目标检测算法在交通场景中应用综述

肖雨晴,杨慧敏
东北林业大学 工程技术学院,哈尔滨 150040

摘 要:目标检测是计算机视觉领域的重要研究任务,在机器人、自动驾驶、工业检测等方面应用广泛。在深度学习理论的基础上,系统性总结了目标检测算法的发展与研究现状,对两类算法的特点、优缺点和实时性进行对比。以交通场景中三类典型物体(非机动车、机动车和行人)为目标,从传统检测方法、目标检测算法、目标检测算法优化、三维目标检测、多模态目标检测和重识别六个方面分别论述和总结目标检测算法检测识别交通场景目标的研究现状与应用情况,重点介绍了各类方法的优势、局限性和适用场景。归纳了常用目标检测和交通场景数据集及评价标准,比较分析两类算法性能,展望目标检测算法在交通场景中应用研究的发展趋势,为智能交通、自动驾驶提供研究思路。

关键词:目标检测;深度学习;交通场景;计算机视觉;自动驾驶
文献标志码:A **中图分类号:**TP391 doi:10.3778/j.issn.1002-8331.2011-0361

Research on Application of Object Detection Algorithm in Traffic Scene

XIAO Yuqing, YANG Huimin
College of Engineering and Technology, Northeast Forestry University, Harbin 150040, China

Abstract: Object detection is an important research task in the field of computer vision. It is widely used in robotics, automatic vehicles, industrial detection and other fields. On the basis of deep learning theory, the development and research status of object detection algorithm are firstly systematically summarized and the characteristics, advantages, disadvantages and real-time performance of the two categories of algorithms are compared. Next to the three kinds of typical targets (non-motor vehicles, motor vehicles and pedestrians) as objects in the traffic scene, the research status and application of object detection algorithm for detecting and identifying objects are discussed and summarized respectively from six aspects in traffic scene: traditional detection method, object detection algorithm, object detection algorithm optimization, 3d object detection, multimodal object detection and re-identification. And the application of focus on the advantages, limitations and applicable scenario of various methods. Finally, the common object detection and traffic scene data sets and evaluation criteria are summarized, the performance of the two categories of algorithms is compared and analyzed, and the development trend of the application of object detection algorithm in traffic scenes is prospected, providing research ideas for intelligent traffic and automatic vehicles.

Key words: object detection; deep learning; traffic scene; computer vision; autonomous vehicles

目标检测是计算机视觉领域重要的研究分支,是目标识别、跟踪的基础环节,其主要研究内容是在图像中找出感兴趣目标,包括目标定位和分类。其中,交通场景目标检测识别是计算机视觉领域研究的热点问题,其目的是运用图像处理、模式识别、机器学习、深度学习等技术在交通场景中检测识别出车辆、行人等交通场景目标信息,达到智能交通、自动驾驶的目标。

传统目标检测方法通常分为三个阶段:首先在图像中选择一些候选区域,然后在候选区域中提取特征,最后采用训练的分类器进行识别分类。然而,该方法操作复杂,精确度不高且训练速度慢,误检率较高,在实际工程应用中不易实现。因此,在卷积神经网络快速发展的背景下,研究人员提出基于深度学习的目标检测算法,该方法实现了端到端检测识别,具有很好的实际意义。如今基于深度学习的目标检测算法已成为机器人导航、自动驾驶感知领域的主流算法。

1 目标检测算法综述

目标检测算法可以分为基于候选区域(两阶段)和基于回归(一阶段)两类。两者最大的区别是前者通过

基金项目:中央高校业务经费(2572016CB11)。
作者简介:肖雨晴(1997—),女,硕士研究生,研究领域为深度学习、图像处理,E-mail:m13628623707@163.com;杨慧敏(1980—),女,博士,高级工程师,研究领域为缺陷检测、图像处理。
收稿日期:2020-11-23 **修回日期:**2021-01-06 **文章编号:**1002-8331(2021)06-0030-12

子网络辅助生成候选边界框,而后者直接在特征图上生成候选边界框。目标检测算法分类如图1所示。

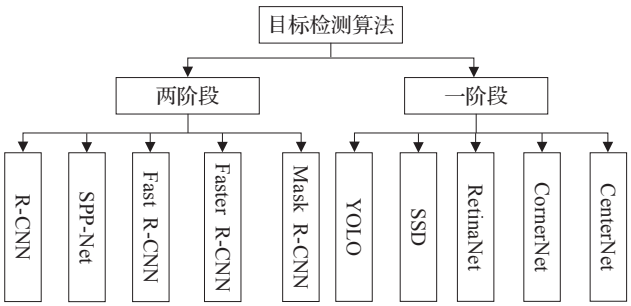

图1 目标检测算法分类

基于候选区域的算法源于2014年Girshick等提出的R-CNN[1],R-CNN首次将深度学习引入目标检测,在Pascal VOC数据集上的mAP值为66.0%。在此基础上,Faster R-CNN[2]、Mask R-CNN[3]等算法相继出现。基于回归的算法源于2016年Redmon等提出的YOLO[4]算法和Liu等提出的SSD[5]算法,该方法将检测转化为回归问题,大幅度提高了检测速度。在此基础上发展的算法包括YOLO v4[6]、RSSD[7]等。具体算法介绍如表1所示。

目标检测算法是近几年计算机视觉领域的热点研究方向,包括基于候选区域和基于回归两类。基于候选区域的算法检测速度普遍较慢,在交通场景中检测的实时性还不能满足,但检测精度在不断提升;基于回归的算法检测速度快、实时性较好,但是检测精度与准确度相对于两阶段的算法还是较差。目前随着研究的深入,各种目标检测算法被提出,未来算法的发展应更多研究检测速度与精度并行且轻量的目标检测算法。

2 目标检测算法在交通场景中的应用

随着城市建设的快速发展,城市人口越来越密集,交通需求量也不断上升,道路交通压力逐步增加。在交通压力增加的同时,道路阻塞、安全事故频发等问题严重影响了人们的出行和生命安全,因此需要将目标检测算法应用到交通场景中实现智能交通、自动驾驶,避免人员伤亡、财产损失。在交通场景中,需要检测的目标主要有非机动车、机动车及行人。

2.1 目标检测算法在非机动车识别的应用

快速、准确识别非机动车对车辆安全行驶具有重要作用,然而非机动车容易受光照强度、天气变化、遮挡等问题影响,这对自动驾驶应用产生了很大的安全风险。因此,在复杂的自然交通场景下,如何实现准确、实时检测识别非机动车是当前需要研究的问题。

2.1.1 传统非机动车识别方法

传统非机动车检测方法主要是人工提取图像中的颜色、形状等特征,然后通过支持向量机、Adaboost等分类器识别,具体传统检测算法比较如表2所示。

表1 目标检测算法比较分析

目标检测算法	实时性	优势	局限性	适用场景
R-CNN	否	首次将深度学习引入目标检测	获取目标区域费时,不共享特征	目标检测
SPP-Net	否	解决输入特征图尺寸不一致问题	各个检测步骤分离,仍需多次训练	目标检测
Fast R-CNN	否	使用感兴趣区域池化层(ROI Pooling layers)结构	使用外部算法来提取目标候选框,比较耗时	目标检测
Faster R-CNN	较差	真正完成端到端检测识别	模型复杂,小目标检测效果不佳	目标检测
R-FCN	较差	定位精度高	模型复杂,计算量大	目标检测、语义分割
Mask R-CNN	较差	分割准确,检测精度高	实例分割代价昂贵	目标检测、实例分割
YOLO v1	优秀	检测转化为回归问题,运行速度加快	产生更多定位误差和精度落后,泛化能力较弱	目标检测
SSD	优秀	结合回归与anchor机制	小目标特征丢失	多尺度目标检测
YOLO v2	优秀	速度进一步提升,召回率提高	小目标检测效果差	目标检测
RSSD	较好	小目标检测效果较好	模型复杂,检测速度一般	目标检测
YOLO v3	优秀	小目标检测精度提高	模型召回率低	多尺度目标检测
YOLO v4	优秀	融合各种调优技巧	检测模型大体未改变	高精度目标检测

