Journal of Computer Applications ISSN 1001-9081 2016-09-10
计算机应用,2016,36(9):2508-2515,2565 CODEN JYIIDU http://www.joca.cn

文章编号:1001-9081(2016)09-2508-08 doi:10.11772/j.issn.1001-9081.2016.09.2508

卷积神经网络研究综述

李彦冬*,郝宗波,雷 航
(电子科技大学 信息与软件工程学院,成都 610054)
(* 通信作者电子邮箱 416840140@qq.com)

摘 要:近年来,卷积神经网络在图像分类、目标检测、图像语义分割等领域取得了一系列突破性的研究成果,其强大的特征学习与分类能力引起了广泛的关注,具有重要的分析与研究价值。首先回顾了卷积神经网络的发展历史,介绍了卷积神经网络的基本结构和运行原理,重点针对网络过拟合、网络结构、迁移学习、原理分析四个方面对卷积神经网络在近期的研究进行了归纳与分析,总结并讨论了基于卷积神经网络的相关应用领域取得的最新研究成果,最后指出了卷积神经网络目前存在的不足以及未来的发展方向。

关键词:卷积神经网络;深度学习;特征表达;神经网络;迁移学习

中图分类号:TP181 **文献标志码**:A

Survey of convolutional neural network

LI Yandong*, HAO Zongbo, LEI Hang
(*School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu Sichuan* 610054, *China*)

Abstract: In recent years, Convolutional Neural Network (CNN) has made a series of breakthrough research results in the fields of image classification, object detection, semantic segmentation and so on. The powerful ability of CNN for feature learning and classification attracts wide attention, it is of great value to review the works in this research field. A brief history and basic framework of CNN were introduced. Recent researches on CNN were thoroughly summarized and analyzed in four aspects: over-fitting problem, network structure, transfer learning and theoretic analysis. State-of-the-art CNN based methods for various applications were concluded and discussed. At last, some shortcomings of the current research on CNN were pointed out and some new insights for the future research of CNN were presented.

Key words: Convolutional Neural Network (CNN); deep learning; feature representation; neural network; transfer learning

图像特征的提取与分类一直是计算机视觉领域的一个基础而重要的研究方向。卷积神经网络(Convolutional Neural Network, CNN)提供了一种端到端的学习模型,模型中的参数可以通过传统的梯度下降方法进行训练,经过训练的卷积神经网络能够学习到图像中的特征,并且完成对图像特征的提取和分类。作为神经网络领域的一个重要研究分支,卷积神经网络的特点在于其每一层的特征都由上一层的局部区域通过共享权值的卷积核激励得到。这一特点使得卷积神经网络相比于其他神经网络方法更适合应用于图像特征的学习与表达。

早期的卷积神经网络结构相对简单,如经典的 LeNet-5 模型[1],主要应用在手写字符识别、图像分类等一些相对单一的计算机视觉应用领域中。随着研究的不断深入,卷积神经网络的结构不断优化,其应用领域也逐渐得到延伸。例如,卷积神经网络与深信度网络(Deep Belief Network, DBN)[2]相结合产生的卷积深信度网络(Convolutional Deep Belief Network, CDBN)[3]作为一种非监督的生成模型,被成功地应用于人脸特征提取[4];AlexNet[5]在海量图像分类领域取得了突破性的成果;基于区域特征提取的 R-CNN(Regions with CNN)[6]在目标检测领域取得了成功;全卷积网络(Fully Convolutional Network, FCN)[7]实现了端到端的图像语义分割,并且在准确率上大幅超越了传统的语义分割算法。近年来,卷积神经网络的结构研究仍然有着很高的热度,一些具有优秀性能的网络结构被提出[8-10]。并且,随着迁移学习理论[11]在卷积神经网络上的成功应用,卷积神经网络的应用领域得到了进一步的扩展[12-13]。卷积神经网络在各个领域不断涌现出来的研究成果,使其成为了当前最受关注的研究热点之一。

1 卷积神经网络的研究历史与意义

1.1 卷积神经网络的研究历史

卷积神经网络的研究历史大致可以分为三个阶段:理论提出阶段、模型实现阶段以及广泛研究阶段。

1)理论提出阶段。20 世纪 60 年代,Hubel 等[14]的生物学研究表明,视觉信息从视网膜传递到大脑中是通过多个层次的感受野(Receptive Field)激发完成的。1980 年,Fukushima 第一次提出了一个基于感受野的理论模型 Neocognitron[15]。Neocognitron 是一个自组织的多层神经网络模型,每一层的响应都由上一层的局部感受野激发得到,对于模式的识别不受位置、较小形状变化以及尺度大小的影响。Neocognitron 采用的无监督学习也是卷积神经网络早期研究

收稿日期:2016-03-30;**修回日期**:2016-04-20。
基金项目:国家科技支撑计划项目(2012BAH44F02);广东省产学研项目(M17010601CXY2011057)。
作者简介:李彦冬(1984—),男,四川泸州人,博士研究生,主要研究方向:机器学习、计算机视觉; 郝宗波(1977—),男,河南新乡人,副教授,博士,主要研究方向:图像理解、视频信息处理; 雷航(1960—),男,四川自贡人,教授,博士,主要研究方向:图像处理。

中占据主导地位的学习方式。

2) 模型实现阶段。1998 年, Lecun 等[1] 提出的 LeNet-5 采用了基于梯度的反向传播算法对网络进行有监督的训练。经过训练的网络通过交替连接的卷积层和下采样层将原始图像转换成一系列的特征图,最后,通过全连接的神经网络针对图像的特征表达进行分类。卷积层的卷积核完成了感受野的功能,可以将低层的局部区域信息通过卷积核激发到更高的层次。LeNet-5 在手写字符识别领域的成功应用引起了学术界对于卷积神经网络的关注。同一时期,卷积神经网络在语音识别[16] 、物体检测[17] 、人脸识别[18] 等方面的研究也逐渐开展起来。

3) 广泛研究阶段。2012 年, Krizhevsky 等[5] 提出的 AlexNet 在大型图像数据库 ImageNet[19] 的图像分类竞赛中以准确度超越第二名 11% 的巨大优势夺得了冠军,使得卷积神经网络成为了学术界的焦点。AlexNet 之后,不断有新的卷积神经网络模型被提出,比如牛津大学的 VGG(Visual Geometry Group) [8] 、Google 的 GoogLeNet[9] 、微软的 ResNet[10] 等,这些网络刷新了 AlexNet 在 ImageNet 上创造的纪录。并且,卷积神经网络不断与一些传统算法相融合,加上迁移学习方法的引入,使得卷积神经网络的应用领域获得了快速的扩展。一些典型的应用包括: 卷积神经网络与递归神经网络(Recurrent Neural Network, RNN) 结合用于图像的摘要生成[20-21] 以及图像内容的问答[22-23] ; 通过迁移学习的卷积神经网络在小样本图像识别数据库上取得了大幅度准确度提升[24] ; 以及面向视频的行为识别模型——3D 卷积神经网络[25] ,等。

