature, this paper reviews the application of the current deep learning model, involves the concept and method framework of deep learning, analyzes the bottlenecks to be broken currently, and points out the future research direction of deep learning.

Key words: deep learning; Convolutional Neural Network(CNN); modeling; face recognition

1 引言

所谓深度学习,就是在神经网络的基础上叠加隐层层数的一种学习网络,深度学习是随着信息时代的发展而诞生的。1998年神经网络因在训练过程中遇到了“梯度爆炸”问题,使神经网络训练中越远离输出层的参数越难以训练,神经网络研究进入了低潮。为了解决神经网络训练问题,2006年Hinton等人提出深度信念网络的概念和新的训练方法,克服了神经网络训练的瓶颈问题,掀起了神经网络研究热潮^[1]。此后随着互联网科技

信息时代的快速发展,深度学习的应用也得到快速扩展。

近年来,深度学习发展迅猛,已经被广泛应用于图像识别、语音识别、视频分析、文本分析和大数据分析等领域,并取得了成功。在图像识别范畴中,近年来将深度学习应用于人脸和自然图像的识别,大大提高了图像识别的准确性^[2]。在语音识别领域中,2012年谷歌的语音识别模型全部更换为深度学习模型之后,其错误率降低了约20%^[3]。苹果公司的Siri语音系统尤为典型,百度也开发出自己的智能音箱,微软公司也利用深度学习

基金项目:北京市教育委员会科技发展计划面上项目(No.KM201410016016);北京建筑大学科学研究基金(No.00331614021);北京建筑大学研究生创新项目。

作者简介:田启川(1971—),男,博士,教授,研究方向为人工智能、模式识别、计算机视觉与图像处理、嵌入式系统与物联网技术等,E-mail:tianqichuan@bucea.edu.cn;王满丽(1995—),女,硕士生,研究方向为模式识别、建筑电气与智能化。

收稿日期:2019-08-23 **修回日期:**2019-09-30 **文章编号:**1002-8331(2019)22-0025-09

CNKI网络出版:2019-10-10, <http://kns.cnki.net/kcms/detail/11.2127.TP.20191010.1107.006.html>

技术开发了同声传译系统,取得技术上的巨大突破^[4]。在文本领域,微软推出了可以写诗的程序,它通过“阅读”大量诗集便可学会写诗^[5]。这些应用成果都让人们相信在不久的将来,以深度学习为基础工具的各个领域的研究一定会给生活带来更大的惊喜。

针对深度学习研究,当前仍然处于比较基础的研究和浅显的应用阶段,文章从深度学习发展现状、典型深度学习网络、深度学习的应用情况以及未来的研究方向这几方面对深度学习进行介绍。

2 深度学习发展现状

2.1 神经网络的曲折发展

在计算机领域,深度学习经常被应用在算法改进中。严格来讲,深度学习的出现是因为神经网络的发展。神经网络的概念最早出现在1943年,计算结构这一概念便是神经网络概念的前身,它可以大致实现模拟人类神经元的工作原理,但是权重参数需要进行手动调节,因此使用起来非常不方便^[6]。神经网络的研究高潮是由感知机带来的。早在1958年,Rosenblatt就提出了感知机模型,这是支持神经网络未来发展和向量机的基础^[7]。感知机模型在计算结构的基础上改进了手动调节权重的问题,升级为自动调节权重值,相比之下便利了很多,但是它解决问题的能力有限,单层的结构特点仅仅能解决线性分类问题,它不能通过其单层结构来训练多层的感知机,因此仍然有一定的局限性^[8]。人们明确单层感知机存在的问题之后,试图通过增加隐层层数来构造多层感知机,这理论上已经是一个非常完美的分类器,但是一直找不到合适的学习算法来完成隐藏层数的权值训练,因此此后的十多年里,神经网络处于进退两难的状态^[9]。

1986年,由Rumelhart教授的团队提出了反向传播算法(Back Propagation, BP),该算法真正意义上解决了感知机模型的问题,使神经网络开始了新的实践时期。BP算法的提出将神经网络研究再一次推向了高潮^[10]。1989年,LeCun基于BP算法提出了真正意义上的适用于深度学习训练的算法——卷积神经网络(Convolutional Neural Network, CNN)。他的实验是用近10 000个由美国邮政系统提供的手写数字作为训练数据来训练神经网络,并在测试样本中获得了错误率仅仅5%的好结果,这一突破将神经网络带入了实用阶段^[7]。但实践发现该算法仍然有局限性,最首要的问题就是“梯度爆炸”问题,即神经网络系统中越远离输出层的参数越难以被训练,且层数越多这个问题越明显。与此同时,LeCun及其同事也采用支持向量机(Support Vector Machine, SVM)进行实验,再次用手写邮票验证,错误率可以降低到0.8%,性能优于神经网络,因此神经网络的发展脚步再次放缓^[7]。

2.2 深度学习的来临

Hinton在2006年发表了一篇对神经网络具有突破性的文章,描述了一种成功的多层神经网络训练方法,称为深度信念网络,采用分层初始化的方法对深度神经网络进行训练,这样容易训练模型,这个训练方法解决了BP算法的瓶颈问题^[11]。因而,神经网络完成了由浅层网络到多层深度网络的转变,深度学习概念也由此诞生,而且越来越广泛地应用在各个领域。就目前的研究成果而言,如果数据集够全面,隐层层数够多,其训练结果都是很好的。近年来,许多研究机构如百度、谷歌、微软等公司都建立了深度学习研究院,掀起了深度学习的研究高潮。

