

DOI:10.3969/j.issn.1000-1565.2017.06.012

卷积神经网络及其研究进展

翟俊海¹, 张素芳², 郝璞¹

(1.河北大学 数学与信息科学学院 河北省机器学习与计算智能重点实验室,河北 保定 071002;

2.中国气象局气象干部培训学院 河北分院,河北 保定 071002)

摘要:深度学习是目前机器学习领域最热门的研究方向,轰动全球的 AlphaGo 就是用深度学习算法训练的.卷积神经网络是用深度学习算法训练的一种模型,它在计算机视觉领域应用广泛,而且获得了巨大的成功.本文的主要目的有 2 个:一是帮助读者深入理解卷积神经网络,包括网络结构、核心概念、操作和训练;二是对卷积神经网络的近期研究进展进行综述,重点综述了激活函数、池化、训练及应用 4 个方面的研究进展.另外,还对其面临的挑战和热点研究方向进行了讨论.本文将为从事相关研究的人员提供很好的帮助.

关键词:机器学习;深度学习;卷积神经网络;计算机视觉;训练算法

中图分类号:TP181 **文献标志码:**A **文章编号:**1000-1565(2017)06-0640-12

Convolutional neural network and its research advances

ZHAI Junhai¹, ZHANG Sufang², HAO Pu¹

(1. Key Laboratory of Machine Learning and Computational Intelligence, College of Mathematics and Information Science, Hebei University, Baoding 071002, China; 2. Hebei Branch of Meteorological Cadres Training Institute, China Meteorological Administration, Baoding 071002, China)

Abstract: Deep learning is the most popular research topic in the field of machine learning, AlphaGo which overwhelmingly impacts the world is trained with deep learning algorithms. Convolution neural network (CNN) is a model trained with deep learning algorithm, CNN is widely and successfully applied in computer version. The main purpose of this paper includes two aspects: one is to provide readers with some insights into CNN including its architecture, related concepts, operations and its training; the other is to present a comprehensive survey on research advances of CNN, mainly focusing on 4 aspects: activation functions, pooling, training and applications of CNN. Furthermore, the emerging challenges and hot research topics of CNN are also discussed. This paper can be very helpful to researchers in related field.

Key words: machine learning; deep learning; convolutional neural network; computer version; training algorithms

收稿日期:2017-09-09

基金项目:国家自然科学基金资助项目(71371063);河北省自然科学基金资助项目(F2017201026);河北大学自然科学研究计划项目(799207217071)

第一作者:翟俊海(1964—),男,河北易县人,河北大学教授,博士,主要从事机器学习和数据挖掘方向研究.

E-mail:mczjh@126.com

通信作者:张素芳(1966—),女,河北蠡县人,中国气象局气象干部培训学院河北分院副教授,主要从事机器学习方向研究.

E-mail:mczsf@126.com

美国心理学家 McCulloch 和 Pitts 于 1943 年首次提出了人工神经元模型,即著名的 M-P 模型^[1],开启了人工神经网络的研究.在这 70 多年的时间里,神经网络研究曾经几起几落.Rosenblatt 于 1958 年提出的感知机(Perceptron)模型^[2],标志着神经网络研究迎来了第 1 次热潮,这次研究热潮持续了近 10 年,直到 1969 年,Minsky 和 Papert 从数学的角度证明了单层神经网络逼近能力有限^[3],甚至连简单的异或问题都不能解决,使神经网络研究陷入了第 1 次低潮.神经网络研究迎来第 2 次热潮的起点是 Werbos 在其博士论文中提出了误差反向传播的思想^[4],但当时并没有引起研究人员的关注,直到 1986 年 Rumelhart 等成功实现了用反向传播算法^[5](即著名的 BP 算法)训练多层神经网络,神经网络研究才真正迎来第 2 次研究的热潮,此后近 10 年,BP 算法始终占据统治地位.但是 BP 算法也有自身的缺陷,例如容易产生过拟合、梯度消失、局部最优等问题.1995 年 Vapnik 和 Cortes 提出了支持向量机^[6](SVM: support vector machine),由于 SVM 具有坚实的理论基础,在应用中也表现出了比神经网络更好的效果,所以 SVM 成为热点研究内容,而神经网络研究则不冷不热.人类进入 21 世纪后,特别是随着大数据时代的到来,神经网络研究又迎来了一次研究热潮,这次研究热潮的起点是 2006 年加拿大多伦多大学 Hinton 教授和他的学生 Salakhutdinov 提出的深度学习思想^[7].这次热潮的标志就是深度学习,从某种意义上讲,深度学习是训练深度模型(如深度神经网络)的算法,这些模型包括:卷积神经网络、受限波尔兹曼机、信念网络和自动编码器^[8].文献^[9]对深度学习进行了全面深入的综述,具有很高的参考价值.

卷积神经网络(CNN)^[10]是一种著名的深度学习模型,其名称的由来是因为卷积运算被引入到了这种模型中.CNN 可以归类为多层前馈神经网络模型,但与传统的多层前馈神经网络不同,CNN 的输入是二维模式(如图像),其连接权是二维权矩阵(也称为卷积核),基本操作是二维离散卷积和池化(Pooling).由于 CNN 可以直接处理二维模式,所以它在计算机视觉领域得到了非常广泛的应用.例如,CNN 已成功应用于图像分类、目标检测和目标跟踪等许多领域.

实际上,卷积神经网络模型并不是 2006 年以后才提出的,早在 1998 年,LeCun 等^[10]就提出了一种称为 LeNet 的卷积神经网络模型,并用于手写数字识别.只是由于当时缺乏大规模的训练数据,计算机的计算能力也有限,所以 LeNet 在解决复杂问题(例如大规模的图像和视频分类问题)时,效果并不好.2006 年以后,特别是随着大数据时代的到来,在高性能的计算平台(如高性能的 PC 机、图形工作站、云计算平台等)上用大规模的数据集训练复杂的模型成为可能.正是在这种背景下,研究人员提出了许多卷积神经网络模型.下面首先以 LeNet 为例介绍卷积神经网络的结构,并综述相关的研究;然后综述卷积神经网络的训练,包括训练的加速机制和几种常用的开源框架;接下来综述卷积神经网络的应用;最后总结了未来几年卷积神经网络研究的热点和面临的挑战.

1 卷积神经网络的结构

卷积神经网络是一种多层前馈神经网络,其基本构成要素包括卷积层、池化层(也称为采样层)和全连接层.图 1 是卷积神经网络 LeNet 的结构图,它有 3 个卷积层,2 个采样层,1 个全连接层,加上输入层和输出层共包含 8 层.从图 1 可以看出,卷积神经网络 LeNet 的输入层是二维模式(如图像);隐含层由多个卷积层和采样层交替构成,卷积层用于提取不同的图像特征,采样层用于降维.卷积层和采样层的输出都称为特征映射图(二维模式);全连接层一般是一个分类器(如单隐含层的神经网络).