1.2　卷积神经网络的研究意义

卷积神经网络领域目前已经取得了许多令人瞩目的研究成果,但是随之而来的是更多的挑战,其研究意义主要体现在三个方面: 理论研究挑战、特征表达研究、应用价值。

1) 理论研究挑战。卷积神经网络作为一种受到生物学研究启发的经验方法,学术界普遍采用的是以实验效果为导向的研究方式。比如 GoogLeNet 的 Inception 模块设计、VGG 的深层网络以及 ResNet 的 short connection 等方法都通过实验证实了其对于网络性能改善的有效性; 但是,这些方法都存在缺乏严谨的数学验证问题。造成这一问题的根本原因是卷积神经网络本身的数学模型没有得到完善的数学验证与解释。从学术研究的角度来说,卷积神经网络的发展没有理论研究的支持是不够严谨和不可持续的。因此,卷积神经网络的相关理论研究是当前最为匮乏也是最有价值的部分。

2) 特征表达。图像的特征设计一直是计算机视觉领域的一个基础而重要的课题。在以往的研究中,一些典型的人工设计特征被证明取得了良好的特征表达效果,如 SIFT (Scale-Invariant Feature Transform) [26] 、HOG (Histogram of Oriented Gradient) [27] 等。但是,这些人工设计特征也存在缺乏良好的泛化性能问题。卷积神经网络作为一种深度学习[28-29] 模型,具有分层学习特征的能力[24] 。研究[30-31] 表明,通过卷积神经网络学习得到的特征相对于人工设计特征具有更强的判别能力和泛化能力。特征表达作为计算机视觉的研究基础,如何利用卷积神经网络学习、提取、分析信息的特征表达,从而获得判别性能更强,泛化性能更好的通用特征,将对整个计算机视觉乃至更广泛的领域产生积极的影响。

3) 应用价值。卷积神经网络经过多年的发展,从最初较为简单的手写字符识别[1] 应用,逐渐扩展到一些更加复杂的领域,如: 行人检测[32] 、行为识别[25,33] 、人体姿势识别[34] ,等。近期,卷积神经网络的应用进一步向更深层次的人工智能发展,如: 自然语言处理[35-36] 、语音识别[37] ,等。最近,由 Google 开发的人工智能围棋程序 Alphago[38] 成功利用了卷积神经网络分析围棋盘面信息,并且在挑战赛中接连战胜了围棋欧洲冠军和世界冠军,引起了广泛的关注。从当前的研究趋势来看,卷积神经网络的应用前景充满了可能性,但同时也面临着一些研究难题,比如: 如何改进卷积神经网络的结构,以提高网络对于特征的学习能力; 如何将卷积神经网络以合理的形式融入新的应用模型中。

2　卷积神经网络基本原理

2.1　卷积神经网络的基本结构

如图 1 所示,典型的卷积神经网络主要由输入层、卷积层、下采样层(池化层) 、全连接层和输出层组成。

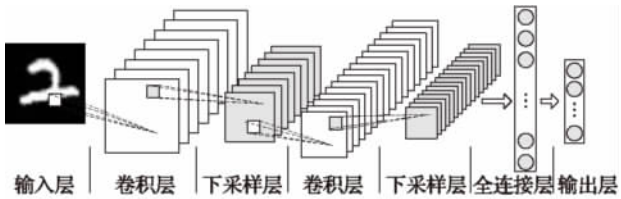

图 1　卷积神经网络的典型结构

卷积神经网络的输入通常为原始图像 $\boldsymbol{X}$。本文用 $\boldsymbol{H}_i$ 表示卷积神经网络第 i 层的特征图($\boldsymbol{H}_0 = \boldsymbol{X}$) 。假设 $\boldsymbol{H}_i$ 是卷积层, $\boldsymbol{H}_i$ 的产生过程可以描述为:

$$\boldsymbol{H}_i = f(\boldsymbol{H}_{i-1} \otimes \boldsymbol{W}_i + \boldsymbol{b}_i) \tag{1}$$

其中: $\boldsymbol{W}_i$ 表示第 i 层卷积核的权值向量; 运算符号“$\otimes$”代表卷积核与第 $i-1$ 层图像或者特征图进行卷积操作,卷积的输出与第 i 层的偏移向量 $\boldsymbol{b}_i$ 相加,最终通过非线性的激励函数 $f(x)$ 得到第 i 层的特征图 $\boldsymbol{H}_i$。

下采样层通常跟随在卷积层之后,依据一定的下采样规则[39] 对特征图进行下采样。下采样层的功能主要有两点: 1) 对特征图进行降维; 2) 在一定程度上保持特征的尺度不变特性。假设 $\boldsymbol{H}_i$ 是下采样层:

$$\boldsymbol{H}_i = subsampling(\boldsymbol{H}_{i-1}) \tag{2}$$

经过多个卷积层和下采样层的交替传递,卷积神经网络依靠全连接网络对针对提取的特征进行分类,得到基于输入的概率分布 $\boldsymbol{Y}$(l_i 表示第 i 个标签类别) 。如式(3) 所示,卷积神经网络本质上是使原始矩阵($\boldsymbol{H}_0$) 经过多个层次的数据变换或降维,映射到一个新的特征表达($\boldsymbol{Y}$) 的数学模型。

$$\boldsymbol{Y}(i) = P(L = l_i \mid \boldsymbol{H}_0; (\boldsymbol{W}, \boldsymbol{b})) \tag{3}$$

卷积神经网络的训练目标是最小化网络的损失函数 $L(\boldsymbol{W}, \boldsymbol{b})$ 。输入 $\boldsymbol{H}_0$ 经过前向传导后通过损失函数计算出与期望值之间的差异,称为“残差”。常见损失函数有均方误差(Mean Squared Error, MSE) 函数,负对数似然(Negative Log Likelihood, NLL) 函数等[40] :

$$MSE(\boldsymbol{W}, \boldsymbol{b}) = \frac{1}{|\boldsymbol{Y}|} \sum_{i=1}^{|\boldsymbol{Y}|} (\boldsymbol{Y}(i) - \hat{\boldsymbol{Y}}(i))^2 \tag{4}$$

$$NLL(\boldsymbol{W}, \boldsymbol{b}) = -\sum_{i=1}^{|\boldsymbol{Y}|} \log \boldsymbol{Y}(i) \tag{5}$$

为了减轻过拟合的问题,最终的损失函数通常会通过增加 L_2 范数以控制权值的过拟合,并且通过参数 λ(weight decay) 控制过拟合作用的强度:

$$E(\boldsymbol{W},\boldsymbol{b}) = L(\boldsymbol{W},\boldsymbol{b}) + \frac{\lambda}{2}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} \tag{6}$$

