深度学习的典型目标检测算法研究综述

许德刚[1,2],王 露[1,2],李 凡[1,2]
1.河南工业大学 粮食信息处理与控制教育部重点实验室,郑州 450001
2.河南工业大学 信息科学与工程学院,郑州 450001

摘 要:目标检测是计算机视觉的一个重要研究方向,其目的是精确识别给定图像中特定目标物体的类别和位置。近年来,深度卷积神经网络(Deep Convolutional Neural Networks,DCNN)所具有的特征学习和迁移学习能力,在目标检测算法特征提取、图像表达、分类与识别等方面取得了显著进展。介绍了基于深度学习目标检测算法的研究进展、常用数据集特点以及性能指标评价的关键参数,对比分析了双阶段、单阶段以及其他改进算法的网络结构和实现方式。阐述了算法在人脸、显著目标、行人、遥感图像、医学图像、粮虫等检测领域的应用进展,结合当前存在的问题和挑战,展望分析了其未来的研究方向。

关键词:深度学习;目标检测;迁移学习;特征提取;计算机视觉
文献标志码:A **中图分类号:**TP391.41;TP183 doi:10.3778/j.issn.1002-8331.2012-0449

Review of Typical Object Detection Algorithms for Deep Learning

XU Degang[1,2], WANG Lu[1,2], LI Fan[1,2]
1.Key Laboratory of Grain Information Processing and Control, Ministry of Education, Henan University of Technology, Zhengzhou 450001, China
2.School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China

Abstract: Object detection is an important research direction of computer vision, its purpose is to accurately identify the category and location of a specific target object in a given image. In recent years, the feature learning and transfer learning capabilities of deep convolutional neural networks have made significant progress in target detection algorithm feature extraction, image expression, classification and recognition. This paper introduces the research progress of target detection algorithm based on deep learning, the characteristics of common data sets and the key parameters of performance index evaluation, compares and analyzes the network structure and implementation mode of target detection algorithm formed by two-stage, single-stage and other improved algorithms. Finally, the application progress of the algorithm in the detection of human faces, salient targets, pedestrians, remote sensing images, medical images, and grain insects is described. Combined with the current problems and challenges, the future research directions are analyzed.

Key words: deep learning; object detection; transfer learning; feature extraction; computer vision

目标检测作为计算机视觉领域的研究方向之一,能够为图像和视频的语义理解提供有价值的信息。其本质是对目标进行定位和分类,主要任务是准确高效地找出给定图像中所有感兴趣的目标,利用矩形边界框来定位被检测目标的位置和大小,与目标分类、语义分割、实例分割有一定联系。目标检测在科学发展和实际工业中扮演着重要的角色,例如人脸识别、文本检测、标志检测等。在目标检测过程中,由于图像中各类目标物体的外观、姿态、形状、数量各异,还受到光照、遮挡等各种因素的干扰,导致目标发生畸变,使目标检测难度增加[1],许多学者对目标检测算法进行了研究探索。

目标检测算法主要分为传统检测算法和基于深度学习的检测算法。传统检测方法是建立在手工制作特征和浅层可训练架构上的,这种方法从目标检测器和场景分类器中结合大量低水平图像特征和高水平语义信息构建成复杂系统的性能不高。传统目标检测算法主

基金项目:国家重点研发计划(2017YFD0401003-4)。
作者简介:许德刚(1978—),男,博士,副教授,硕士生导师,研究方向为系统工程理论、智能优化;王露(1995—),通信作者,女,硕士研究生,CCF会员,研究方向为人工智能、机器学习,E-mail:wanglu_lyxa@163.com;李凡(1997—),男,硕士研究生,研究方向为人工智能、智能优化。
收稿日期:2020-12-24 **修回日期:**2021-01-25 **文章编号:**1002-8331(2021)08-0010-16

要包括DPM(Deformable Parts Model)[2]、选择性搜索(Selective Search)[3]、Oxford-MKL[4]和NLPR-HOGLBP[5]等,其基本结构主要包括以下三个部分:

(1)区域选择器。首先对给定图像设置不同大小和比例的滑动窗口,将整个图像从左到右、从上到下进行遍历以框出待检测图像中的某一部分作为候选区域。

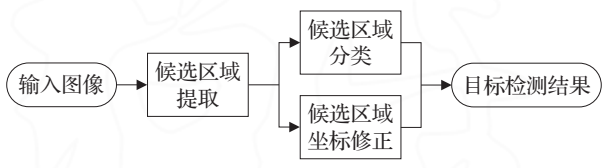
(2)特征提取。提取候选区域的视觉特征,例如在人脸和普通目标检测中常用的尺度不变特征变换(Scale-Invariant Feature Transform,SIFT)[6]、Haar[7]、方向梯度直方图(Histogram of Oriented Gradient,HOG)[8]等特征,对每个区域进行提取特征。