深度学习本质上就是一个不断进行特征描述的过程,即分层级的特征提取过程,特征提取得越多、越准确,则容易达到过拟合(过分依赖训练数据),特征提取得越少、越模糊,则容易达到欠拟合(未能充分学习到训练数据之间的关系)。因此如何利用海量数据来训练模型参数,使得其训练结果达到准确度的峰值便是深度学习的难点所在。

近年来国内外的深度学习领域研究出现较多的关键词和发表的文献趋势如图1、图2所示。

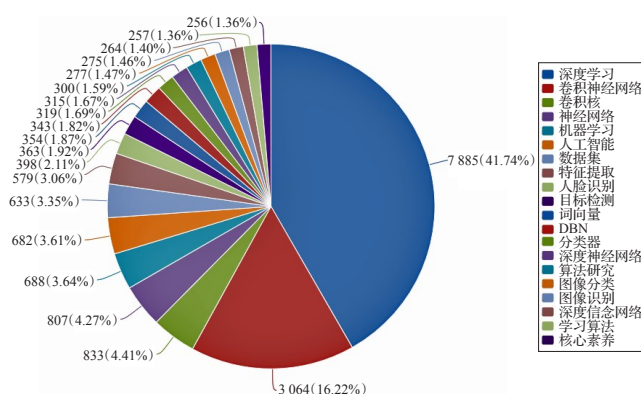


图1 深度学习研究关键词占比

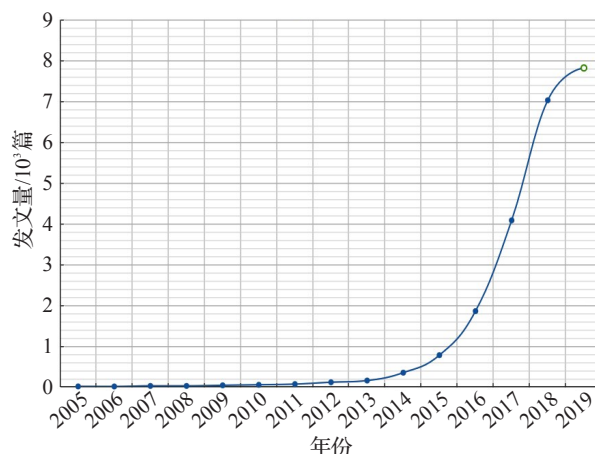


图2 深度学习主题论文发表数量趋势

由图1可见,迄今为止,以深度学习为主题词的研究论文中关键词出现频率由高到低为深度学习

(41.74%)、卷积神经网络(16.12%)、卷积核(4.41%)、神经网络(4.27%)、机器学习(3.64%)、人工智能(3.61%)、数据集(3.35%)、特征提取(3.06%)、人脸识别(2.11%)、目标检测(1.92%)、词向量(1.87%)、DBN(1.82%)、分类器(1.69%)、深度神经网络(1.67%)、算法研究(1.59%)、图像分类(1.47%)、图像识别(1.46%)、深度信念网络(1.40%)、学习算法(1.36%)和核心素养(1.36%),涉及到人工智能的各个领域范畴,推动了人工智能的发展。

由图2可见,2005年到2010年文献发表较少,2011年到2013年有一个较小的斜率呈上升趋势,平均一年发表大约100篇文献。2013年到2018年的文献数量呈指数函数图像形状上升,其中斜率最大的点在2017年。发文量反映了研究状况越来越热,深度学习的成功案例和一些比赛取得的成果吸引了大量的研究人员,因此文献成果猛增。但现在发文量有趋于平稳的趋势,说明研究人员人数基本饱和,增量趋于平稳,深度学习研究有条不紊地进行。

3 主流的深度学习网络

针对不同的应用目的和应用领域,人们提出了一些不同的神经网络。20世纪80年代提出的卷积神经网络(CNN)是经典深度学习网络,在计算机飞速发展和海量数据的背景下,人们不断尝试用不同的方法来训练CNN。特别地,Krizhevsky等人在2012年提出了一种经

典CNN架构,证明了该结构在特征提取问题上的潜力,在图像领域取得了重要突破^[11]。2014年Girshick等人提出了R-CNN模型,它最早是被用来进行物体检测的一种深度学习算法^[12]。下面分别介绍几种具有代表性的神经网络深度学习框架,其汇总对比见表1。

3.1 卷积神经网络

最典型的深度学习模型就是1989年提出的卷积神经网络,其网络结构如图3所示,卷积神经网络的基本架构通常包括特征抽取器和分类器^[13]。在卷积神经网络中,特征抽取器通常由若干个卷积层和池化层共同叠加组成,卷积和池化过程可以连续地缩小特征图,同时也会导致特征图数目的增加。分类器连接在特征抽取器的后面,通常由多层感知机组成。在最后一个特征抽取器后面,将所有的特征图展开并排列成一个一维的特征向量,该特征向量作为输入连接分类器。

至于参数选择,以卷积核选取为例,在达到相同感受野的情况下,采用较小的卷积核和增加卷积层比采用较大的卷积核和较少卷积层所需的参数和计算量少,但是卷积核必须大于1才能有提升视野的作用。卷积模板通常为奇数,因为偶数大小的卷积核不能保证锚点刚好在中间,这样进行滑动卷积时容易产生位置偏移。一般选择大小为3×3的卷积核。卷积层数需要根据数据特点、尺寸、模板大小以及训练效果进行优化。