训练过程中,卷积神经网络常用的优化方法是梯度下降方法。残差通过梯度下降进行反向传播,逐层更新卷积神经网络的各个层的可训练参数($\boldsymbol{W}$ 和 $\boldsymbol{b}$)。学习速率参数(η)用于控制残差反向传播的强度:

$$\boldsymbol{W}_i = \boldsymbol{W}_i - \eta\frac{\partial \mathrm{E}(\boldsymbol{W},\boldsymbol{b})}{\partial \boldsymbol{W}_i} \tag{7}$$

$$\boldsymbol{b}_i = \boldsymbol{b}_i - \eta\frac{\partial \mathrm{E}(\boldsymbol{W},\boldsymbol{b})}{\partial \boldsymbol{b}_i} \tag{8}$$

2.2 卷积神经网络的工作原理

基于 2.1 节的定义,卷积神经网络的工作原理可以分为网络模型定义、网络训练以及网络的预测三个部分:

1) 网络模型定义。网络模型的定义需要根据具体应用的数据量以及数据本身的特点,设计网络深度、网络每一层的功能,以及设定网络中的超参数,如:λ、η 等。针对卷积神经网络的模型设计有不少的研究,比如模型深度方面[8,10]、卷积的步长方面[24,41]、激励函数方面[42-43]等。此外,针对网络中的超参数选择,也存在一些有效的经验总结[44]。但是,目前针对网络模型的理论分析和量化研究相对还比较匮乏。

2) 网络训练。卷积神经网络可以通过残差的反向传播对网络中的参数进行训练。但是,网络训练中的过拟合以及梯度的消逝与爆炸等问题[45]极大影响了训练的收敛性能。针对网络训练的问题,一些有效的改善方法被提出,包括:基于高斯分布的随机初始化网络参数[5];利用经过预训练的网络参数进行初始化[8];对卷积神经网络不同层的参数进行相互独立同分布的初始化[46]。根据近期的研究趋势,卷积神经网络的模型规模正在迅速增大,而更加复杂的网络模型也对相应的训练策略提出了更高的要求。

3) 网络的预测。卷积神经网络的预测过程就是通过对输入数据进行前向传导,在各个层次上输出特征图,最后利用全连接网络输出基于输入数据的条件概率分布的过程。近期的研究表明,经过前向传导的卷积神经网络高层特征具有很强的判别能力和泛化性能[30-31];而且,通过迁移学习,这些特征可以被应用到更加广泛的领域。这一研究成果对于扩展卷积神经网络的应用领域具有重要的意义。

3 卷积神经网络研究进展

经过数十年的发展,卷积神经网络从最初的理论原型,到能够完成一些简单的任务,再到近期取得大量研究成果,成为了一个受到广泛关注的研究方向,其发展的推动力量主要来源于以下四个方面的基础研究:1) 卷积神经网络过拟合问题的相关研究提高了网络的泛化性能;2) 卷积神经网络结构的相关研究提高了网络拟合海量数据的能力;3) 卷积神经网络的原理分析指导着网络结构的发展,同时也提出了全新的具有挑战性的问题;4) 基于迁移学习的卷积神经网络相关研究拓展了卷积神经网络的应用领域。

3.1 卷积神经网络的过拟合问题

过拟合(over-fitting)[40]是指学习模型在训练过程中参数过度拟合训练数据集,从而影响到模型在测试数据集上的泛化性能的现象。卷积神经网络的结构层次比较复杂,目前的研究针对卷积神经网络的卷积层、下采样层以及全连接层的过拟合问题均有涉及。当前主要的研究思路是通过增加网络的稀疏性以及随机性,以改善网络的泛化性能。

Hinton 等[47]提出的 Dropout 通过在训练过程中随机地忽略一定比例的节点响应,减轻了传统全连接神经网络的过拟合问题,有效地提高了网络的泛化性能。但是,Dropout 对于卷积神经网络的性能改善并不明显,其主要原因是卷积神经网络由于卷积核的权值共享特性,相比于全连接的网络大大减少了训练参数的数量,本身就避免了较为严重的过拟合现象。因此,作用于全连接层的 Dropout 方法对于卷积神经网络整体的去过拟合效果不够理想。

基于 Dropout 的思想,Wan 等[48]提出了 DropConnect 的方法。与 Dropout 忽略全连接层的部分节点响应不同,DropConnect 随机地将神经网络卷积层一定比例的连接断开。对于卷积神经网络,作用于卷积层的 DropConnect 相比作用于全连接层的 Dropout 具有更强的去过拟合能力。

与 DropConnect 类似,Goodfellow 等[42]提出了作用于卷积层的 Maxout 激励函数。不同于 DropConnect 的是,Maxout 只保留神经网络的上一层节点往下一层的激励最大值。并且,Goodfellow 等[42]证明了 Maxout 函数可以拟合任意凸函数,在减轻过拟合问题的基础上还具有强大的函数拟合能力。

如图 2 所示,Dropout、DropConnect 和 Maxout 三种方法虽然具体实现机制有所差别,但是其根本原理都是通过增加网络连接的稀疏性或者随机性以达到消除过拟合,提高网络泛化能力的目的。

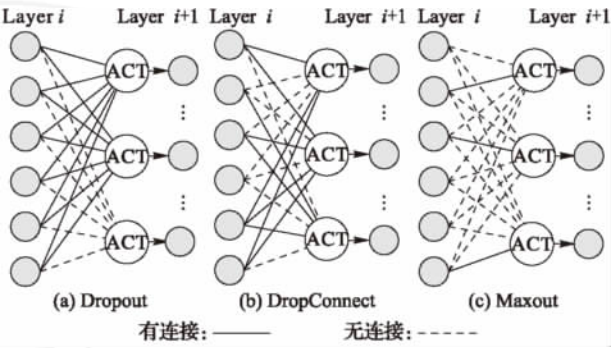

图 2 Dropout、DropConnect、Maxout 原理示意图

Lin 等[43]指出了卷积神经网络中的全连接网络易过拟合的问题以及 Maxout 激活函数只能够拟合凸函数的局限,提出了一种 NIN(Network in Network)的网络结构。一方面,NIN 放弃了利用全连接网络进行特征图到概率分布的映射,采用了直接针对特征图进行 Global average pooling 的方法获取到最终的概率分布,在减少网络中的参数数量的同时也避免了全连接网络的过拟合问题;另一方面,NIN 使用“微神经网络”(micro neural network)取代传统的激励函数(如:Maxout)。理论上,微神经网络突破了传统激励函数的局限,可以拟合任意的函数,使网络具有了更好的拟合性能。

此外,针对卷积神经网络的下采样层,Zeiler 等[39]提出了一种随机下采样的方法(Stochastic pooling)来改善下采样层的过拟合问题。与传统的 Average pooling 和 Max pooling 分别指定了下采样区域的均值和最大值进行下采样的方式不同,Stochastic pooling 依据概率分布进行随机的下采样操作,给下采样的过程引入了随机性。实验表明,这种随机性能够有效提高卷积神经网络的泛化性能。