表2 传统目标检测算法比较

方法	简单描述	优点	缺点	适用场景
SIFI	在空间寻找极值点,提取目标信息	独特性好,信息丰富;具有不变性和稳定性	计算量大;对目标边缘图像失效	图像匹配、三维建模
LBP	相邻与中心点像素比较,反映纹理信息	具有旋转不变性;运算速度快	对方向信息敏感	图像分类、行人检测
ORB	通过关键点创建特征向量	检测速度快;不受噪点和图像变换影响	不确定特征点较多	图像识别
VIOLA-JONES	积分图像+特征选择+级联检测	检测效率高	特征代表性不突出	图像分类、人脸检测
HOG-SVM	计算局部区域梯度方向直方图构建特征	具有不变性;对刚性物体检测效果好	特征维度大;计算量大	轮廓信息捕获、行人检测
DPM	使用多组件提取HOG特征	运算速度快;适应目标变形	性能一般;工作量大	人体姿态检测

1999年,Lowe等[8]提出SIFI(Scale Invariant Feature Transform)算法,通过将原图像与目标图像特征匹配获得关键点。SIFI算法对噪声、视角改变具有一定的鲁棒性,但复杂度高、检测速度慢,对模糊图像不敏感。2001年,Viola等[9]提出采用积分图的形式表现图像特征,之后采用级联Adaboost分类器选择代表性特征对人脸检测识别。该方法可以实现实时检测,但准确率一般、鲁棒性不足。2002年,Ojale等[10]提出LBP(Local Binary Patterns)纹理特征,该特征计算量小并且可以有效检测大量旋转和尺度不同的纹理信息,但稳定性较差。2005年,Dalal等[11]利用HOG(Histogram of Oriented Gradient)特征对行人检测,在INRIA行人数据集上实验表明该方法具有较高的检测识别率,尤其对道路行人有着特别突出的表现。2008年,Felzenszwalb等[12]提出DPM(Deformable Part-based Model)算法,该算法采用多组件策略,首先在不同分辨率上提取特征形成融合特征,然后采用SVM分类回归获得目标位置。DPM算法计算简单、运算速度快、适用变形目标,但特征是人为设计、工作量大,性能一般、稳定性差。在此基础上,Girshick等引入混合模型、负例挖掘、边界盒回归对DPM算法进行改进,加快了检测速度。2011年,Rublee等[13]提出ORB(Oriented FAST and Rotated BRIEF)算法,采用FAST算法检测特征点,然后利用BRIEF算法描述特征点,最后通过特征匹配获得目标。该方法计算速度快、占用内存小、具有很高的效率,但不具备尺度与旋转不变性且对噪声敏感。

传统非机动车检测方法计算量小、易实现,但人工设计的特征对目标的多样性没有很好的鲁棒性,常会出现窗口冗余等问题,并且在真实交通场景中非机动车会因遮挡、占用像素较少等因素导致难以提取特征,所以传统方法很难满足实际应用的需求。

2.1.2 目标检测算法识别方法

近几年随着深度学习的发展,目标检测算法成为非机动车检测识别的主流方法。目标检测算法由于对几何变换、形变等具有一定程度的不变性,有效克服了非机动车外观多变带来的检测识别困难,并且在样本中可自适应构建特征,避免了人工构建特征不全、遗漏等情况。

2006年,Hinton[14]首次提出深度学习概念,开启了深度研究的热潮。2012年,AlexNet[15]模型在ILSVRC分类比赛中获得冠军,在计算机视觉领域获得突破性成果。此后,深度学习研究不断深入。在卷积神经网络的基础上,目标检测算法也随之取得突破性进展。2014年,R-CNN算法提出将候选区域与CNN结合对目标检测识别。2016年,Faster R-CNN算法实现端到端识别。YOLO算法、SSD算法实现了速度的进一步加快。对于目标检测算法在非机动车检测识别上的应用,Ahmad等[16]采用Faster R-CNN为基础网路,利用SVM+MOG(背景提取)方法提取运动车辆信息。Chen等[17]提出混合深度卷积神经网络(HDNN)对卫星图像车辆目标检测,该算法将最后卷积层和池化层的映射分为多个可变感受野,获取可变尺度特征。叶佳林等[18]通过设计特征融合结构和采用GIOU损失函数改进YOLO v3,降低非机动车漏检率,提高定位准确度。曹伟等[19]采用多尺度融合SSD算法检测车辆目标,并利用Camshift跟踪和Kalman滤波算法实现了目标实时跟踪。更多目标检测算法在非机动车检测识别上的应用如表3所示。这些方法可以检测识别出非机动车目标,但在实际应用中需要大规模的数据集训练模型,并且由于非机动车体积小、行驶相对密集,检测识别的准确度和实时性还比较差。

很多研究已经表明,基于深度学习的目标检测算法比传统检测识别方法具有更好的检测效果,并在mAP值上有很好的体现。采用目标检测算法检测交通场景中的非机动车目标,可以避免传统人工提取特征的局限性,更加有效提取特征,准确检测出非机动车目标,但在实际应用中,非机动车体积小、常会相互遮挡,所以检测识别还有一些困难。

2.1.3 非机动车检测识别小结

目前,将目标检测算法应用在非机动车检测识别方面的研究不多。但近几年外卖、非接触配送行业快速发展,非机动车在交通场景中的占比越来越大,由于非机动车数量大、分布广,所以对非机动车目标检测识别存在一定难度。对于非机动车体积小、遮挡严重等问题,目标检测算法在非机动车方面的检测识别应主要研究小目标、多尺度等问题。

2.2 目标检测算法在机动车识别的应用

机动车主要指车辆,在交通场景中检测识别车辆目标的主要困难是算法的精度和实时性。在传统车辆检测方法中人工特征提取占主导地位,在特征提取阶段提取的特征图优劣极大地影响检测效果,存在一定局限性。

与传统检测方法不同,目标检测算法不需要人工提取特征,一定程度上解决了人工提取特征缺失、可移植

表3 目标检测算法在非机动车检测中的应用

文献	具体算法	优势	局限性	适用场景
[20]	Lightened-Alexnet	使用卷积神经网络提取特征	网络结构简单;识别精度低	图像识别、图像分类
[21]	Canny+CNN	特征算子与神经网络结合,检测效果提升	未实现端到端分类识别;计算量增加	图像识别
[22]	edageBoxes+Fast R-CNN	准确获得目标区域,识别准确率提高	各个检测步骤分离	目标检测、图像识别
[18]	YOLO v3	检测速度大幅提高	需要大规模数据集训练模型	目标检测
[23]	Resnet+LSTM	融合序列信息,提高小目标检测效果	计算量大,模型复杂	目标检测
[24]	Faster R-CNN	速度快,精度高	模型复杂	目标检测、姿态估计
[25]	SegNet	节约时间和内存	图像分割代价昂贵	图像分割、目标检测

性差等问题。并且,近几年在计算机硬件和GPU发展和完善的背景下,目标检测算法速度比过去大大提高,从而被越来越多研究者应用在交通场景车辆检测识别中。

2.2.1 目标检测算法的优化

随着目标检测算法研究的深入,需要面对的困难与挑战也逐渐增多,比如检测准确率提高但随之速度下降、小目标检测效果差等问题。常规的目标检测算法越来越不能满足交通场景目标检测识别应用的需求,因此需要对常规目标检测算法优化改进。目前,目标检测算法的优化主要是特征增强、引入上下文信息、锚点框设计、非极大值抑制算法和损失函数五个方面。下面从这五方面分别论述目标检测算法的优化研究。