表1 主流网络结构对比

名称	适用范围	优点	缺点
CNN	大多数领域都可以,在大型图像处理应用中表现出色	1. 模型泛化能力强; 2. 不对输入数据的平移不变性做高要求	1. 容易出现梯度消散情况; 2. 空间关系辨识度差; 3. 物体大幅度旋转后识别能力下降
RNN	适用于处理序列数据,如音频、语言等	可以对序列内容建模,如语音、视频	1. 参数太多; 2. 容易出现梯度消散、梯度爆炸问题; 3. 不具备特征学习能力
GAN	适用于需要生成样本的场景,如图像建模	1. 能够产生更好的样本; 2. 可以训练任何一种生成器网络; 3. 避开了复杂的马尔科夫过程	1. 模型收敛性差; 2. 训练过程中容易出现崩溃; 3. 模型过于自由不可控
FC	适用于特征简单的场景,如房价预测、广告推荐等	当训练集和测试集数据有较大差异时,能保证较大的模型有良好的迁移能力	1. 参数太多,训练速度慢; 2. 图像变大,色彩数变多; 3. 不便处理高维度数据

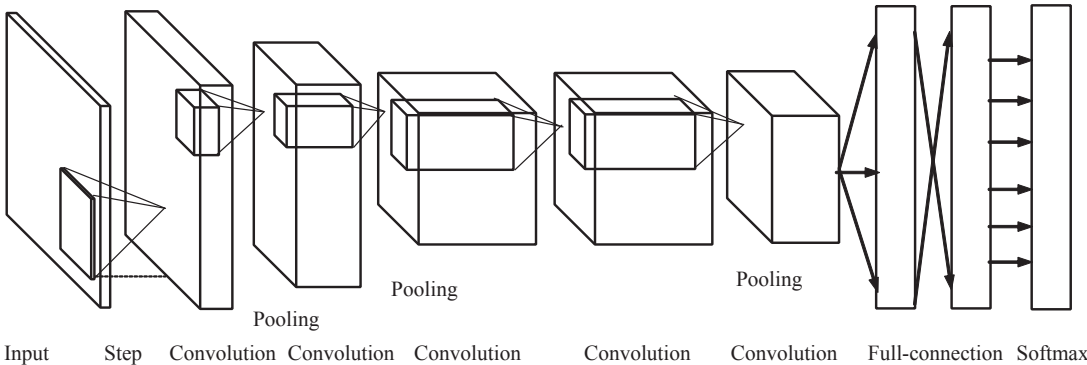


图3 卷积神经网络原理图

3.2 循环神经网络

循环神经网络(Recurrent Neural Network, RNN), 也称作递归神经网络, 是1986年提出的一种神经网络结构, 其结构原理如图4所示, 主要用于序列数据的特征提取, 例如音频分析和语言识别等^[14]。

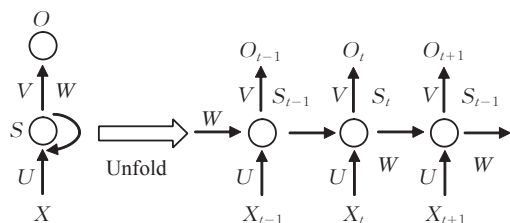


图4 循环神经网络原理图

相比于其他网络, 循环神经网络在机器翻译和语音识别领域有更加出色的表现^[15]。其主要功能是进行特征提取, 优点是可以对序列内容建模, 缺点是训练参数较多, 调节不方便, 可能会出现梯度爆炸和梯度消散的情况, 而且该网络不具备特征学习的能力。

3.3 生成对抗网络

2014年, Goodfellow等人提出了一个通过对抗过程估计生成模型的框架——生成对抗网络(Generate Adversarial Network, GAN)。该框架需要训练两个模型: 用来捕获数据分布的生成模型G(Generative Model), 以及用来估计样本来自于训练数据的概率的判别模型D(Discriminative Model)^[16]。其网络结构如图5所示。

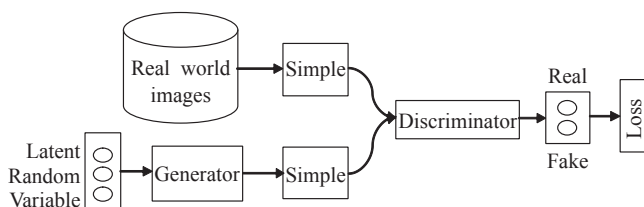


图5 生成对抗网络原理图

生成对抗网络的主要功能是生成对抗样本, 与其他模型相比, 它的优点是可以生成较好的样本, 能够训练任何一种生成器网络, 能有效避免马尔科夫过程和近似计算概率的问题^[17]。缺点是模型的收敛性较差, 在训练过程中, 容易出现崩溃的问题, 因此存在一定的不可控性。

3.4 全连接网络结构

全连接网络结构(Full Connected, FC)是深度学习网络中最基本的方法之一, 早期它常用来进行特征提取和分类识别, 但它的相邻层关系紧密, 且全连接层结构的参数最多, 对存储空间和计算空间需求量大^[18]。参数过多导致了全连接网络结构很少会应用到复杂的场景中, 多数应用于特征简单的场景, 如房价估算等, 其网络原理图如图6所示。

3.5 基于CNN的改进网络模型

一种名为LeNet5的卷积神经网络, 1994年由誉为

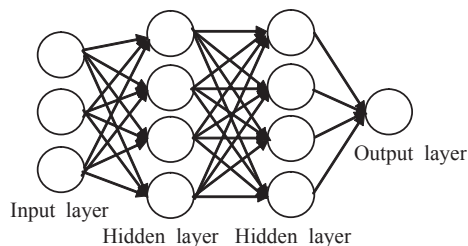


图6 全连接网络结构原理图

“卷积神经网络之父”的LeCun提出, LeNet5认为图像具有很强的空间相关性, 每个像素作为一个大型多层神经网络的输入, 通过不同卷积提取空间的相关性特征^[19]。LeNet5的诞生推动了深度学习的研究进展。