(3)分类器分类。使用训练好的分类器对特征进行目标类别识别,如常用的形变部件模型(DPM)、Adaboot[9]支持向量机(Support Vector Machines,SVM)[10]等分类器。

传统目标检测方法取得了一定的成果,但也暴露了其固有的弊端。首先,采用滑动窗口进行区域选择会导致较高的时间复杂度和窗口冗余。其次,外观形态的多姿性、光照变化的不定性和背景的多样性导致人工手动设计特征的方法鲁棒性不好[11],泛化性差,繁杂的算法步骤导致检测效率慢、精度不高[12]。传统的检测方法已经难以满足人们对目标检测高性能效果的需求。

近年来,深度学习得到了快速发展,引入了一些能学习语义、高水平、深层次特征的工具来解决传统体系结构中存在问题,使模型在网络架构、训练策略和优化功能方面性能提升。Hinton等[13]在2012年首次将深层卷积神经网络AlexNet[14]应用于大规模图像分类中,在目标检测数据集ImageNet[15]大规模视觉识别挑战的分类任务中获得冠军。随着深度卷积神经网络获得的重大研究进展,DCNN逐渐成为人们关注的焦点,也成为目标检测新的研究方向。

主流的深度学习目标检测算法主要分为双阶段检测算法和单阶段检测算法,如图1所示。双阶段检测算法是以R-CNN系列为代表的基于候选区域的目标检测算法;单阶段检测算法是以YOLO、SSD为代表的基于回归分析的目标检测算法。



(a)双阶段检测算法

(b)单阶段检测算法

图1 基于深度学习的目标检测算法

1 目标检测算法性能评价

1.1 相关数据集

数据集在目标检测研究中扮演着十分重要的角色,在进行目标检测任务时,适用性强的数据集可以有效测试和评估算法的性能,并且推动相关领域的研究发展。基于深度学习的目标检测任务中应用最广泛的数据集有:PASCAL VOC2007[16]、PASCAL VOC2012[17]、Microsoft COCO[18]、ImageNet[15]和OICOD(the Open Image Challenge Object Detection)[19]等。

PASCAL VOC(The PASCAL Visual Object Classification)数据集在2005年首次发布,至2012年更新了多个版本,主要用于图像分类、图像识别、目标检测任务。目前主流的是PASCAL 2007和PASCAL 2012两个版本,包含20个类别,数据集图像都有对应的格式标签文件对目标位置和类别进行标注。PASCAL VOC2007包含9 963张标注过的图片,标注出24 640个目标物体;PASCAL VOC2012包含11 540张图片,一共标注出27 450个目标物体。

Microsoft COCO数据集在2015年首次发布,是一种基于日常复杂场景的常见目标数据库,具有小目标物体多、单幅图片目标多等特点,一共包含30多万张完全分割的图像,平均每张图像含有7个目标实例,共标注出250万个目标物体,包含91个物体类别。主要用于图像标题生成、目标检测、实例分割等任务,该数据集是图像分割领域最大的数据集,极大推动了该领域的发展。

ImageNet数据集于2010年发布,增加了类和图像的数量,提高了目标检测任务的训练和评估标准,主要用于图像分类、目标定位、目标检测、场景分类和场景解析等任务。该数据集的图像数量增加到1 200万多张,2.2万个类别,约103万张图片进行了目标物体和类别标注,包含200个对象类别。ImageNet数据集是图像识别最大的可视化数据库,推动了深度学习的浪潮;另一方面,由于数据集大,类别数目多,导致训练所用的计算量很大,目标检测难度增加。

OICOD数据集是在OpenImageV4基础上提出的最大的公共使用数据集,与ILSVRC(ImageNet Large Scale Visual Recognition Challenge)和MS COCO目标检测数据集不同,它提供了更多的类、图像、边界框、实例分割分支和大量的注释过程,OICOD为目标实例提供了人工确认的标签。OpenImageV4使用分类器注释图像,并且只使用人工验证得分很高的标签。ILSVRC和MS COCO是一个完整的注释数据集。