1.1 卷积层

一般地,每一个卷积层都由若干个结点构成.卷积层通过卷积运算提取图像的不同特征,卷积层结点的结构如图 2 所示.

在图 2 中,卷积层结点的输入是特征映射图,因为从形式上特征映射图是二维矩阵,所以全 1 矩阵也可以看作一个特征映射图. b_j^l 是 l 层结点 j 的偏置,是一个标量. \mathbf{k}_{js}^l ($s=1,2,\dots$)是可学习的卷积核矩阵,矩阵中的元素就是权值.每一个卷积核矩阵 \mathbf{k}_{js}^l 与其对应的输入特征映射图做卷积运算,得到一个二维模式(矩阵),这样共得到 m_i 个二维模式(矩阵).然后,对这些矩阵求和,再加上偏置,作为激活函数的输入,最后得到

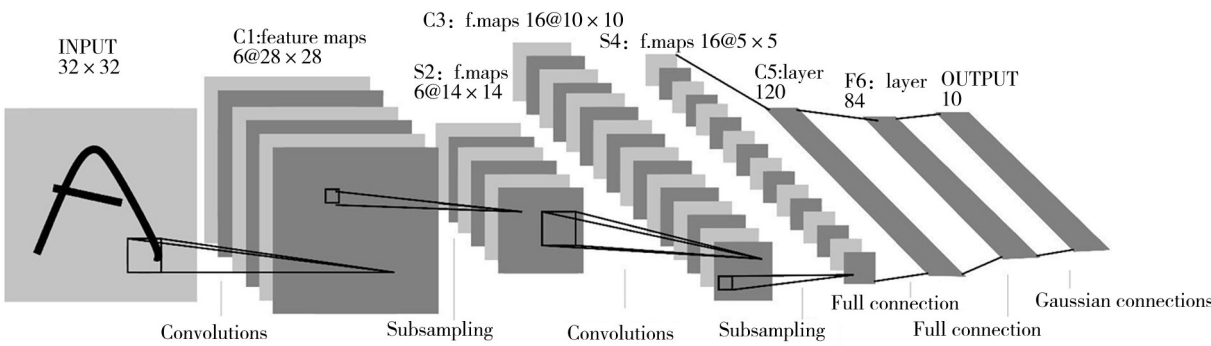


图 1 卷积神经网络 LetNet 的结构^[10]

Fig.1 Architecture of convolutional neural network LetNet^[10]

结点的输出特征图.卷积层结点的输出可用下面的公式表示：

$$\mathbf{x}_j^l = f\left(\sum_{i=1}^{m_i} \mathbf{x}_j^{l-1} * \mathbf{k}_{ji}^l + b_j^l \mathbf{I}\right), \tag{1}$$

其中 \mathbf{x}_j^{l-1} 和 \mathbf{x}_j^l 分别为第 $l-1$ 层和第 l 层的特征图, \mathbf{I} 是元素全是 1 的矩阵, “ $*$ ”是卷积运算符.

- 说明：
- 1)一般地,卷积核(也称为滤波器矩阵)是 3×3 或 5×5 的小矩阵;
 - 2)在做卷积运算时,卷积核(滤波器矩阵)与对应的特征图(或图像)的一个局部区域相连接,这个局部区域称为卷积核的局部感受域.整个特征图(或整幅图像)共享卷积核权值,这称为权值共享.

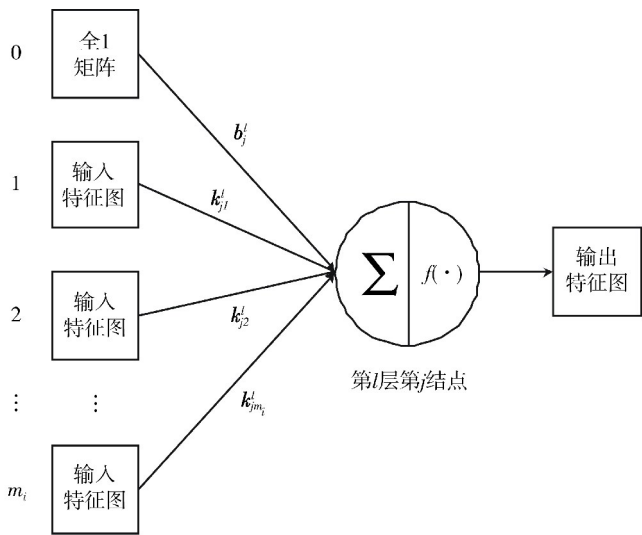


图 2 第 l 个卷积层第 j 个结点的结构

Fig.2 Architecture of j th node in l th convolutional layer

关于卷积层,研究的重点主要在激活函数上.激活函数在卷积神经网络中起着重要作用,它将非线性性质引入到网络中,非线性性质是保证网络具有一致逼近能力的重要因素.另外,因为大多数前馈神经网络(也包括卷积神经网络)的训练算法都是基于误差反向传播的思想,激活函数的性质对梯度的计算具有很大的影响.例如,激活函数的输出饱和程度就严重影响着网络的收敛性,输出饱和程度越高,网络的收敛性越差.因此激活函数对卷积神经网络的训练速度或收敛性具有很大的影响.神经网络训练的稳定性一般通过将网络中的变量(包括网络结点的输入和输出)限定在一定的范围内.而激活函数就有对网络变量限界的作用,所以激活函数对网络训练的稳定性也有重要的影响.总之,激活函数对网络的收敛性、稳定性和一致逼近能力都有重要的影响.

卷积神经网络中的激活函数大致可分为 2 类: Sigmoid 型的和非 Sigmoid 型的. Sigmoid 型的激活函数具有连续性和可微性, 它们将输入变量变换到一个有限区间. 最常用的 Sigmoid 型激活函数包括 logistic 函数和 tanh 函数. logistic 函数将输入变量变换到区间 $[0, 1]$, 其定义为

$$f(x) = \frac{1}{1 + e^x}, \tag{2}$$

tanh 函数将输入变量变换到区间 $[-1, 1]$, 其定义为

$$f(x) = \tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}, \tag{3}$$

与 logistic 函数相比, tanh 函数的梯度具有更好的渐变性, 而这在神经网络的训练中是更期望的. 另外, tanh 函数关于原点是奇对称的, 这种性质可使网络收敛更快. 然而, tanh 函数在两侧都具有比较大的饱和区域, 这样难以逼近其边界值. 因此研究人员通过在 tanh 函数中引入幅度参数和倾斜参数, 提出了伸缩 tanh 函数^[10], 可以克服上述缺点. 伸缩 tanh 函数的定义为

$$f(x) = \alpha \tanh(\beta x) = \frac{\alpha(1 - e^{-2\beta x})}{1 + e^{-2\beta x}}, \tag{4}$$

在 LeCun 等^[10]提出的卷积神经网络 LetNet 中, 使用的就是这种激活函数, 而且 LeCun 等通过实验研究发现: $\alpha = 1.715\ 9, \beta = 0.666\ 7$ 时, LetNet 的性能最优.