目前针对卷积神经网络过拟合问题的研究,主要还存在以下问题:1) 针对过拟合现象的量化研究和评价标准的缺

失,使得当前的研究都只能通过实验对比来证明新的方法对于过拟合问题的改善,而这种改善的程度和通用性都需要更为统一且通用的评价标准来进行衡量; 2) 针对卷积神经网络,过拟合问题在各种层次(如: 卷积层、下采样层、全连接层) 中的严重程度、改善空间及改进方法还有待进一步的探索。

3.2　卷积神经网络的结构

Lecun 等[1] 提出的 LeNet-5 模型采用了交替连接的卷积层和下采样层对输入图像进行前向传导,并且最终通过全连接层输出概率分布的结构是当前普遍采用的卷积神经网络结构的原型。LeNet-5 虽然在手写字符识别领域取得了成功,但是其存在的缺点也比较明显,包括: 1) 难以寻找到合适的大型训练集对网络进行训练以适应更为复杂的应用需求; 2) 过拟合问题使得 LeNet-5 的泛化能力较弱; 3) 网络的训练开销非常大,硬件性能支持的不足使得网络结构的研究非常困难。以上三大制约卷积神经网络发展的重要因素在近期的研究中取得了突破性的进展是卷积神经网络成为一个新的研究热点的重要原因。并且,近期针对卷积神经网络的深度和结构优化方面的研究进一步提升了网络的数据拟合能力。

针对 LeNet-5 的缺陷, Krizhevsky 等[5] 提出了 AlexNet。AlexNet 有 5 层卷积网络,约 65 万个神经元以及 6 000 万个可训练参数,从网络规模上大大超越了 LeNet-5。另外,AlexNet 选择了大型图像分类数据库 ImageNet[19] 作为训练数据集。ImageNet 提供了 1 000 个类别共 120 万张图片进行训练,图片的数量和类别都大幅度超越了以往的数据集。在去过拟合方面,AlexNet 引了 dropout,一定程度上减轻了网络过拟合问题。在硬件支持方面,AlexNet 使用了 GPU 进行训练,相比传统的 CPU 运算,GPU 使网络的训练速度提高了十倍以上。AlexNet 在 ImageNet 的 2012 图像分类竞赛中夺得冠军,并且相比于第二名的方法在准确度上取得了高出 11% 的巨大优势。AlexNet 的成功使得卷积神经网络的研究再次引起了学术界的关注。

Simonyan 等[8] 在 AlexNet 的基础上,针对卷积神经网络的深度进行了研究,提出了 VGG 网络。VGG 由 3 × 3 的卷积核构建而成,通过对比不同深度的网络在图像分类应用中的性能,Simonyan 等证明了网络深度的提升有助于提高图像分类的准确度。然而,这种深度的增加并非没有限制,在恰当的网络深度基础上继续增加网络的层数,会带来训练误差增大的网络退化问题[49]。因此,VGG 的最佳网络深度被设定在了 16 ~ 19 层。

针对深度网络的退化问题, He 等[10] 分析认为如果网络中增加的每一个层次都能够得到优化的训练,那么误差是不应该会在网络深度加大的情况下提高的。因此,网络退化问题说明了深度网络中并不是每一个层次都得到了完善的训练。He 等提出了一种 ResNet 网络结构。ResNet 通过 short connections 将低层的特征图 x 直接映射到高层的网络中。假设原本网络的非线性映射为 $F(x)$,那么通过 short connection 连接之后的映射关系就变为了 $F(x) + x$。He 等提出这一方法的依据是 $F(x) + x$ 的优化相比 $F(x)$ 会更加容易。因为,从极端角度考虑, 如果 x 已经是一个优化的映射, 那么 short connection 之间的网络映射经过训练后就会更趋近于 0。这就意味着数据的前向传导可以在一定程度上通过 short connection 跳过一些没有经过完善训练的层次,从而提高网络的性能。实验证明,ResNet 虽然使用了和 VGG 同样大小的卷积核,但是网络退化问题的解决使其可以构建成为一个 152 层的网络,并且 ResNet 相比 VGG 有更低训练误差和更高的测试准确度。虽然 ResNet 在一定程度上解决了深层网络退化的问题,但是关于深层网络的研究仍然存在一些疑问: 1) 如何判断深度网络中哪些层次未能得到完善的训练; 2) 是什么原因导致深度网络中部分层次训练的不完善; 3) 如何处理深层网络中训练不完善的层次。

在卷积神经网络深度的研究以外,Szegedy 等[9] 更关注通过优化网络结构从而降低网络的复杂程度。他们提出了一种卷积神经网络的基本模块称为 Inception。如图 3 所示, Inception 模块由 1 × 1,3 × 3,5 × 5 的卷积核组成,小尺度卷积核的使用主要有两大优点: 1) 控制了整个网络中的训练参数数量,降低了网络的复杂度; 2) 不同大小的卷积核在多尺度上针对同一图像或者特征图进行了特征提取。实验表明,使用 Inception 模块构建的 GoogLeNet 的训练参数数量只有 AlexNet 的 1/12,但是在 ImageNet 上的图像分类准确度却高出 AlexNet 大约 10%。

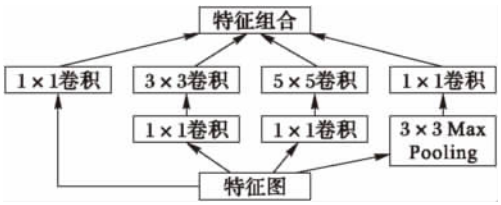

图 3　Inception 模块原理

此外,Springenberg 等[50] 对卷积神经网络下采样层存在的必要性提出了质疑,并设计了不含下采样层的“完全卷积网络”。“完全卷积网络”在结构上相比于传统的卷积神经网络结构更加简单,但是其网络性能却不低于带有下采样层的传统模型。

卷积神经网络结构方面的研究是一个开放的问题,基于当前的研究状况,目前的研究主要形成了两大趋势: 1) 增加卷积神经网络的深度; 2) 优化卷积神经网络的结构,降低网络的复杂度。在卷积神经网络的深度研究方面,主要依赖于进一步分析深层次网络存在的潜在隐患(如: 网络退化) ,以解决深层网络的训练问题(如: VGG、ResNet) 。而在优化网络结构方面,目前的研究趋势是进一步加强对于当前网络结构的理解和分析,以更简洁高效的网络结构取代当前的结构,进一步地降低网络复杂度并且提升网络的性能(如: GoogLeNet、完全卷积网络) 。

3.3　卷积神经网络的原理分析

卷积神经网络虽然在众多应用领域已经取得了成功,但其原理的分析和解释一直都是备受质疑的一个弱点。近期的一些研究开始采用可视化的方法对卷积神经网络的原理进行了分析,直观地比较了卷积神经网络的学习特征与传统人工设计特征的差异,展现了网络从低层到高层的特征表达过程。

