

文章编号:1671-1637(2021)04-0048-24

# 卷积神经网络及其在智能交通系统中的应用综述

马永杰<sup>1</sup>,程时升<sup>1</sup>,马芸婷<sup>1</sup>,马义德<sup>2</sup>

(1. 西北师范大学 物理与电子工程学院,甘肃 兰州 730070; 2. 兰州大学 信息科学与工程学院,甘肃 兰州 730030)

**摘 要:**从特征传输方式、空间维度、特征维度 3 个角度,论述了近年来卷积神经网络结构的改进方向,介绍了卷积层、池化层、激活函数、优化算法的工作原理,从基于值、等级、概率和转换域四大类总结了近年来池化方法的发展,给出了部分具有代表性的激活函数对比、梯度下降算法及其改进型和自适应优化算法的工作原理和特点;梳理了卷积神经网络在车牌识别、车型识别、交通标志识别、短时交通流预测等智能交通领域中的应用和国内外研究现状,并将卷积神经网络算法与支持向量机、差分整合移动平均回归模型、卡尔曼滤波、误差反向传播神经网络、长短时记忆网络算法从优势、劣势和在智能交通领域的主要应用场景三方面进行了对比;分析了卷积神经网络在智能交通领域面临的鲁棒性不佳和实时性较差等问题,并从算法优化、并行计算层面和有监督学习到无监督学习方向研判了卷积神经网络的发展趋势。研究表明:卷积神经网络在视觉领域具有较强优势,在智能交通系统中主要应用于交通标志、车牌、车型识别、交通事件检测、交通状态预测;相比其他算法,卷积神经网络所提取的特征更加全面,有效地提高了识别准确度与速度,具有较大的应用价值;卷积神经网络未来将通过网络结构的优化、算法的改进、算力的提升以及基准数据集的增强,为智能交通带来新的突破。

**关键词:**交通信息;深度学习;卷积神经网络;智能交通;网络结构;图像识别;研究进展

**中图分类号:**U495 **文献标志码:**A **DOI:**10.19818/j.cnki.1671-1637.2021.04.003

## Review of convolutional neural network and its application in intelligent transportation system

MA Yong-jie<sup>1</sup>, CHENG Shi-sheng<sup>1</sup>, MA Yun-ting<sup>1</sup>, MA Yi-de<sup>2</sup>

(1. College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, Gansu, China;

2. School of Information Science and Engineering, Lanzhou University, Lanzhou 730030, Gansu, China)

**Abstract:** From the perspectives of feature transmission mode, spatial dimension and feature dimension, the improvement directions of convolution neural network structure in recent years were reviewed. The working principles of the convolution layer, pooling layer, activation function and optimization algorithm were introduced, and the recent developments of pooling methods in terms of value, level, probability, and transformation domain were summarized. The comparison of some representative activation functions, and the working principle and characteristics of the gradient descent algorithm and its improved and adaptive optimization algorithm were given. The application and research status of convolutional neural network in

收稿日期:2021-02-20

基金项目:国家自然科学基金项目(62066041)

作者简介:马永杰(1967-),男,甘肃灵台人,西北师范大学教授,工学博士,从事图像处理、人工智能、测控技术等研究。

引用格式:马永杰,程时升,马芸婷,等.卷积神经网络及其在智能交通系统中的应用综述[J].交通运输工程学报,2021,21(4):48-71.

Citation:MA Yong-jie, CHENG Shi-sheng, MA Yun-ting, et al. Review of convolutional neural network and its application in intelligent transportation system[J]. Journal of Traffic and Transportation Engineering, 2021, 21(4): 48-71.

intelligent transportation fields such as license plate recognition, vehicle type recognition, traffic sign recognition, and short-term traffic flow prediction were reviewed. The convolutional neural network algorithm was compared with the support vector machine, differential integrated moving average regression model, Kalman filter, error back propagation neural network, and long-term and short-term memory network algorithms from the advantages and disadvantages and main application scenarios in the field of intelligent transportation. The issues of poor robustness and poor real-time performance of convolutional neural network in the field of intelligent transportation were analyzed. The development trend of convolutional neural network was evaluated in terms of algorithm optimization, parallel computing, and supervised learning to unsupervised learning. Research results show that the convolutional neural network has strong advantages in the field of vision. It is mainly used for traffic sign, license plate, vehicle type recognition, traffic event detection, and traffic state prediction in intelligent transportation system. Compared with other algorithms, the convolutional neural network can extract more comprehensive features. It can effectively improve the recognition accuracy and speed and has great application value. The convolutional neural network will bring new breakthroughs to intelligent transportation in the future through the optimization of network structure, the improvement of algorithm and computing power, and the enhancement of benchmark data sets. 5 tabs, 3 figs, 146 refs.

**Key words:** traffic information; deep learning; convolutional neural network; intelligent transportation; network structure; image recognition; research progress

**Author resume:** MA Yong-jie(1967-), male, professor, PhD, myjmyj@nwnu.edu.cn.

**Foundation item:** National Natural Science Foundation of China (62066041)

## 0 引言

智能交通系统(Intelligent Transportation System, ITS)在国内起步于20世纪末,随着科学技术水平的不断提高、制造业的快速发展,机器视觉、通信、传感技术等都已经成为ITS的重要技术手段。ITS在实际交通中的广泛应用促进了道路的技术进步,使得车辆以及行人的出行质量都得到了很大的改善,提高了交通工作效率、减少了交通事故的发生、缓解了道路阻塞问题。ITS将人、路、车三者有机地联系在一起,使得人们出行质量得到提高,确保了自身的安全,促进了运输行业的发展。智能交通系统是卷积神经网络(Convolutional Neural Network,

CNN)应用领域中具有发展潜力的研究领域,其应用包括车牌、车型、交通标志的识别研究等。

人工神经网络(Artificial Neural Network, ANN)是一种模仿生物神经网络行为特征的数学模型。通过内部大量神经元之间相互连接、相互作用模拟大脑的非线性、非局限性,构成一个自适应非线性动态网络系统<sup>[1]</sup>,从而完成对信息的处理,并具备自学习的能力。人工神经网络发展路线见图1。

1943年,Mcculloch等<sup>[2]</sup>建立了神经网络的数学模型,形式化地用数学语言描述了神经元和网络结构,证明了单个神经元能执行逻辑功能,开创了人工神经网络的先河。1958年,人工神经网络得到进一步发展,Rosenblatt<sup>[3]</sup>在MP模型的基础上提出

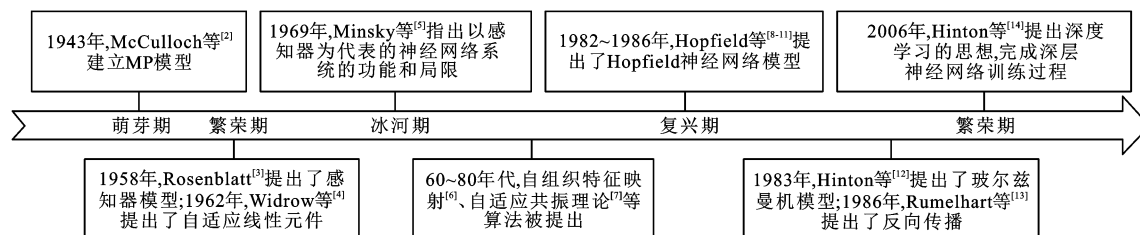


图1 人工神经网络发展路线

Fig. 1 Development route of artificial neural network

了单层感知器模型,并证明了 2 层感知器经过训练后可以对一定的输入进行分类识别,建立了第 1 个真正意义上的神经网络,随后他还提出了带有隐层的 3 层感知器。感知器模型包含了一些现代神经网络的基本原理,其结构符合神经生物学,对神经网络的发展具有重大的推动作用。1962 年,Widrow 等<sup>[4]</sup>提出了自适应线性元件结构和最小均方算法(Least Mean Square, LMS)准则(现在也被称为 Widrow-Hoff 准则),并将其应用于实际工程,促进了神经网络的应用与发展。1969 年,Minsky 等<sup>[5]</sup>从数学的角度仔细分析了以感知器为代表的神经网络系统的功能和局限,指出单层感知器本质上是线性模型,无法处理“异或”问题,应用领域极为有限,加上当时计算机的计算能力无法满足大型神经网络长时间的运行需求,这几乎宣判了感知机模型甚至人工神经网络的死刑,导致后来神经网络研究经历多年的寒冬。

尽管神经网络的发展处于低潮时期,但一些学者仍致力于该领域的研究,提出了自组织特征映射(Self-Organizing Feature Map, SOM)<sup>[6]</sup>、自适应共振理论(Adaptive Resonance Theory, ART)<sup>[7]</sup>等算法,为神经网络的研究和发展奠定了基础。在 1982~1986 年,Hopfield 等<sup>[8-11]</sup>通过总结和吸取前人的成果和经验,引入“计算能量函数”的概念,提出了具有联想记忆的反馈互连网络,称为 Hopfield 模型,该模型在数学层面对人工神经网络信息存储和提取进行了概括,提供了重要公式和参数,促使人工神经网络主要从反馈和多层前馈两方面展开研究,使其构造和学习有了理论指导,激发起大量学者的研究热情,极大的推动了神经网络的发展。1983 年,Hinton 等<sup>[12]</sup>提出了玻尔兹曼机模型,该模型是最早能够学习神经元内部表示的神经网络之一。1986 年,Rumelhart 等<sup>[13]</sup>提出了反向传播(Back Propagation, BP)算法,解决了 2 层神经网络所需要的复杂计算量问题,同时克服了神经网络无法解决异或的问题。但受限于当时数据获取的瓶颈,神经网络只能在中小规模数据上训练,因此,过拟合极大困扰着神经网络型算法,同时各种浅层机器学习模型也在快速发展,致使神经网络的研究再次跌入“谷底”。随着研究者的不懈努力和硬件性能的不断优化,2006 年,Hinton 等<sup>[14]</sup>提出了深度学习的思想,通过逐层预训练的方式有效完成了深层神经网络的训练过程,同时证明了神经网络模型的预测能力相比其他传统机器学习算法的优越性,各种深度学习模型也相继被提出<sup>[15-18]</sup>,其中,卷积神经网络在计算机视觉领域取得了突破性

的进展。卷积神经网络和全连接神经网络结构较为相似,其输入输出以及训练过程也基本一致,不同的是卷积神经网络相邻两层之间只是部分节点相连,而全连接(Fully Connected, FC)网络相邻 2 层之间的节点都有连接,使网络参数急剧增多,不仅减缓计算速度,还极易产生过拟合问题,因此,卷积神经网络更适用于处理图像数据。其实在 1998 年,Lecun 等<sup>[19]</sup>提出了基于梯度学习的 CNN 网络模型,并成功应用于手写数字识别,但受限于当时的数据处理能力,因此,未能得到发展。直到 2017 年,Krizhevsky 等<sup>[17]</sup>提出的 AlexNet 模型在 ImageNet<sup>[20]</sup>大规模视觉识别挑战(ImageNet Large Scale Visual Recognition Challenge, ILSVRC)竞赛中获得了第 1 名,从此,AlexNet 成为 CNN 领域内具有重要历史意义的一个网络模型。相比传统机器学习方法,CNN 实现了类似人脑的视野域,依靠网络本身学习和生成特征,在图像处理中逐渐体现出无可替代的优越性,同时,其在交通领域的应用也推动了智能交通的进一步发展。

## 1 卷积神经网络

20 世纪 60 年代,加拿大神经科学家 Hubel 等<sup>[21]</sup>通过对猫脑视觉皮层的研究,发现视觉信息从视觉皮层传递到大脑是通过多个层次的感受野逐层激发实现。1980 年,日本科学家 Fukushima 等<sup>[22]</sup>根据 Hubel-Wiesel 的层级模型,模拟生物视觉系统提出了与之结构相似的神经认知机(Neocognitron),神经认知机被认为是卷积神经网络的前身,它采用 S 型细胞和 C 型细胞交替堆叠组成,其中,S 型细胞能够最大程度地响应感受野内的特定边缘刺激,提取其输入层的局部特征,C 型细胞对来自确切位置的刺激具有局部不敏感性,这与卷积神经网络的卷积层和池化层一一对应。随后 1998 年,Lecun 等<sup>[19]</sup>在 Fukushima 研究工作基础上,采用 BP 算法和随机梯度下降法设计并训练了第一个真正意义上的 CNN 模型(该模型称为 LeNet-5),LeNet-5 吸取了 Neocognitron 稀疏交互的优点,并在生成特征图时,每个感受局部图像区域的神经元用同一个卷积核去卷积图像,这一特性便是权值共享,很大程度上减少了参数量。LeNet-5 通过局域感受野、权值共享和池化实现识别位移、缩放和扭曲不变性<sup>[19]</sup>,为 CNN 后来的发展奠定了坚实的基础。

随着大规模训练数据的出现(如 ImageNet 等<sup>[20]</sup>)以及计算机硬件的飞速发展(特别是 GPU 的出现),使得深度卷积神经网络得到进一步发展。从



2010年开始,每年都有ILSVRC竞赛来供各种算法挑战分类极限,其所用数据集为ImageNet的子集,共有1 000个类别、上百万张图像。继AlexNet后,ZFNet<sup>[23]</sup>、VGGNet<sup>[24]</sup>、GoogleNet<sup>[25]</sup>、ResNet<sup>[26]</sup>、DensNet<sup>[27]</sup>等算法相继被提出,使得深度卷积神经网络模型逐渐强大。从结构方面来看,CNN发展方向是加深网络和增强卷积模块功能。表1从空间维度、特征维度、特征传输方式三方面总结了网络结构的差异。

表1 基于空间维度、特征维度、特征传输方式的网络结构

Tab. 1 Network structure based on spatial dimensions, characteristic dimensions, and characteristic transmission methods

分类标准	方法	代表性网络
空间维度	(1)用多个小卷积核代替大卷积核	VGGNet <sup>[24]</sup>
	(2)多尺度非线性,单一尺寸卷积核用多尺寸卷积核代替,使用 $1 \times 1$ 卷积核	Inception系列 <sup>[25,28-30]</sup>
	(3)固定形状卷积核趋于使用可变形卷积核	可变形卷积网络 <sup>[31]</sup>
	(4)空洞卷积	FCN <sup>[32]</sup>
	(5)反卷积	ZFNet <sup>[23]</sup>
特征维度	(1)深度可分离卷积,每个通道使用不同卷积操作	Xception <sup>[33]</sup>
	(2)使用分组卷积	AlexNet <sup>[17]</sup>
	(3)分组卷积前对通道随机分组	ShuffleNet <sup>[34]</sup>
	(4)通道加权计算	SE-Net <sup>[35]</sup>
特征传输方式	(1)使用跳跃连接,让模型更深	ResNet <sup>[26]</sup>
	(2)密集连接,使每一层都融合上其他层的特征输出	DenseNet <sup>[27]</sup>

VGGNet通过加深网络层数来提高识别精度,采用 $3 \times 3$ 尺寸的卷积核代替原较大卷积核和更小的步长来堆叠网络,从而加深整个神经网络的层级,证明了增加网络层数有利于提高图像分类的准确度。GoogleNet同时增加了网络的深度和宽度,增强了特征提取功能。