实际上, 在 2012 年以后提出的卷积神经网络模型中, 大多用的都不是 Sigmoid 型的激活函数. 例如, 在著名的 AlexNet 模型中^[11], 使用的激活函数是 ReLU (rectified linear unit). 常用的非 Sigmoid 型的激活函数除 ReLU 外, 还包括 LReLU (Leaky ReLU)、BiFire (Bi-firing) 等.

ReLU 激活函数^[11]是近年来最著名的非饱和激活函数之一, ReLU 的定义为

$$ReLU(x) = \max(x, 0), \tag{5}$$

Krizhevsky 等的工作显示, 具用 ReLU 激活函数的卷积神经网络, 即便没有预训练也能有效训练^[11].

ReLU 的不足之处是当神经元结点没有激活时, 其梯度为 0. 这样可能导致初始没有激活的神经元结点, 在梯度优化过程中, 它们的权值不会得到调整. 另外, 由于零梯度还可能降低网络的训练速度. 针对这一问题, Mass 等^[12]提出了 LReLU (Leaky ReLU). 其定义为

$$LReLU(x) = \max(x, 0) + \lambda \min(x, 0), \tag{6}$$

其中, $\lambda \in$ 是用户预定义的参数. 如果将 λ 改为可学习的参数, 则得到 PReLU (Parameteric ReLU)^[12-13].

BiFire 是 Li 等提出的另一种新的非 Sigmoid 型激活函数^[14], 它能消除梯度扩散现象, 其定义为

$$\text{BiFire}(x) = \begin{cases} x - \frac{A}{2} & x > A, \\ -x - \frac{A}{2} & x < -A, \\ \frac{x^2}{2A} & \text{其他}, \end{cases} \tag{7}$$

其中, A 是一个光滑参数.

1.2 池化层

卷积层通过不同的卷积核提取图像的不同特征, 而池化层 (也称为采样层) 通过不同的池化操作对图像进行降维, 并提高图像特征的变换不变特性^[15]. 常用的池化操作包括最大池化 (max pooling) 和平均池化 (average pooling). 池化操作和卷积操作有类似的地方, 即池化窗口按着一定规则在输入特征图中, 按从上到下, 从左到右顺序的移动. 最大池化对窗口所覆盖的子矩阵求其元素的最大值, 作为池化输出特征图 (一个矩阵) 中的一个元素.

近几年, 关于池化层的研究主要集中在新的池化操作及将其他技术 (例如, Dropout 技术、学习技术、自适应技术等) 引入到池化中. 在最大池化和平均池化的基础上, 提出的新池化操作包括混合池化、 L_p 范数池化、空域金字塔池化等.

Yu 等^[16]将最大池化和平均池化结合起来,提出了混合池化方法.混合池化可用公式(8)描述.

$$y_{ij} = \lambda \max_{(m,n) \in R_{ij}} \alpha_{m,n} + (1 - \lambda) \frac{1}{|R_{ij}|} \sum_{(m,n) \in R_{ij}} \alpha_{m,n}, \quad (8)$$

其中, λ 是取值为 0 和 1 的随机变量, R_{ij} 是池化区域或池化窗口, $|R_{ij}|$ 表示 R_{ij} 中所含元素个数, $\alpha_{m,n}$ 表示窗口中的元素.

Gulcehre 等^[17]将最大池化和平均池化的思想进行了推广,提出了 Lp 范数池化(Lp Pooling),其定义如下:

$$y_{ij} = \left[\sum_{(m,n) \in R_{ij}} (\alpha_{m,n})^p \right]^{\frac{1}{p}}. \quad (9)$$

在文献[18]中,Estrach 等对 Lp 范数池化进行了理论分析,得出了结论:“与最大池化和平均池化相比,Lp 范数池化能获得更好的泛化性能”.

He 等^[19]提出的空域金字塔池化(spatial pyramid pooling)将不同尺度的池化特征图组合在一起,可以得到固定长度的输出特征图,不论输入特征图的大小是多少.在空域池化研究中,研究人员还提出了一些新的方法.例如,Xie 等^[20]提出了异构特征空域池化方法, Lee 等^[21]提出了沿时间轴的池化方法, Perlaza 等^[22]对频域池化和空域池化进行了比较研究.

近几年,研究人员还将其他一些技术等引入到池化中,提出了相应池化方法.例如, Wu 等^[23]将 Dropout 技术引入到池化中,提出了 Dropout 最大池化方法. Wang 等^[24]将自适应技术引入到池化中,提出了自适应池化方法. Sun 等^[25]将学习机制引入到池化运算中,提出了一种基于训练误差最小化的学习池化方法.

1.3 新的卷积神经网络模型

在 LetNet 模型的基础上,最近几年,研究人员提出了许多新的卷积神经网络模型.代表性的模型包括以下几种:

1) Krizhevsky 等^[11]于 2012 年提出的 AlexNet 模型.该模型以巨大的优势获得了当年 ImageNet 竞赛(也称为 ILSVRC(ImageNet large scale visual recognition challenge)竞赛)的冠军, top-5 的错误率为 16.4%, 第 2 名的 top-5 的错误率为 26.2%.从结构上来讲, AlexNet 和 LeNet 是类似的,但具有更宽和更深的结构.在 AlexNet 模型中, Krizhevsky 等首次成功地应用了 ReLU、Dropout 和 LRN(local response normalization)等技术,同时还使用了 GPU(graphics processing unit)加速技术,并开源了他们在 GPU 上训练卷积神经网络的 CUDA(compute unified device architecture)代码^[26-27]. AlexNet 共有 8 个需要训练参数的层,不包括池化层和 LRN 层,前 5 层为卷积层,后 3 层为全连接层,最后一个全连接层是具有 1 000 个输出的 softmax. AlexNet 的成功,确立了其在计算机视觉中的统治地位.

2) Lin 等^[28]提出的 NIN(Network In Network). NIN 模型由堆叠在一起的 3 个 MLP(MultiLayer Perceptron)卷积层和一个全局平均池化层构成. MLP 卷积层由 1×1 的卷积层连接一个 MLP 层, MLP 层的作用是增加单个卷积特征的有效性.从 NIN 的结构可以看出,其思想其实很简单.然而,其后的许多模型,例如, Google Inception Net 和 ResNet,都用到了 NIN 的基本思想.