Donahue 等[30] 提出了利用 t-SNE[51] 的方法来分析卷积神经网络提取的特征。t-SNE 的原理是将高维特征降低到二维,然后在二维空间直观地展示特征。利用 t-SNE, Donahue 等将卷积神经网络特征与传统的人工设计特征 GIST(GIST 的含义是能够激发记忆中场景类别的抽象场景) [52] 和 LLC (Locality-constrained Linear Coding) [53] 进行了比较,发现判别能力更强的卷积神经网络特征在 t-SNE 的可视化结果中表现

出了更好的区分度,证明了特征判别能力与 t-SNE 可视化结果的一致性。但是,Donahue 等的研究仍然遗留下来了以下问题: 1) 未能解释卷积神经网络提取的特征到底是什么; 2) Donahue等挑选了卷积神经网络部分层次的特征进行可视化,但是对于这些层次之间的关系并没有进行分析; 3) t-SNE 算法本身存在一定的局限性,对于特征类别过多的情况并不能很好地反映类别间的差异。

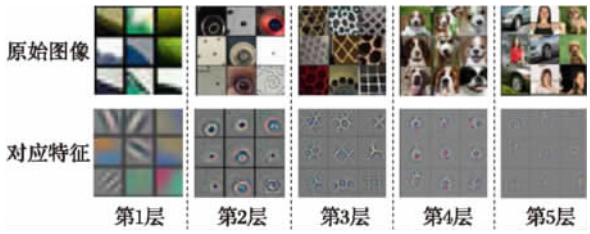

图 4 卷积神经网络特征与图像对比

Zeiler 等[24] 的研究较好地解决了 t-SNE 的遗留问题。他们通过构建 DeConvNet[54] ,对卷积神经网络中不同层次的特征进行反卷积,展示了各个层次提取的特征状况。图 4 选取了卷积神经网络各个层次的部分较强特征可视化结果,并且与像素空间的原始图像的对应像素块进行了对比。可以发现: 卷积神经网络较低的第一和第二层主要提取了边缘、颜色等低层特征,第三层开始出现了较为复杂的纹理特征,而第四层和第五层开始出现了较为完整的个体轮廓和形状特征。通过可视化各个层次的特征,Zeiler 等改进了 AlexNet 的卷积核大小和步长,提升了网络性能。并且,他们还利用可视化特征对卷积神经网络的图像遮挡敏感性、物体部件相关性以及特征不变性进行了分析。Zeiler 等的研究体现了卷积神经网络的原理研究对于改进卷积神经网络的结构与性能具有重大的指导意义。

Nguyen 等[55] 对卷积神经网络提取特征的完备性提出了质疑。如图 5 所示,Nguyen 等通过进化算法[56] 将原始图像处理成在人类看来根本无法识别和解释的一种形式,但是卷积神经网络对于这些转换后的图像形式却给出了非常确切的物体类别判断。Nguyen 等的研究并没有针对出现这一现象的原因作出明确的解释,只是证明了卷积神经网络虽然具有分层的特征提取能力,但在图像的识别机理上并不是与人类完全一致。这一现象表明了当前的研究对于卷积神经网络的原理认知与分析还存在很大的不足。

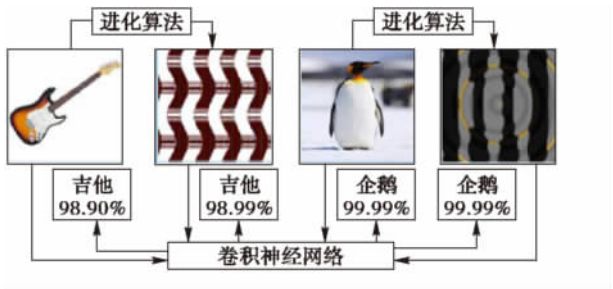

图 5 卷积神经网络的“欺骗”现象

总体来说,目前针对卷积神经网络原理的研究与分析还相当地不足,主要存在的问题包括: 1) 与传统的人工设计特征不同,卷积神经网络的特征受到特定的网络结构、学习算法以及训练集等多种因素影响,其原理的分析与解释相比人工设计特征更加地抽象和困难; 2) Nguyen 等[55] 的研究展示了卷积神经网络出现的被“欺骗”现象引起了人们对于其完备性的关注。虽然卷积神经网络是基于仿生学的研究而来,但是如何解释卷积神经网络与人类视觉的差异,如何使卷积神经网络的识别机制更加完备,仍然是有待解决的问题。

3.4 卷积神经网络的迁移学习

迁移学习的定义是: “运用已存有的知识对不同但相关领域问题进行求解的一种机器学习方法”[57] ,其目标是完成知识在相关领域之间的迁移[11] 。对于卷积神经网络而言,迁移学习就是要把在特定数据集上训练得到的“知识”成功运用到新的领域之中。如图 6 所示,卷积神经网络的迁移学习的一般流程是: 1) 在特定应用之前,先利用相关领域大型数据集(如 ImageNet) 对网络中的随机初始化参数进行训练; 2) 利用训练好的卷积神经网络,针对特定应用领域的数据(如 Caltech) 进行特征提取; 3) 利用提取后的特征,针对特定应用领域的数据训练卷积神经网络或者分类器。

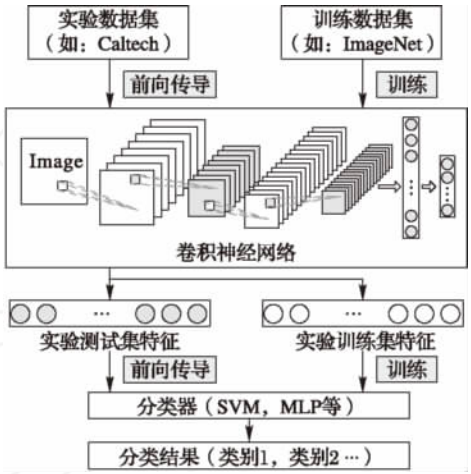

图 6 卷积神经网络的迁移学习流程

相比直接在目标数据集上训练网络的传统方法,Zeiler 等[24] 让卷积神经网络在 ImageNet 数据集上进行预训练,然后再将网络分别在图像分类数据集 Caltech-101[58] 和 Caltech-256[59] 上进行迁移训练和测试,其图像分类准确度提高了约 40%。但是,ImageNet 和 Caltech 都属于物体识别数据库,其迁移学习的领域相对比较接近,对于跨度更大领域的研究还存在不足。于是,Donahue 等[30] 采用了与 Zeiler 类似的策略,通过基于 ImageNet 的卷积神经网络预训练,成功地将卷积神经网络的迁移学习应用到了与物体识别差异更大的领域,包括: domain adaption、subcategory recognition 以及 scene recognition 等。

除了卷积神经网络在各个领域的迁移学习研究,Razavian 等[31] 还对卷积神经网络不同层次特征的迁移学习效果进行了探索,发现卷积神经网络的高层特征相对于低层特征具有更好的迁移学习能力。