后来, 对于不同的分类问题, LeNet5不断被改进, 其中比较成功的模型有AlexNet^[20]。它所做的改进是选择ReLU作为激活函数, 采用Dropout技术、数据增强技术、多GPU平行训练技术等, 有效避免过拟合现象, 网络模型的收敛速度稳定, 能避免或抑制网络训练时的梯度消失现象, 模型训练速度较快, 因为其具有更深的网络结构, 所以具有更多的参数, 计算量增大。AlexNet证明了CNN在复杂模型下的有效性, 利用CPU能够在可接受的时间范围内得到训练结果。

VGGNet模型是2014年诞生的, 意在使用更深的网络, 从而得到更好的训练结果^[21]。VGGNet模型使用 1×1 和 3×3 的小卷积核。小卷积核使判决函数更具有判决性, 具有更少的参数, 增加了非线性表达能力。不同于LeNet5和AlexNet, VGGNet使用了更小的滤波器, 多次卷积达到更大的感受野(如 5×5)类似效果, 以提取更多的复杂特征。这种思想也被许多新生网络所采纳, 如ResNet。

同期出现的还有GoogLeNet模型, 相比于VGGNet, 它充分展现了“没有最深, 只有更深”的道理^[22]。GoogLeNet引入Inception结构代替了单纯的卷积+激活的传统操作技术, 引入Inception结构, 使用 1×1 卷积核来降维, 解决计算量大的问题, 中间层使用LOSS单元作为辅助更新网络。全连接层全部替换为简单的全局平均pooling, 虽然网络深, 但参数只是AlexNet的1/12。后续也有很多网络在它的基础上进行改进, 如GoogleNet-PNN, 将卷积神经网络与概率神经网络结合应用, 不需要人工设计和手工提取特征, 提高了分类的效率, 并且避免了深度神经网络在小数据集上产生的过拟合现象, 提高了深度学习模型的识别精度和泛化能力, 得到了较高的识别率^[23]。

ResNet模型是2015年提出的, 通过引入残差块、平均池化、引入残差单元、直接将输入信息绕道传到输出, 保护了信息的完整性; 整个网络只需要学习输入、输出存在差别的部分, 简化了学习目标和难度, 在一定程度上解决了信息传递时存在信息丢失、损耗的问题; 同时

解决了梯度消失或者梯度爆炸导致很深的网络无法训练等问题^[24]。

基于CNN改进模型的对比分析详见表2。

4 深度学习的应用

深度学习提出后被广泛应用于人工智能研究的各个领域,并且取得了不错的成效,特别是在图像、语音、视频、文本和数据分析这五个领域的应用效果尤为明显,但是深度学习所涉及的领域远不止这些。下面重点介绍深度学习在图像、语音和文本领域的应用情况。

4.1 图像领域

深度学习网络具有特征提取能力强、识别精度高的特性,特别是在图像分类、人脸识别研究中表现了其优异的性能。传统识别方法模型结构简单,计算量小,不需要大量图像样本作基础就可以完成识别,但传统识别方法的识别结果受环境影响大,识别率不稳定^[25]。通过深度学习方法可以获得图像的多描述特征,受外界干扰

因素影响小,识别准确率高且稳定。深度学习方法需要有大量图像作为训练样本进行特征学习,计算量大,模型训练时间长,对硬件设备要求高,资源消耗大,模型训练过程中容易发生“梯度消散”和坍塌。

LeCun 和他的同事最先将CNN应用在手写邮政编码的小图像上进行识别,取得了令人满意的效果,但是在大数据上的效果不是很理想,因此当时深度学习未能在图像领域引起重视^[7]。自从2006年Hinton发表关于深度网络训练的文章后,深度学习在图像识别领域开始不断有新的算法被提出,研究者在网络训练中通过引入权重衰减的概念,有效防止了过拟合。此外,在图像领域能有大的突破的关键因素是计算机性能的提升,这样才可以支持训练更多数据。2013年,ImageNet大赛中前20名算法都采用了深度学习,可见深度学习在图像领域中取得了绝对的优势。

深度学习在图像领域中的应用有很多,比如医疗诊断、动作识别^[26]等,最典型的应用就是人脸识别。人脸

表2 基于CNN改进模型对比分析

模型名称	改进方式	优缺点	意义
LeNet5	1. 卷积神经网络使用三个层作为一个序列:卷积、池化、非线性; 2. 使用卷积提取空间特征; 3. 池化层的采样方式为映射到空间均值下采样; 4. 使用双曲正切或S型形式的非线性池化函数; 5. 使用多层神经网络作为最后的分类器; 6. 层与层之间的稀疏连接矩阵避免了高额的计算成本	网络结构简单,模型深度较浅,计算量小,图像特征提取能力一般,训练过程中容易出现过拟合	LeNet5 标志着 CNN 的真正问世,是近几年来大量网络架构的起源,为现代深度学习领域的发展做了重要铺垫
AlexNet	1. 使用 Dropout 技术按照一定概率随机丢弃单个神经元,避免模型过拟合; 2. 使用效果更好的有重叠的最大池化代替或避免平均池化; 3. 使用数据增强方式增加训练样本; 4. 设计了 LRN(Local Response Normalization)层,利用邻近的数据做归一化; 5. 使用多 CPU 并行计算,大幅度减少了训练时间,反过来允许使用更大的数据集和更大的图像进行训练	有效避免过拟合现象,网络模型的收敛速度会相对稳定,能避免或抑制网络训练时的梯度消失现象,模型训练速度较快;具有更深的网络结构,计算量增大,具有更多的参数	AlexNet 证明了 CNN 在复杂模型下的有效性,并利用 GPU 使得训练时间在可接受的范围内。AlexNet 的成功掀起了一场卷积神经网络的研究热潮,极大地促进了卷积神经网络的研究和发展
VGGNet	1. 使用更小的滤波器(如 3×3); 2. 依次采用多个卷积,能够达到与更大的感受野类似的效果	小卷积核使判决函数更具有判决性,具有更少的参数,增加了非线性表达能力,网络结构更深,计算量更大,可以提取更多复杂特征和这些特征的组合	VGGNet 继承了 AlexNet 的理念,意在用更深的网络来获取更好的训练结果。通过使用多个卷积来提取更复杂的特征这一思想后来被许多新生网络采纳
GoogLeNet	1. GPU 平行训练技术; 2. 引入 Inception 结构代替了单纯的卷积+激活的传统操作技术; 3. 引入 1×1 卷积层降维减少通道; 4. 加入辅助分类器	解决计算量大的问题,中间层使用 LOSS 单元作为辅助更新网络,全连接层全部替换为简单的全局平均池化,参数更少,虽然网络深,但参数只是 AlexNet 的 1/12	GoogLeNet 进一步阐释了“没有最深,只有更深”的道理,一共 22 层的网络参数却只是 AlexNet 的 1/12,为其他网络提供了新思路
ResNet	引入残差单元,通过直接将输入信息绕道传到输出,保护信息的完整性,整个网络只需要学习输入、输出差别的那一部分,简化学习目标和难度	在一定程度上解决了信息传递时或多或少会存在信息丢失、损耗等问题,也解决了很深的网络无法训练等问题	ResNet 的结构设计使得训练误差会随着层数的增大反而减小,可用于训练上百乃至近千层的卷积神经网络,对深度学习发展研究有着深远意义