可变形卷积<sup>[31]</sup>认为卷积核的形状可以是变化的,变形的卷积核能让它只看感兴趣的图像区域,这样识别出来的特征更佳。反卷积<sup>[23]</sup>是卷积的逆运算,可得到和原始输出图像尺寸相同的图像。高鑫等<sup>[36]</sup>提出了DF-RCNN模型来提高遥感图像中密集区域的检测精度。该模型在VGGNet基础上,首先利用反卷积将深浅层特征图统一到较高分辨率并进行融合,多层融合特征比单层特征具有更强的表达能力;其次引入可变形卷积和可变形RoI池化模块来学习密集区域目标的形变信息,增强网络对空

间几何变换的学习能力。试验表明,DF-RCNN模型在遥感图像中具有良好的检测性能,有效的提高了密集区域与非密集区域的检测精度。

空洞卷积在图像分割领域被提出<sup>[37]</sup>,池化操作可增大感受野、降低特征维度,上采样操作可增大图像尺寸。图像经过上采样操作恢复了之前的大小,但是在经过池化操作时易丢失很多细节信息,导致特征的不完善,空洞卷积是在原卷积核中插一些0的权值,即在原卷积核“挖洞”,达到扩大感受野的目的,同时避免信息的损失,空洞卷积在图像需要全局信息问题中得到广泛使用。Google团队所提出的DeepLab系列<sup>[38-41]</sup>均采用了空洞卷积,来获取更多的上下文信息,DeepLab v3+还引入Xception作为编译器的基础网络,使得网络更加快速和鲁棒,在PASCAL VOC 2012数据集上平均交并比达到89%。Kaiming等<sup>[42]</sup>提出的Mask R-CNN网络将经典的目标检测算法Faster R-CNN<sup>[43]</sup>和语义分割算法全卷积网络(Fully Convolutional Networks, FCN)有效结合,可以实现目标分类、目标检测、语义分割和实例分割等多个任务。Liu等<sup>[44]</sup>按照信息在神经网络中的传播方式,基于Mask R-CNN,提出了一种路径聚合网络(Path Aggregation Network, PANet),将所有特征层的特征进行汇聚,缩短底层和顶层特征层之间的距离,从而可靠地传递信息,在COCO 2017挑战实例分割任务中获得冠军。Huang等<sup>[45]</sup>认为Mask R-CNN使用分类置信度来衡量Mask的分数,不能很好地反应分割质量,因此提出了一种对实例分割进行评分的网络,称为Mask Scoring R-CNN,通过在COCO数据集上进行评估,Mask Scoring R-CNN表现较好,并且优于Mask R-CNN网络。

深度可分离卷积<sup>[33]</sup>可看做是Inception的延续,对输入的每一个通道进行各自的卷积操作,得到每一个通道的特征图,之后再对这些特征图进行 $1 \times 1$ 跨通道卷积操作,即对分离的通道进行融合。因为深度可分离卷积对每一个通道都进行了学习,且每个通道对应一个不同的过滤器,因此,可得到的更具表达力的图像特征,此外,相比标准的卷积操作,它还降低了参数量。

分组卷积最早出现在AlexNet中,受限于当时的硬件资源,在训练网络时Krizhevsky等<sup>[17]</sup>将特征图分给多个GPU分别进行处理,最后把多个GPU的结果进行融合。但只在最后一层才融合不同组之间的特征,不利于模型的泛化性。为了解决

这个问题,Zhang 等<sup>[34]</sup>提出 ShuffleNet,该网络在每一次层叠分组卷积前,都先进行一次通道随机选择,随机将不同通道分配到不同组当中。经过通道随机选择之后,分组卷积输出的特征能考虑到更多通道,且 ShuffleNet 中的分组卷积操作采用深度卷积,因此,输出的特征表达能力更强。但在连接通道产生特征时,ShuffleNet 忽略了不同特征通道对模型的作用是不相等的问题,Hu 等<sup>[35]</sup>提出的 SE-Net 通过学习的方式来自动获得每个特征通道的重要程度,然后依照此程度提升有用的特征并削弱用处较小的特征,即对特征通道进行加权计算,该网络在 ILS-VRC 2017 竞赛中将前 5 项错误率降至 2.251%。

Sarigül 等<sup>[46]</sup>提出了一种新的卷积技术,称为差分卷积和更新的误差反向传播算法。提出的卷积技术通过考虑相邻像素之间的方向变化来改进标准卷积,旨在将包含方向激活差异的特征映射传输到下一层。差分卷积通过应用预定义的常量滤波器生成表示信号变化的新特征图,这些新特征图可检测到模式的巨大差异,这些差异是分类问题的有价值信息,它充分考虑了特征映射中卷积特征如何变化。平铺卷积通过增加滤波器的数量来保持特征映射的数量不变,而差分卷积在不增加滤波器数量的情况下增加特征图的数量。空洞卷积试图增加滤波器的感受野,而差分卷积增加了滤波器的效率而不改变感受野大小。具有新特征映射的差分卷积需要改进反向传播算法,在反向传播期间,误差以相反方向传输到每个特征映射。每个额外特征图上的误差乘以相应的固定值滤波器权重,并添加到第一特征图的相应误差上,此误差矩阵用于训练相关滤波器,并向后传播。试验结果证明了差分卷积技术的有效性和适应性,它保留了像素激活中的方向变化量,提取图方向的有价值信息,增加特征图的数量而不增加滤波器的数量,并通过邻域激活提供误差馈送。此外,它易于适应所有 CNN 结构,能提升网络性能并加快误差最小化。

增加网络深度可提高网络性能,但不得不面对深层次网络所带来的网络退化问题,仅线性增加网络层数会造成梯度弥散,反而使得网络性能下降,如何构建一个更深层次的网络成为 CNN 的研究热点。He 等<sup>[26]</sup>通过分析网络训练问题,认为深度网络中并非每一层都得到优化训练,随着层数的增加,误差也会逐层增大,引起网络退化和梯度消失问题,为了解决这一问题,提出了深度残差网络 ResNet。ResNet 通过在浅层网络的基础上叠加恒等映射层,

将低层特征图直接映射到高层网络中,这种结构被称为跳跃连接。ResNet 将网络加深至 152 层,并且使用了和 VGGNet 同样大小的卷积核,使其有更高的测试准确度和更低训练误差。DensNet 通过参考 ResNet 以及 GoogleNet 将前面所有层的输出都传递给下一层,建立了不同层之间的连接关系,最大化信息的传播,重复利用特征,进一步减轻了梯度消失问题,使得深度网络得到较好的训练。

以上网络框架尽管都通过优化网络结构使得性能越来越好,计算量也得到很大程度减少,但都是趋向于往更深层次的方向发展,这意味着需要训练的参数还是很多,需要的存储空间与计算时间也会增大,这在实际应用中,仍会有一些阻碍。如何进一步给网络“瘦身”并且提高性能,将成为目前及未来的研究热点。

### 1.1 卷积层

卷积神经网络是一类特殊的人工神经网络,其最主要的特点是卷积运算操作,因此,CNN 在诸多领域特别是图像相关任务上表现优异。卷积层通过对图像进行卷积运算提取图像特征。在一个卷积层上将得到若干个特征图,每个特征图提取一类特征,同一特征图的神经元共享权值(卷积核)。2D 卷积的数学公式为

$$S(i,j) = (X * K)(i,j) = \sum_h \sum_n x(i+h,j+n)k(h,n) \quad (1)$$

式中: $S(i,j)$ 为第  $i$  行第  $j$  列元素的卷积结果; $X$ 为输入的二维图像; $*$ 为卷积运算; $K$ 为相应的二维卷积核; $x(i+h,j+n)$ 为第  $i+h$  行  $j+n$  列的图像元素; $k(h,n)$ 为第  $h$  行  $n$  列的卷积核元素。

卷积层的参数通过反向传播算法进行训练更新。设滑动步长为 1,卷积操作过程见图 2, $k_1 \sim k_9$  为卷积核元素,卷积核在原图上进行滑动来遍历整张图片,通过计算原图与卷积核上的数值来提取图像特征。

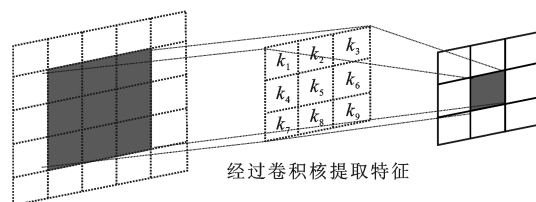


图 2 卷积操作

Fig. 2 Convolution operation

### 1.2 池化层

为了防止特征维度过高,卷积层后通常会跟一个池化层进行特征降维。池化操作在加快计算速率

的同时避免了过拟合问题,提高所提取特征的鲁棒性。传统的池化方式中最常见的有最大池化<sup>[47]</sup>和平均池化<sup>[48]</sup>2种,图3所示为最大和平均池化操作过程,池化步长为2,对邻域内特征点取最大或平均

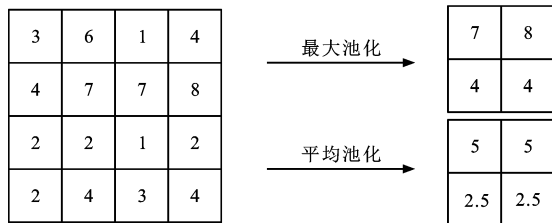


图3 最大池化与平均池化操作

Fig. 3 Operations of max-pooling and average-pooling

表2 基于值、等级、概率和转换域的池化方法

Tab. 2 Pooling methods based on values, rank, probabilities and transformed domains

分类标准		代表性方法
基于值的池化方法	基于显著性特征	最大池化 <sup>[47]</sup> 、平均池化 <sup>[48]</sup> 、混合池化 <sup>[50]</sup> 、全局平均池化 <sup>[51]</sup> 、跳跃池化 <sup>[52]</sup> 、空间金字塔池化 <sup>[53]</sup> 、内核池化 <sup>[54]</sup> 、动态相关池化 <sup>[55]</sup> 、多激活池化 <sup>[56]</sup> 、联合池化 <sup>[57]</sup> 、细节保留池化 <sup>[58]</sup> 、同心圆池化 <sup>[59]</sup>
	基于补丁	补丁子类池化 <sup>[60]</sup> 、系列多池化 <sup>[61]</sup> 、部分平均池化 <sup>[62]</sup>
	多采样方法	多采样棋盘池化 <sup>[63]</sup> 、并行网格池化 <sup>[64]</sup>
基于等级的池化方法		多级池化 <sup>[65]</sup> 、顺序池化 <sup>[66]</sup> 、全局加权排名池化 <sup>[67]</sup>
		基于排名的平均池化 <sup>[68]</sup> 、基于排名的加权合并池化 <sup>[68]</sup> 、基于等级的随机池化 <sup>[68]</sup>
基于概率的池化方法		随机空间采样池化 <sup>[69]</sup> 、混合池化 <sup>[70]</sup> 、随机池化 <sup>[71]</sup>
		丢弃最大池化 <sup>[72]</sup> 、稀疏随机池化 <sup>[73]</sup>
基于转换域的池化方法		基于时域的池化 <sup>[74]</sup> 、基于频域的池化 <sup>[75]</sup> 、基于小波域的池化 <sup>[76]</sup>

### 1.2.1 基于值的池化方法

基于值的池化对图像区域的采样,可进一步分为基于显著性特征、基于补丁和多采样方法。

在基于显著性特征池化方法中,池化操作将池化区域中的信息汇总,然后根据池化方法的标准将其转换为单个值,例如最大池化和平均池化。一些研究者选择结合最大和平均池化来提高CNN的正则化能力并解决平均池化和最大池化的瓶颈<sup>[47,77]</sup>,称为混合池化<sup>[50]</sup>。白琮等<sup>[78]</sup>结合2种池化方式优点,将其分别以权重为1相加作为新的最大-平均池化方法,对比单一的最大值池化,最大-平均池化方法保留了更重要的图像特征,对外界干扰具有更高的鲁棒性;Eom等<sup>[79]</sup>提出了一种新的池化方法,引入一个定义池化类型的可训练参数 $\vartheta$ ,由训练数据通过反向传播训练得到,收敛的 $\vartheta$ 可自动决定池的类型,模型可以从给定任务的训练数据中找到最优的池化方法;刘万军等<sup>[80]</sup>通过对比不同的池化模型对网络性能的影响,在最大池化算法的基础上,根据插值原理,设计出一种动态自适应的改进池化算法。这种模型可以根据不同的特征图以及池化域的内

值作为该邻域池化后的值。

平均值池化是选取池化域内所有值的平均值作为池化后特征图的特征值,可以减小因邻域大小受限造成的估计值方差增大,对微小变形具有鲁棒性;最大值池化则是选取池化域内所有值的最大值作为池化后特征图的特征值,可以减小因卷积层参数误差造成的估计均值偏移,对图像纹理信息保留更好。但这两种方式均不能很好的进行特征提取,因此,许多研究者对其进行改进。池化方法可分为基于值、等级、概率和转换域四大类<sup>[49]</sup>,如表2所示。基于值的池化方法又分为基于显著性特征、基于补丁和多采样的池化方法。

容,动态调整其池化过程和权值,分别在手写数字、车牌、人脸3种数据集下进行试验验证,与最大池化法、平均池化法以及中间值池化法(对平均池化法和最大池化法的一种折中算法)相比,动态自适应池化模型均为更优模型,且有效的提升了CNN的收敛性、准确率以及学习性能。

全局平均(Global Average, GA)池化<sup>[51]</sup>取代了FC层,其基本思想是对每个特征图的平均值求和,然后将此向量输入分类层。与CNN的其他层相比,FC层需要约90%的参数,因此,没有FC层将极大程度减少参数数量和计算时间,同时避免了输入特征图的过拟合问题。全局平均池化保持分类任务类别数和特征图数一致,使特征图和类别信息直接映射,对最后一层卷积层的每个特征图求均值并连接成向量,再经过SoftMax分类,求均值结合了空间信息,对输入的空间变换具有鲁棒性。Zhang等<sup>[81]</sup>对GA池化进行改进,使用对数平均指数函数来提取特征。

细节保留池化(Detail Preserving Pooling, DPP)<sup>[58]</sup>采用反向双边滤波器来保存特征图的重要



细节,有不对称和对称两种变体。对称结构增强了所有细节,而非对称结构增强的特征高于平均激活。DPP 的计算开销较小,其性能类似于最大或平均池化,或者在中间函数的非线性连续体上,DPP 可以与随机池化<sup>[55]</sup>方法相结合,细节保存和正则化相辅相成,进一步提高了精度。

空间金字塔池化(Spatial Pyramid Pooling, SPP)<sup>[53]</sup>将最后一层卷积层输出的特征图分别进行 1、4、16 等分,然后在每个分块进行不同窗口大小和移动步长的池化操作(如最大池化),可得到固定大小的特征。将 SPP 层放在 FC 层前,特征图通过 SPP 层可输出固定长度的数据,解决了网络对图片输入尺寸的限制,同时,不同级别等分的特征对应的原始图像区域不同,多个对应区域意味着物体的多个尺度,由于把一个特征图从不同的角度进行特征提取再融合,使特征包含图像的整体与局部信息,更具表达能力。SPP 的另一种变体是金字塔池化<sup>[82]</sup>,它将空间信息集成到特征向量中,在不丢失重要信息的情况下降低了维度。