Zhou 等[60] 利用了大型的图像分类数据库(ImageNet) 和场景识别数据库(Places[60]) 分别对两个相同结构的卷积神经网络进行了预训练,并且在一系列的图像分类和场景识别数据库上进行了迁移学习效果的验证。实验结果显示,经过 ImageNet 和 Places 预训练的网络分别在各自领域的数据库上取得的迁移学习效果更好,这一事实说明了领域的相关性对于卷积神经网络的迁移学习具有一定的影响。

关于卷积神经网络迁移学习的研究,其意义包括: 1) 解决卷积神经网络在小样本条件下的训练样本不足问题; 2) 对

于卷积神经网络的迁移利用,能大幅度减少网络的训练开销; 3) 利用迁移学习能进一步扩大卷积神经网络的应用领域。

卷积神经网络迁移学习还有待进一步研究的内容包括: 1) 训练样本的数量对于迁移学习效果的影响,以及迁移学习对于拥有不同训练样本数量的应用的效果还有待进一步的研究; 2) 基于卷积神经网络本身的结构,进一步分析卷积神经网络体系中各个层次的迁移学习能力; 3) 分析领域间相关性对于迁移学习的作用,寻找优化的跨领域迁移学习策略。

4　卷积神经网络的应用分析

随着网络性能的提升和迁移学习方法的使用,卷积神经网络的相关应用也逐渐向复杂化和多元化发展。总体来说,卷积神经网络的应用主要呈现出以下四大发展趋势:

1) 随着卷积神经网络相关研究的不断推进,其相关应用领域的精度也得到了迅速的提高。以图像分类领域的研究为例,在 AlexNet 将 ImagNet 的图像分类准确度大幅提升到 84.7% 之后,不断有改进的卷积神经网络模型被提出并刷新了 AlexNet 的纪录,具有代表性的网络包括: VGG[8]、GoogLeNet[9]、PReLU-net[46] 和 BN-inception[61] 等。最近,由微软提出的 ResNet[10] 已经将 ImageNet 的图像分类准确度提高到了 96.4%,而 ResNet 距离 AlexNet 的提出,也仅过去了四年的时间。卷积神经网络在图像分类领域的迅速发展,不断提升已有数据集的准确度,也给更加大型的图像应用相关数据库的设计带来了迫切的需求。

2) 实时应用领域的发展。计算开销一直是卷积神经网络在实时应用领域发展的阻碍。但是,近期的一些研究展现了卷积神经网络在实时应用中的潜力。Gishick 等[6,62] 和 Ren 等[63] 在基于卷积神经网络的物体检测领域进行了深入的研究,先后提出了 R-CNN[6]、Fast R-CNN[62] 和 Faster R-CNN[63] 模型,突破了卷积神经网络的实时应用瓶颈。R-CNN 成功地提出了利用 CNN 在 region proposals[64] 的基础上进行物体检测。R-CNN 虽然取得了很高的物体检测准确度,但是过多的 region proposals 使得物体检测的速度非常缓慢。Fast R-CNN 通过在 region proposals 之间共享卷积特征,大幅减少了大量 region proposals 带来的计算开销,在忽略产生 region proposals 的时间情况下,Fast R-CNN 取得了接近实时的物体检测速度。而 Faster R-CNN 则是通过利用端到端的卷积神经网络[7] 提取 region proposals 取代了传统的效率较低的方法[64],实现了卷积神经网络对于物体的实时检测。随着硬件性能的不断提高,以及通过改进网络结构带来的网络复杂度的降低,卷积神经网络在实时图像处理任务领域逐渐展现出了应用前景。

3) 随着卷积神经网络性能的提升,相关应用的复杂程度也随之提高。一些具有代表性的研究包括: Khan 等[65] 通过利用两个卷积神经网络分别学习图像中的区域特征和轮廓特征,完成了阴影检测任务; 卷积神经网络在人脸检测和识别的应用中也取得了巨大的进步,取得了接近人类的人脸识别效果[66-67]; Levi 等[68] 利用卷积神经网路学习到的人脸细微特征,进一步实现了对人的性别和年龄进行预测; Long 等[7] 提出的 FCN 结构实现了图像与语义的端到端映射; Zhou 等[60] 研究了利用卷积神经网络进行图像识别与更为复杂的场景识别任务之间的相互联系; Ji 等[25] 利用了 3D 卷积神经网络实现了行为识别。目前,卷积神经网络的性能和结构仍然处于高速的发展阶段,其相关的复杂应用在接下来的一段时间内都将保持其研究热度。

4) 基于迁移学习以及网络结构的改进,卷积神经网络逐渐成为了一种通用的特征提取与模式识别工具,其应用范围已经逐渐超越了传统的计算机视觉领域。比如,AlphaGo 成功地利用了卷积神经网络对围棋的盘面形势进行判断[38],证明了卷积神经网络在人工智能领域的成功应用; Abdel-Hamid 等[37] 通过将语音信息建模成符合卷积神经网络的输入模式,并结合隐马尔可夫模型(Hidden Markov Model, HMM),将卷积神经网络成功地应用到了语音识别领域; Kalchbrenner 等[35] 利用卷积神经网络提取了词汇和句子层面的信息,成功地将卷积神经网络应用于自然语言处理; Donahue 等[20] 结合了卷积神经网络和递归神经网络,提出了 LRCN(Long-term Recurrent Convolutional Network) 模型,实现了图像摘要的自动生成。卷积神经网络作为一种通用的特征表达工具,逐渐表现出了在更加广泛的应用领域中的研究价值。

从目前的研究形势来看,一方面,卷积神经网络在其传统应用领域的研究热度不减,如何改善网络的性能仍有很大的研究空间; 另一方面,卷积神经网络良好的通用性能使其应用领域逐渐扩大,应用的范围不再局限于传统的计算机视觉领域,并且向应用的复杂化、智能化和实时化发展。

5　卷积神经网络的缺陷与发展方向

目前,卷积神经网络正处于研究热度非常高的阶段,该领域仍然存在的一些问题以及发展方向,包括:

1) 完备的数学解释和理论指导是卷积神经网络进一步发展过程中无法回避的问题。作为一个基于实证的研究领域,卷积神经网络的理论研究目前还相对比较滞后。卷积神经网络的相关理论研究对卷积神经网络的进一步发展具有非常重要的意义。

2) 卷积神经网络的结构研究还具有很大的空间。目前的研究表明,仅仅通过简单地增加网络的复杂程度,会遇到一系列的瓶颈,如: 过拟合问题,网络退化问题等。卷积神经网络性能的提升需要依靠更加合理的网络结构设计。

3) 卷积神经网络的参数众多,但是目前的相关设置大多基于经验和实践,参数的量化分析与研究是卷积神经网络的一个有待解决的问题。

4) 卷积神经网络的模型结构不断改进,旧有的数据集已经不能满足当前的需求。数据集对于卷积神经网络的结构研究和迁移学习研究等都具有重要意义。数量和类别更多、数据形式更为复杂是当前相关研究数据集的发展趋势。

5) 迁移学习理论的应用,有助于进一步拓展卷积神经网络向更为广阔的应用领域发展; 并且,基于任务的端到端卷积神经网络的设计(如: Faster R-CNN, FCN 等) 有助于提升网络的实时性,是目前的发展趋势之一。

6) 虽然卷积神经网络在众多应用领域取得了优异的成绩,但是关于其完备性的相关研究与证明仍然是目前较为匮乏的部分。卷积神经网络的完备性研究有助于进一步理解卷积神经网络与人类视觉系统之间的原理差异,并且帮助发现和解决当前网络结构存在的认知缺陷。

6　结语

本文对卷积神经网络的历史、原理进行了简要的介绍,重点从卷积神经网络的过拟合问题、结构研究、原理分析、迁移

学习共四个方面对卷积神经网络当前的发展状况进行了综述。另外,本文还对于目前卷积神经网络已经取得的一些应用成果进行了分析,指出了当前卷积神经网络相关研究的一些缺陷及发展方向。卷积神经网络是当下一个具有很高热度的研究领域,具有广阔的研究前景。

参考文献:

[1] 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.