识别是用来解决日常生活中普遍面临的一个身份识别的技术,已大量应用在门禁系统、安全监控和电子商务等各个方面。人脸识别过程是由人脸图像采集、人脸图像定位、人脸图像预处理和搜索识别等步骤组成的^[27]。在深度学习出现之前,人脸识别一直以PCA(Principal Component Analysis)、LDA(Linear Discriminant Analysis)、LBP(Local Binary Pattern)算法或者其改进算法为主,直接在图像上提取人脸的描述特征,并用于身份识别。由于上述传统识别图像算法的局限性和主观性,很多研究中无法达到应用要求,因此深度学习在图像识别应用后很快便应用于人脸识别当中。针对在非约束条件下许多算法的人脸识别性能差、特征判别能力弱等问题,郑健等人提出一种基于深度学习的改进人脸识别算法,实现了非约束图像的人脸检测和人脸归一化,减少了模型的干扰,提高了训练图像的人脸信息^[28]。实验结果表明该深度学习算法在LFW标准测试集上实现了较高的识别率,但是其测试数据过于理想,未考虑到实际应用中人脸识别情况存在的光照、遮挡等问题,因此该训练模型的鲁棒性还不足以让它应用到实际场景中。若需符合实际场景应用,应增加训练样本达到多样化。

杨恢先等人针对传统的归一化指数损失函数(即Softmax损失)识别特征能力低下,无法识别人脸特征的问题,提出了一种聚合判别多任务学习算法^[29]。该算法通过去除无关人脸识别的区域、人脸检测和人脸对齐操作,再利用深层卷积神经网络和多任务学习算法进行人脸特征提取,并将采集到的人脸特征向量分解为学习类内向量和判别类向量,增强了对类内特征^[30]的约束,提高了类间特征可分性,从卷积神经网络提取的特征向量分解为两个分量,MNIST数据集的深度可视化证明了所提方法的有效性,探讨超参数、网络结构变化和特征规范化对验证精准度的影响^[31],最后分别采用十折交叉验证方法和最近邻分类器进行人脸验证和人脸识别。实验结果表明:在LFW人脸数据库中,聚合判别多任务学习算法的验证精度达到了99.68%,比传统的归一化指数损失函数的验证精度提高了2%左右,在AgeDB人脸库和CFP_FP人脸库中验证精度分别为97.29%和92.71%,均提高了3%左右,人脸识别的性能明显提高^[32]。在表情、光照和年龄变化等方面具有良好的鲁棒性,该方法可有效地应用于人脸识别的工程实践。在CFP_FP数据集,虽然准确率也提高到了92.71%,但是相比于其他两个数据集的准确率就逊色得多,由此可以分析出该算法对人脸姿势变化的鲁棒性较差。

2019年,曾接贤等人提出一种改进的卷积神经网络用于单幅图像超分辨率重建方法^[33],设计了密集残差网

络和反卷积网络组成的新型网络。与经典超分辨率重建方法相比,该算法重建的图像在整体清晰度和边缘锐度方面更好,峰值信噪比也有所提高,图像整体的重建质量得到提升,但是重建图像的边缘细节不够好,因此比较适用于重建倍数小的场景。刘鹏飞等人针对浅层结构在处理内部结构复杂数据时的表征能力不足的问题,提出了一个基于特征转移的八层卷积神经网络来实现图像超分辨率重建。经过大量实验证明,该算法的峰值信噪比最高,网络收敛速度快,在精细度方面有所提高^[34],为深度学习在图像领域应用提供了新思路。总体来看,在图像领域,数据库的丰富、训练样本多样化、模型的鲁棒性都是待改进的地方。