在基于补丁的方法中,首先检测对象,并在这些补丁上应用池化操作,补丁池化方法有子类池化(SubClass Pooling, SCP)<sup>[60]</sup>、系列多池化<sup>[61]</sup>和部分平均池化<sup>[62]</sup>等。SCP 是一个 3 层结构,解决了部分训练数据的双重阻塞问题。该方法保留了空间信息,抑制了遮挡等噪声,提高了整体性能。首先,将局部特征汇集到一起,根据空间区域将空间相关性保存到子类中。测试阶段采用模糊最大池化,克服障碍物区域不稳定的局部特征。最终采用平均池化通过对每个子类进行常规加权来增强健壮性。该方法对随机模式下的各种遮挡具有较强的鲁棒性。系列多池化<sup>[61]</sup>方法是受 SPP<sup>[53]</sup>的启发,并与所选的特征图补丁相关联。该方法创建了多尺度的特征,并通过扩展补丁区域来提取丰富的特征。

池化层的下采样操作可学习到空间不变特征并减少计算成本,下采样丢失了大约 75% 的信息,而在二维 CNN 中分辨率按比例缩小。为了克服这些问题,Sadigh 等<sup>[63-64]</sup>提出了 2 种类型的采样方法以及池化方法,即方格二次采样和并行网格池化。这种多采样方法在提高输出特征图的空间分辨率的同时,保留了传统子采样层的优势,例如增加接收域并降低计算成本,这样可产生更高分辨率的特征图,在训练过程中对深层特征进行更好的梯度更新以及通过减少膨胀卷积的需求来简化 CNN 设计,通过简单地应用多次采样而不使用数据增强来提高图像分

类的准确性。

### 1.2.2 基于等级的池化方法

在基于等级的池化方法中,池化区域中特征映射的激活具有不同的权重,通过加权的方法将其组合在一起,这些权重是在基于梯度的优化或训练过程中学习得到的。顺序池化网络(Ordinal Pooling Network, OPN)<sup>[66]</sup>和池化操作之间的一个关键区别是,典型的池化一次作用于一个特征图,而 OPN 由每个特征图的不同权重集组成,因此,池化特性来自所有特征图。

多级池化<sup>[65]</sup>将特征投影到一个新的空间,然后通过累积的二分排序方法对它们进行评分。特征选择算子通过从训练集学习多部分的排序方案,在一个池化窗口中选择信息量最大、得分最高的循环特征。受随机池化的启发,在每个窗口中,根据其评分函数响应选择更高级别的激活。在不同基准上进行的试验证实,所提出的多级池化策略提高了深度卷积网络的性能,优于其他具有相同评估协议的池化策略(平均、最大和随机池化)。

### 1.2.3 基于概率的池化方法

基于概率的池化方法以平均池化和最大池化之间的概率来权衡两者之间的关系。通过引入混合机制,将最大池化和平均池化的特性反映在池化过程中。基于概率的池化方法有助于提高错误率,防止过拟合,此方法有随机法、最大丢包法、混合法、感兴趣区域法等。随机空间采样池化(Stochastic Spatial Sampling Pooling, S3Pool)<sup>[77]</sup>是使用随机向下采样的两步方法。合并窗口( $2 \times 2$ )在特征图上滑动,单位步幅值向下采样,下采样以均匀和确定的方式从非重叠池化区域中选取单个值,随机和 S3Pool 的混合通过在特征图中引入失真,实现正则化和数据增强。

基于稀疏的随机池化<sup>[73]</sup>融合了最大、平均和随机池化的思想,引入一定程度的稀疏性来获取池化区域的优化特征,特征值从平均值扩展到最大值,优化后的特征值用于正态分布中激活的概率权重分配。该方法采用加权随机抽样,以保留随机池化的优点,图像特征的非平稳性和池化区域的随机性提高了池化性能<sup>[68]</sup>。该方法在不同基准数据集上进行了测试,结果显示有效地提高了识别精度。

### 1.2.4 基于变化域的池化方法

变换域方法使用不同的域,如时间域<sup>[83]</sup>、空间域<sup>[84]</sup>、频率域<sup>[75]</sup>和小波域<sup>[76]</sup>。基于频域变换的池化方法,利用频谱池化通过截断 CNN 在频域中的

表示来进行降维。这种变换是通过使用离散傅立叶变换(Discrete Fourier Transform, DFT)<sup>[75]</sup>、快速傅立叶变换(Fast Fourier Transform, FFT)<sup>[85]</sup>、Hartley变换<sup>[86]</sup>和离散余弦变换<sup>[79]</sup>等变换来实现的。小波池化<sup>[76]</sup>采用二级小波分解,丢弃一级子带,以更紧凑的结构方式降低特征维数。

在CNN中,池化层将池核特征转化为最显著的特征,保留关键信息。与后一层相比,低层的特征图保留了更重要的信息,可产生较高的精度,但需要更多的内存空间和计算时间,低层使用最大池化和高层使用平均池化是一个较好的选择。基于值的池化方法依赖于基于特定条件选择激活值,大多数传统的基于值的池化方法没有考虑池化内核中噪声的影响,由于噪声值在后续特征图中累积并传播,因此,可能会导致不希望的输出,噪声的存在会影响特征图中的大多数局部特征。基于等级的方法考虑了所有具有加权值的激活,因此,此类噪声的影响将通过其他激活来减弱,但它需要更多的参数,以权重形式存储。此外,基于变换域的池化方法可以很容易地滤除由于频率变换而产生的噪声,但与其他池化方法相比,此方法所需的处理时间可能要长得多,这主要是由于从空间域到其他域和原始域的转换而引起的。基于概率的方法通过评估最显著特征的概率,从而避免了这种计算时序问题,对不同层的概率值进行细化可以获得更好的性能。

### 1.3 激活函数

神经网络每一层输出时需要进行一次非线性变换,这个起到非线性变换作用的函数称为激活函数,若没有激活函数,则网络中每一层都是线性变换,而实际上数据通常是线性不可分的,线性模型不足以表达数据,引入激活函数是为了增加整个网络的非线性因素,增强其表达能力。激活函数可分为3类,第1类是饱和非线性函数,如Sigmoid、双曲正切(Tanh)函数,两者都为S型饱和函数,当训练数据规模较大时,Sigmoid和Tanh激活函数容易造成梯度弥散的问题。研究发现,饱和的函数会造成梯度弥散现象;为了解决此类问题,深度卷积神经网络中出现了非饱和激活函数,如ReLU(Rectified Linear Unit)<sup>[87]</sup>、Softplus<sup>[88]</sup>及其改进型,此为第2类;第3类为组合函数,为充分发挥2种激活函数的优势,弥补彼此之间的不足,研究者们将两者有效地结合起来,构成一个新的激活函数来提高网络性能。

激活函数的选择对于卷积神经网络性能有着重要的影响,下面介绍常用的激活函数ReLU及代表

性的改进方法、其他代表性的改进方法。表3给出部分具有代表性的激活函数对比, $f_1(x)$ 为Sigmoid函数; $f_2(x)$ 为Tanh函数; $f_3(x)$ 为ReLU函数; $f_4(x)$ 为Softplus函数; $f_5(x)$ 为Maxout函数,其中 $G$ 为隐含层神经元数量, $s_{qg}$ 为第 $q$ 个输入神经元和第 $g$ 个隐含层神经元的单元输出, $N$ 为输入神经元数量, $w_{qg}$ 为第 $q$ 个输入神经元和第 $g$ 个隐含层神经元的连接权重, $x_q$ 为第 $q$ 个神经元的输入, $v_g$ 为第 $g$ 个隐含层的偏置; $f_6(x)$ 为ELU函数,其中 $\alpha_6$ 为预定义的常数,表示负数部分的斜率; $f_7(x)$ 为Leaky ReLU函数,其中 $\alpha_7 \in (0, 1)$ 为预定义的斜率; $f_8(x)$ 为PReLU函数,其中 $\alpha_8$ 取高斯分布中一随机数; $f_9(x)$ 为SeLU函数,其中 $\alpha_9$ 和 $\lambda$ 均为超参数。

#### 1.3.1 ReLU及代表性的改进方法

在AlexNet模型之前,Sigmoid函数是使用最多的激活函数,但近年来,已逐渐被ReLU函数及其改进型所取代。Sigmoid激活函数易产生收敛速度慢和梯度消失的问题,使用ReLU得到的随机梯度下降法的收敛速度比Sigmoid和Tanh快很多。杨观赐等<sup>[97]</sup>分析了当前卷积神经网络模型特征提取过程中存在的问题:(1)训练数据量较大时,Sigmoid激活函数使训练变得困难;(2)CNN对于过拟合问题,通常采用提早结束训练与正则化策略来缓解,但网络加深时,解决过拟合问题的表现会变差;(3)训练过程中,CNN常采用梯度下降策略,虽然减少了训练时间,但会造成累计误差的增加。Kingma等<sup>[98]</sup>提出了基于Dropout与Adam(Adaptive Moment)优化器的改进CNN算法,设计了二次卷积神经网络结构,采用ReLU代替传统激活函数,以防止梯度弥散,加快收敛速度;在网络中引用了Dropout技术,并采用Adam优化方法替代传统的随机梯度下降法(Stochastic Gradient Descent, SGD)<sup>[99]</sup>,以不同的学习率优化交叉熵以缓解过拟合现象。SGD优化方法通过随机抽取一个样本来计算误差,并以此为依据来更新参数,在初始化和学习率设定合适的情况下,结果更可靠,但是训练时间较长,即收敛速度慢。Adam优化方法通过梯度一阶矩估计和二阶矩估计,来动态调整参数的学习率,收敛速度较快,若训练的网络层数较多较复杂时,使用这种学习率自适应的优化方法表现会更好。

由于ReLU在负值的梯度恒为零,神经元在训练过程中可能发生死亡现象<sup>[90]</sup>。2013~2015年期间,部分学者提出Leaky ReLU<sup>[94]</sup>、Maxout<sup>[91]</sup>、ELU<sup>[93]</sup>、



表 3 部分具有代表性的激活函数对比

Tab. 3 Contrast of some representative activation functions

激活函数	数学表达式	优点	缺点
Sigmoid	$f_1(x) = \frac{1}{1+e^{-x}}$	具有连续可微分性的特征,曾是激活函数的最优选择	不具有稀疏特性,容易梯度弥散,不关于原点对称,收敛速度较慢,且计算复杂
Tanh	$f_2(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	关于原点对称,比 Sigmoid 函数收敛速度快	无法解决梯度弥散的问题,且计算复杂
ReLU <sup>[87]</sup>	$f_3(x) = \max\{0, x\}$	解决了梯度弥散问题,收敛速度快,更加接近生物神经元的激活模型 <sup>[89]</sup> ,且计算简单	稀疏性过强,易出现神经元死亡现象 <sup>[90]</sup>
Softplus <sup>[88]</sup>	$f_4(x) = \ln(1+e^x)$	避免 ReLU 强制性稀疏的缺点,对全部数据进行非线性映射,避免有价值的信息丢失	欠缺对样本模型的表达能力,收敛速度较慢
Maxout <sup>[91]</sup>	$f_5(x) = \max_{g \in [1, G]} s_{qg}$ $s_{qg} = \sum_{q=1}^N w_{qg} x_q + v_g$	克服了 ReLU 的缺点	增加了参数量 <sup>[92]</sup>
ELU <sup>[93]</sup>	$f_6(x) = \begin{cases} x & x > 0 \\ \alpha_6(e^x - 1) & x \leq 0 \end{cases}$	缓解神经元死亡现象,具有一定抗干扰能力,输出均值接近于 0,收敛速度更快	会发生梯度弥散,计算复杂
Leaky ReLU <sup>[94]</sup>	$f_7(x) = \begin{cases} x & x > 0 \\ \alpha_7 x & x \leq 0 \end{cases}$	缓解 ReLU 的神经元死亡现象,具有一定稀疏性	欠缺对样本模型的表达能力,需调整超参数 $\alpha$
PReLU <sup>[95]</sup>	$f_8(x) = \begin{cases} x & x > 0 \\ \alpha_8 x & x \leq 0 \end{cases}$	继承了 Leaky ReLU 的优点,且较 Leaky ReLU 效果更好	欠缺对样本模型的表达能力,需调整超参数 $\alpha$ ,模型训练难度有所提高
SeLU <sup>[96]</sup>	$f_9(x) = \lambda \begin{cases} x & x > 0 \\ \alpha_9 e^x - \alpha_9 & x \leq 0 \end{cases}$ $\alpha_9 \approx 1.673\ 3, \lambda \approx 1.050\ 7$	确保训练过程中梯度不会爆炸或消失,自动将样本分布归一化到 0 均值和单位方差	

PReLU<sup>[95]</sup>等改进函数,缓解了神经元死亡问题。ELU 会发生梯度弥散,且导数计算复杂度为指数级。Goodfellow 等<sup>[91]</sup>引入一种称为丢失的近似模型平均技术,提出名为 Maxout 的激活函数,并且表明 Maxout 可拟合各种维度的凸函数,而 Sigmoid、ReLU 等激活函数只能拟合二维函数。Maxout 函数是在每个隐含层节点的  $k$  个延伸节点中,选取最大值作为该节点的输出,其中  $k$  为 Maxout 函数参数,其思想类似于 Dropout 技术,旨在通过丢失来进行优化。试验表明,Maxout 效果要优于 ReLU 函数,并且通过将 Maxout 和 Dropout 结合的试验,验证该结合方式更易优化网络,因为 Dropout 技术改变了 Maxout 每次训练的输入数据,使得模型更加多样化。

### 1.3.2 其他代表性的改进方法

如表 3 所示,单一的激活函数既有其优点,也有一些不足,这些不足也会严重影响到卷积神经网络的性能,因此,一些研究者根据不同激活函数的特点,将两种激活函数组合起来,保留两者的优点,弥补不足。石琪<sup>[100]</sup>将 ReLU 和 Softplus 两种激活函数有效的结合起来,提出了 ReLU-Softplus 激活函数,为非线性非饱和的激活函数。ReLU 函数收敛速度快,但在小于 0 的部分梯度被强制为 0,不能有

效地学习到数据特征,Softplus 是 ReLU 近似平滑的一种表达形式,可对全部数据进行非线性映射,但其收敛速度较慢。ReLU-Softplus 函数数学表达式为  $f(x) = \max\{\ln(1+e^x) - \ln(2), x\}$ ,即在  $x$  负半轴使用 Softplus 函数,并将其函数曲线向下平移  $\ln(2)$  个单位,在  $x$  正半轴使用 ReLU 函数,这样既保留了 ReLU 收敛速度快以及 Softplus 防止有价值信息丢失的优点,也解决了神经元死亡的问题。通过在手写字符数据集中的试验表明,改进的激活函数训练效果优于单一的 ReLU、Softplus 函数。但研究发现,使用该激活函数时模型只能使用 0.000 1 或更小的学习率,从而导致收敛速度降低。王红霞等<sup>[101]</sup>针对此问题,提出了一种新的组合激活函数 ReLU-Softsign,是由 ReLU 和 Softsign 函数在  $x$  轴正、负半轴组成的一个分段函数,其中 Softsign 函数由 Tanh 函数改进得到,与 Tanh 函数相比,Softsign 函数在缓解收敛速度慢及梯度弥散的问题上具有一定的积极作用。在  $x$  负半轴中 Softsign 函数的导数能够保持变化,逐渐减小,而 Softplus 的导数比 Softsign 更容易趋向零,当选用较大的学习率时,Softplus 的导数更容易降为零,使神经元坏死。此外,当输入数据在零附近时,Softsign 的导数具有更快的变化,说明收