[2] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527 - 1554.

[3] LEE H, GROSSE R, RANGANATH R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations [C]// ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning. New York: ACM, 2009: 609 - 616.

[4] HUANG G B, LEE H, ERIK G. Learning hierarchical representations for face verification with convolutional deep belief networks [C]// CVPR '12: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2012: 2518 - 2525.

[5] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 1106 - 1114.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 580 - 587.

[7] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 3431 - 3440.

[8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2015-11-04]. http://www.robots.ox.ac.uk:5000/~vgg/publications/2015/Simonyan15/simonyan15.pdf.

[9] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 1 - 8.

[10] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [EB/OL]. [2016-01-04]. https://www.researchgate.net/publication/286512696_Deep_Residual_Learning_for_Image_Recognition.

[11] PAN S J, YANG Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345 - 1359.

[12] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch [J]. Journal of Machine Learning Research, 2011, 12(1): 2493 - 2537.

[13] OQUAB M, BOTTOU L, LAPTEV I, et al. Learning and transferring mid-level image representations using convolutional neural networks [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1717 - 1724.

[14] HUBEL D H, WIESEL T N. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex [J]. Journal of Physiology, 1962, 160(1): 106 - 154.

[15] FUKUSHIMA K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position [J]. Biological Cybernetics, 1980, 36(4): 193 - 202.

[16] WAIBEL A, HANAZAWA T, HINTON G, et al. Phoneme recognition using time-delay neural networks [M]// Readings in Speech Recognition. Amsterdam: Elsvier, 1990: 393 - 404.

[17] VAILLANT R, MONROCQ C, LE CUN Y. Original approach for the localization of objects in images [J]. IEE Proceedings—Vision, Image and Signal Processing, 1994, 141(4): 245 - 250.

[18] 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.

[19] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database [C]// Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2009: 248 - 255.

[20] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 2625 - 2634.

[21] VINYALS O, TOSHEV A, BENGIO S, et al. Show and tell: a neural image caption generator [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 3156 - 3164.

[22] MALINOWSKI M, ROHRBACH M, FRITZ M. Ask your neurons: a neural-based approach to answering questions about images [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 1 - 9.

[23] ANTOL S, AGRAWAL A, LU J, et al. VQA: visual question answering [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 2425 - 2433.

[24] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]// Proceedings of European Conference on Computer Vision, LNCS 8689. Berlin: Springer, 2014: 818 - 833.

[25] JI S, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221 - 231.

[26] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91 - 110.

[27] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2005: 886 - 893.

[28] LECUN Y, BENGIO Y, HINTON G E. Deep learning [J]. Nature, 2015, 521(7553): 436 - 444.

[29] 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. (SUN Z J, XUE L, XU Y M, et al. Overview of deep learning [J]. Application Research of Computers, 2012, 29(8): 2806 - 2810)

[30] DONAHUE J, JIA Y, VINYALS O, et al. DeCAF: a deep convolutional activation feature for generic visual recognition [J]. Computer Science, 2013, 50(1): 815 - 830.

[31] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: an astounding baseline for recognition [EB/OL]. [2015-11-22]. http://www.csc.kth.se/~azizpour/papers/ha_cvpr14w.pdf.

[32] SERMANET P, KAVUKCUOGLU K, CHINTALA S, et al. Pedestrian detection with unsupervised multi-stage feature learning [C]// CVPR '13: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2013: 3626 - 3633.

[33] KARPATHY A, TODERICI G, SHETTY S, et al. Large-scale video classification with convolutional neural networks [C]// CVPR '14: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1725 - 1732.

[34] TOSHEV A, SZEGEDY C. DeepPose: human pose estimation via deep neural networks [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1653 - 1660.

[35] KALCHBRENNER N, GREFENSTETTE E, BLUNSOM P. A convolutional neural network for modelling sentences [EB/OL]. [2016-01-07]. http://anthology.aclweb.org/P/P14/P14-1062.pdf.

[36] KIM Y. Convolutional neural networks for sentence classification [EB/OL]. [2016-01-07]. http://anthology.aclweb.org/D/D14/D14-1181.pdf.

[37] ABDEL-HAMID O, MOHAMMED A, JIANG H, et al. Convolutional neural networks for speech recognition [J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2014, 22(10): 1533 - 1545.

[38] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search [J]. Nature, 2016, 529(7587): 484 - 489.

[39] ZEILER M D, FERGUS R. Stochastic pooling for regularization of deep convolutional neural networks [EB/OL]. [2016-01-11]. http://www.matthewzeiler.com/pubs/iclr2013/iclr2013.pdf.

[40] MURPHY K P. Machine Learning: A Probabilistic Perspective [M]. Cambridge, MA: MIT Press, 2012: 82 - 92.

[41] CHATFIELD K, SIMONYAN K, VEDALDI A, et al. Return of the devil in the details: delving deep into convolutional nets [EB/OL]. [2016-01-12]. http://www.robots.ox.ac.uk/~vedaldi/assets/pubs/chatfield14return.pdf.

[42] GOODFELLOW I J, WARDE-FARLEY D, MIRZA M, et al. Maxout networks [EB/OL]. [2016-01-12]. http://www-etud.iro.umontreal.ca/~goodfeli/maxout.pdf.

[43] LIN M, CHEN Q, YAN S. Network in network [EB/OL]. [2016-01-12]. http://arxiv.org/pdf/1312.4400v3.pdf.

[44] MONTAVON G, ORR G, MÜLLER K R. Neural Networks: Tricks of the Trade [M]. 2nd ed. London: Springer, 2012: 49 - 131.

[45] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult [J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157 - 166.

[46] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 1026 - 1034.

[47] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaption of feature detectors [R/OL]. [2015-10-26]. http://arxiv.org/pdf/1207.0580v1.pdf.

[48] WAN L, ZEILER M, ZHANG S, et al. Regularization of neural networks using dropconnect [C]// Proceedings of the 2013 International Conference on Machine Learning. New York: ACM Press, 2013: 1058 - 1066.