3) Simonyan 等^[29]提出的 VGGNet.该模型获得了 2014 年 ILSVRC 竞赛的亚军.与 AlexNet 模型相比, VGGNet 除了包含更多的层(19 层)之外,所有的卷积层都使用 3×3 同样大小的卷积核和 2×2 同样大小的池化窗口,池化操作采用最大池化.

4) Szegedy 等^[30]提出的 Google Inception Net. Inception 结构是这种模型的基本构成单元,其基本思想是找出最优的局部稀疏结构. Google Inception Net 是一个大家族,由 4 个模型构成: Inception V1-V4.在参考文献[30]中提出的模型是 Inception V1,它获得了 2014 年 ILSVRC 竞赛的冠军, top-5 错误率为 6.67%.该模型最大的特点是在控制计算量和参数量的同时,获得了非常好的性能. Inception V1 模型包括 22 层,比 AlexNet(8 层)和 VGGNet(19 层)更深.但是, Inception V1 的计算量和参数量却比上述 2 种网络模型少很多.例如,其参数量是 AlexNet 的 $\frac{1}{12}$.可以说 Inception V1 是一个性能非常好的模型,其计算量和参数量少的原因包括 2 点:①借鉴 NIN 模型,用全局平均池化取代了全连接层;②精心设计的 Inception 结构极大地提

高了参数的利用率.Inception V2 模型是 2015 年提出的^[31],它借鉴了 VGGNet,用 2 个 3×3 的小卷积核代替了 5×5 的大卷积核.此外,在该模型中,还提出了著名的 BN(batch normalization)模型加速机制,top-5 错误率降至 4.8%.BN 的基本思想是在每次 mini-batch 反向传播之后,重新对参数进行 0 均值 1 方差标准化.这样可以使用更大的学习速率,参数初始化也变得简单高效.实际上,BN 充当着正则化、减少甚至消除掉 Dropout 的作用.Inception V3 模型也是 2015 年提出的^[32],该模型主要从 2 个方面进行了改造:一是将较大的二维卷积核拆分成 2 个较小的一维卷积核;二是优化了 Inception 结构^[26].Inception V4 是 2016 年提出的^[33],与 Inception V3 相比,在 Inception V4 中,主要引入了下面将要介绍的 ResNet(Residual Neural Network)模型的思想.

5)He 等提出的 ResNet(Residual Neural Network)^[34].该模型在 2015 年的 ILSVRC 竞赛中获得了冠军,top-5 的错误率为 3.57%.ResNet 包含 152 层,是一个很深的网络,但是其参数量却比 VGGNet 低,性能非常优越.ResNet 的基本构成模块是残差学习单元,包括 2 种:2 层残差学习单元和 3 层残差学习单元.在 2 层残差学习单元中,卷积核大小都是 3×3 的;在 3 层残差学习单元中,包括 2 个大小为 1×1 的卷积核和一个 3×3 的卷积核, 3×3 的卷积核在 2 个 1×1 的卷积核之间.ResNet 和普通 CNN 的最大区别在于,ResNet 有很多旁路的支线将输入直接连接到后面的层,使得后面的层可以直接学习残差,简化了学习目标和学习难度^[26].此外,ResNet 通过直接将输入绕道传到输出,保护了信息的完整性,从某种程度上解决了普通 CNN 的信息丢失问题.

除了上面这些模型外,比较有代表性模型还有 Zeiler 等提出的 ZFNet^[35],Iandola 等提出的 DenseNet^[36]和 SqueezeNet^[37].由于篇幅所限,对于这 3 种模型不再展开综述,有兴趣的读者可参考相关论文.

2 卷积神经网络的训练

2.1 损失函数

针对具体应用,用卷积神经网络解决实际问题时,选择合适的损失函数至关重要.除传统的均方误差损失函数外,常用的损失函数包括以下几种.

1)软最大化损失函数

给定训练集, $D = \{(\mathbf{x}_i, y_i) | i = 1, 2, \dots, n, y_i \in \{1, 2, \dots, k\}\}$,对于第 i 个样例,预测为第 j 类的后验概率用下面的软最大化函数进行计算:

$$p_i^j = \frac{e^{z_i^j}}{\sum_{l=1}^k e^{z_i^l}},$$

其中, $z_i^j = \mathbf{w}_j^T \mathbf{a}_i b_j$.

软最大化损失函数定义如下^[11]:

$$L = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k 1\{y_i = j\} \log p_i^j.$$

2)对比损失函数

给定第 i 对数据, $(\mathbf{x}_a^{(i)}, \mathbf{x}_\beta^{(i)})$,假定 $(\mathbf{z}_a^{(i,l)}, \mathbf{z}_\beta^{(i,l)})$ 表示网络第 l 层的输出.在文献[38]和[39]中,Chopra 等针对卷积神经网络的训练,定义了如下的对比损失函数:

$$L = \frac{1}{2n} \sum_{i=1}^n [y d^{(i,l)} + (1 - y) \max(m - d^{(i,l)}, 0)],$$

其中, $d^{(i,l)} = \|\mathbf{z}_a^{(i,l)} - \mathbf{z}_\beta^{(i,l)}\|_2$, m 是一个影响非匹配对的间隔参数.如果 $(\mathbf{x}_a^{(i)}, \mathbf{x}_\beta^{(i)})$ 是匹配对,那么 $y = 1$;否则, $y = 0$.

3)K-L 散度损失函数

K-L 散度是 2 个概率分布 $p(x)$ 和 $q(x)$ 之间的距离度量,其定义为

$$D_{KL}(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)}.$$

在文献[40]中,Kingma 等针对自动编码器的训练,提出了一种基于 K - L 散度的损失函数,其定义如下:

$$L = E_{z \sim q_{\varphi}(z|x)} \log p_{\theta}(x|z) - D_{KL}(q_{\theta}(z|x) \parallel p(z)),$$

其中, φ 是编码器参数, θ 是解码器参数,该损失函数也适用于卷积神经网络.

2.2 卷积神经网络的训练方法及加速策略

因为从宏观上看 CNN 属于多层前馈神经网络,所以可以用反向传播算法(BP 算法)及其各种改进的版本^[41](例如,共轭梯度反向传播算法、递归最小二乘反向传播算法、Levenberg-Marquardt 反向传播算法等)进行训练.例如,在 LetNet 中用的是基于随机梯度下降^[42]的随机对角线 Levenberg-Marquardt 反向传播算法^[10].权参数更新公式为

$$w_k = w_{k-1} - \epsilon_k \frac{\partial E^p}{\partial w_k},$$

其中, w 是权参数, E^p 是相对于模式 p 的瞬时均方误差损失函数, ϵ_k 是学习率参数.