(1)特征增强。特征增强的目的是生成高质量的特征表示,以提升对目标的检测效果。特征增强的主要方法有优化基础网络、多尺度特征融合和引入注意力机制。

① 基础网络优化。早期目标检测算法的基础网络大多使用VGG[26]网络,该网络结构清晰,通过卷积层和池化层的反复堆叠以提升特征提取能力。然而,该网络只有19层,提取的特征表达能力有限。若仅通过加深网络的方法提取深层特征则会发生梯度消失和退化等问题,因此He等提出Resnet[27]结构,通过短连接(shortcut)融合浅层与深层特征信息提高网络性能。利用ResNet网络基本思路,DenseNet[28]提出密集连接机制,同时拼接不同层的特征图,增加了不同层之间的联系。STDN[29](Scale Transferrable Object Detection)算法在DenseNet网络基础上引入尺寸转换层,将特征图不加参数转为大尺寸特征图,提高了检测精度与速度。但随着网络加深带来参数的增加是成倍的,因此采用深度可分离卷积、向量化卷积与通道及模块化卷积多种方式轻量化网络结构、减少参数量,代表网络有SqueezeNet[30]、MoblieNet[31]、Xception[32]等。轻量化网络结构可以缩小模型占用内存、加快模型训练速度,但模型的检测精度和准确率也会有所下降。

② 多尺度特征融合。很多文献表明卷积层蕴涵大量特征信息,多层卷积层可以学习不同层次的图像特征。多尺度特征融合将浅层特征与深层特征相互融合,构建具有细粒度特征和丰富语义特征的特征表示,提高目标检测算法的鲁棒性。

对于基于候选区域的算法,HyperNet[33]算法融合多层卷积层特征图,获得具有浅层几何信息和高层语义信息的Hyper特征图。Lin等[34]提出特征金字塔网络(FPN),将多尺度特征融合应用在目标检测算法中。Singh等[35]提出图像金字塔尺度归一化(SNIP)方法,生成三种不同分辨率的输入图像,高分辨率检测小目标,中分辨率检测中目标,低分辨率检测大目标。对于基于回归的算法,Jeong等提出RSSD[7]算法,通过池化将不同卷积层特征级联。Li等提出FSSD[36]算法,将多尺度特征层卷积后通过上采样级联后再次卷积。Cui等提出MDSSD[37]算法,将浅层与高层特征图逐元素相加,构建丰富特征表示。多尺度特征融合是增强特征表示的常用方法,该方法可以不加辅助特征模块大幅提高算法检测准确率,但计算量也同时增大。

③ 引入注意力机制。注意力机制是近几年的热点,本质是聚焦局部信息变化,抑制无用信息。该机制分为空间注意力、通道注意力和空间通道混合注意力。

空间注意力机制的代表模型是STN[38](Spatial Transformer Network)和DCN[39](Dynamic Capacity Networks)网络。前者通过学习输入图像确定和修正目标位置;后者则采用两个子网络,低性能网络处理全图、定位感兴趣区域,高性能网络对感兴趣区域精细化处理。通道注意力机制的代表模型是SENet、SKNet网络。SENet[40]将卷积后特征在空间维度上压缩,然后建模特征通道间的相关性,基于特定的任务学习不同通道的重要性。SKNet[41]将通道加权思想与Inception多分支网络结构结合,获得明显性能提升。空间通道混合注意力机制的的代表模型为CBAM[42](Convolutional Block Attention Module),该模型同时在空间和通道上进行特征融合,文献表明加入CBAM模块比基准模型具有更好的性能,更关注目标本身。注意力机制核心目标是在众多信息中选择对当前任务目标更关键的信息,该机制可以直接获取全局与局部信息的联系,但不能学习序列中的顺序关系,常与位置信息结合研究。

特征增强是目标检测算法优化的主流方法。特征优秀的表现力是检测和识别的基础,同时也是提升算法鲁棒性的关键。

(2)引入上下文信息。在目标检测任务中,融入目标附近的上下文信息有利于在复杂的背景中区分出目标物体。该方法可以分为全局上下文信息和局部上下文信息。前者是基于注意的循环模型在整张图像上获取上下文信息,后者是在特定建议目标框之外,利用内部与外部上下文信息来增强特征表示。

对于全局上下文信息,Bell等提出ION(Inside-Outside Network)[43]网络,应用空间关联信息分析每个特征的附近信息。Ouyang等提出DeepID[44]网络,融合学习特征(上下文信息)与目标特征。Guan等[45]提出语义上下文感知网络,通过金字塔结构融合全局上下文信息。对于局部上下文信息,Cai等[46]利用多尺度网络提取多尺度特征信息,同时引入上下文信息,提高对小目标的检测性能。Chen等[47]提出空间记忆网络,保留与替换上下文特征。Zeng等[48]提出双向门卷积网络(Gated Bi-dircetional CNN,GBDNet),筛选有用的上下文信息以获得更好的目标特征。Zhu等[49]提出CoupleNet网络,通过融合全局、局部与上下文信息提高小目标检测精度。

在目标检测任务中引入上下文信息有利于丰富特征表示、区分小目标,提高检测精度。上下文信息也常被用在显著性目标检测中,对于该任务常考虑与实例分割结合研究。

(3)锚点框设计。基于候选区域的算法通过经验设计先验锚点框大小与比例,这种方式会导致先验候选框对不同目标适应性较差。设置密集候选框可以保证目标定位的准确率,但也会引入更多参数、增大计算量。因此设置合理先验候选框是必要的。

Krishna等[50]通过公示推导计算先验候选框尺寸,提高候选框定位准确率。YOLO v2[51]采用 K-means算法对训练目标真实框聚类分析,生成合适候选框。Xie等[52]将锚点在维度上分解,使用锚点字符串机制匹配目标尺寸,以解决特殊比例目标的检测。Wang等[53]提出Guided-Anchoring方法,通过图像特征指导先验候选框的生成。但是生成候选框的方式存在大量参数且会导致正负样本不均等问题,因此基于anchor-free的目标检测算法相继被提出,比如CornerNet[54]、CenterNet[55]等。上述anchor-free算法是基于关键点、分割的思想来解决检测问题,避免了anchor相关的复杂计算和参数设计,使得训练过程占用内存更低。但是,该算法未解决训练时正负样本不平衡等难题,且常会出现语义模糊性(两个目标中心点重叠无法识别)等问题。

合理设计锚点框是目标准确定位的关键,总之锚点框的设计应遵循几点原则:① 符合数据集特点,根据检测目标设计相匹配尺度。② 对于小目标适当增大锚点框密集密度,对于大目标适当降低锚点框密集密度。③ 与特征图网络中心点位置尽量重合。

(4)非极大值抑制算法优化。非极大值抑制(NMS)算法通过交并比(IoU)方式选择置信度最高的候选框,然而IoU方法剔除候选框粗暴、会产生漏检、错检等问题。

因此,Bodla等[56]提出Soft-NMS算法,通过降低重叠大于阈值边界框的置信度来提高模型的召回率。Ning等[57]提出Weighted-NMS算法,认为最大得分框未必精确,冗余框也可能包含精确位置信息,通过对坐标加权平均获得目标框。Zheng等[58]提出DIoU-NMS算法,通过框中心的距离判别冗余框。Zheng等[59]提出Cluster-NMS算法,融合惩罚机制、中心点距离、加权平均法,通过聚类减少迭代次数、提高整体推理速度。