收敛速度也更快,因此,相比 ReLU-Softplus, ReLU-Softsign 函数可使模型使用更高的学习率,从而提高收敛速度。

激活函数的改进还有很大的发展空间,未来可根据在不同优化器上以及不同的数据集上的适应程度,进一步改进激活函数,提高网络收敛性能。

#### 1.4 卷积神经网络的基本工作原理

卷积神经网络总体来说是一种层次模型,原始数据作为输入,通过卷积操作、池化操作和非线性激活函数映射等一系列操作的层层堆叠,从原始数据输入层中将高层语义信息逐层抽取出来,这一过程称为“前馈运算”。最终,卷积神经网络的最后一层将其目标任务(分类、回归等)形式化为目标函数,计算预测值与真实值之间的误差或损失,通过反向传播算法将误差(损失)由最后一层逐层向前反馈,更新每层参数,并在更新参数后再次前馈,如此往复,直到网络模型收敛,从而达到模型训练的目的。训练过程如下。

步骤 1:给定一个样本集,包含  $m$  个样例,初始化权值偏置,通过前馈计算每层的状态和激活值,直到最后一层,即

$$z_l = W_l f_l(z_{l-1}) + b_l \quad (2)$$

式中: $z_l$  为  $l$  层神经元的状态; $W_l$  为  $l-1$  层到  $l$  层的权重矩阵; $f_l(\cdot)$  为  $l$  层神经元的激活函数; $b_l$  为  $l-1$  层到  $l$  层的偏置。

步骤 2:计算每一层的误差  $\delta_l$ ,将误差由最后一层逐层向前反馈,即

$$\delta_l = f'_l(z_l) [(W_{l+1})^T \delta_{l+1}] \quad (3)$$

式中: $(W_{l+1})^T$  为  $l+1$  层的权重转置矩阵,通过  $(W_{l+1})^T$  向前反馈到第  $l$  层。

步骤 3:计算每层参数的导数,即

$$\frac{\partial J(W, b; x_m, y_m)}{\partial W_l} = \delta_l (a_{l-1})^T \quad (4)$$

$$\frac{\partial J(W, b; x_m, y_m)}{\partial b_l} = \delta_l \quad (5)$$

$$a_{l-1} = f_l(z_l)$$

式中: $a_{l-1}$  为  $l-1$  层神经元的活性值矩阵; $J(W, b; x_m, y_m)$  为单个样例计算得到的方差代价函数; $W$  为权重矩阵; $b$  为偏置矩阵; $(x_m, y_m)$  为第  $m$  个样例。

步骤 4:对参数  $W$  和  $b$  进行更新,即

$$\begin{aligned} W_l &= W_l - r \frac{\partial J(W, b)}{\partial W_l} \\ b_l &= b_l - r \frac{\partial J(W, b)}{\partial b_l} \end{aligned} \quad (6)$$

$$\begin{aligned} \frac{\partial J(W, b)}{\partial W_l} &= \left[ \frac{1}{m} \sum_m \left( \frac{\partial J(W, b; x_m, y_m)}{\partial W_l} \right) \right] + \beta W_l \\ \frac{\partial J(W, b)}{\partial b_l} &= \frac{1}{m} \sum_m \left( \frac{\partial J(W, b; x_m, y_m)}{\partial b_l} \right) \end{aligned} \quad (7)$$

式中: $J(W, b)$  为整体样本代价函数; $r$  为学习速率; $\beta$  为权重衰减参数。

步骤 5:在更新参数后再次进行前馈运算,如此反复,直到网络模型收敛,从而完成网络训练。

#### 1.5 优化算法

优化问题是计算数学中最为重要的研究方向之一,是指将模型预测值和数据之间的误差减小问题转换为求一个目标函数的最小值问题。在深度学习领域,优化算法的选择也十分重要,不同的优化算法很可能产生截然不同的训练效果,常用的优化算法可分为梯度下降算法及其改进和自适应优化算法,如表 4 所示。

表 4 优化算法类别

Tab. 4 Classification of optimization algorithms

类别	基本算法	自适应学习率算法
代表性算法	BGD、SGD、MBGD、动量 <sup>[102]</sup> 、 基于 NAG 的动量方法 <sup>[103]</sup>	Adagrad <sup>[104]</sup> 、Adadelata <sup>[105]</sup> 、 RMSprop <sup>[106]</sup> 、Adam <sup>[98,107]</sup> 、 Nadam <sup>[108]</sup> 、AdaBound <sup>[109]</sup> 、 AMSBound <sup>[109]</sup>
特点	需人工设计学习率,在适合的 初始化和学习率调度方案 的情况下,结果更可靠	自适应地为每个参数分配 不同学习率,性能更稳定, 适用于训练较深较复杂的 网络以及稀疏数据

##### 1.5.1 基本算法

梯度下降最常见的 3 种变形为批量梯度下降法 (Batch Gradient Descent, BGD)、随机梯度下降法 (Stochastic Gradient Descent, SGD) 和小批量梯度下降法 (Mini-Batch Gradient Descent, MBGD),这 3 种形式的区别取决于使用多少数据来计算目标函数的梯度。BGD 采用整个训练数据集计算梯度,对于凸函数可以收敛到全局极小值,对于非凸函数可以收敛到局部极小值,但遇到大规模数据集时计算速度非常慢,而且不能加入新的数据实时更新模型。SGD 是通过每个样本来迭代更新一次,速度较快并且可以新增数据,但 SGD 因为更新比较频繁,易造成损失函数的震荡。MBGD 每一次利用一小批样本进行计算,可以降低参数更新时的方差,收敛更稳定,并且可以充分利用高度优化的矩阵操作来更有效的计算梯度,但该算法对所有的参数更新都使用相同的学习率,不能满足个别要求,且容易陷入局部最优。

动量方法<sup>[102]</sup>在 MBGD 基础上模拟物理学动量的概念改进更新策略,即在每次更新时一定程度的保留之前的更新方向,同时利用当前批次的梯度向最终更新方向微调。这样做使收敛更加稳定、快速,并且提高了摆脱局部最优的能力。Sutskever 等<sup>[103]</sup>利用斯特洛夫加速梯度(Nesterov's Accelerated Gradient, NAG)<sup>[84]</sup>对动量方法进行了改进,在梯度更新时做一个校正,提高了梯度更新灵敏度,对实际情况更有针对性。

### 1.5.2 自适应学习率算法

梯度下降算法及其改进算法都需要人工来设置学习率,总会出现不适合的情况。Adagrad (Adaptive Gradient)<sup>[104]</sup>对学习率进行了约束,能够自适应地为不同参数分配不同的学习率,无需手动调节。Adagrad 对低频参数更新使用较大学习率,对高频参数使用较小学习率,因此,它对于稀疏的数据表现很好,比 MBGD 更具鲁棒性,但同时该方法也需要设置一个全局学习率,且在前期梯度很大的情况下,该方法对梯度调节过大,中后期使得学习率过小导致训练提前结束,因此,效果并不理想。Adadelta<sup>[105]</sup>在计算上做了简化,Adagrad 会将之前所有的梯度平方累加,而 Adadelta 只考虑固定大小的项,并且采用近似计算对应的平均值,解决了 Adagrad 算法学习率单调递减、手工设置全局学习率等问题。

为了进一步优化损失函数在更新中存在摆动幅度过大的问题以及加快函数收敛速度,RMSprop<sup>[106]</sup>对权重  $\mathbf{W}$  和偏置  $\mathbf{b}$  的梯度使用了微分平方加权平均,并且为了防止分母为 0,使用了一个很小的数值来进行平滑,一般取  $10^{-8}$ ,RMSprop 更适用于处理非平稳目标。Adam 方法本质上是带有动量项的 RMSprop,它利用梯度的一阶矩估计和二阶矩估计动态调整每个参数的学习率,主要优点是经过偏置校正后,每一次迭代学习率都有确定的范围,使得参数比较平稳。Nadam<sup>[108]</sup>类似于带有 Nesterov 动量项的 Adam,对学习率有了更强的约束。自适应优化算法旨在通过调整学习率来加快收敛速度,但它们与 SGD 相比泛化性能较差,甚至会由于不稳定和极端的学习率导致无法收敛。Luo 等<sup>[109]</sup>提出 AdaBound 和 AMSBound 方法,对学习率使用动态边界,以实现从自适应方法到 SGD 的逐步平滑过渡,为收敛提供理论依据。通过在多个流行任务和模型上进行试验,结果表明所提方法能够弥补自适应方法和 SGD 之间的泛化差距,同时在训练早期保持更高的学习速度。

## 2 卷积神经网络在智能交通的研究及应用

近年来,卷积神经网络在图像识别、目标检测、图像分割等领域取得了突破性的进展,同时随着计算机硬件水平的提升,复杂算法所带来的实时计算困难的问题已逐渐解决,极大地推动了 CNN 在智能交通系统的应用,在车牌、车型、交通标志的识别以及交通状态预测等领域取得了许多研究成果。

### 2.1 车牌识别

车牌识别(License Plate Recognition, LPR)技术是现代智能交通系统重要组成部分,通过一些后续处理技术可以实现停车场出入口收费管理、盗抢车辆管理、高速公路超速自动化管理、闯红灯电子警察、公路收费管理等功能,对于维护交通安全和城市治安,防止交通堵塞,实现交通全自动化管理有着现实意义。车牌识别方法可分为 2 类,一类是通过识别 IC 卡或条形码中所存储的车牌信息来识别车牌,该方法准确度高,运行可靠,但实现成本较大,对硬件设备要求较高,不利于推广。另一类是基于图像的方法,该方法有模板匹配法、特征匹配法和神经网络的方法。近几年来,基于神经网络的车牌识别方法已逐渐发展成熟,其中卷积神经网络应用最为广泛。

车牌识别方法通常由车牌定位、字符分割和字符识别 3 个处理步骤组成。然而,在光照不均匀、天气条件恶劣、背景复杂、车牌不清晰、监控摄像头质量较差等复杂环境下,该识别方案的效果并不理想。Wang 等<sup>[110]</sup>为了解决这些问题,提出了一种新的车牌识别方法,通过一种端到端的深度学习结构,直接对车牌进行识别。该网络是一种混合的深度结构,它由用于提取基本特征的残差网络、用于提取多尺度特征的多尺度网络、用于定位板块和字符的回归网络和用于识别的分类网络组成。同时提出了一种有效的基于批量归一化的训练方案,提高了学习过程中的训练速度。试验结果表明,该方法具有良好的识别精度,与现有的复杂环境识别方法相比,具有更强的鲁棒性和更高的识别效率。李祥鹏等<sup>[111]</sup>采用 Faster R-CNN 算法和改进的 AlexNet 网络对车牌目标进行定位和识别,解决了复杂环境下车牌定位困难的问题。

Lin 等<sup>[112]</sup>提出了一个级联框架,该框架在整个图像上以滑动窗口方式使用一个 4 层的 CNN 分类器,区分 37 类(10 位数字,26 个大写字母以及非字



符类别)目标,基于字符的聚类特性,提取出类文本区域。然后采用另一种车牌/非车牌的二分类 CNN 分类器来剔除误报,区分车牌和普通文本。这种方法的主要优点是不需要分割,通过挖掘上下文信息,避免了分割带来的误差,具有较高的识别精度。Xiang 等<sup>[113]</sup>为了提高车牌在复杂环境中的检测性能,提出了一种用于车牌检测的高效轻量级全卷积网络。该网络缩小了输入图像的比例,从而大大加快了计算速度,降低了计算成本,同时结合了多层次、多尺度的视觉特征,采用密集连接和膨胀方法。为了进一步提高预测精度,在训练中加入了融合损失结构。试验分别在 3 977 张不同场景不同条件下拍摄的图像和加州理工学院车牌数据集上进行,结果表明,该方法在精度和计算量之间取得了较好的平衡。

对于多车道场景下的跨国车牌检测,现有方法大多易受光照变化大、背景复杂、车牌边缘弱的影响。Asif 等<sup>[114]</sup>提出了一种新的光照不变性方法来处理不同颜色和风格的跨国车牌。红色电晕最初用于检测车辆的尾灯,又因为车牌在其尾灯附近,便于确定感兴趣的区域。为了检测每个感兴趣区域内的车牌,利用启发式能量图对提取出的垂直边缘进行有效的预处理,保留车牌边缘提取出高边缘密度区域,最后通过 AlexNet 卷积神经网络提取的检测区域高层特征进行分类,验证检测到的区域为车牌区域。该方法已在 6 个不同国家的 2 900 多幅不同天气和照明条件下拍摄的图像上进行了测试,准确率为 91.5%,优于最新的方法,同时平均每张图像只需 0.16 s 即可达到实时性能。

目前许多车牌识别技术中大多数只能在单行牌照上工作,在实际应用场景中,也存在较多的多行牌照,传统的方法需要对多行牌照的原始输入图像进行分割,这是复杂场景中的一个非常困难的问题。Puarungroj 等<sup>[115]</sup>对泰国摩托车牌照进行了研究,提出了一种两步训练识别方法,这种类型的车牌为三行牌照,从上到下分别表示类别、注册地和注册号。该方法首先利用 MobileNets 和 Inception-v3 模型对每个车牌的 3 行数据进行分割和训练;然后对车牌上的泰语字母和数字进行分割和训练。通过对不同方位的车牌图像进行测试,3 行不同类别识别正确率为 96.94%,字符识别正确率为 91.76%。Cao 等<sup>[116]</sup>提出了一种用于单行和双行车牌识别的端到端神经网络。该模型的网络结构是专门为多行车牌识别而设计的。利用深度 CNN 直接将每幅图像作为一个

单元在特征图上进行处理,然后对得到的特征图进行重组,每个特征图都表示为所有单行和双行牌照的一个类似于序列的特征图,得到的特征图采用双向长短时记忆进行序列标签的计算,通过对这些序列标签信息进行组合,得到最终的车牌识别结果。该方法具有训练方便、测试方便等优点,可应用于多种车牌识别系统,大量试验结果表明,提出的方法与目前最先进的方法相比,取得了显著的改进。现有算法对双行车牌的识别率较低,并且很少涉及对包含中文字符车牌的自动识别。赵汉理等<sup>[117]</sup>提出一种基于 CNN 的双行车牌分割算法。该算法首先利用 CNN 提取的车牌特征训练多标签分类模型,将双行车牌分割为两个单行车牌,然后采用卷积循环神经网络对车牌字符识别。基于所构建的数据集的试验结果表明,文中算法自动分割双行车牌的准确率较高,但文中所采用的特征提取网络层数较少,且分割和识别采用两种网络进行,有进一步改进的空间。