此外,图像识别还应用在医疗诊断方面。文献[35]中提出了一种全自动肝脏衰减估计方法,利用深度学习和形态学操作手段,结合CT结果在五分钟之内就可以了解到肝脏衰减情况,有助于快速诊断非酒精性脂肪性肝病。文献[36]中提出了一种CNN模型,用来诊断癌症肿瘤,鉴别结肠直肠息肉和肺结核的平均AUC分别为0.86和0.71。它改进的地方是不直接将CT图像作为训练数据输入CNN,而是将额外的图像特征输入CNN网络来进一步提高分类精度。因此在小的、不均匀分布的数据集中可以显现出更优秀的性能。但是笔者认为,CT图像都是2D的,而患者做医疗检查的时候医学图像大多是3D的,3D转化成2D后图像信息必然会缺失一部分,其结果可靠度有待商榷。再者,把深度学习输出结果直接作为医学检查的参考不具有说服力,因为深度学习网络目前还是一种黑箱操作模式,具体的内部操作人们不得而知,在医学这样严谨的领域中是否能独当一面还不能过早下定论。最后,医学领域的图像资料受法律和道德约束,不会全部拿出来作训练数据,以至于深度学习应用时其训练数据不够全面,结果可能会出现欠拟合的情况,因此不太具有说服力。但是随着深度学习技术的不断成熟,它能快速地过滤图像信息辅助筛选,笔者相信这将会对医疗应用方面有很大的帮助,深度学习在医疗领域的前景是广阔的。

4.2 语音领域

语音识别的流程主要是语音输入、语音预处理和特征提取、建立声学模型、声学模型和语言模型对比进行解码搜索、语音识别等步骤。语音识别过程中最重要的两个步骤为特征提取和对比,最重要的两个问题就是语音的高质量和可读性^[37]。传统的语音识别方法在大语音数据中难以取得较好的结果。最近几年深度学习方法为其提供了新思路,虽然深度学习最先在图像领域内尝试,但是却在语音识别领域率先获得了成功。

2009年,深度学习的概念被引入到语音领域。人们开始尝试在语音领域找到深度学习应用的切入点。

2011年,微软研究院的邓立、俞栋和Hinton合作的产品发布,使用深度学习技术击败了传统的高斯混合模型(Gaussian Mixture Model, GMM),在语音识别准确度和快速性上都得到了大幅度改善,取得一定成果。2012年,谷歌公司的语音识别模型全部由GMM模型换为深度学习模型后,成功将语音识别的错误率降低了20%,这个幅度已经比过去几年的总和还多。

深度学习网络之所以取得巨大突破的主要原因是传统的高斯混合模型是一种浅层网络,其建模数据特征维数较小,特征的状态空间分布和特征之间的相关性不能被充分描述。在采用深度网络后,可以自动在数据中提取更多更复杂且有效的特征,样本数据特征间相关性信息得到充分表示,将连续的特征信息结合构成高维特征,通过高维特征样本对深度神经网络模型进行训练^[3]。随后,各大公司都开始了语音领域产品的研发,苹果公司Siri系统的语音输入功能,支持包括中文在内的20多种语言^[4];微软公司也基于深度学习开发了同声传译系统,实现了技术突破;国内的百度公司也不停努力,2016年百度公司的语音识别准确率高达96%,该成果被美国权威科技杂志《麻省理工评论》列为2016年十大突破技术之一。

语音识别中不仅涉及文字的识别,还有拼音的识别。目前深度学习已经成功应用在词汇识别、声韵母识别和身份识别中^[38],在环境安静的情况下语音识别准确率非常高,但是有噪声时识别率便大大降低,这也是未来需要进一步研究的方向。语音识别应用的深度学习网络主要是RNN网络,针对实际应用需要不同的性能,目前也基于RNN改进了很多不同的模型网络,如长短时记忆网络(Long Short-Term Memory, LSTM)模型,解决了RNN存在的“梯度爆炸”问题,能够学习长期依赖关系,缺点是网络结构复杂,训练时间过长。GRU(Gated Recurrent Unit)模型在LSTM模型的基础上,参数更少,结构更简单,但是实际应用中训练时间并没有缩短太多,没有较大突破。双向RNN网络,该模型从前向后保留目标词前面的重要信息,同时从后向前保留目标词后面的重要信息,优点是训练时间快,模型复杂度低,缺点是仍存在“梯度爆炸”问题。

虽然这些改进的网络对语音识别的准确率都有所提高,但是在环境嘈杂、普通话不标准(或其他语言)、说话断句等情况下的识别精准度还是不尽人意,因此依靠深度学习针对不同方言的语言识别也是重要的应用研究领域。此外,语音数据库还不够全面,语音数据比较特殊,无法向图像数据那样通过旋转、剪裁等操作来自行丰富。世界语言千万种,我国方言就有九大种,因此在没有大数据支撑的情况下,语言识别是无法有效实现的。

4.3 文本领域

文本领域是深度学习应用的继图像和语音之后的又一重要领域,最开始鉴于人类语言的复杂度很高,机器很难理解语义,因此在文本领域的成果迟迟没有太大的突破。20世纪50年代,出现专家规则,80年代,出现知识工程建立专家系统,90年代,出现人工特征学习结合浅层网络分类模型。这些分类方式基本不考虑词序信息,常常忽略上下文的联系,词语之间都是独立的,无法表征语义信息^[5]。直到2016年,深度学习在文本领域中应用才逐渐成熟起来。深度学习在文本领域的主要应用有情感分析、文本生成、语言翻译、聊天机器人等。

近年来,对文本领域的研究上,张立民等构建了基于浅层结构文本分析模型的深度玻尔兹曼机模型,结合了新的交叉熵稀疏惩罚因子,经过名为20-newsgroups文档的数据集进行训练和测试后,分类准确率更高,证明了该模型在大型文本分析上的可行性^[39]。但是它的特征提取方式主要依赖词频,虽然不需要对特征词和类别词之间关系的性质做出任何假设,但是容易受到词边缘概率的影响,而且参数多,训练时间较长。针对这些问题有人改进了卷积神经网络,以topoc2vec词向量模型作为模型的输入特征用于文本的分类问题^[40]。基于知乎实验数据的测试结果表明,该模型对于短文本的分类准确度达到93.27%,对于长文本的分类准确度高达98.06%。而且训练时间有所减少。