非极大值抑制算法去除多余边界框、精准定位目标,是目标检测模型中常用算法。关于NMS算法的改进,针对不同的任务和场景应设计不同的NMS算法。当目标较大且稀疏、背景简单时,优化NMS算法几乎对模型鲁棒性没有改变。当目标较小且相对密集时,NMS算法优化能有效提升检测性能。

(5)损失函数的优化。目标检测算法的损失函数大多使用分类和定位损失函数的加权求和。

对于分类损失函数,Lin等[60]提出Focal Loss函数,在分类函数的基础上添加两个平衡因子,用来平衡正负样本不均问题。Chen等[61]提出AP(Average Precision)Loss函数,对每个预测框排序,用排序后的序号设计Loss。Cui等[62]提出Class-Balanced Loss函数,用样本数量调节损失函数缓解样本不均衡。对于定位损失函数,Li等[63]提出GHM-R(Gradient Harmonized Mechanism)Loss函数,通过利用计算损失前的特征梯度信息,对原损失函数进行规范化。Yu等[64]提出IoU Loss函数建立坐标值间联系。在IoU Loss的基础上,GIoU Loss[65]函数增加了不重叠预测框的损失,DIoU Loss[58]函数不仅考虑了边界框间的距离,也考虑了框的尺度。

损失函数量化了算法的表现形式,设计合适的损失函数可以提高算法的鲁棒性。损失函数的优化应注意:① 对于分类损失函数,应全面考虑不同种类样本的贡献。② 对于定位损失函数,应选取合适的决策变量并进行合理修正。③ 总损失函数的权重应根据数据集的特点或具体任务实验获得。

目标检测算法的优化主要是以上五个方面,具体在车辆目标检测识别上的应用如表4所示。交通场景背景复杂多变,通过优化方法提升目标检测算法鲁棒性可以实现目标检测算法更好的应用。但是只从目标检测模型本身出发实现算法的应用是单一的并且检测效果提升不显著,因此还需要与其他方法结合深入研究。

2.2.2 三维目标检测算法

将优化的目标检测算法应用在实际交通场景中还是会出现很多问题,比如光照变化、恶劣天气、无法全面感知实际场景中立体目标等。因此,为提高目标检测算法的应用性、更好保障驾驶人员的安全,研究人员开始采用激光雷达或视觉信息与目标检测算法结合的方法识别交通场景车辆目标,主要研究内容是通过目标检测算法对采集的激光点云数据或视觉信息数据检测识别,一些实验研究表明激光点云数据检测识别的效果最好。

表4 优化目标检测算法在车辆检测中的应用

目标检测算法的优化		简单描述	相关算法	优势	局限性	适用场景
特征增强	基础网络	替换为更深、更轻量的卷积神经网络	DenseNet	模型占用内存小;检测时间缩短	深层网络参数增多;轻量网络检测性能下降	任何检测任务
	多尺度特征融合	卷积层特征融合	HyperNet	丰富特征表示	增加时间成本	任何检测任务
	引入注意力模型	加入注意力模块	SENet	深层特征提取能力提高	增大模型计算量	任何检测任务
引入上下文信息		考虑目标周围信息	GBDNet	联系全局与局部特征信息	模型复杂;增大计算量	小目标、复杂场景
锚点框设计		合理设计锚点框;Anchor free模型	CenterNet	精准表现目标尺寸	参数量和计算成本增加	目标密集场景
非极大值抑制算法		精确去除多余边界框	Soft-NMS	精准定位目标位置	增加计算成本	目标密集、背景复杂
损失函数		分类与回归损失	DIoU	增强模型鲁棒性	增加计算量	复杂密集场景

对于激光点云三维目标检测,一般需要对点云数据处理,主要包括间接处理、直接处理和融合处理3类基本方法。间接处理点云的方法主要是对点云数据进行体素化或降维后再投入已有的深度神经网络进行处理。Beltaran等[66]为提高方法对不同线束激光雷达的普适性提出BirdNet算法,采用Faster RCNN为基础网络,正则化处理每个点云通道。Zeng等[67]提出RT3D(Real Time 3D)算法,通过R-FCN网络检测体素化后的车辆点云栅格信息。此外,为提高计算与检测效率,Shi等[68]提出Part-A^2 Net算法,对每一个点云栅格提取特征,利用类似U-Net的主干网络输出标签与位置。PV-RCNN[69]算法将3D特征图转为俯视图,高度变为通道,使用每个特征块生成两个候选框。间接处理点云方法有效利用了完善的二维检测网络,但也忽视了目标的三维空间信息,检测精度不高、计算量大。

直接处理点云的方法主要是重新设计针对三维点云数据的深度神经网络对点云进行处理,如PointNet系列、YOLO 3D等。Qi等提出针对点云数据的三维目标检测算法PointNet[70],该算法输入为点云数据集合,通过与转换矩阵相乘保证模型的不变性,利用多层感知机(MLP)生成全局特征,最后实现分类与分割任务。PointNet算法直接在点云数据上应用深度学习模型,充分利用点云三维信息,但不能很好捕捉点云局部信息。受CNN启发,PointNet++[71]算法通过点距离构建局部区域提取特征。Frustum-PointNet[72]算法通过PointNet对生成的点云视锥进行实例分割,然后对3D边界框回归获得最终输出。STD[73](Sparse-to-Dense)算法提出一种球形锚点机制,使用PointNet++生成特征和标签得分。VoteNet[74]算法引入霍夫投票机制,不依赖彩色图像、使用纯几何信息。直接处理点云方法可以很好地获得点云的局部或全局特征,但主要难点是设计的神经网络架构是否符合点云数据的特点。

融合处理点云的方法则是融合图像和点云的检测结果再进一步处理。Chen等提出MV3D[75](Multi-View 3D)算法,将点云与图像作为输入,通过点云栅格化构建俯视图和前视图,以实现自动驾驶三维目标检测。Xu等提出PointFusion[76]算法,分别使用ResNet和PointNet提取特征进行融合,然后预测目标的3D边界框。Ku等提出一种用于自动驾驶的目标检测算法,AVOD[77](Aggregate View Object Detection Network)设计了一个生成多模态高分辨率特征映射的RPN网络,以预测场景中目标的大小、方向和类别。对于传感器无法同步等问题,RoarNet-3D[78]算法使用RoarNet-2D估计物体的三维姿态并获得候选区域作为输入,然后深度推断候选区域获得最终姿态。融合处理点云方法检测识别精度高,效果最好。目前,融合处理点云的方法是点云处理的主要技术且检测效果优势明显,但是融合处理计算量大、采集数据困难。具体三维检测车辆目标应用见表5。

表5 三维目标检测在车辆检测中的应用

信息	传感器	优势	局限性
视觉信息	深度相机	数据易获取;成本低	数据精度不高
激光点云	激光雷达	采集信息丰富	设备昂贵
融合信息	相机、雷达	融合信息特征丰富	坐标配准困难

二维目标检测可以很好识别图像中的目标物体,但在自然交通场景中,车辆、行人等为三维目标,因此需要获得场景中三维数据检测识别。目前,激光雷达采集三维数据效果最好,将激光点云数据与目标检测算法结合为三维目标检测是研究的热点。但这种方法仍面临许多困难,如间接处理点云数据导致数据特征失真、直接处理点云数据的新算法设计难度大、融合处理图像与点云数据对计算机硬件要求较高等。

2.2.3 机动车检测识别小结

目标检测算法识别交通场景中车辆目标主要是优化目标检测算法和三维目标检测两方面。目标检测算法的优化可以提高算法检测识别的准确度和精度,三维目标检测可以全面感知交通场景中的目标信息,然而优化目标检测算法检测效果提升不高,在实际应用中易受环境因素影响,三维目标检测设备昂贵,对计算机硬件要求很高。