## 2.2 交通标志识别

交通标志识别(Traffic Sign Recognition, TSR)属于智能交通系统的外部感知信息,将采集道路中的标志图像,通过相关技术检测、识别出交通标志图像。交通标志识别的准确度和实时性是分类的重要指标,快速准确的识别可以为自动驾驶、行车安全以及交通的运行效率提供保障。卷积神经网络在图像分类领域取得巨大的进展,但同时需要大量的数据集作为支撑,为了推动卷积神经网络在交通标志识别领域的研究,各类交通标志数据集也得以开源,其中主要有德国交通标志识别集、德国交通标志检测数据集、清华-腾讯 100 k、瑞典交通标志检测数据集(Sweden Traffic Sign Detection Dataset, STSD)等。

利用深度学习的方法进行交通标志识别,基本流程包括图像预处理、交通标志特征提取和分类识别。环境因素对交通标志的分类精度影响明显,张秀玲等<sup>[118]</sup>将三角形和圆形交通标志使用 K-means 算法进行二聚类操作,卷积神经网络作用是将切割模板切割的感兴趣区域的 HOG 特征进行过滤和降维,其中切割预处理操作的目的是降低环境因素对信息的干扰,进行过滤和降维后的二次特征经过 Squeeze 网络进行重要性标定,增强数据之间的关系,将 HOG 特征与卷积特征进行加权运算后,训练神经网络模型。该模型不仅提高了分类精度并且模型的收敛速度和泛化能力都有所增强,与 BP 网络、支持向量机(Support Vector Machines, SVM)分类器及

CNN 模型对比,分类精度达到了 98.58%,训练时间上大致相同,缺陷是在反光、遮挡环境影响下无法自动定位检测,这也是目前需要突破的交通标志分类的重点问题。

交通标志所包含的信息丰富,包括颜色、形状和文本等语义信息,人类理解该信息简单,但这些信息的类间差异小并且种类多,不利于机器分析理解。王方石等<sup>[119]</sup>为了解决该问题,在 HSV 颜色空间的 3 个分量形成的灰度图上提取最大稳定极值区域,首先得到交通标志候选区域,之后将卷积神经网络加入交通标志的形状、颜色、图案内容 3 种信息的属性学习约束,用于识别其是否为交通标志区域,如果为交通标志区域便对进行分类,实现交通标志的检测。该方法充分考虑了交通标志特征的多样性,针对颜色、形状、文本的特征,同时进行属性学习和分类学习,将交通标志的多样视觉属性特征作为学习约束,通过试验表明,在复杂环境下能很快的识别出交通标志区域和非交通标志区域,检测精度较高,具有一定的实际应用价值。

交通标志检测的难点在于影响检测效率的干扰较多,一类为交通标志的遮挡、污损、模糊、褪色、变形等自身干扰,除此之外,同一种交通标志在不同的时间、天气、角度、光照以及分辨率的情况下检测效果也不相同。交通标志检测主要服务于智能交通中的自动驾驶技术,属于实景检测项目,因此,对于检测的实时性和准确性要求较高,李旭东等<sup>[120]</sup>将 YOLO V3-tiny 检测算法作为基础网络,提出了一种三尺度嵌套残差结构的交通标志快速检测算法。首先构成两个一大一小的残差结构,通过在 YOLO V3-tiny 网络上采用逐像素相加的跨层连接构成小残差结构,利用同样的连接方式,在原基础网络输出的基础上增加了 1 层空间分辨率更高的预测输出,增强了该尺度输出的空间信息,构成大残差结构。然后将构成的两个残差网络进行嵌套,使其形成一个三尺度预测的嵌套残差网络模型,将 YOLO V3-tiny 基础网络放入两个残差结构中,可以进行 3 次调参。通过在德国交通标志检测数据集上进行试验其总  $F_1$  分数为 91.77%、检测时间为 5 ms;长沙理工大学中国交通标志检测数据集上进行试验,指示、禁令、警告三大类交通标志  $F_1$  分数分别为 92.41%、93.91%、92.03%,检测时间为 5.0 ms。 $F_1$  分数是机器学习中为综合考虑精确率和召回率而引入的评价指标, $F_1$  分数越高,说明模型性能越好。

交通标志图像的特殊性在于一个大类交通标志

(如禁令标志、指示标志等)中各个小类的区别体现在其中心区域的局部差异。将交通标志图像的全局特征和局部纹理特征相结合,多尺度融合提取交通标志特征引起了学者的关注。宋青松等<sup>[121]</sup>提出了一种改进的多尺度卷积神经网络的交通标志识别算法,首先为了降低环境因素的影响,对交通标志图像进行预处理,经过预处理比选试验后,最终确定限制对比度自适应直方图均衡化方法作为图像预处理方法,然后设计了一种多尺度 CNN 模型,将交通标志的全局特征和局部特征融合,得到多尺度融合特征,最后将得到的融合特征输入 SoftMax 分类器对交通标志进行分类。文章在设计多尺度模型之前首先对网络参数的确定做了比选试验,确定了适合交通标志识别的卷积核尺寸、全连接层神经元个数、不同的权重初始化方式、激活函数、优化器、Dropout 的参数,该文多尺度主要体现在将第 1、2 池化层的特征再经过一次池化,然后与第 3 池化层连接,最后将融合后的特征送入分类器进行分类。通过在德国交通标志基准数据库试验,其分类精度达到了 98.82%,单张交通标志的识别速度为 0.1 ms。CNN 属于一种层级结构,通过层层映射得到图像的特征,高层和底层特征都代表图像的不同属性,底层特征主要提取图像的粗粒度特征,包含对象的纹理、边缘等信息,高层特征可以得到图像细粒度的语义信息。利用好图像的多层特征属性,对于图像分类精度的提升有一定帮助。孙伟等<sup>[122]</sup>在对交通标志图像进行分类时,利用了 CNN 多层特征属性特点,在卷积神经网络提取交通标志特征时,将后 3 层特征提取出来,经过一个多尺度池化层,下采样后得到具有多层属性的特征向量,分类器选择将传统的 SoftMax 分类器用极限学习机分类器替换,完成交通标志的识别。选择极限学习机(Extreme Learning Machine, ELM)是因为融合的多层特征维度较高,而 ELM 是一种基于单隐层前馈神经网络的监督型算法,训练过程中不需要反复调整所有的参数,降低计算量,加快训练速度,通过试验表明结合 CNN 和 ELM 的识别算法的鲁棒性更强,在 GTSRB 交通标志数据库上分类精度达到了 96.93%,单张交通标志图像的分类时间为 8.2 ms,具有较高的准确性和实时性。

交通标志在图片中所占比例较小,因此,检测难度更大,一般通过图像分块技术来提高小目标的相对尺寸比例,降低小目标检测难度。张淑芳等<sup>[123]</sup>针对单尺度分块提升效果有限的问题,提出多尺度分块方法,并构造了更加丰富的卷积层,进一步提高对



不同尺寸目标的检测能力。同时采取由粗到精的策略来加快检测速度,对纯背景图像块不再预测,先利用中等大小图像块来确定交通标志目标的范围,再对范围内的其他图像块进行检测,最后通过非极大值抑制实现交通标志的精准识别。该方法在不同的环境下进行测试,均表现出良好的检测性能和鲁棒性。Liu等<sup>[124]</sup>提出了一种基于反卷积区域的卷积神经网络来对交通标志进行检测,该方法在卷积层的输出中加入一个反卷积层和一个归一化层。将不同层次的特征拼接成融合的特征图,为小交通标志的检测提供充分的信息。为了提高训练的有效性,提出了一种针对区域建议网络和反卷积区域卷积神经网络两阶段自适应分类损失函数,通过该设计来完成对交通标志目标的检测,试验利用清华-腾讯100 k数据集评估算法的性能,通过比对试验,分析了融合特征图和两阶段分类损失函数的有效性,最后的试验结果证明了该方法在检测交通标志方面的优越性。

交通标志识别一般应用于车载系统,首先通过安装在车辆上的摄像头采集实际交通中的图像,再输入车载系统进行数据分析,大型卷积神经网络可实现较高的实时性、准确性和鲁棒性,但同时存储空间和计算资源需求也较高,车载计算机很难达到要求。张建明等<sup>[125]</sup>提出了一种基于压缩卷积神经网络的交通标志分类算法,采用基于泰勒展开的通道剪枝技术删除冗余的特征图通道,并对剪枝后的网络模型采用三值量化方法进行参数量化,试验验证了通道剪枝、参数量化和组合压缩的有效性,使得组合压缩后的网络参数数量减少为原始模型的9%,提高了模型的处理速度与精度,具有一定的实用价值。在移动平台上部署深度神经网络具有很大的挑战性,因为它具有较长的延迟和巨大的计算成本。以往的研究工作多数集中在计算精度上,如何在计算量和精度之间取得平衡是移动计算领域需要解决的重要问题。Song等<sup>[126]</sup>提出了一种高效的卷积神经网络,在CNN模型中构建了2个子网络,从低层的细粒度特征出发,通过增加特征图和下采样的经验值,对小目标的特征进行识别。该检测子网络经过良好的训练,可以同时对小交通标志进行分类和检测,另一方面,使用了卷积分解、冗余层裁剪和全连接变换3种策略来压缩模型,降低了计算成本,提高了小目标检测的效率。试验结果表明该方法网络性能优于Faster R-CNN和Fast R-CNN,准确率提高了10%。可以在精度损失较小的前提下,网络运

行时间和内存的计算消耗有所减少,显著的减少冗余,减少参数,提高网络的处理速度。

### 2.3 车型识别

车型分类分为2种,车辆类型分类和车辆型号分类,车辆类型一般包含客车、卡车、轿车等大类,车辆型号包含信息较为细致,例如,汽车的品牌以及车辆颜色。随着深度学习在智能交通领域研究不断进步,开源的车辆数据集也为其发展提供了帮助。北京理工大学制作的BIT-Vehicle数据集,其车辆图片全部来自高速监控卡口,包含客车、小型客车、小型货车、轿车、越野车以及卡车6种类型共9 850张车辆图片。香港中文大学制作的CompCars数据集,由监控卡口和互联网收集两部分构成,其中,互联网场景图片包含163个品牌、1 716个车型,共136 727张整车图片,27 618张车辆局部图片;监控卡口的数据都为车辆的正脸图像,共有44 481张图像,分为281种类型。Stanford Cars数据集由斯坦福大学所收集,包含196类,共16 185张图片,其中8 144张为训练数据,8 041张为测试数据。每个类别按照年份、制造商、型号等进行区分。

在智能交通系统中,车辆颜色自动识别在视频监控中具有重要意义。在光照条件不断变化的复杂交通场景中,车辆颜色的正确识别是非常困难的。Zhang等<sup>[127]</sup>针对传统的神经网络参数较大,并且只使用图像的高层特征的问题,提出了一种基于轻量级卷积神经网络的车辆颜色识别方法。该轻量级的CNN网络架构是专门为车辆颜色识别任务而设计的,它包含5层,其中3个卷积层,一个全局池化层和一个完全连接层。与现有CNN网络仅使用最后一层输出特征进行识别的方法不同,该卷积特征提供了层次化的表示,将3个卷积层的特征图全部用于识别,采用空间金字塔匹配(Spatial Pyramid Matching, SPM)策略对特征图进行划分,并对每个SPM子区域进行编码生成特征表示向量。将卷积层的特征表示向量和全局池化层输出的特征向量归一化,并级联为一个整体特征向量,最后将得到的特征向量输入支持向量机进行分类。与常规方法相比,该方法主要有两个优点:首先是设计一种特殊的轻量级卷积神经网络来完成车辆颜色识别任务,减少网络参数量,降低对计算和存储资源的需求;其次是应用SPM策略将空间信息嵌入到卷积特征图中,将低层到高层的特征进行组合,以提高深度特征表示能力。最终试验结果表明,该方法识别率达到了95.41%,而特征向量的维数仅为18.00%,



存储占用空间只有 0.50%,同时兼顾了识别率和模型的体量。Fu 等<sup>[128]</sup>提出了一种新的网络结构,基于残差学习的多尺度综合特征融合卷积神经网络。首先,利用 MCFF-CNN 网络提取车辆的深颜色特征;然后使用支持向量机来获得最终的颜色识别结果,并建立了一个实际交通场景中车辆颜色鲁棒识别系统,从不同的角度对算法进行了评价,并与现有的算法进行了比较,试验结果表明该算法具有较好的性能。

深度学习算法在目标检测领域得到了广泛的应用,设计合适的深度学习算法进行车辆检测具有重要的研究意义。Li 等<sup>[129]</sup>研究了卷积神经网络在车辆目标类型检测与识别中的应用,将 Faster R-CNN 模型应用于实际的交通环境。通过设计改进的模型以及使用 MIT、Caltech 车辆数据集和网络图片的自建数据,着重训练 ZF 和 VGG16 网络以检测交通场景中的车辆类型,来检测和识别 3 种类型的车辆交通场景,试验结果表明,与传统的机器学习方法相比,该模型在平均目标检测精度和检测率上均得到了改进。

细粒度车型识别是指对任意角度及场景下的车辆外观图像进行识别,包括生产厂家、品牌型号、年款等信息,在智能交通、交通安全管理等方面有重要作用。目前细粒度车型识别由传统的手工提取特征向 CNN 自主提取特征过渡。细粒度车型识别面临许多挑战,其中包括:(1)如何将车辆特征联合优化而不是单独处理;(2)如何自主学习车辆特征而不是使用手工制作;(3)如何感知和定位车辆区域,而不是识别整个图像。鉴于此问题,Bin 等<sup>[130]</sup>提出了一种基于卷积神经网络的细粒度车型识别方法。多任务 CNN 在第一阶段对车辆进行定位,在第二阶段对子类进行识别,从而能够处理背景混乱的样本,以及车辆无法填充大部分图像的问题。通过从每个阶段多个任务中的协作学习特性,来处理下属级别的细微类间差异,并在 CNN 训练中开发了新的面向车辆的数据增强策略。为了推进以车辆为中心的任务研究,他们还发布了一个新的车辆数据集,并提供了语义标签和边界框,该数据集可用于车辆定位、识别、视点估计等,相比基于视频的数据集,提供了更多的线索,可作为视频序列车辆分类的辅助数据集。在细粒度车型识别中,由于具有区分度特征的绝对位置的不稳定变化,较大的车类差异会严重影响最终的预测指标,感兴趣区域定位不能完全解决这一问题。Ye 等<sup>[131]</sup>提出了一种基于全局相对位置空间池化的

方法,将绝对位置信息替换为全局相对位置信息,从而有效地减少了类内的差异。其中,包含绝对位置信息和外观信息的鉴别特征由卷积神经网络提取,然后将绝对位置转换为全局相对位置,并对全局相对位置空间中离散位置的概率进行估计,据此计算出不同类型特征在不同代表位置的联合概率,构成池向量,最后用支持向量机进行分类训练。最终结果表明该方法可以有效改善各类卷积神经网络,并在车型精细识别中取得了理想的分类效果。

车辆型号精细识别的关键是提取有区分性的细节特征。余烨等<sup>[132]</sup>提出了一种基于残差网络特征重用的深度卷积神经网络模型 FR-ResNet。该方法的核心为特征重用,先以 ResNet 残差结构为基础网络,使用多尺度输入降低网络退化和易陷入局部最优问题,低层特征在高层中实现特征重用,低层网络部分采用特征图权重学习策略抑制冗余特征的比重。该方法可以高效的利用特征,并降低了网络参数,在 CompCars 和 Stanford Cars 数据集上进行试验,并与其他网络模型进行对比,试验结果表明,该方法鲁棒性较好,分类精度高。