因为在卷积神经网络中存在权值共享,所以上面公式中的偏导数用下面的公式计算:

$$\frac{\partial E^p}{\partial w_k} = \sum_{(i,j) \in V_k} \frac{\partial E^p}{\partial u_{ij}},$$

其中, u_{ij} 是结点 j 到结点 i 的连接权, V_k 是具有权值共享关系的结点编号对 (ij) 的集合,即结点 i 和结点 j 之间的连接共享权值 w_k .换句话说,对于任意的结点编号对 (i,j) ,如果 $(i,j) \in V_k$,则 u_{ij} 和 w_k 之间有如下关系:

$$u_{ij} = w_k.$$

一般地,学习速率 ϵ_k 不是一个常数.在 LetNet 中, ϵ_k 是损失函数沿方向 w_k 的二阶导数

$$\epsilon_k = \frac{\mu}{\mu + h_{kk}},$$

其中, μ 是一个常数, h_{kk} 是损失函数 E 关于 w_k 的二阶导数的估计. h_{kk} 的计算公式如下:

$$h_{kk} = \sum_{(i,j) \in V_k} \sum_{(l) \in V_k} \frac{\partial^2 E}{\partial u_{ij} \partial u_{kl}},$$

去掉 Hessian 矩阵中的非对角线上的元素,上式变为

$$h_{kk} = \sum_{(i,j) \in V_k} \frac{\partial^2 E}{\partial u_{ij}^2},$$

而 $\frac{\partial^2 E}{\partial u_{ij}^2}$ 是训练集上二阶导数的平均

$$\frac{\partial^2 E}{\partial u_{ij}^2} = \frac{1}{N} \sum_{p=1}^N \frac{\partial^2 E^p}{\partial u_{ij}^2},$$

其中, N 是训练集中的样例个数.

随着深度学习的发展,现在的卷积神经网络模型越来越复杂,训练复杂网络需要的数据集越来越大.这样,上述训练 LetNet 的反向传播算法已不能满足需要.近几年,研究人员提出了一些新的训练卷积神经网络的方法,这些方法都是设法加速网络的训练.根据加速机制,这些方法可大致分为 2 类:基于 GPU 或 CPU (多核和众核)的硬加速方法和基于加速算法的软加速方法.

基于 GPU 或 CPU(多核和众核)的硬加速方法是将训练卷积神经网络的算法用 GPU 编程实现,以提高训练的深度,其中,基于 GPU 的加速方法应用最普遍.在深度卷积神经网络发展中,具有重大影响的 Alex-Net^[11]使用 Dropout 技术和 GPU 相结合的方法加速网络的训练.Dropout 用于有选择地删除某些神经元,以避免模型过拟合,使用 GPU 加速训练,以减少训练时间.实际上,其他几种著名的卷积神经网络模型,例如 VGGNet、GoogleNet、ResNet 等,都是用 GPU 实现加速.其他代表性的工作包括:Li 等^[43]提出的 2 种基于 GPU 的加速卷积神经网络的训练方法.一种是用于加速卷积神经网络前向计算的图像组方法,另一种是基于 GPU 的低内存消耗的训练卷积神经网络的方法,该方法可以训练任意大的网络.Mathieu 等^[44]提出的基于傅里叶变换的快速卷积神经网络训练方法.该方法对卷积层结点的输入特征图首先做二维离散快速傅

里叶变换,然后在频域做卷积运算,并用 GPU 编程实现.因为相同的特征图傅里叶变换可以反复使用,所以可以加快卷积神经网络的训练.在基于多核和众核的加速方法中,代表性的工作包括:Rajeswar 等^[45]提出的多核处理器(Multi-core CPUs)和图形处理器(GPUs)的可扩展的卷积神经网络训练方法,与相应的串行训练方法相比,基于多核处理器的训练方法可提高 10 倍的训练速度,而基于图形处理器的训练方法可提高 12 倍的训练速度.Zlateski 等^[46]提出了 2 种卷积神经网络并行化训练方法,一种是基于多核处理器的并行化训练方法,另一种是基于众核处理器的并行化训练方法.前者能实现与几乎同等物理处理器核的加速效果,而后者可实现 90 倍加速比.另外,近几年研究人员还提出了基于云计算平台的加速方法.例如,Morcel 等^[47]提出了一种基于阿帕奇 SPARK 云计算环境的深度卷积网络训练方法,该方法对卷积操作的加速效果明显,在不同的数据集上可达到 40~250 倍的加速比.

软加速通过设计加速算法来提高卷积神经网络的训练速度.在这类方法中,除前面 Inception V2 模型中基于 BN 的加速方法外^[31],研究人员还提出了许多其他的软加速方法.Girshick 等^[48-49]提出了一种基于区域的卷积神经网络快速训练方法.该方法采用了多任务损失函数,单阶段训练能更新所有网络层参数,具有很快的训练速度.Gusmao^[50]提出了一种基于核召回的卷积神经网络快速训练方法.该方法的基本思想是开始用低分辨率核和输入特征图预训练卷积神经网络,然后利用卷积运算的空域扩展性质和全分辨率核精细调整网络参数.与其他方法相比,该方法可减少 20% 的训练时间,但没有任何精度损失.Cong 等^[51]将 Strassen 矩阵乘法的思想应用于卷积运算,提出了一种降低卷积神经网络计算量的快速方法.文献[51]的实验结果显示,计算量能最多减少 47%.另外,文献[52]也是用矩阵快速计算方法加速卷积神经网络的训练.Zhang 等^[53]提出了一种基于响应重构的深度卷积神经网络加速训练方法,该方法通过求解一个具有低秩约束的非线性优化问题来加速深度卷积网络的训练,设计了求解该优化问题的广义奇异值分解方法.Korytkowski 等^[54]提出了一种基于多特征输入中间缓存技术的卷积神经网络快速训练方法.在该方法中,将多特征看作输入到各个网络层的 n 个通道图像,这样可加速卷积操作运算,提高网络的训练速度.受支持向量机加速计算的压缩技术和 LASSO 中筛选技术的启发,Zheng 等^[55]提出了一种用压缩和召回技术加速训练网络的方法.该方法不仅适用于加速训练卷积神经网络,也适用于其他的深度学习模型,如深度受限波尔兹曼机、深度信念网络、深度自动编码器.Kim 等^[56]将 CNN 和 ELM 结合起来,提出了一种卷积神经网络快速训练方法.在该方法中,在卷积层和池化层之间加入了一个附加层.在 CNN 的全连接层之后的附加层参数用 ELM 算法优化,其他的附加层参数用基于 ELM 的改进的反向传播学习规则进行优化.KIM 等提出的方法有 2 个优点:1)在保持高测试精度的前提下,能极大地降低训练时间;2)用中小型数据集进行训练,也能得到理想的结果.Grinsven 等^[57]将样例选择的思路引入到卷积神经网络中,提出了一种用选择的样例加速卷积神经网络训练的方法.