文本领域的研究已经渗透到文本感情分析当中。最开始李寿山等人提出“感情字典”,其主要是看句子中的积极或消极的词语,如“美”“好”“优”“差”“坏”等,再去数据库中匹配来完成感情分析^[41-43]。这种方法过于片面,忽略上下文联系和近义词等,因此很快被基于机器学习方法的感情分析方法取代。李阳辉等人则利用降噪自动编码器模型在深度学习框架theano下展开对文本的情感分析研究,证实了降噪自动编码器模型在文本信息分析情感上的优越性,但是它仍然存在训练速度慢的问题^[44]。针对机器学习存在的训练速度慢、测试输出结果慢等问题,翟东海等人提出并行处理框架算法。该算法采用“分而治之”的方法对数据集进行分块处理,迅速计算每一个数据块的误差,最后在缓冲区取得数据块误差并用于计算目标优化函数,再调整参数直至目标函数收敛,大量数据测试表明训练速度明显提高^[45]。文献[46]设计了一种双卷积神经网络模型和基于知识理解模型相结合的方法来进行文本分析,使用知识库的词嵌入作为CNN的输入,它可以充分利用知识库中的监督信息来进行词义分辨。实验表明该方法在句子分类方面确实有更优秀的性能,但是其功能单一,日后可以尝试拓展到文本理解方面的研究。文献[47]通过深

度卷积神经网络学习高级特征,并将低级特征结合到深层模型中以提高精准度,针对特定场景中的语义表征判断问题,引入了铰链损耗SVM检测器,能更好区分文本边框内的文字内容。然而该算法缺乏对全局和局部表征的联系研究,这将是未来的重点工作。

文本领域的难点不仅仅在于网络的不断改进,还有一个难点便是随着网络时代的发展,新兴语言出现得特别快,例如“明学”“康康”等高频网络词语,因此数据库如何更新完善、如何确定语义表征方式等将是下一步重要研究工作。

5 结束语

深度学习在图像、视频、语音、文本和大数据处理方面展现了很多优秀性能,但仍存在很多问题影响其实际应用。因此,未来的深度学习算法研究重点应集中在以下几个方面:

(1)无监督的学习问题。就上述学习网络架构而言,由于中间隐层层数变得越来越多,人们对大规模的有效数据和高性能的计算能力的需求也越来越大,进而对未来研究深度学习的硬件设备也提出了新要求。同时,深度学习依赖大数据,传统的人工搜集标签数据投入的人力物力过大导致成本过高,现在已经逐渐被淘汰,如何自动采集数据建立更强数据库也成为未来研究的新方向,基于部分标签或无标签的无监督的学习方式正逐渐成为主流。

(2)训练算法。目前训练算法都是采用异步SGD算法提高模型的训练速度,通过多CPU和GPU可以提高训练速度,对硬件有较高的要求,优化训练算法的研究仍具有很大的使用价值。文中介绍了很多改进的算法,但针对不用的应用情况性能不尽相同,因此开发高效可扩展的训练算法尤为重要,深度学习模型在训练过程中往往需要较长时间,在保证精准度的情况下,如何降低复杂性并快速训练得到模型也是非常重要的。

(3)模型参数选择。上述各种网络应用于不同的任务所要面临的关键性问题就是如何选择最优的模型训练参数,例如隐层层数、卷积核大小、池化层层数、优化函数的设定等,目前主要靠技术人员的经验和不断尝试。深度学习模型依然以黑箱方式运作,人们还难以理解其工作原理和工作方式,因此发现参数与训练结果的规律也是未来研究比较重要的一个方向。

(4)深度学习网络内部可视化。目前深度学习仍以黑箱方式运作,因此在严谨的领域(如医疗诊断方面)中不具有解释性和说服力。为了更好更可靠地将深度学习应用到更多领域,发挥其最大价值,早日将深度学习的内部结构原理可视化也是至关重要的。

参考文献:

- [1] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [2] Nazaré T S, da Costa G B P, Contato W A, et al. Deep convolutional neural networks and noisy images[C]// LNCS 10657: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2018: 416-424.
- [3] 王彦哲, 张立民, 张兵强, 等. 改进卷积输入的端到端普通话语音识别[J]. 计算机工程与应用, 2019, 55(17): 143-149.
- [4] 王勇. 基于混沌特征向量的动态纹理识别[D]. 上海: 上海交通大学, 2014.
- [5] 冯岭, 谢世博, 刘斌. 基于多层感知机的技术创新人才发现方法[J]. 计算机应用与软件, 2019, 36(7): 26-31.
- [6] 高强. 基于深度卷积网络学习算法及其应用研究[D]. 北京: 北京化工大学, 2015.
- [7] 樊峻畅. 红外图像中基于卷积神经网络的车辆检测[D]. 西安: 西安电子科技大学, 2017.
- [8] 路玉君. 基于RNN的陆空通话语义描述与度量方法[D]. 天津: 中国民航大学, 2017.
- [9] 周记. 基于循环卷积神经网络的图像分类算法研究[D]. 西安: 陕西师范大学, 2017.
- [10] Rumelhart D E, Hinton G E, Williams R J. Learning internal representation by back-propagation of errors[J]. Nature, 1986, 323(323): 533-536.
- [11] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems, 2012: 1097-1105.
- [12] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [13] 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(1): 48-59.
- [14] 赵志雄. 基于OpenCV的实时人脸识别系统设计[D]. 西安: 西安工程大学, 2015.
- [15] 郑健, 王志明, 张宁. 一种基于深度学习的改进人脸识别算法[J]. 计算机与现代化, 2018(12): 90-95.
- [16] Goodfellow I J, Erhan D, Carrier P L, et al. Challenges in representation learning: a report on three machine learning contests[J]. Neural Networks, 2015, 64: 59-63.
- [17] Szegedy C, Toshev A, Erhan D. Deep neural networks for object detection[C]// Advances in Neural Information Processing Systems, 2013.
- [18] 李帅奇. 基于高维特征的人脸认证方法研究[D]. 成都: 电子科技大学, 2016.
- [19] Xi Hailong, Liu Haiyan, Zhang Yu. Recognition and opti-