细粒度车型识别类间差异较小,并且差异多数集中在关键的语义部分,深度卷积神经网络模型虽然在分类识别上取得了较好的成绩,但对于语义信息的表达不足,杨娟等<sup>[133]</sup>提出基于语义信息融合的深度卷积神经网络模型及细粒度车型识别模型。该网络模型由定位和识别两部分组成,首先使用 Faster R-CNN 定位网络定位出车辆位置和关键语义部件位置,将定位出的目标通过识别网络进行特征提取;再将各个位置的特征通过小卷积核进行特征拼接和融合,最终进行分类识别。试验在 Stanford BMW-10 数据集和 Stanford Cars-197 数据集中进行,分类精度分别为 78.74% 和 98.27%,迁移学习模型在 BMVC Car-Types 数据集中进行,分类精度为 98.27%,比该数据集目前最好的识别效果提高 3.77%;该模型解决了深度卷积神经网络在语义信息表达不足的问题,并具有较高的分类精度和普适性。

细粒度车型识别类别信息丰富,在进行特征提取时会产生冗余信息,蒋行国等<sup>[134]</sup>提出了一种基于奇异值分解与中心度量的细粒度车型识别算法。在处理冗余特征时利用奇异值分解卷积神经网络的全连接层权重矩阵,之后重新赋值并进行微调操作,得到车型的细粒度特征,为了使车型特征类间距离更小,选择将中心距离损失和分类损失进行加权融合以降低损失。使用 ResNet 框架在 Cars-196 细粒

度车型数据集上测试,分类精度为 93.02%,并通过扩展试验表明该方法的适用性更好,在其他网络框架下仍满足实际场景的要求。细粒度目标识别关键在于区域目标局部的特征信息,而卷积神经网络必须先知道车辆的具体位置。因此,杨娟等<sup>[135]</sup>提出基于区域建议网络的细粒度识别方法,区域建议网络的目的是为了提供区域候选,该网络本质上是一种全卷积神经网络,先要进行特征提取,然后在卷积特征上滑窗产生区域候选,将候选区域的特征进行分类和回归操作,得到目标的类别概率和位置信息,通过目标检测网络获取其具体类别及目标的精确位置,使用非极大值抑制算法得到最终识别结果。分别在 Stanford BMW-10 数据集和 Stanford Cars-196 数据集上进行试验,得到准确率分别为 76.38% 和 91.48%,相比于传统的人工特征提取有了大幅领先,与目前常用算法识别性能相当,并在实景试验中也取得了优异的识别效果。区域建议网络在目标检测时可以提供目标的具体位置信息,在进行细粒度目标识别时可以提供具有区分度的局部区域,克服了对目标位置信息的依赖性,具备一图多车的车型细粒度识别能力,具有更好的鲁棒性及实用性。

#### 2.4 短时交通流预测

短时交通流预测是智能交通系统的重要分支,主要预测模型有时间序列统计模型、非参数预测模型和混合模型。随着硬件条件和大数据处理的迅速发展,基于卷积神经网络的交通流预测研究也获得良好的进展。CNN 稳定性较好,预测性能不易受外部环境的干扰,提取数据时时空特性较为明显。因此,罗文慧等<sup>[136]</sup>利用 CNN 和支持向量回归分类器(Support Vector Regression, SVR)的各自优势,将 2 种方法结合用于预测短时交通流,先使用 CNN 在网络底层提取交通流特征,然后利用 SVR 预测流量,通过试验结果可知,该模型预测性能较好,相对于传统模型性能提升 11%,可以实时准确的预测交通流。

准确预测交通拥堵状况可以提高出行效率,而在短时交通拥堵预测中受到的干扰因素较多,其中主要包括临近街道的交通拥堵情况、天气状况、节假日人流变化的影响,这些因素会导致车速数据波动较大且异常值较多,在预测中不仅需要考虑常态数据,异常时的数据特征对于预测结果同样重要,因此需要有效的方法来提取这些数据的时空特征信息。近年来深度学习研究进展迅速,越来越多的学者通过改进深度学习的预测模型来获取和利用数据的时空特征。石敏等<sup>[137]</sup>通过研究发现,卷积神经网络对

短时交通拥堵情况进行预测时,卷积池化等操作易使目标的位置信息丢失和分辨率持续下降,导致在预测时性能降低。为了解决该问题,提出了一种空洞卷积结合稠密网络的空洞-稠密神经网络模型。利用空洞卷积和稠密网络的连接,获得较少的参数和更大的感受野,并通过稠密网络和下采样的等值映射,解决因为参数增加而导致的网络退化问题。通过采集实际交通道路信息进行试验,结果表明,该模型相比于卷积神经网络模型,预测平均绝对误差降低 3%~23%。

Deng 等<sup>[138]</sup>提出基于 CNN 的随机子空间学习方法,将数据转换成图像的形式并进行时空特征信息的提取,将时间序列分析问题转换为图像分析问题,利用类似图像的数据形式,可以使用二维卷积算子探索时空关系,所提出的模型对于不完整的数据也可以实现很好的预测,之后提出了一种基于均匀设计的改进随机搜索顺序优化深度卷积神经网络,进一步提高了模型的预测能力;An 等<sup>[139]</sup>通过利用残差结构来增加卷积的层数,并联合不同时间间隔输入建立起一个模糊卷积神经网络深度模型,为了提取交通流数据的时空特征,将整个区域划分为  $32 \times 32$  的小块,并构造了 3 个具有流入和流出类型的趋势序列,使得数据在时间上的特征信息能被充分利用。

由于相邻道路之间潜在的内在关系,空间信息在一定程度上影响交通流预测。同时,时间特征对交通流数据的周期信息也有积极的影响。Han 等<sup>[140]</sup>分别从空间特征和时间特征两方面考虑,提出了一种并行的高速公路交通流短期预测的时空深度学习网络,该模型利用卷积神经网络提取交通流的空间特征,利用长短时记忆提取交通流的时间特征,反映了较强的交通流预测性能。以上海市内环高架路为例,对 591 个传感器进行了 6 个月的预测。试验结果表明,并行时空深度学习网络的总体性能优于其他先进的方法。针对交通流明显的时空相关性,Zhang 等<sup>[141]</sup>提出了一种基于时空分析和 CNN 深度学习模型的短期交通流预测方法,将交通流预测方法分为数据预处理和预测模型 2 个阶段,在数据预处理阶段,将交通流数据转换为包含交通时空特征的矩阵形式,并使用时空特征选择算法确定最佳输入数据。在预测模型阶段利用 CNN 提取交通流的时空特征并基于所选的最佳输入矩阵数据进行预测。将该方法与六种基准算法进行了比较。利用美国西雅图 I-5 高速公路的交通流量数据进行试验,所提出的模型

在不同的预测间隔下获得了良好的预测性能。

注意力模型和卷积神经网络的结合被广泛应用于交通流预测。Guo 等<sup>[142]</sup>针对现有的大多数交通流预测方法缺乏对交通数据时空相关性建模的能力,提出了一种基于注意力的时空图卷积网络(Attention based Spatial-Temporal Graph Convolution Network, ASTGCN)模型,应用空间注意对不同位置之间的复杂空间相关性进行建模,应用时间注意以捕获不同时间之间的动态时间相关性。为建立交通数据的时空相关性模型,设计了一种新的时空卷积模型,它包括获取空间特征的图形卷积和描述邻近时间片依赖关系的时间维度卷积。该模型可以直接对原始的基于图形的交通网络上的交通数据进行处理,有效地捕捉到交通网络的动态时空特征。在 Caltrans 性能测量系统的 2 个真实数据集上进行了试验,与现有基准算法相比,所提出的 ASTGCN 模型获得了更好的预测性能。为了提高交通流预测的精度,Wu 等<sup>[143]</sup>提出了一种基于深度神经网络的交通流预测模型。该模型充分利用了交通流的周期性和时空特性,受机器学习的启发,引入注意力机制,自主学习以确定过去交通流的重要性。利用卷积神经网络挖掘交通流的空间特征,利用递归神经网络挖掘交通流的时间特征,利用 PeMS 数据在长期水平预测任务中验证,所提出的模型优于现有的

基准方法。

尽管基于卷积神经网络的方法具有较高的准确率,但它在交通流预测中的应用还需要进一步的研究。如何在计算资源使用和精度性能之间保持平衡一直是一个难题,因此,用于大规模交通流预测的高级深度学习架构值得深入研究。

## 2.5 其他应用

现代交通运输行业发展迅速,带来的交通治理、道路交通安全的问题较为突出。交通事故的发生严重威胁着人类生命财产安全,但是车载网络传输慢,现有算法复杂度高造成计算时延,都极大的影响着预测交通风险的实时性和准确性。赵海涛等<sup>[144]</sup>提出一种基于深度学习的车联边缘网络交通事故风险预测算法,首先利用边缘服务器中建立的卷积神经网络将车载自组织网络中采集的大量交通数据进行多维特征提取,并对特征进行归一化、去均值等预处理得到新的变量,再将新变量通过卷积核下采样操作,最终在全连接层输出判别值,以此来预测交通事故发生的风险,通过试验可得,该算法具有较好的交通事故预测能力,相较于传统的机器学习算法,具有更高的预测精度和更低的损失。

## 2.6 卷积神经网络与 ITS 常用算法的比较

卷积神经网络与 ITS 常用的其他 5 种算法的优势、劣势和应用方向的对比,如表 5 所示。

表 5 卷积神经网络与其他 ITS 常用算法对比

Tab. 5 Comparison among convolutional neural network and other ITS common algorithms

算法名称	算法优势	算法劣势	应用方向
卷积神经网络(CNN)	自主学习进行特征提取、分类,识别率高	训练数据集庞大、对硬件要求较高	交通标志识别、车牌识别、车型识别、交通事件检测、交通状态预测
支持向量机(SVM)	可以解决高维数据和非线性问题	处理多分类问题以及数据量较大时效果不佳、核函数参数确定困难	交通标志识别
差分整合移动平均回归模型(Autoregressive Integrated Moving Average Model, ARIMA)	模型体量较小,便于计算	时间序列需要稳定、非线性问题预测性能差、对数据敏感、数据不可缺失稀疏	客流量预测等回归问题
卡尔曼滤波(Kalman Filtering)	非线性问题处理性能佳、精度高	计算量大、模型复杂、易受噪声数据干扰	客流量预测、交通标志识别
误差反向传播神经网络(Back Propagation Feed-Forward Neural Network, BPNN)	模型收敛较快,数据的内部规律不需深入了解	随机初始化权值参数、易陷入局部最优	客流量预测、交通状态判别
长短期记忆网络(Long Short-Term Memory Networks, LSTM)	很好地利用了数据中心的时间特征、鲁棒性较强	训练数据集庞大、对数据敏感、数据不可缺失稀疏、结构复杂、收敛速度较慢	客流量预测等回归问题

由表 5 可以看出,CNN、SVM、Kalman Filtering 在交通标志、车牌、车型的识别与检测方面应用较

多,传统的识别、检测方法提取的特征存在局限性,产生候选区域的方法需要大量的计算开销,检测精



度和速度很难达到实际应用的要求,CNN通过自主学习来提取特征,因此所提特征更加全面,虽然在训练时对硬件和数据集规模要求较高,但随着技术的发展,这些问题已逐渐缓解。在流量预测方面,CNN、ARIMA、Kalman Filtering、BPNN 和 LSTM 应用较多,CNN 被用来提取交通流的空间特征,通常和 LSTM 相结合,通过时空特征信息的相关性实现短时车流量预测。ARIMA 和 Kalman Filtering 实现简单,模型体量较小,但对于数据的时间信息较为敏感。

### 3 结 语

智能交通在中国的引入时间虽然不长,但由于其有效地解决了很多交通问题,越来越受到人们的重视。在实际道路交通环境中,具有许多不确定性与细节性,车辆行驶过程中车速较快、车辆的颜色和形状各不相同、车流量较大时互相遮挡使得车辆特征不明显等自身因素的影响,以及容易受到光照(天气)强度变化、道路交通背景多样复杂、拍摄角度变化等外界因素的影响,增加了智能计算的难度。近年来 CNN 凭借其优良特性受到了研究者的关注,尤其是在图像分类、目标检测等领域,它利用卷积层、池化层、激活函数等基本模块就可以实现对图像端到端的学习,避免了传统方法的手工提取特征,且对大数据的处理是其他传统算法无法相比的。虽然 CNN 技术已经很大程度上提高了车辆、车牌、交通标志等领域的识别准确度与速度,但距离实现鲁棒、精准的识别和无人驾驶的普遍应用,仍具有一定的距离,未来智能交通随着 CNN 算法的改进、高效算力以及基准数据集等多方面的推进,将取得新的突破。

(1)网络结构的进一步优化。从 AlexNet、VGG、GoogleNet、ResNet 以及 DensNet 来看,一个典型的发展趋势是网络深度越来越深,网络深度加深可提高网络性能,但这意味着需要训练的参数也在增加,虽然相比以前,计算量降低了很多,但网络仍是大规模的,需要大量的存储空间与计算时间,这会限制其实际应用,使得它们无法部署在资源有限的移动平台上,因此,需要进一步优化网络结构,尽可能的在降低网络计算量的同时提高其性能。此外,对于一个实际特定问题,如何设计一个更有效的网络结构仍是一个问题,目前只能是根据数据反复试验来获取。因此需要更详细的知识来确定合适的网络层数、每一层所使用的神经元数,以及训练时合理参数(如学习率等)的选择。

(2)开发并行训练算法。道路交通对于实时性的要求较高,随着网络日趋复杂和数据集规模逐渐增大,训练一个网络所需要的代价越来越高,因此,开发并行训练算法成为提升其应用时效性的重要方法。深度神经网络目前主要从模型并行和数据并行两方面实现并行化,模型并行是指将网络模型分配到多台设备上,通过设备之间的共同协作来完成训练任务,数据并行是指将训练数据分到多个模型中去训练,最后通过参数服务器完成参数交换<sup>[145]</sup>。为了加快训练过程,已经有一些异步 SGD 算法通过使用 CPU 和 GPU 集群得到好的结果,并且深度神经网络开源框架也都在并行化上做了相似工作,网络在 CPU 上训练时采用了高性能多线程库,在 GPU 上都支持 cuDNN<sup>[146]</sup>库。随着异构计算平台的快速发展以及有效且可扩展的并行训练算法的开发,深度神经网络将会应用到更多的领域中。

(3)无监督学习。现在应用较为广泛的都是有监督的学习,使卷积神经网络十分依赖于交通数据集,随着卷积神经网络的结构不断改进,以前的交通数据集已经不足以满足目前的需求,需要进一步从数据集数量、类别以及形式着手,研究相关数据集。在实际应用中,有标注的数据要远少于无标注的数据,且标注大量的数据需要耗费较大的人力。因此,未来的发展趋势将可能是无监督训练深度卷积神经网络。

未来智能交通随着不同领域的交叉融合将会得到更大的发展。卷积神经网络在视觉领域展现的强大能力,将使得智能交通的研究成果能够全面服务于大众,提高道路的智能信息化水平。