目前,许多流行的深度学习开源框架,都提供了训练卷积神经网络的开源代码,有兴趣的读者可参考文献[26]和[27].

3 卷积神经网络的应用

因为卷积神经网络能直接处理图像数据,所以它在计算机视觉领域的应用最广泛.计算机视觉领域中的各种任务都有卷积神经网络大量成功的应用,包括图像分类、目标跟踪、目标检测等.另外,卷积神经网络在语音识别、文本分类中也有成功的应用.

3.1 图像分类

卷积神经网络在图像分类中的应用最有影响的工作是 Krizhevsky 等提出的 AlexNet^[11],该模型在 2012 年的 ImageNet^[58]大规模图像分类竞赛(ILSVRC)中以超过第二名 10 个百分点的成绩获得冠军.在这一工作的激励下,随后几年提出的几种卷积神经网络模型在图像分类中都取得了非常好的成绩.例如,在 2014 年进行的 ILSVRC 竞赛中,VGGNet 和 GoogLeNet 的 Top-5 精度分别达到了 92.7% 和 93.3%.在 2015 年,ResNet 在 1 000 类的图像分类中,Top-5 精度达到了 96.43%,超过人类的分类能力.近几年,研究人员还提出了许多图像分类新方法.例如,Han 等^[59]针对卷积神经网络的密集型计算和密集型内存需求问题,提出

了一种深度压缩方法.该方法包括 3 个阶段:剪枝、训练量化和哈夫曼编码.在大型图像数据库 ImageNet 上用 AlexNet 模型进行实验,在没有精度损失的前提下,内存需求减低了约 35 倍.用 VGG-16 进行实验,在没有精度损失的前提下,内存需求减低了约 49 倍.该论文获得了著名国际会议 ICLR2016 的最佳论文奖,它在将基于深度卷积神经网络的应用移植到嵌入式或移动设备上,具有重要的参考价值.

3.2 目标跟踪

针对可视对象的在线跟踪问题,Li 等^[60]利用单卷积神经网络,提出了一种鲁棒而且有效的跟踪算法.该算法用截断结构损失函数作为目标函数,通过鲁棒的样例选择机制,增强随机梯度下降算法训练卷积神经网络的效率.在文献[61]中,Fan 等将目标跟踪问题建模为一个机器学习问题.在给定行人前一个位置和步幅的前提下,通过学习来估计行人当前的位置和步幅.在该方法中,卷积神经网络用于从视频的相邻 2 帧中一起学习行人的空间和时间特征.在文献[62]中,Ma 等研究了用卷积神经网络进行可视目标跟踪的相关滤波器的设计与选择问题.针对不同的网络,提出了相应的设计与选择方法.

3.3 目标检测

目标检测是一个更加具有挑战性的计算机视觉任务,目标检测是指在目标图像中把对象用矩形框框起来.近几年,研究人员先后提出了 R-CNN^[63]、Fast R-CNN^[64]、Faster R-CNN^[49]等代表性的方法.在计算机视觉著名数据集 PASCAL VOC 上的检测平均精度分别达到了 53.3%、68.4%和 75.9%.检测速度也越来越快,从 R-CNN 模型处理 1 张图片需要 2 s 多,到 Faster R-CNN 提高到了 198 ms/张.针对目标检测问题,在这些工作的基础上,Zhang 等^[65]系统研究了具有很深的卷积神经网络的加速问题.Tome 等^[66]针对行人检测特定任务,提出了一种性能优于传统模型的卷积神经网络结构,能获得很好的检测效果,而且计算复杂度低.最近研究人员提出了 2 种目标检测的新方法:YOLO^[67]和 SSD^[68],这 2 种方法能够通过单一通道检测直接预测类别.YOLO 将目标检测问题转化为一个回归问题,进行求解.而 SSD 将包围盒的输出离散化为一组默认的包围盒.2 种方法都取得了良好的检测效果.

3.4 其他应用

在计算机视觉领域之外,卷积神经网络应用较多的是语音识别.卷积神经网络能够减少语音信号中的谱变化和谱相关,Sainath 等^[69]利用这一特性,用卷积神经网络解决大规模语音识别问题,取得了良好的效果.在文献[70]中,Qian 等研究了基于卷积神经网络的带噪声语音识别问题,通过选择合适的滤波器、池化,补零机制以及输入特征图的大小,来确定适于带噪声语音识别的卷积网络结构,以增加模型的鲁棒性.除了上面这些经典的应用之外,研究人员还研究了卷积神经网络在其他方面的应用,但这些应用相对较少.例如,Dong 等^[71]将卷积神经网络应用于从低分率图像生成高分辨率图像的方法.

4 结束语

对卷积神经网络及其研究进展进行了全面的讨论.本文可使读者:1)快速掌握卷积神经网络的核心概念(如权值共享、局部感受域、池化等),也能快速掌握卷积神经网络的结构、运算和训练.2)快速了解卷积神经网络近几年的发展现状、研究热点和应用情况.作者认为,未来几年卷积神经网络研究面临的挑战主要在以下 3 个方面:①卷积神经网络在多模态大数据(特别是多模态视频大数据)分类中的应用研究;②卷积神经网络在非平衡视频跟踪大数据中的应用研究;③卷积神经网络在云平台上并行化实现研究.热点研究方向包括:自动确定卷积神经网络参数的研究;与其他神经网络如:LSTM、RNN 等结合的研究;与强化学习结合的研究;卷积神经网络非梯度下降训练研究.

参 考 文 献:

- [1] MCCULLOCH W S,PITTS W.A logical calculus of the ideas immanent in nervous activity[J].Bulletin of Mathematical Biology,1943,52(4):99-115.DOI:10.1007/BF02478259.DOI:10.1007/BF02478259.
- [2] ROSENBLATT F.The perception: a probabilistic model for information storage and organization in the brain[J].Psycho-