2.3 目标检测算法在行人识别的应用

行人检测是目标检测的重要研究任务之一,主要内容是通过计算机判断图像中是否存在行人目标,如存在则标出检测目标在图像中的类别与位置。传统行人检测方法常会出现特征遗漏、检测精度不高等问题,操作复杂、需要大量人力物力。近年来,目标检测算法由于良好的检测性能被应用在行人检测中。

2.3.1 多模态目标检测

目前,在通用的行人图像数据集中,目标检测算法有着良好的表现力。然而,自然场景中人员伤亡、财产损失的事故主要发生在夜晚、恶劣天气下,如何在夜晚及恶劣天气条件下检测识别出行人目标是当前研究的难点。研究人员采用多种方法,其中较好的方法是多模态目标检测。

多模态目标检测采用不同传感器采集数据信息,融合信息检测识别目标。Wang等[79]提出一种CIMDL(Correlated and Individual Multi-Modal)方法,输出为两个模态信息特征和一个融合特征,在充分融合特征信息的基础上,保留了各自模态的特有信息。Liu等[80]改进Faster R-CNN网络,融合彩色图像和多光谱图像特征信息对行人目标检测识别。Park等[81]认为仅用两个模态融合是不够的,通过概率模型考虑每个模态特征信息,并采用通道加权融合有选择使用信息。Guan等[82]提出一种光照感知加权机制以学习不同光照条件下的多光谱行人特征,将光照信息与多光谱数据综合实现行人检测的多任务学习和语义分割。Zhou等[83]将毫米波雷达与摄像机信息融合,利用时空同步关联多传感器数

据,最后改进YOLO v2算法实现深度融合对交通场景目标检测识别。这些研究是基于多种传感器采集场景信息,目前常用的传感器为RGB相机、激光雷达、深度相机、多光谱相机等。除此之外,高精地图、雷达、毫米波雷达也同样被应用在自动驾驶目标检测中。多模态检测方法比较如表6所示[84]。

表6 多模态检测方法比较

传感器	模态	信息	检测目标
RGB相机	图像	RGB信息	2D
深度相机	图像	RGB信息、深度	2D&3D
红外相机	图像	红外图像	2D
多光谱相机	图像	多光谱图像	2D
激光雷达	点云	深度、反射强度	2D&3D
雷达	点云	深度、径向速度	2D&3D
毫米波雷达	点云	深度、径向速度	2D
高精地图	地图	地图先验信息	2D&3D

采用多种传感器融合场景信息,避免单一模态感知信息缺陷,提高模型鲁棒性是目标检测发展的趋势。多模态成像不受光线条件影响、可以获得全面场景信息,因此在复杂环境下也可对目标检测识别。目前,多模态目标检测是研究热点方向,然而多模态图像融合坐标配准困难、占用内存大且缺少相应数据集。

2.3.2 行人重识别

除此之外,行人重识别也是行人检测的重要研究分支。行人重识别主要研究内容是判断某个摄像头中的某个行人是否曾经出现在其他的摄像头中,即需要将某个行人特征与其他行人特征进行对比,判断是否属于同一个行人。

目前,行人重识别主要是传统方法、强监督深度学习方法和无监督方法,传统方法主要通过特征提取和度量学习方法,大部分无监督方法也是基于传统方法的研究。2005年,Zajdel等[85]探讨了如何在多个摄像头中将行人轨迹关联等问题,该文献采用贝叶斯网络度量相似行人特征。2006年,Gheissari等[86]首次在CVPR上提出行人重识别概念,掀起重识别研究热潮。2007年,Gray[87]提出VIPeR行人重识别数据库,为行人重识别深入研究奠定基础。2016年,Zheng等[88]将行人重识别定义为行人检测与重识别综合,首先对原始视频帧行人检测,再相似度度量行人检测后与待检测图像特征。传统行人重识别方法首先通过特征提取学习不同摄像头下行人变化特征,然后将学习到的特征映射到新的空间度量学习,最后根据图像特征间距离进行排序,获得检索结果。该方法依赖手工特征,不能适应大环境行人重识别应用的需求。

当前随着深度学习的发展,研究人员考虑采取深度学习方法对行人重识别研究,深度学习不仅可以提取丰富的特征表示,还为度量学习带来革新。Yan等[89]首先获取图像的颜色特征和LBP特证,然后通过LSTM(长短期记忆网络)获得基于序列的特征,充分利用图像特征和序列特征。Yi等[90]采用siamese网络学习行人颜色特征、纹理特征和度量,针对行人外观的巨大变化,利用二项式偏差法进行评估。Mclaughlin等[91]结合CNN网络与RNN网络,在CNN基础上获得每个行人外貌特征,在RNN基础上获得时空信息,两者联合进行调参。Zheng等[92]提出Market-1501数据集,该数据集规模为当时最大且自动标注行人边界框,每个行人有多个摄像头多张影像,目前依然是具有挑战性的数据集。行人重识别主要应用于刑侦工作、图像检索等方面,将深度学习方法与行人重识别结合可以提高行人重识别的准确度、充分利用图像特征,具体目标检测算法在行人检测上的应用如表7所示。

近几年在深度学习的基础上,行人重识别取得高速发展,但还依然面临许多挑战。目前,现有数据集是处理后的高质量图像,然而在自然场景环境下,行人重识别会遇到目标遮挡、特征近似和不同摄像头下行人外观发生巨大变化等困难。

2.3.3 行人检测识别小结

目前,行人检测在公开数据集上已经有了非常高的精度和识别准确度,但是针对复杂、密集的交通场景,行人检测还有很长一段路要走。当前,行人重识别是行人检测领域研究的重点,如何在复杂自然环境下准确识别遮挡行人目标仍然是研究的难点。

3 相关数据集及评价标准

3.1 目标检测数据集

当前,在目标检测领域常用的数据集有Pascal VOC[93]、Microsoft COCO[94]、ImageNet[95]、Open Images[96]等,相关交通场景数据集如表8所示。

表7 目标检测算法在行人检测中的应用

具体应用	简单描述	优势	局限性	适用场景
多模态检测	RGB+多光谱	获得夜间环境下具有区分度图像	计算量大、融合信息操作复杂	夜间目标检测
	RGB+点云	全面感知场景信息	设备昂贵、点云具有稀疏性	三维建模
	RGB+红外	不受光照影响	缺少相应数据集	夜间、恶劣天气
行人重识别	表征学习	训练时间短、易收敛;模型鲁棒性强	不适合大规模数据集	小规模数据集
	度量学习	训练大型数据集;性能良好	参数量大;训练时间长	大规模数据集
	局部特征+全局特征	一定程度可以解决行人姿态多样问题	需要规范化图像;需要额外姿态估计模型	需要帧率较高的应用
	视频序列	包含信息丰富	计算效率低	监控、视频等场景
	GAN网络	图片生成与转换	训练耗时	数据难获取、危险场景