### 参考文献:

#### References :

- [1] 周飞燕,金林鹏,董 军. 卷积神经网络研究综述[J]. 计算机学报,2017,40(6):1229-1251.  
ZHOU Fei-yan, JIN Lin-peng, DONG Jun. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251. (in Chinese)
- [2] MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity[J]. Bulletin of Mathematical Biophysics, 1943, 5: 115-133.
- [3] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological Review, 1958, 65(6): 386-408.
- [4] WIDROW B, HOFF M E. Associative storage and retrieval of digital information in networks of adaptive "neurons"[J]. Biological Prototypes and Synthetic Systems, 1962: 160-166.
- [5] MINSKY M L, PAPERT S A. Perceptrons[M]. Cambridge:

- MIT Press, 1969.
- [6] KOHONEN T. Self-organized formation of topologically correct feature maps[J]. *Biological Cybernetics*, 1982, 43: 59-69.
- [7] CARPENTER G A, GROSSBERG S. The ART of adaptive pattern recognition by a self-organizing neural network[J]. *IEEE Computer*, 1988, 21(3): 77-88.
- [8] HOPFIELD J J. Neural networks and physical systems with emergent collective computational abilities[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1982, 79(8): 2554-2558.
- [9] HOPFIELD J J. Neurons with graded response have collective computational properties like those of two-state neurons[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1984, 81(10): 3088-3092.
- [10] HOPFIELD J J, TANK D W. "Neural" computation of decisions in optimization problems[J]. *Biological Cybernetics*, 1985, 52: 141-152.
- [11] HOPFIELD J J, TANK D. Computing with neural circuits: a model[J]. *Science*, 1986, 233(4764): 625-633.
- [12] HINTON G E, SEJNOWSKI T J. Optimal perceptual inference[C]//IEEE. 2007 IEEE International Conference on Image Processing. New York: IEEE, 1983: 448-453.
- [13] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. *Nature*, 1986, 323: 533-536.
- [14] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. *Science*, 2006, 313(5786): 504-507.
- [15] BENGIO Y, LAMBLIN P, DAN P, et al. Greedy layer-wise training of deep networks[C]//NeurIPS. 20th Annual Conference on Neural Information Processing Systems. San Diego: NeurIPS, 2006: 153-160.
- [16] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[J]. *Journal of Machine Learning Research*, 2010, 11: 3371-3408.
- [17] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [18] 张晓男, 钟 兴, 朱瑞飞, 等. 基于集成卷积神经网络的遥感影像场景分类[J]. *光学学报*, 2018, 38(11): 1128001.
- ZHANG Xiao-nan, ZHONG Xing, ZHU Rui-fei, et al. Scene classification of remote sensing images based on integrated convolutional neural networks[J]. *Acta Optica Sinica*, 2018, 38(11): 1128001. (in Chinese)
- [19] LECUN Y L, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [20] DENG Jia, DONG Wei, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//IEEE. 2009 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2009: 248-255.
- [21] 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: 106-154.
- [22] 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: 193-202.
- [23] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//Springer. 13th European Conference on Computer Vision. Berlin: Springer, 2014: 818-833.
- [24] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//ICLR. 3rd International Conference on Learning Representations. La Jolla: ICLR, 2015: 1-14.
- [25] SZEGEDY C, LIU Wei, JIA Yang-qing, et al. Going deeper with convolutions[C]//IEEE. 2015 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2015: 1-9.
- [26] HE Kai-ming, ZHANG Xiang-yu, REN Shao-qing, et al. Deep residual learning for image recognition[C]//IEEE. 2016 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770-778.
- [27] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 2261-2269.
- [28] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//ICML. 32nd International Conference on Machine Learning. San Diego: ICML, 2015: 448-456.
- [29] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//IEEE. 2016 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 2818-2826.
- [30] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]//AAAI. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2017: 4278-4284.
- [31] DAI Ji-feng, QI Hao-zhi, XIONG Yue-wen, et al. Deformable convolutional networks[C]//IEEE. 2017 IEEE International Conference on Computer Vision. New York: IEEE, 2017: 764-773.
- [32] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640-651.
- [33] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 1800-1807.
- [34] ZHANG Xiang-yu, ZHOU Xin-yu, LIN Meng-xiao, et al.

- ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//IEEE. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 6848-6856.
- [35] HU Jie, SHEN Lin, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [36] 高鑫,李慧,张义,等.基于可变形卷积神经网络的遥感影像密集区域车辆检测方法[J].电子与信息学报,2018,40(12): 2812-2819.
- GAO Xin, LI Hui, ZHANG Yi, et al. Vehicle detection in remote sensing images of dense areas based on deformable convolution neural network[J]. Journal of Electronics and Information Technology, 2018, 40(12): 2812-2819. (in Chinese)
- [37] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[C]//ICLR. 4th International Conference on Learning Representations. La Jolla: ICLR, 2016: 1-13.
- [38] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[C]//ICLR. 3rd International Conference on Learning Representations. La Jolla: ICLR, 2015: 1-14.
- [39] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [40] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. ArXiv E-Print, 2017, DOI: arXiv:1706.05587.
- [41] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Springer. 15th European Conference on Computer Vision. Berlin: Springer, 2018: 833-851.
- [42] KAIMING H, GEORGIA G, PIOTR D, et al. Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 386-397.
- [43] REN S, HE K, 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, 39(6): 1137-1149.
- [44] LIU Shu, QI Lu, QIN Hai-fang, et al. Path aggregation network for instance segmentation[C]//IEEE. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 8759-8768.
- [45] HUANG Zhao-jin, HUANG Li-chao, GONG Yong-chao, et al. Mask scoring R-CNN[C]//IEEE. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 6402-6411.
- [46] SARIGUL M, OZYILDIRIM B M, AVCI M. Differential convolutional neural network[J]. Neural Networks, 2019, 116: 279-287.
- [47] ZEILER M D, FERGUS R. Stochastic pooling for regularization of deep convolutional neural networks[C]//ICLR. 1st International Conference on Learning Representations. La Jolla: ICLR, 2013: 1-9.
- [48] FEI Jian-chao, FANG Hu-sheng, YIN Qin, et al. Restricted stochastic pooling for convolutional neural network[C]//ACM. 10th International Conference on Internet Multimedia Computing and Service. New York: ACM, 2018: 1-4.
- [49] AKHTAR N, RAGAVENDRAN U. Interpretation of intelligence in CNN-pooling processes: a methodological survey[J]. Neural Computing and Application, 2020, 32(3): 879-898.
- [50] YU D, WANG H, CHEN P, et al. Mixed pooling for convolutional neural networks[C]//Springer. 9th International Conference on Rough Sets and Knowledge Technology. Berlin: Springer, 2014: 364-375.
- [51] LIN Min, CHEN Qiang, YAN Shui-cheng. Network in network[C]//ICLR. 2nd International Conference on Learning Representations. La Jolla: ICLR, 2014: 1-10.
- [52] SUN Man-li, SONG Zhan-jie, JIANG Xiao-heng, et al. Learning pooling for convolutional neural network[J]. Neurocomputing, 2017, 224: 96-104.
- [53] HE Kai-ming, ZHANG Xiang-yu, REN Shao-qing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [54] CUI Yin, ZHOU Feng, WANG Jiang, et al. Kernel pooling for convolutional neural networks[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 3049-3058.
- [55] CHEN Jun-feng, HUA Zhou-dong, WANG Jing-yu, et al. A convolutional neural network with dynamic correlation pooling[C]//IEEE. 13th International Conference on Computational Intelligence and Security (CIS). New York: IEEE, 2017: 496-499.
- [56] ZHAO Qi, LYU Shu-chang, ZHANG Bo-xue, et al. Multiactivation pooling method in convolutional neural networks for image recognition[J]. Wireless Communications and Mobile Computing, 2018, 2018: 8196906.
- [57] ZHANG Jian-ming, HUANG Qian-qian, WU Hong-lin, et al. A shallow network with combined pooling for fast traffic sign recognition[J]. Information, 2017, 8(2): 45.
- [58] SAEEDAN F, WEBER N, GOESELE M, et al. Detail-preserving pooling in deep networks[C]//IEEE. 2018 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 9108-9116.
- [59] QI Kun-lun, GUAN Qing-feng, YANG Chao, et al. Concentric circle pooling in deep convolutional networks for remote sensing scene classification[J]. Remote Sensing, 2018, 10(6): 934.
- [60] LONG Yang, ZHU Fan, SHAO Ling, et al. Face recognition with a small occluded training set using spatial and statistical pooling[J]. Information Sciences, 2018, 430/431: 634-644.



- [61] WANG Feng, HUANG Si-wei, SHI Lei, et al. The application of series multi-pooling convolutional neural networks for medical image segmentation [J]. *International Journal of Distributed Sensor Networks*, 2017, 13(12): 1-10.
- [62] ZHI Tian-cheng, DUAN Ling-yu, WANG Yi-tong, et al. Two-stage pooling of deep convolutional features for image retrieval[C]//IEEE. 23rd IEEE International Conference on Image Processing. New York: IEEE, 2016: 2465-2469.
- [63] SADIGH S, SEN P. Improving the resolution of CNN feature maps efficiently with multisampling [J]. *ArXiv E-Print*, 2018, DOI: arXiv:1805.10766.
- [64] TAKEKI A, IKAMI D, IRIE G, et al. Parallel grid pooling for data augmentation [J]. *ArXiv E-Print*, 2018, DOI: arXiv:1803.11370.
- [65] SHAHRIARI A, PORIKLI F. Multipartite pooling for deep convolutional neural networks [J]. *ArXiv E-Print*, 2017, DOI: arXiv:1710.07435.
- [66] KUMAR A. Ordinal pooling networks : for preserving information over shrinking feature maps[J]. *ArXiv E-Print*, 2018, DOI: arXiv:1804.02702.
- [67] KOLESNIKOV A, LAMPERT C H. Seed, expand and constrain: three principles for weakly- supervised image segmentation[C]//Springer. 21st ACM Conference on Computer and Communications Security. Berlin: Springer, 2016: 695-711.
- [68] SHI Zeng-lin, YE Yang-ding, WU Yun-peng. Rank-based pooling for deep convolutional neural networks[J]. *Neural Networks*, 2016, 83: 21-31.
- [69] ZHAI Shuang-fei, WU Hui, KUMAR A, et al. S3Pool: pooling with stochastic spatial sampling[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 4003-4011.
- [70] TONG Zhi-qiang, AIHARA K, TANAKA G. A hybrid pooling method for convolutional neural networks[C]//Springer. International Conference on Neural Information Processing. Berlin: Springer, 2016: 454-461.
- [71] TURAGA S C, MURRAY J F, JAIN V, et al. Convolutional networks can learn to generate affinity graphs for image segmentation[J]. *Neural Computation*, 2010, 22(2): 511-538.
- [72] WU Hai-bing, GU Xiao-dong. Max-pooling dropout for regularization of convolutional neural networks[C]//Springer. 22nd International Conference on Neural Information Processing. Berlin: Springer, 2015: 46-54.
- [73] SONG Zhen-hua, LIU Yan, SONG Rong, et al. A sparsity-based stochastic pooling mechanism for deep convolutional neural networks[J]. *Neural Networks*, 2018, 105: 340-345.
- [74] WANG P, LI W, GAO Z, et al. Depth pooling based large-scale 3D action recognition with convolutional neural networks[J]. *IEEE Transactions on Multimedia*, 2018, 20(5): 1051-1061.
- [75] RIPPEL O, SNOEK J, ADAMS R P. Spectral representations for convolutional neural networks[C]//NeurIPS. 29th Annual Conference on Neural Information Processing Systems. San Diego: NeurIPS, 2015: 2449-2457.
- [76] WILLIAMS T, LI R. Wavelet pooling for convolutional neural networks[C]//ICLR. 6th International Conference on Learning Representations. La Jolla: ICLR, 2018: 1-12.
- [77] SAINATH T N, KINGSBURY B, MOHAMED A R, et al. Improvements to deep convolutional neural networks for LVCSR[C]//IEEE. 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. New York: IEEE, 2013: 315-320.
- [78] 白 琮, 黄 玲, 陈佳楠, 等. 面向大规模图像分类的深度卷积神经网络优化[J]. *软件学报*, 2018, 29(4): 1029-1038.
- BAI Cong, HUANG Ling, CHEN Jia-nan, et al. Optimization of deep convolutional neural network for large scale image classification[J]. *Journal of Software*, 2018, 29(4): 1029-1038. (in Chinese)
- [79] EOM H, CHOI H. Alpha—integration pooling for convolutional neural networks [J]. *ArXiv E-Print*, 2018, DOI: arXiv: 1811.03436.
- [80] 刘万军, 梁雪剑, 曲海成. 不同池化模型的卷积神经网络学习性能研究[J]. *中国图象图形学报*, 2016, 21(9): 1178-1190.
- LIU Wan-jun, LIANG Xue-jian, QU Hai-cheng. Learning performance of convolutional neural networks with different pooling models[J]. *Journal of Image and Graphics*, 2016, 21(9): 1178-1190. (in Chinese)
- [81] ZHANG Bo-xue, ZHAO Qi, FENG Wen-quan, et al. AlphaMEX: a smarter global pooling method for convolutional neural networks[J]. *Neurocomputing*, 2018, 321: 36-48.
- [82] JOSE A, LOPEZ R D, HEISTERKLAUS I, et al. Pyramid pooling of convolutional feature maps for image retrieval[C]//IEEE. 25th IEEE International Conference on Image Processing. New York: IEEE, 2018: 480-484.
- [83] WAIBEL A, HANAZAWA T, HINTON G, et al. Phoneme recognition using time-delay neural networks[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1989, 37(3): 328-339.
- [84] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [85] WANG Ze-long, LAN Qiang, HUANG Da-fei, et al. Combining FFT and spectral-pooling for efficient convolution neural network model[C]//Advances in Intelligent Systems Research. 2nd International Conference on Artificial Intelligence and Industrial Engineering. Paris: Atlantis Press, 2016: 203-206.
- [86] SMITH J S, WILAMOWSKI B M. Discrete cosine transform spectral pooling layers for convolutional neural networks[C]//Springer. 17th International Conference on Artificial Intelligence and Soft Computing. Berlin: Springer, 2018: 235-246.
- [87] NAIR V, HINTON G E. Rectified linear units improve restricted Boltzmann machines [C]//ICML. 27th International Conference on Machine Learning. San Diego: ICML, 2010: 807-814.
- [88] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[J]. *Journal of Machine Learning Research*,