[49] HE K, SUN J. Convolutional neural networks at constrained time cost [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 5353 - 5360.

[50] SPRINGENBERG J T, DOSOVITSKIY A, BROX T, et al. Striving for simplicity: the all convolutional net [EB/OL]. [2015-12-24]. http://arxiv.org/pdf/1412.6806.pdf.

[51] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE [EB/OL]. [2015-12-24]. http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf.

[52] OLIVA A, TORRALBA A. Modeling the shape of the scene: a holistic representation of the spatial envelope [J]. International Journal of Computer Vision, 2001, 42(3): 145 - 175.

[53] WANG J, YANG J, YU K. Locality-constrained linear coding for image classification [C]// Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2010: 3360 - 3367.

[54] ZEILER M D, TAYLOR G W, FERGUS R. Adaptive deconvolutional networks for mid and high level feature learning [C]// ICCV '11: Proceedings of the 2011 International Conference on Computer Vision. Piscataway, NJ: IEEE, 2011: 2018 - 2025.

[55] NGUYEN A, YOSINSKI J, CLUNE J, et al. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 427 - 436.

[56] FLOREANO D, MATTIUSSI C. Bio-inspired Artificial Intelligence: Theories Methods and Technologies [M]. Cambridge, MA: MIT Press, 2008: 1 - 97.

[57] 庄福振,罗平,何清,等. 迁移学习研究进展[J]. 软件学报, 2015,26(1):26 - 39. (ZHUANG F Z, LUO P, HE Q, et al. Survey on transfer learning research [J]. Journal of Software, 2015, 26(1): 26 - 39.)

[58] LI F, FERGUS R, PERONA P. One-shot learning of object categories [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4):594 - 611.

[59] GRIFFIN B G, HOLUB A, PERONA P. The Caltech-256 [R/OL]. [2016-01-03]. http://xueshu.baidu.com/s?wd=paperuri%3A%28699092e99ad6f96f8696507d539a51c8%29&filter=sc_long_sign&tn=SE_xueshusource_2kduw22v&sc_vurl=http%3A%2F%2Fciteseer.ist.psu.edu%2Fshowciting%3Fcid%3D11093943&ie=utf-8&sc_us=16824823650146432853.

[60] ZHOU B, LAPEDRIZA A, XIAO J, et al. Learning deep features for scene recognition using places database [C]// Proceedings of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press. 2014:487 - 495.

[61] LOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [EB/OL]. [2016-01-06]. http://jmlr.org/proceedings/papers/v37/ioffe15.pdf.

[62] GIRSHICK R B. Fast R-CNN [EB/OL]. [2016-01-06]. http://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Girshick_Fast_R-CNN_ICCV_2015_paper.pdf.

(下转第 2565 页)

for Video Technology, 2014, 24(9): 1522 – 1540.

[6] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254 – 1259.

[7] HOU X, ZHANG L. Saliency detection: a spectral residual approach [C]// Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2007: 1 – 8.

[8] JUDD T, EHINGER K, DURAND F, et al. Learning to predict where humans look [C]// Proceedings of the IEEE 12th International Conference on Computer Vision. Piscataway, NJ: IEEE, 2009: 2106 – 2113.

[9] KIM J S, SIM J Y, KIM C S. Multiscale saliency detection using random walk with restart [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(2): 198 – 210.

[10] CHENG M M, ZHANG G X, MITRA N J, et al. Global contrast based salient region detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 569 – 582.

[11] KRAHENBUHL P. Saliency filters: contrast based filtering for salient region detection [C]// CVPR '12: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2012: 733 – 740.

[12] YANG C, ZHANG L, LU H, et al. Saliency detection via graph-based manifold ranking [C]// CVPR '13: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2013: 3166 – 3173.

[13] LIANG Z, WANG M, ZHOU X, et al. Salient object detection based on regions [J]. Multimedia Tools and Applications, 2014, 68(3): 517 – 544.

[14] SHI J, YAN Q, XU L, et al. Hierarchical image saliency detection on extended CSSD [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(4): 717 – 729.

[15] LIU T, YUAN Z, SUN J, et al. Learning to detect a salient object [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(2): 353 – 367.

[16] ZHAO R, OUYANG W, LI H, et al. Saliency detection by multi-context deep learning [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 1265 – 1274.

[17] TONG N, LU H, RUAN X, et al. salient object detection via bootstrap learning [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 1884 – 1892.

[18] LI C, YUAN Y, CAI W, et al. Robust saliency detection via regularized random walks ranking [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 2710 – 2717.

[19] QIN Y, LU H, XU Y, et al. Saliency detection via cellular automata [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 110 – 119.

[20] WEI Y, WEN F, ZHU W, et al. Geodesic saliency using background priors [C]// Proceedings of the 2012 12th European Conference on Computer Vision, LNCS 7574. Berlin: Springer, 2012: 29 – 42.

[21] ZHOU D, WESTON J, GRETTON A, et al. Ranking on data manifolds [EB/OL]. [2015-11-08]. http://www.kyb.mpg.de/fileadmin/user_upload/files/publications/pdfs/pdf2334.pdf.

[22] WANG B, PAN F, HU K M, et al. Manifold-ranking based retrieval using k-regular nearest neighbor graph [J]. Pattern Recognition, 2012, 45(4): 1569 – 1577.

[23] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels [EB/OL]. [2015-12-11]. http://islab.ulsan.ac.kr/files/announcement/531/SLIC_Superpixels.pdf.

[24] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection [C]// CVPR 2009: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2009: 1597 – 1604.

Background

This work is partially supported by the National Key Technology Research and Development Program of China (2011BAH25B04).

ZHU Zhengyu, born in 1959, Ph. D., professor. His research interests include Web intelligent retrieval, data mining, image processing.

WANG Mei, born in 1992, M. S. candidate. Her research interests include image processing.

(上接第 2515 页)

[63] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [EB/OL]. [2016-01-06]. http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf.

[64] UIJLINGS J, SANDE K, GEVERS T, et al. Selective search for object recognition [J]. International Journal of Computer Vision, 2013, 104(2): 154 – 171.

[65] KHAN S H, BENNAMOUN M, SOHEL F, et al. Automatic feature learning for robust shadow detection [C]// CVPR '14: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1939 – 1946.

[66] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: closing the gap to human-level performance in face verification [C]// CVPR '14: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1701 – 1708.

[67] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 815 – 823.

[68] LEVI G, HASSNER T. Age and gender classification using convolutional neural networks [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington, DC: IEEE Computer Society, 2015: 34 – 42.

Background

This work is partially supported by the National Key Technology R&D Program (2012BAH44F02), Project on the Integration of Industry, Education and Research of Guangdong Province (M17010601CXY2011057).

LI Yandong, born in 1984, Ph. D. candidate. His research interests include machine learning, computer vision.

HAO Zongbo, born in 1977, Ph. D., associate professor. His research interests include image understanding, video processing.

LEI Hang, born in 1960, Ph. D., professor. His research interests include image processing.