表8 常用交通场景数据集

目标	类别	数据集名称	数据集介绍	来源
车辆	二维	MIT-CBCL Car	用于车辆检测识别,共有526张128×128格式为ppm的图像	麻省理工学院(MIT)
		UA-DETRAC	在北京和天津的道路过街天桥拍摄,并手动标注8 250个车辆和121万目标对象框	纽约州立大学奥尼尔分校研究IT小组
		BDD100K	当前最大规模、内容最具多样性的公开驾驶数据集	伯克利大学AI实验室(BAIR)
	三维	KITTI	用于车辆检测、车辆追踪、语义分割等	德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合创办
		CityScapes	城市景观数据集,包含2 975张图片,来源于50个不同城市街道场景	梅塞德斯-奔驰公司
		ApolloScape	用于自动驾驶、语义分割,包含密集3D点、立体视频和全景图像	百度公司
		nuScenes	大规模自动驾驶数据集,由1 000个场景组成	nuTonomy公司
行人	行人检测	USC	包含三组数据集,A组来自网络(正面或背面角度、无遮挡),B组来自CAVIAR视频库(多角度、有遮挡),C组来自网络(多角度、无遮挡)	南加利福尼亚大学(USC)
		Caltech	分为Caltech101和Caltech256两类	加州理工学院(Caltech)
	行人重识别	VIPeR	两个摄像头采集,包含632个行人的1 264张图像,图像大小为128×48	加利福尼亚大学圣克鲁兹分校(UCSC)
		Market-1501	训练集有751人,包含12 936张图像;测试集有750人,包含19 732张图像	清华大学(THU)

Pascal VOC数据集用于图像分类和目标检测,Pascal VOC 2007和Pascal VOC 2012为主要流行数据集。Pascal VOC数据集包含20个类别,其中Pascal VOC 2007共有9 963张图片24 640个目标;Pascal VOC 2012共有23 080张图片54 900个目标,每张图片都有对应的xml文件。

Microsoft COCO数据集用于目标检测、人体关键点和语义分割等方面,包含91个种类。对于目标检测领域,该数据集来源于真实的自然场景,是挑战性最大的数据集之一,每张图片对应JSON格式的标注文件。

ImageNet数据集用于图像分类、目标检测和场景识别等,包含2.2万个类别,1 420万张图片。对于目标检测任务,它具有200个目标类别,每张图片的标注以Pascal VOC格式保存在XML文件中。

Open Images数据集用于目标检测、语义分割等,于2017年发布。该数据集包含约900万张标注图片,6 000个类别的标签,每张图像平均有8个标签,其分为包含9 011 219张图像的训练集、41 620张图像的验证集和125 436张图像的测试集,是具有目标位置标注的最大现有数据集。

3.2 评价标准

目标检测算法常用的评价标准主要有准确率、召回率、平均精确率和平均精确率均值。其中,准确率(Precision,P)表示在全部已识别样本中正样本被正确识别为正样本的比率,召回率(Recall,R)表示在正样本中被正确识别为正样本的比率。通常情况下,准确率和召回率呈负相关,即召回率越高,准确率越低。将召回率(P)和准确率(R)分别作为横、纵坐标,选择合适的阈值,获得的曲线为P-R曲线,平均精确率(Average Precision,AP)是指P-R曲线下的面积,平均精确率均值(mean Average Precision,mAP)是指每个类别的平均AP值。

除此之外,检测速度也是评价目标检测算法性能好坏的标准之一。衡量目标检测算法检测速度的标准为每秒帧率(Frame Per Second,FPS),即每秒内处理图片的数量,一般来说,FPS越大实时性越好。目标检测的评价标准是衡量目标检测算法性能的关键,表9列出了

表9 目标检测算法性能对比

类别	目标检测算法	基础网络	输入尺寸	FPS/(帧·s^{-1})	mAP50(VOC07+12)/%	mAP50(COCO)/%
两阶段	Faster R-CNN	ResNet-101	~1 000×600	7.0	76.4	45.3
	Mask R-CNN	ResNet-101	1 300×800	4.8	—	60.3
	Cascade R-CNN	ResNet-101	1 300×800	8.0	—	62.1
一阶段	YOLO v1	VGG-16	448×448	45	63.4	—
	SSD	VGG-16	300×300	46	74.3	41.2
	YOLOv2	Darknet-19	416×416	67	76.8	44.0
	RetinaNet	ResNet-101-FPN	800×800	—	—	59.1
	YOLOv3	Darknet-53	416×416	34	—	55.3
	TridentNet	Resnet-101	321×321	—	—	63.6
	YOLOv4	CSPDarknet-53	512×512	23	—	64.9

相关目标检测算法的性能对比。可以看出,基于候选区域算法的检测精度和准确率在不断上升,但在检测速度上明显比基于回归算法差,在应用上不能满足实时性;基于回归算法的检测速度比较快,已可以达到实时性,但是检测精度和准确率比基于候选区域算法较差,目前YOLO v4在现有实时目标检测算法中检测精度最高,一阶段目标检测算法是研究的重点。

4 总结与展望

目标检测是十分重要的研究领域,具有广泛的应用前景。本文详细综述目标检测算法的发展历程及研究现状,包括基于候选区域和基于回归两大类算法。在此基础上,以非机动车、机动车和行人三类典型交通场景物体为目标,从传统检测方法、目标检测算法、目标检测算法优化、三维目标检测、多模态目标检测和重识别六个方面分别论述和总结目标检测算法检测识别交通场景目标的研究现状和应用情况。最后,给出常用目标检测和交通场景数据集及评价标准,对两大类目标检测算法的性能进行比较分析。

总体来看,目标检测算法在机动车和行人检测识别上应用较多,在非机动车上应用较少。不同目标检测任务对模型的要求不同,应根据具体场景和任务特点对模型进行相应改进。具体来说,对于目标检测模型增强特征表示和引入上下文信息的改进方法几乎对任何场景和任何任务都是有利的,具有普适性。当交通场景中目标密集、相互遮挡时,改进非极大值抑制算法、合理设计边界框可以有效缓解目标漏检、误检等问题。当交通场景相对复杂、背景多变时,损失函数的改进可以提升模型的训练效果,进而提高模型的鲁棒性。

当前,在公开交通场景数据集中,目标检测算法已具有良好的表现力,但应用在具体实际交通场景中还存在一些问题,对此提出几点研究趋势:

(1)研究更符合目标检测任务的特征提取网络。当前目标检测算法的特征提取网络主要为分类网络,分类与检测任务的网络设计原则不同,数据集间的差异也导致目标检测存在问题,因此需要从目标检测模型的本身出发,构建符合目标检测任务的特征提取网络,提高目标物体的检测性能。

(2)获得更加丰富的图像语义信息。对于复杂交通场景的小目标检测,仅提取小目标的特征信息是不够的,因此需要利用上下文关联信息、场景信息、语义信息构建丰富特征表示。目前,主流的方法主要有生成高清特征表示和利用语义信息,丰富特征表示是目标检测的关键,值得深入研究。

(3)三维目标检测。实现三维目标检测是自动驾驶技术应用的关键,目前三维检测相较于二维算法在精度和实时性等关键指标方面还有较大提升空间。对于三维目标检测必须有效对原始点云数据处理,提升检测的效率和精度。此外,如何解决遮挡、远距离的小目标检测也是亟需解决的关键问题。

(4)多模态目标检测。数据融合是实现目标检测应用任务的重要趋势,尽管针对多模态目标检测的算法不断被提出,但主要还是基于图像,当光照变化幅度较大时,会导致相机记录失真、无法感知场景信息。因此,应考虑利用多模态数据的互补性来提升模型的鲁棒性,例如融合图像、音频、文本信息等。

(5)弱监督目标检测模型。目前,目标检测算法一般基于监督学习,监督学习需要大量已标注的数据。对于数据的标注,需要大量的人工成本,因此利用弱监督学习、少样本学习等方法在标注数据缺失的情况下建立弱监督目标检测模型是研究的热点。

(6)提高模型的可解释性。目标检测模型通过复杂的深层网络模型从海量数据中学习特征并进行分类与定位,这种模型内部的复杂性使人们难以理解模型的决策结果,导致模型的不可解释性。模型的不可解释性存在很多安全风险,在不同领域部署会受到极大的限制。因此,需要深入研究模型内部的复杂过程,提高模型的可解释性,从而进一步实现模型应用。

参考文献:

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

[2] REN S, HE K, GIRSHICK R, et al.Faster R-CNN: towards real-time object detection with region proposal networks[C]// Conference on Neural Information Processing Systems, 2015:91-99.