- 2011, 15: 315-323.
- [89] LI Y, DING P, LI B. Training neural networks by using power linear units (PoLUs)[J]. ArXiv E-Print, 2018, DOI: arXiv:1802.00212.
- [90] DOLEZEL P, SKRABABEK P, GAGO L. Weight initialization possibilities for feed forward neural network with linear saturated activation functions[J]. IFAC—Papers on Line, 2016, 49(25): 49-54.
- [91] GOODFELLOW I J, WARDE-FARLEY D, MIRZA M, et al. Maxout networks[C]//ICML. 30th International Conference on Machine Learning. San Diego: ICML, 2013: 2356-2364.
- [92] CASTANEDA G, MORRIS P, KHOSHGOFTAAR T M. Evaluation of maxout activations in deep learning across several big data domains[J]. Journal of Big Data, 2019, DOI: 10.1186/s40537-019-0233-0.
- [93] CLEVERT D A, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by exponential linear units (ELUs)[C]//ICLR. 4th International Conference on Learning Representations. La Jolla: ICLR, 2016: 1-14.
- [94] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models[C]//ACM. 30th International Conference on Machine Learning. New York: ACM, 2013: 456-462.
- [95] HE Kai-ming, ZHANG Xiang-yu, REN Shao-qing, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification[C]//IEEE. 15th IEEE International Conference on Computer Vision. New York: IEEE, 2015: 1026-1034.
- [96] KLAMBAUER G, UNTERTHINER T, MAYR A, et al. Self-normalizing neural networks[C]//NeurIPS. 31st Annual Conference on Neural Information Processing Systems. San Diego: NeurIPS, 2017: 972-981.
- [97] 杨观赐, 杨静, 李少波, 等. 基于 Dropout 与 ADAM 优化器的改进 CNN 算法[J]. 华中科技大学学报(自然科学版), 2018, 46(7): 122-127.
- YANG Guan-ci, YANG Jing, LI Shao-bo, et al. Modified CNN algorithm based on Dropout and ADAM optimizer[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2018, 46(7): 122-127. (in Chinese)
- [98] KINGMA D P, BA J L. Adam: a method for stochastic optimization[C]//ICLR. 3rd International Conference on Learning Representations. La Jolla: ICLR, 2015: 1-15.
- [99] ROBBINS H, MONRO S. A stochastic approximation method[J]. The Annals of Mathematical Statistics, 1951, 22(3): 400-407.
- [100] 石琪. 基于卷积神经网络图像分类优化算法的研究与验证[D]. 北京: 北京交通大学, 2017.
- SHI Qi. Research and verification of image classification optimization algorithm based on convolutional neural network[D]. Beijing: Beijing Jiaotong University, 2017. (in Chinese)
- [101] 王红霞, 周家奇, 辜承昊, 等. 用于图像分类的卷积神经网络中激活函数的设计[J]. 浙江大学学报(工学版), 2019, 53(7): 1363-1373.
- WANG Hong-xia, ZHOU Jia-qi, GU Cheng-hao, et al. Design of activation function in CNN for image classification[J]. Journal of Zhejiang University (Engineering Science), 2019, 53(7): 1363-1373. (in Chinese)
- [102] POLYAK B T. Some methods of speeding up the convergence of iteration methods[J]. USSR Computational Mathematics and Mathematical Physics, 1964, 4(5): 1-17.
- [103] SUTSKEVER I, MARTENS J, DAHL G, et al. On the importance of initialization and momentum in deep learning[C]//ACM. 30th International Conference on Machine Learning. New York: ACM, 2013: 2176-2184.
- [104] DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12: 2121-2159.
- [105] ZEILER M D. Adadelta: an adaptive learning rate method[J]. ArXiv E-Print, 2012, DOI: arXiv:1212.5701.
- [106] JI Shi-hao, VISHWANATHAN S V N, SATISH N, et al. BlackOut: speeding up recurrent neural network language models with very large vocabularies[C]//ICLR. 4th International Conference on Learning Representations. La Jolla: ICLR, 2016: 1-4.
- [107] LOUIZOS C, WELLING M, KINGMA D P. Learning sparse neural networks through  $L_0$  regularization[C]//ICLR. 6th International Conference on Learning Representations. La Jolla: ICLR, 2018: 1-13.
- [108] DOZAT T. Incorporating nesterov momentum into adam[C]//ICLR. 4th International Conference on Learning Representations. La Jolla: ICLR, 2016: 1-4.
- [109] LUO Liang-chao, XIONG Yuan-hao, LIU Yan, et al. Adaptive gradient methods with dynamic bound of learning rate[C]//ICLR. 7th International Conference on Learning Representations. La Jolla: ICLR, 2019: 1-19.
- [110] WANG Di, TIAN Yu-min, GENG Wen-hui, et al. LPR-Net: recognizing Chinese license plate in complex environments[J]. Pattern Recognition Letters, 2020, 130: 148-156.
- [111] 李祥鹏, 闵卫东, 韩清, 等. 基于深度学习的车牌定位和识别方法[J]. 计算机辅助设计与图形学学报, 2019, 31(6): 979-987.
- LI Xiang-peng, MIN Wei-dong, HAN Qing, et al. License plate location and recognition based on deep learning[J]. Journal of Computer-Aided Design and Computer Graphics, 2019, 31(6): 979-987. (in Chinese)
- [112] LIN Hui, WANG Peng, YOU Chun-hua, et al. Reading car license plates using deep neural networks[J]. Image and Vision Computing, 2018, 72: 14-23.
- [113] XIANG Han, ZHAO Yong, YUAN Yu-le, et al. Lightweight fully convolutional network for license plate detection[J]. Optik, 2019, 178: 1185-1194.
- [114] ASIF M R, QI Chun, WANG Tie-xiang, et al. License plate detection for multi-national vehicles: an illumination invariant approach in multi-lane environment[J]. Computers and Electrical Engineering, 2019, 78: 132-147.

- [115] PUARUNGROJ W, BOONSIRISUMPUN N. Thai license plate recognition based on deep learning[J]. Procedia Computer Science, 2018, 135: 214-221.
- [116] CAO Yu, FU Hui-yuan, MA Hua-dong. An end-to-end neural network for multi-line license plate recognition[C]// IEEE. 24th International Conference on Pattern Recognition. New York: IEEE, 2018: 3698-3703.
- [117] 赵汉理, 刘俊如, 姜 磊, 等. 基于卷积神经网络的双行车牌分割算法[J]. 计算机辅助设计与图形学学报, 2019, 31(8): 1320-1329.
- ZHAO Han-li, LIU Jun-ru, JIANG Lei, et al. Double-row license plate segmentation with convolutional neural networks[J]. Journal of Computer-Aided Design and Computer Graphics, 2019, 31(8): 1320-1329. (in Chinese)
- [118] 张秀玲, 张逞逞, 周凯旋. 基于感兴趣区域的 CNN-Squeeze 交通标志图像识别[J]. 交通运输系统工程与信息, 2019, 19(3): 48-53.
- ZHANG Xiu-ling, ZHANG Cheng-cheng, ZHOU Kai-xuan. Traffic sign image recognition via CNN-Squeeze based on region of interest [J]. Journal of Transportation Systems Engineering and Information Technology, 2019, 19(3): 48-53. (in Chinese)
- [119] 王方石, 王 坚, 李 兵, 等. 基于深度属性学习的交通标志检测[J]. 吉林大学学报(工学版), 2018, 48(1): 319-329.
- WANG Fang-shi, WANG Jian, LI Bing, et al. Deep attribute learning based traffic sign detection[J]. Journal of Jilin University (Engineering and Technology Edition), 2018, 48(1): 319-329. (in Chinese)
- [120] 李旭东, 张建国, 谢志鹏, 等. 基于三尺度嵌套残差结构的交通标志快速检测算法[J]. 计算机研究与发展, 2020, 57(5): 1022-1036.
- LI Xu-dong, ZHANG Jian-ming, XIE Zhi-peng, et al. A fast traffic sign detection algorithm based on three-scale nested residual structures[J]. Journal of Computer Research and Development, 2020, 57(5): 1022-1036. (in Chinese)
- [121] 宋青松, 张 超, 田正鑫, 等. 基于多尺度卷积神经网络的交通标志识别[J]. 湖南大学学报(自然科学版), 2018, 45(8): 131-137.
- SONG Qing-song, ZHANG Chao, TIAN Zheng-xin, et al. Traffic sign recognition based on multi-scale convolutional neural network[J]. Journal of Hunan University (Natural Sciences), 2018, 45(8): 131-137. (in Chinese)
- [122] 孙 伟, 杜宏吉, 张小瑞, 等. 基于 CNN 多层特征和 ELM 的交通标志识别[J]. 电子科技大学学报, 2018, 47(3): 343-349.
- SUN Wei, DU Hong-ji, ZHANG Xiao-rui, et al. Traffic sign recognition method based on multi-layer feature CNN and extreme learning machine[J]. Journal of University of Electronic Science and Technology of China, 2018, 47(3): 343-349. (in Chinese)
- [123] 张淑芳, 朱 彤. 基于残差单发多框检测器模型的交通标志检测与识别[J]. 浙江大学学报(工学版), 2019, 53(5): 940-949.
- ZHANG Shu-fang, ZHU Tong. Traffic sign detection and recognition based on residual single shot multibox detector model[J]. Journal of Zhejiang University (Engineering Science), 2019, 53(5): 940-949. (in Chinese)
- [124] LIU Zhi-gang, LI Dong-yu, GE Shu-zhi, et al. Small traffic sign detection from large image[J]. Applied Intelligence, 2020, 50(1): 1-13.
- [125] 张建明, 王 伟, 陆朝铨, 等. 基于压缩卷积神经网络的交通标志分类算法[J]. 华中科技大学学报(自然科学版), 2019, 47(1): 103-108.
- ZHANG Jian-ming, WANG Wei, LU Chao-quan, et al. Traffic sign classification algorithm based on compressed convolutional neural network[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2019, 47(1): 103-108. (in Chinese)
- [126] SONG Shi-jin, QUE Zhi-qiang, HOU Jun-jie, et al. An efficient convolutional neural network for small traffic sign detection[J]. Journal of Systems Architecture, 2019, 97: 269-277.
- [127] ZHANG Qiang, ZHOU Li, LI Jia-feng, et al. Vehicle color recognition using multiple-layer feature representations of lightweight convolutional neural network[J]. Signal Processing, 2018, 147(7): 146-153.
- [128] FU Hui-yuan, MA Hua-dong, WANG Gao-ya, et al. MCFF-CNN: multiscale comprehensive feature fusion convolutional neural network for vehicle color recognition based on residual learning[J]. Neurocomputing, 2020, 395: 178-187.
- [129] LI Su-hao, LIN Jin-zhao, LI Guo-quan, et al. Vehicle type detection based on deep learning in traffic scene[J]. Procedia Computer Science, 2018, 131: 564-572.
- [130] HU Bin, LAI Jian-huang, GUO Chun-chao. Location-aware fine-grained vehicle type recognition using multi-task deep networks[J]. Neurocomputing, 2017, 243: 60-68.
- [131] XIANG Ye, FU Ying, HUANG Hua. Global relative position space based pooling for fine-grained vehicle recognition [J]. Neurocomputing, 2019, 367: 287-298.
- [132] 余 焯, 傅云翔, 杨昌东, 等. 基于 FR-ResNet 的车辆型号精细识别研究[J]. 自动化学报, 2021, 47(5): 1125-1136.
- YU Ye, FU Yun-xiang, YANG Chang-dong, et al. Fine-grained car model recognition based on FR-ResNet[J]. Acta Automatica Sinica, 2021, 47(5): 1125-1136. (in Chinese)
- [133] 杨 娟, 曹浩宇, 汪荣贵, 等. 基于语义 DCNN 特征融合的细粒度车型识别模型[J]. 计算机辅助设计与图形学学报, 2019, 31(1): 141-157.
- YANG Juan, CAO Hao-yu, WANG Rong-gui, et al. Fine-grained car recognition model based on semantic DCNN features fusion[J]. Journal of Computer-Aided Design and Computer Graphics, 2019, 31(1): 141-157. (in Chinese)
- [134] 蒋行国, 万今朝, 蔡晓东, 等. 奇异值分解与中心度量的细粒度车型识别算法[J]. 西安电子科技大学学报, 2019, 46(3): 82-88.
- JIANG Xing-guo, WAN Jin-zhao, CAI Xiao-dong, et al. Algorithm for identification of fine-grained vehicles based on singular value decomposition and central metric[J]. Journal of Xidian University, 2019, 46(3): 82-88. (in Chinese)



- [135] 杨娟,曹浩宇,汪荣贵,等. 区域建议网络的细粒度车型识别[J]. 中国图象图形学报, 2018, 23(6): 837-845.  
YANG Juan, CAO Hao-yu, WANG Rong-gui, et al. Fine-grained car recognition method based on region proposal networks[J]. Journal of Image and Graphics, 2018, 23(6): 837-845. (in Chinese)
- [136] 罗文慧,董宝田,王泽胜. 基于CNN-SVR混合深度学习模型的短时交通流预测[J]. 交通运输系统工程与信息, 2017, 17(5): 68-74.  
LUO Wen-hui, DONG Bao-tian, WANG Ze-sheng. Short-term traffic flow prediction based on CNN-SVR hybrid deep learning model [J]. Journal of Transportation Systems Engineering and Information Technology, 2017, 17(5): 68-74. (in Chinese)
- [137] 石敏,蔡少委,易清明. 基于空洞-稠密网络的交通拥堵预测模型[J]. 上海交通大学学报, 2021, 55(2): 124-130.  
SHI Min, CAI Shao-wei, YI Qing-ming. A traffic congestion prediction model based on dilated-dense network [J]. Journal of Shanghai Jiaotong University, 2021, 55(2): 124-130. (in Chinese)
- [138] DENG Shao-jiang, JIA Shu-yuan, CHEN Jing. Exploring spatial-temporal relations via deep convolutional neural networks for traffic flow prediction with incomplete data[J]. Applied Soft Computing Journal, 2019, 78: 712-721.
- [139] AN Ji-yao, FU Li, HU Meng, et al. A novel fuzzy-based convolutional neural network method to traffic flow prediction with uncertain traffic accident information[J]. IEEE Access, 2018, 12: 2169-3536.
- [140] HAN Dong-xiao, CHEN Juan, SUN Jian. A parallel spatiotemporal deep learning network for highway traffic flow forecasting[J]. International Journal of Distributed Sensor Networks, 2019, 15(2): 1-12.
- [141] ZHANG Wei-bin, YU Ying-hao, QI Yong, et al. Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning[J]. Transportmetrica A: Transport Science, 2019, 15(2): 1688-1711.
- [142] GUO Sheng-nan, LIN You-fang, FENG Ning, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting[C]// AAAI. 33rd AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2019: 922-929.
- [143] WU Yuan-kai, TAN Hua-chun, QIN Ling-qiao, et al. A hybrid deep learning based traffic flow prediction method and its understanding[J]. Transportation Research Part C: Emerging Technologies, 2018, 90: 166-180.
- [144] 赵海涛,程慧玲,丁仪,等. 基于深度学习的车联边缘网络交通事故风险预测算法研究[J]. 电子与信息学报, 2020, 42(1): 50-57.  
ZHAO Hai-tao, CHENG Hui-ling, DING Yi, et al. Research on traffic accident risk prediction algorithm of edge internet of vehicles based on deep learning [J]. Journal of Electronics and Information Technology, 2020, 42(1): 50-57. (in Chinese)
- [145] 朱虎明,李佩,焦李成,等. 深度神经网络并行化研究综述[J]. 计算机学报, 2018, 41(8): 1861-1881.  
ZHU Hu-ming, LI Pei, JIAO Li-cheng, et al. Review of parallel deep neural network [J]. Chinese Journal of Computers, 2018, 41(8): 1861-1881. (in Chinese)
- [146] CHETLUR S, WOOLLEY C, VANDERMERSCH P, et al. cuDNN: efficient primitives for deep learning [J]. arXiv e-Print, 2012, DOI: arXiv:1410.0759.