- logical Review, 1958, 65(6): 386-408. DOI: 10.1037/h0042519.
- [3] MINSKY M, PAPERT S. Perceptrons[M]. Oxford: MIT Press, 1969.
- [4] WERBOS P. Beyond regression: New tools for prediction and analysis in the behavioral sciences[D]. Boston: PhD Thesis, Harvard University, 1974.
- [5] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536. DOI: 10.1038/323533a0.
- [6] CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3): 273-297. DOI: 10.1007/BF00994018.
- [7] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507. DOI: 10.1126/science.1127647.
- [8] LECUN Y, BENGIO Y, HINTON G E. Deep learning[J]. Nature, 2015, 521: 436-444. DOI: 10.1038/nature14539.
- [9] 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. DOI: 10.7544/issn1000-1239.2013.20131180.
- YU K, JIA L, CHEN Y Q, et al. Deep learning: Yesterday, Today and Tomorrow [J]. Journal of Computer Research and Development, 2013, 50(9): 1799-1804. DOI: 10.7544/issn1000-1239.2013.20131180.)
- [10] 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. DOI: 10.1109/5.726791.
- [11] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [Z]. International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, USA, 2012. DOI: 10.1145/3065386.
- [12] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models [Z]. The 30th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013.
- [13] HE K M, ZHANG X Y, REN S Q, et al. Delving deep into rectifiers: surpassing human-level performance on imagenet classification [Z]. IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015. DOI: 10.1109/ICCV.2015.123.
- [14] LI J C, NG W W Y, YEUNG D S, et al. Bi-firing deep neural networks[J]. International Journal of Machine Learning & Cybernetics, 2014, 5(1): 73-83. DOI: 10.1007/s13042-013-0198-9.
- [15] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. Massachusetts: MIT Press, 2016.
- [16] YU D, WANG H, CHEN P, et al. Mixed pooling for convolutional neural networks [Z]. The 9th international conference on rough sets and knowledge technology, Shanghai, China, 2014.
- [17] GULCEHRE C, CHO K, PASCANU R, et al. Learned-norm pooling for deep feedforward and recurrent neural networks [Z]. European Conference on Machine Learning and Knowledge Discovery in Databases, Nancy, France, 2014. DOI: 10.1007/978-3-662-44848-9_34.
- [18] ESTRACH J B, SZLAM A, LECUN Y. Signal recovery from Pooling Representations [Z]. The 31st International Conference on Machine Learning, Beijing, China, 2014.
- [19] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9): 1904-16. DOI: 10.1109/TPAMI.2015.2389824.
- [20] XIE L, TIAN Q, WANG M, et al. Spatial pooling of heterogeneous features for image classification[J]. IEEE Transactions on Image Processing, 2014, 23(5): 1994-2008. DOI: 10.1109/TIP.2014.2310117.
- [21] LEE H, KIM G, KIM H G, et al. Deep CNNs along the time axis with intermap pooling for robustness to spectral variations[J]. IEEE Signal Processing Letters, 2016, 23(10): 1310-1314. DOI: 10.1109/LSP.2016.2589962.
- [22] PERLAZA S M, FAWAZ N, LASAULCE S, et al. From spectrum pooling to space pooling: opportunistic interference alignment in MIMO cognitive networks[J]. IEEE Transactions on Signal Processing, 2010, 58(7): 3728-3741. DOI: 10.1109/TSP.2010.2046084.
- [23] WU HBB, GU X D. Towards dropout training for convolutional neural networks[J]. Neural Networks, 2015, 71: 1-10. DOI: 10.1016/j.neunet.2015.07.007.
- [24] WANG J Z, WANG W M, WANG R G, et al. CSPS: An adaptive pooling method for image classification[J]. IEEE Transactions on Multimedia, 2016, 18(6): 1000-1010. DOI: 10.1109/TMM.2016.2544099.
- [25] SUN M L, SONG Z J, JIANG X H, et al. Learning pooling for convolutional neural network[J]. Neurocomputing, 2017, 224: 96-104. DOI: 10.1016/j.neucom.2016.10.049.
- [26] 黄文坚, 唐源. TensorFlow 实战[M]. 北京: 电子工业出版社, 2017.

- [27] 乐毅,王斌.深度学习-Caffe 之经典模型详解与实战[M].北京:电子工业出版社,2017.
- [28] LIN M,CHEN Q,YAN J,et al.Network in network [J/OL].[2014-03-04].<https://arxiv.org/abs/1312.4400v3>.
- [29] SIMONYAN K,ZISSERMAN A.Very deep convolutional networks for large-scale Image recognition [J/OL].[2014-09-15].<http://arxiv.org/abs/1409.1556v2>.
- [30] SZEGEDY C,LIU W,JIA Y,et al.Going deeper with convolutions[Z].IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015),Boston,MA,USA,2015.DOI:10.1109/CVPR.2015.7298594.
- [31] IOFFE S,SZEGEDY C.Batch normalization: accelerating deep network training by reducing internal covariate shift[J/OL].[2015-03-02].<https://arxiv.org/abs/1502.03167v3>.
- [32] SZEGEDY C,VANHOUCKE V,IOFFE S,et al.Rethinking the inception architecture for computer vision[Z].IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016),Las Vegas,NV,United States,2016.DOI:10.1109/CVPR.2016.308.
- [33] SZEGEDY C,IOFFE S,VANHOUCKE V,et al.Inception-v4,inception-ResNet and the impact of residual connections on learning [J/OL].[2016-08-23].<https://arxiv.org/abs/1602.07261>.
- [34] HE K M,ZHANG X Y,REN S Q,et al.Deep residual learning for image recognition[Z].IEEE Conference on Computer Vision and Pattern Recognition(CVPR2016),Las Vegas,NV,United States,2016.DOI:10.1109/CVPR.2016.90.
- [35] ZEILER M D,FERGUS R.Visualizing and Understanding Convolutional Networks[C]//Computer Vision-ECCV 2014, Lecture Notes in Computer Science,2014,8689:818-833.DOI:10.1007/978-3-319-10590-1_53.
- [36] IANDOLA F,MOSKEWICZ M,KARAYEV S,et al.DenseNet: implementing efficient ConvNet descriptor pyramids [J/OL].[2014-04-07].<https://arxiv.org/abs/1404.1869>.
- [37] IANDOLA F N,HAN S,MOSKEWICZ M W,et al.SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[J/OL].[2016-09-04].<https://arxiv.org/abs/1602.07360v4>.
- [38] CHOPRA S,HADSELL R,LECUN Y.Learning a similarity metric discriminatively,with application to face verification [Z].IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005),San Diego,California,2005.DOI:10.1109/CVPR.2005.202.
- [39] HADSELL R,CHOPRA S,LECUN Y.Dimensionality reduction by learning an invariant mapping[Z].IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006),New York,2006.DOI:10.1109/CVPR.2016.435.
- [40] KINGMA D P,WELLING M.Auto-encoding variational Bayes [J/OL].[2014-05-01].<https://arxiv.org/abs/1312.6114>.
- [41] HAYKIN S.神经网络与机器学习(影印版)[M].北京:机械工业出版社,2009.
- [42] BOTTOU L.Large-scale machine learning with stochastic gradient descent [Z].19th International Conference on Computational Statistics,Paris France,2010.DOI:10.1007/978-3-7908-2604-3_16.DOI:10.1016/j.neucom.2016.11.046.
- [43] LI S J,DOU Y,NIU X,et al.A fast and memory saved GPU acceleration algorithm of convolutional neural networks for target detection[J].Neurocomputing,2017,230:48-59.
- [44] MATHIEU M,HENAFF M,LECUN Y.Fast training of convolutional networks through PPTs [J/OL].[2014-03-06].<https://arxiv.org/abs/arXiv:1312.5851v5>.
- [45] RAJESWAR M S,SANKAR A R,BALASUBRAMANIAM V N,et al.Scaling up the training of deep CNNs for human action recognition [Z].IEEE International Parallel and Distributed Processing Symposium Workshop (IPDPSW2015),Hyderabad,INDIA,2015.DOI:10.1109/IPDPSW.2015.93.
- [46] ZLATESKI A,LEE K,SEUNG H S.ZNN-A fast and scalable algorithm for training 3D convolutional networks on multi-core and many-core shared memory machines [Z].IEEE International Parallel and Distributed Processing Symposium (IPDPS2016),Chicago,IL,USA,2016.DOI:10.1109/IPDPS.2016.119.
- [47] MORCEL R,EZZEDDINE M,AKKARY H.FPGA-based accelerator for deep convolutional neural networks for the SPARK environment [Z].IEEE International Conference on Smart Cloud (SMART-CLOUD2016),New York,USA,2016.DOI:10.1109/SmartCloud.2016.31.
- [48] GIRSHICK R.Fast R-CNN [Z].2015 IEEE International Conference on Computer Vision(ICCV),Santiago,Chile,2015.
- [49] REN S Q,HE K M,GIRSHICK R,et al.Faster R-CNN: towards real-time object detection with region proposal networks [J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,online first,DOI:10.1109/TPAMI.2016.2577031.
- [50] GUSMÃO P P B.Fast training of convolutional neural networks via kernel recalling [J/OL].[2016-10-12].<https://arxiv.org/abs/1610.03623>.