[3] HE K M, GKIOXARI G, DOLLAR P, et al.Mask R-CNN[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018(42):386-397.

[4] REDMON J, DIVVALA S, GIRSHICK R, et al.You only look once: unified, real-time object detection[C]//Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition,2016:779-788.

[5] LIU W, ANGUELOV D, ERHAN D, et al.SSD: single shot multi box detector[C]//European Conference on Computer Vision and Pattern Recognition,2016:21-37.

[6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M.YOLO v4: optimal speed and accuracy of object detection[EB/OL].[2020-11-23].https://arxiv.org/pdf/2004.10934.pdf.

[7] JEONG J, PARK H, KWAK N.Enhancement of SSD by concatenating feature maps for object detection[C]//British Machine Vision Conference,2017.

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

[9] VIOLA P, JONES M J.Robust real-time face detection[J].

International Journal of Computer Vision,2001,57(2):137-154.
[10] OJALA T,PIETIKAINEN M,MAENPAA T.Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2002,24(7):971-987.
[11] DALAI N,TRIGGS B.Histograms of oriented gradients for human detection[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition,2005.
[12] FELZENSZWALB P,MCALLESTER D,RAMA-NAN D.A discriminatively trained,multiscale,deformable part model[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition,2008:1-8.
[13] RUBLEE E,RABAUD V,KONOLIGE K,et al.ORB:an efficient alternative to SIFT or SURF[C]//Proceedings of IEEE International Conference on Computer Vision,2012:2564-2571.
[14] HINTON G E,SALAKHUTDINOV R R.Reducing the dimensionality of data with neural networks[J].Science,2006,313(5786):504.
[15] KRIZHEVSKY A,SUTSKEVER I,HINTON G E.ImageNet classification with deep convolutional neural networks[J].Neural Information Processing System,2012,141(5):1097-1105.
[16] AHMAD A,JAKA A P,ARLAN A G,et al.Detection and classification of vehicles for traffic video analytics[J].Procedia Computer Science,2018,144.
[17] CHEN X,XIANG S,LIU C L,et al.Vehicle detection in satellite images by hybrid deep convolutional neural networks[J].IEEE Geoscience and Remote Sensing Letters,2014,11(10):1797-1801.
[18] 叶佳林,苏子毅,马浩炎,等.改进YOLOv3的非机动车检测与识别方法[J].计算机工程与应用,2021,57(1):194-199.
[19] 曹伟.基于SSD的车辆检测与跟踪算法研究[D].合肥:安徽大学,2018.
[20] 陈煌.面向车载应用的非机动车识别系统硬件加速设计[D].合肥:中国科学院大学,2018.
[21] 吕浩.基于实时视频对非机动车违法行为的自动判别关键技术研究[D].杭州:浙江工商大学,2020.
[22] 路雪,刘坤,程永翔.一种深度学习的非机动车辆目标检测算法[J].计算机工程与应用,2019,55(8):182-188.
[23] 梁光胜,谢东.基于深度学习的机动车违规行为监测方法[J].计算机应用,2020,40(S1):195-198.
[24] 李晓飞.基于深度学习的行人及骑车人车载图像识别方法[D].北京:清华大学,2016.
[25] 华中科技大学.一种基于深度学习的交通违规分析方法及装置:CN201910940889.4[P].2020-01-20.
[26] SIMONYAN K,ZISSERMAN A.Very deep convolutioal networks for large scale image recognition[C]//International Conference on Learning Representations,2014.
[27] HE K,ZHANG X,REN S,et al.Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:770-778.
[28] HUANG G,LIU Z,VAN D M,et al.Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,Honolulu,HI,USA,2017:2261-2269.
[29] ZHOU P,NI B B,GENG C,et al.Scale transferrable object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018.
[30] IANDOLA F N,HAN S,MOSKEWICZ M,et al.SqueezeNet:AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[J].arXiv:1602.07360,2016.
[31] HOWARD A G,ZHU M,CHEN B,et al.Mobilenets:efficient convolutional neural networks for mobile vision applications[J].arXiv:1704.04861,2017.
[32] CHOLLET F.Xception:deep learning with depthwise separable convolutions[C]//Proceeding of the IEEE Conferencce on Computer Vision and Pattern Recognition,Honolulu,2017:1800-1807.
[33] KONG T,YAO A,CHEN Y,et al.HyperNet:towards accurate region proposal generation and joint object detection[C]//Proceeding of the IEEE Conferencce on Computer Vision and Pattern Recognition,Las Vegas,NV,USA,2016:845-853.
[34] LIN T Y,DOLLAR P,GIRSHICK R,et al.Feature pyramid networks for object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,USA,2017:936-944.
[35] SINGH B,DAVIS L S.An analysis of scale invariance in object detection-snip[C]//Proceeding of IEEE Conference on Computer Vision and Pattern Recognition,USA,2018:3578-3587.
[36] LI Z,ZHOU F.FSSD:feature fusion single shot multibox detector[J].arXiv:1712.00960,2017.
[37] CUI L,MA R,LV P,et al.MDSSD:multi scale deconvolutional single shot detector for small objects[J].Science in China Series F(Information Sciences),2020,63(2):98-100.
[38] JADERBERG M,SIMONYAN K,ZISSERMAN A.Spatial transformer networks[C]//Advances in Neural Information Processing Systems,2015:2017-2025.
[39] ALMAHAIRI A,BALLAS N,COOIJMANS T,et al.Dynamic capacity networks[C]//International Conference on Machine Learning,2016:2549-2558.
[40] HU J,SHEN L,SUN G.Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,Salt Lake City,UT,USA,2018:7132-7141.
[41] LI X,WANG W,HU X,et al.Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020.

[42] WOO S,PARK J,LEE J Y,et al.CBAM:convolutional block attention module[C]//European Conference on Computer Vision.Cham:Springer International Publishing,2018:3-19.

[43] BELL S,LAWRENCE Z C,BALA K,et al.Inside-outside net:detecting objects in context with skip pooling and recurrent neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,Las Vegas,NV,USA,2016:2874-2883.

[44] OUYANG W L,LUO P,ZENG X Y,et al.Deepid-net:multi-stage and deformable deepconvolutional neural networks for object detection[J].arXiv:1409.3505,2014.

[45] GUAN L,WU Y,ZHAO J,et al.SCAN:semantic context aware network for accurate small object detection[J].International Journal of Computational Intelligence Systems,2018,11(1):936-950.

[46] CAI Z,FAN Q,FERIS R S,et al.A unified multiscale deep convolutional neural network for fast object detection[C]//European Conference on Computer Vision,Amsterdam,NL,2016:354-370.

[47] CHEN X L,GUPTA A.Spatial memory for context reasoning in object detection[C]//Proceedings of the IEEE Conference on Computer Vision,Honolulu,HI,USA,2017:4086-4096.

[48] ZENG X Y,OUYANG W L,YANG B,et al.Gated bidirectional CNN for object detection[C]//Proceedings of the 14th European Conference on Computer Vision.Amsterdam,The Netherlands:Springer,2016:354-369.

[49] ZHU Y,ZHAO C,WANG J,et al.CoupleNet:coupling global structure with local parts for object detection[C]//International Conference on Computer Vision,Venice,IT,2017:4146-4154.

[50] KRISHNA H,JAWAHAR C V.Improving small object detection[C]//IAPR Asian Conference on Pattern Recognition,2017:340-345.

[51] REDMON J,FARHADI A.YOLO9000:better,faster,stronger[C]//Computer Vision and Pattern Recognition,Honolulu,HI,USA,2017:6517-6525.