- mization algorithm of MNIST dataset based on LeNet5 network structure[C]//2018 International Conference on Transportation & Logistics, Information & Communication, Smart City, 2018:322-328.
- [20] 邹鲁,赵永新,王西超,等.基于深度卷积神经网络的集装箱锁销识别研究[J].上海电机学院学报,2019,22(4):193-197.
- [21] Liu Kunhua, Zhong Peisi, Zheng Yi, et al. P_VggNet: a convolutional neural network (CNN) with pixel-based attention map[J]. PLoS One, 2018, 13(12):1-11.
- [22] Tang Pengjie, Wang Hanli, Sam K.G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition[J]. Neurocomputing, 2017, 225:188-197.
- [23] 张红霞,王灿,刘鑫,等.复杂背景图像的字符识别算法研究[J].计算机测量与控制,2019,27(8):162-166.
- [24] 李梁,董旭彬,赵清华.改进Mask R-CNN在航拍灾害检测的应用研究[J].计算机工程与应用,2019,55(21):167-176.
- [25] Hu Haifeng, Liao Zhongke, Xiao Xiang. Action recognition using multiple pooling strategies of CNN features[J]. Neural Processing Letters, 2019, 50(1):379-396.
- [26] Khagi B, Kwon G R, Lama R. Comparative analysis of Alzheimer's disease classification by CDR level using CNN, feature selection, and machine learning techniques[J]. International Journal of Imaging Systems and Technology, 2019, 29(3):297-310.
- [27] 熊才华,巩言丽,廉华,等.基于ResNet-50改进的Faster R-CNN手势识别算法[J].计算机时代,2019(9):1-4.
- [28] 郑健,王志明,张宁.一种基于深度学习的改进人脸识别算法[J].计算机与现代化,2018(12):90-95.
- [29] 杨恢先,陈凡,甘伟发.基于多任务学习的深层人脸识别算法[J].激光与光电子学进展,2019,56(18):1-9.
- [30] Aghamaleki J A, Chenarlogh V A. Multi-stream CNN for facial expression recognition in limited training data[J]. Multimedia Tools and Applications, 2019, 78(16):22861-22882.
- [31] Zhang Dacheng, Lei Weimin, Zhang Wei, et al. A novel in-loop filtering mechanism of HEVC based on 3D sub-bands and CNN processing[J]. Signal, Image and Video Processing, 2019, 13(6):1045-1053.
- [32] 周南,杜军平,姚旭,等.基于卷积神经网络的微博话题内容搜索方法[J].计算机科学与探索,2019,13(5):753-764.
- [33] 曾接贤,倪申龙.改进的卷积神经网络单幅图像超分辨率重建[J].计算机工程与应用,2019,55(13):1-7.
- [34] 刘鹏飞,赵怀慈,刘明第.基于卷积神经网络的图像超分辨率重建[J].计算机工程与应用,2019,55(9):197-202.
- [35] Huo Yuankai, Terry J G, Wang Jiachen, et al. Fully automatic liver attenuation estimation combining CNN segmentation and morphological operations[J]. Medical Physics, 2019, 46(8):3508-3519.
- [36] Zhang Shu, Han Fangfang, Liang Zhengrong, et al. An investigation of CNN models for differentiating malignant from benign lesions using small pathologically proven datasets[J]. Computerized Medical Imaging and Graphics, 2019, 77:1884-2020.
- [37] 鞠维欣,赵希梅,魏宾,等.深度学习模型GoogleNet-PNN对肝硬化的识别[J].计算机工程与应用,2019,55(5):112-117.
- [38] Jiang C, Jiang C, Lu H, et al. CNN for saliency detection with low-level feature integration[J]. Neurocomputing, 2017, 226(C):212-220.
- [39] 张立民,刘凯.基于深度玻尔兹曼机的文本特征提取研究[J].微电子学与计算机,2015,32(2):142-147.
- [40] Hua Chengcheng, Wang Hong, Chen Jichi, et al. Novel functional brain network methods based on CNN with an application in proficiency evaluation[J]. Neurocomputing, 2019, 359:153-162.
- [41] 李寿山,黄居仁.基于Stacking组合分类方法的中文情感分类研究[J].中文信息学报,2010,24(5):56-61.
- [42] 李寿山,黄居仁.基于词边界分类的中文分词方法[J].中文信息学报,2010,24(1):3-7.
- [43] 李寿山.基于词边界分类的中文分词方法[C]//中国中文信息学会.中国计算机语言学研究前沿进展(2007—2009), 2009:3-7.
- [44] 李阳辉,谢明,易阳.基于降噪自动编码器及其改进模型的微博情感分析[J].计算机应用研究,2017,34(2):373-377.
- [45] 翟东海,侯佳林,刘月.基于深度学习的文本情感分析算法并行化研究[J].西南交通大学学报,2019,54(3):647-654.
- [46] Li Jun, Huang Guimin, Chen Jianheng, et al. Dual CNN for relation extraction with knowledge-based attention and word embeddings[J]. Computational Intelligence and Neuroscience, 2019:1-10.
- [47] Li Hongyang, Chen Jiang, Lu Huchuan, et al. CNN for saliency detection with low-level feature integration[J]. Neurocomputing, 2017, 226:212-220.