- [51] CONG J S,XIAO B J.Minimizing computation in convolutional neural networks [Z].The 24th International Conference on Artificial Neural Networks,Hamburg,Germany,2014.DOI:10.1007/978-3-319-11179-7_36.
- [52] LAVIN A,GRAY S.Fast algorithms for convolutional neural networks [Z].2016 IEEE Conference on Computer Vision and Pattern Recognition,Las Vegas,Nevada,USA,2016.DOI: 10.1109/CVPR.2016.435.
- [53] ZHANG X Y,ZOU J H,HE K M,et al.Accelerating very deep convolutional networks for classification and detection[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2016,38(10):1943-1955.DOI:10.1109/TPAMI.2015.2502579.
- [54] KORYTKOWSKI M,STASZEWSKI P,WOLDAN P.Fast computing framework for convolutional neural networks [Z].2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud),Social Computing and Networking (SocialCom),Sustainable Computing and Communications (SustainCom),Atlanta,GA,USA,2016.DOI: 10.1109/BD-Cloud-SocialCom-SustainCom.2016.28.
- [55] ZHENG S,VISHNU A,DING C.Accelerating Deep Learning with Shrinkage and Recall [J/OL].[2016-09-19].http://arxiv.org/abs/1605.01369v2.
- [56] KIM J,KIM J,JANG G J,et al.Fast learning method for convolutional neural networks using extreme learning machine and its application to lane detection[J].Neural Networks,2017,87:109-121.DOI: 10.1016/j.neunet.2016.12.002.
- [57] GRINSVEN M J J P V,GINNEKEN B V,HOYNG C B,et al.Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images[J].IEEE Transactions on Medical Imaging,2016,35(5):1273-1284.DOI: 10.1109/TMI.2016.2526689.
- [58] DENG J,DONG W,SOCHER R,et al.ImageNet: A large-scale hierarchical image database [Z].IEEE Conference on Computer Vision and Pattern Recognition,2009(CVPR 2009),Miami,FL,USA,2009.DOI: 10.1109/CVPR.2009.5206848.
- [59] HAN S,MAO H,DALLY W J.Deep compression: compressing deep neural networks with pruning,trained quantization and huffman coding[Z].International Conference on Learning Representations 2016 (ICLR 2016),San Juan,Puerto Ri 2016.DOI: 10.1109/TIP.2015.2510583.
- [60] LI H,LI Y,PORIKLI F.DeepTrack: learning discriminative feature representations online for robust visual tracking[J].IEEE Transactions on Image Processing,2016,25(4):1834-1848.DOI: 10.1109/TIP.2015.2510583.
- [61] FAN J L,XU W,WU Y,et al.Human tracking using convolutional neural networks[J].IEEE Transactions on Neural Networks,2010,21(10):1610-1623.DOI: 10.1109/TNN.2010.2066286.
- [62] MA C,XU Y,NI B B,et al.When correlation filters meet convolutional neural networks for visual tracking[J].IEEE Signal Processing Letters,2016,23(10):1454-1458.DOI: 10.1109/LSP.2016.2601691.
- [63] GIRSHICK R,DONAHUE J,DARRELL T,et al.Region-based convolutional networks for accurate object detection and segmentation[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2016,38(1):142-158.DOI: 10.1109/TPAMI.2015.2439281.
- [64] GIRSHICK R.Fast R-CNN [Z].2015 IEEE International Conference on Computer Vision (ICCV2015),Santiago,Chile,2015.DOI: 10.1109/ICCV.2015.169.
- [65] ZHANG X Y,ZOU J H,HE K M,et al.Accelerating very deep convolutional networks for classification and detection[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,38(10):1943-1955.DOI: 10.1109/TPAMI.2015.2439281.
- [66] TOME D,MONTI F,BAROFFIO L,et al.Deep convolutional neural networks for pedestrian detection[J].Signal Processing: Image Communication,2016,47:482-489.DOI:10.1016/j.image.2016.05.007.
- [67] REDMON J,DIVVALA S,GIRSHICK R,et al.You only look once: unified,real-time object detection[Z].2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),Las Vegas,NV,United States,2016.DOI:10.1109/CVPR.2016.91.
- [68] LIU W,ANGUELOV D,ERHAN D,et al.SSD: single shot multiBox detector[Z].European Conference on Computer Vision (ECCV 2016),Amsterdam,The Netherlands,2016.DOI: 10.1007/978-3-319-46448-0_2.
- [69] SAINATH T N,KINGSBURY B,SAON G,et al.Deep convolutional neural networks for large-scale speech tasks[J].Neural Networks,2015,64:39-48.DOI:10.1016/j.neunet.2014.08.005.
- [70] QIAN Y M,BI M X,TAN T,et al.Very deep convolutional neural networks for noise robust speech recognition[J].IEEE/ACM Transactions on Audio,Speech,and Language Processing,2016,24(2):2263-2276.DOI: 10.1109/TASLP.2016.2602884.
- [71] DONG C,LOY C C,HE K M,et al.Image super-resolution using deep convolutional networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,38(2):295-307.DOI: 10.1109/TPAMI.2015.2439281.