[52] XIE L L,LIU Y L,JIN L W,et al.DeRPN:taking a further step toward more general object detection[C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence,Honolulu,Hawaii,USA,2019:9046-9053.

[53] WANG J Q,CHEN K,YANG S,et al.Region proposal by guided anchoring[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,LongBeach,CA,USA,2019:2965-2974.

[54] LAW H,DENG J.Cornernet:detecting objects as paired key points[C]//European Conference on Computer Vision,Munich,Germany,2018:765-781.

[55] ZHOU X,WANG D,PHILIPP K.Objects as points[EB/OL].(2019)[2020-11-25].https://arxiv.org/pdf/1904.07850.

[56] BODLA N,SINGH B,CHELLAPPA R,et al.Soft-NMS improving object detection with one line of code[C]//Proceedings of the IEEE International Conference on Computer Vision,Honolulu,HI,USA,2017:5561-5569.

[57] NING C,ZHOU H,SONG Y,et al.Inception single shot multibox detector for object detection[C]//Proceedings of the IEEE International Conference on Multimedia & Expo Workshops,2017.

[58] ZHENG Z H,WANG P,LIU W,et al.Distance-IoU Loss:faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence,2019.

[59] ZHENG Z,WANG P,REN D,et al.Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J].arXiv:2005.03572,2020.

[60] LIN T Y,GOYAL P,GIRSHICK R,et al.Focal loss for dense object detection[C]//International Conference on Computer Vision,2017:2999-3007.

[61] CHEN K,LI J G,LIN W Y,et al.Towards accurate one-stage object detection with AP-Loss[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,Long Beach,CA,USA,2019:5119-5127.

[62] CUI Y,JIA M,LIN T Y,et al.Class-balanced loss based on effective number of samples[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2020.

[63] LI B Y,LIU Y,WANG X G.Gradient harmonized single stage detector[C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence,Honolulu,Hawaii,USA,2019:8577-8584.

[64] YU J H,JIANG Y N,WANG Z Y,et al.Unit-Box:an advanced object detection network[C]//Proceedings of the ACM International Conference on Multimedia Amsterdam.Netherlands:ACM,2016:516-520.

[65] REZATOFIGHI H,TSOI N,GWAK J Y,et al.Generalized intersection over union:a metric and a loss for bounding box regression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019.

[66] BELTARAN J,GUINDEL C,MORENO F M,et al.Birdnet:a 3D object detection framework from lidar information[C]//Proceedings of 2018 21st International Conference on Intelligent Transportation Systems,2018:3517-3523.

[67] ZENG Y,HU Y,LIU S,et al.RT3D:Real-time 3D vehicle detection in lidar point cloud for autonomous driving[J].IEEE Robotics and Automation Letters,2018,3(4):3434-3440.

[68] SHI S,WANG Z,WANG X,et al.Part-A^2 Net:3D part-aware and aggregation neural networkfor object detection from point cloud[J].arXiv:1907.03670,2019.

[69] SHI S,GUO C,JIANG L,et al.PV-RCNN:point-voxel feature set abstraction for 3D object detection[J].arXiv:1912.13192,2019.

[70] QI C,SU H,MO K,et al.PointNet:deep learning on

point sets for 3D classification and segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:77-85.

[71] QI C,YI L,SU H,et al.PointNet++:deep hierarchical feature learning on point sets in a metr-icspace[C]//Annual Conference on Neural Information Processing Systems,2017:5100-5109.

[72] QI C,LIU W,WU C,et al.Frustum pointnets for 3D object detection from RGB-D data[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:918-927.

[73] YANG Z,SUN Y,LIU S,et al.STD:sparse-to-dense 3D object detector for point cloud[J].arXiv:1907.10471,2019.

[74] QI C,LITANY O,HE K,et al.Deep hough voting for 3D object detection in point clouds[C]//International Conference on Computer Vision,2019.

[75] CHEN X,MA H,WAN J,et al.Multi-view 3D object detection network for autonomous driving[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:6526-6534.

[76] XU D F,ANGUELOV D,JAIN A.Point fusion:deep sensor fusion for 3D bounding box estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:244-253.

[77] KU J,MOZIFIAN M,LEE J,et al.Joint 3D proposal generation and object detection from view aggregation[C]// International Conference on Intelligent Robots and Systems,2017:5750-5757.

[78] SHIN K,KWON Y P,TOMIZUKA M.RoarNet:a robust 3D object detection based on region approximation refinement[J].arXiv:1811.03818,2018.

[79] WANG Z,LU J,LIN R,et al.Correlated and individual multi-modal deep learning for RGB-D object recognition[J].arXiv:1604.01655,2016.

[80] LIU J,ZHANG S,WANG S,et al.Multispectral deep neural networks for pedestrian detection[C]//British Machine Vision Conference,York,UK,2016:1-13.

[81] PARK K,KIM S,SOHN K.Unified multi-spectral pedestrain detection based on probabilistic fusion networks[J]. Pattern Recognition,2018,80:143-155.

[82] GUAN D,CAO Y,YANG J,et al.Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection[J].Information Fusion,2018, 50:148-157.

[83] ZHOU T,JIANG K,XIAO Z,et al.Object detection using multi-sensor fusion based on deep learning[C]//COTA International Conference of Transportation,Nanjing,China, 2019:5770-5782.

[84] 张新钰,邹镇洪,李志伟,等.面向自动驾驶目标检测的深度多模态融合技术[J].智能系统学报,2020,15(4):758-771.

[85] ZAJDEL W,ZIVKOVIC Z,KROSE B.Keeping track of humans:have I seen this person before?[C]//Proceedings of the IEEE Conference on Robotics and Automation, Barcelona,Spain,2005:2081-2086.

[86] GHEISSARI N,SEBASTIAN T B,HARTLEY R.Person reidentification using spatiotemporal appearance[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2006.

[87] GRAY D,BRENNAN S,TAO H.Evaluating appearance models for recognition,reacquisition,and tracking[C]// Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance,Rio de Janeiro,2007:1-7.

[88] ZHENG L,YANG Y,HAUPTMANN A G.Person reidentification:past,present and future[J].arXiv:1610.02984,2016.

[89] YAN Y,NI B,SONG Z,et al.Person reidentification via recurrent feature aggregation[C]//European Conference on Computer Vision.[S.l.]:Springer International Publishing, 2016:701-716.

[90] YI D,LEI Z,LI S Z.Deep metric learning for practical person re-identification[C]//International Conference on Pattern Recognition,Stockholm Waterfront,Sweden,2014.

[91] MCLAUGHLIN N,RINCON J M,MILLER P.Recurrent convolutional network for video-based person re-identification[C]//IEEE Conference on Computer Vision and Pattern Recognition,Las Vegas,USA,2016:51-58.

[92] ZHENG L,SHEN L Y,TIAN L,et al.Scalable person re-identification:a benchmark[C]//Proceedings of the IEEE Conference on Computer Vision,Santiago,Chile,2015: 1116-1124.

[93] SHETTY S.Application of convolutional neural network for image classification on pascal VOC challenge 2012 dataset[C]//Proceedings of the Conference Computer Vision and Pattern Recognition,2016.

[94] LIN T,MAIRE M,BELONGIE S J,et al.Micorsoft coco: common objects in context[C]//European Conference on Computer Vision,2014:740-755.

[95] RUSSAKOVSKY O,DENG J,SU H,et al.Imagenet large scale visual recognition challenge[J].International Journal of Computer Vision,2015,115(3):211-252.

[96] KUZNETSOVA A,ROM H,ALLDRIN N,et al.The open images dataset-v4:unified image classification,object detection,and visual relationship detection at scale[EB/OL].(2018)[2020-11-23].http://dblp.org/abs/1811.00982.