1.2 目标检测算法性能评价指标

目标检测算法的性能主要通过以下几个参数评估:交并比(Intersection over Union,IoU)[20]、检测速度(Frame Per Second,FPS)、准确率(Accuracy,A)、召回率(Recall,R)、精确度(Precision,P)、平均精确度(Average Precision,AP)和平均精确度均值(mean Average Precision,mAP)[21]。其中AP由P-R曲线和坐标围起来的面积组成,mAP是AP的均值。

目标检测模型的分类和定位能力是其最主要的性能体现,而mAP值是其最直观的表达方式,mAP值越大,表明该模型的精度越高;检测速度代表了目标检测模型的计算性能,用FPS值来体现,FPS值越大,说明算法模型实时性越好。在实际应用中,评估目标检测模型的性能时,一般结合平均精确度均值mAP和检测速度FPS两者来判断。

2 基于深度学习的目标检测算法

2.1 双阶段目标检测算法

双阶段目标检测方法主要通过选择性搜索(Selective Search)或者Edge Boxes[22]等算法对输入图像选取可能包含检测目标的候选区域(Region Proposal)[23],再对候选区域进行分类和位置回归以得到检测结果。典型的算法有R-CNN[24]系列、R-FCN[25]、Mask R-CNN[26]等。

2.1.1 OverFeat算法

Sermanet等[27]改进AlexNet提出了OverFeat算法。该算法结合AlexNet通过多尺度滑动窗口[28]实现特征提取功能,共享特征提取层,并应用于图像分类、定位和目标检测等任务。在ILSVRC 2013[29]数据集上的mAP达24.3%,检测效果相比传统检测方法有显著提升。该算法对深度学习的目标检测算法有启发式意义,但对小目标物体检测效果不好,存在较高错误率。

2.1.2 R-CNN算法

2014年,Girshick等将卷积神经网络CNN应用于目标检测任务,提出了典型的双阶段目标检测算法R-CNN[24],R-CNN将AlexNet与选择性搜索算法相结合,包含区域建议、基于CNN的深度特征提取、分类回归三个模块:(1)使用选择性算法从每张图像中提取2 000个左右可能包含目标物体的区域候选框;(2)对候选区域进行归一化操作缩放成固定大小,进行特征提取;(3)使用AlexNet将候选区域特征逐个输进SVM进行分类,通过使用边界框回归(Bounding Box Regression)和非极大值抑制(Non-Maximum Suppression,NMS)对区域得分进行调整和过滤,在全连接网络进行位置回归。

R-CNN算法在ILSVRC2013数据集上mAP提升至31.4%,在VOC2007数据集上mAP为58.5%,相比传统目标检测算法,性能得到改进。但仍存在以下问题:

(1)训练是一个多阶段通道,缓慢且难以优化,因为每个阶段都必须分别进行培训。

(2)对于SVM分类器和边界框回归器的训练,无论是在磁盘空间还是时间上都是昂贵的,因为CNN特征需要从每幅图像的每个目标建议中提取,这对大规模检测来说是巨大的挑战,尤其是在非常深的网络下,如VGG16。

(3)测试速度慢,因为CNN特征需要在每个测试图像的目标建议提取,没有共享计算。

2.1.3 SPP-Net算法

针对R-CNN对所有候选区域分别提取特征、计算量大的问题,2015年He等[30]提出空间金字塔网络(Spatial Pyramid Pooling Network,SPP-Net)。SPP-Net在最后一个卷积层和全连接层之间加入空间金字塔结构;使用多个标准尺度微调器对图像进行分割,将量化后的局部特征融合生成中层表达,在第五个卷积层的特征图上生成固定长度的特征向量,一次性提取特征避免重复特征提取,打破了固定尺寸输入的束缚。

以上改进使SPP-Net算法比R-CNN算法检测速度提高24~102倍,在PASCAL 2007数据集上mAP提高至59.2%。但以下问题仍需改进:(1)需存储大量的特征,空间消耗大;(2)依然使用SVM分类器,训练步骤繁琐且周期长。

2.1.4 Fast R-CNN算法

为克服SPP-Net存在的问题,2015年Girshick等提出基于边界框和多任务损失分类的Fast R-CNN[31]算法。该算法将SPP层简化,设计出单尺度的ROI Pooling池化层结构;将整张图像的候选区域采样成固定大小,生成特征图后作SVD分解,通过RoI Pooling层得到Softmax的分类得分和BoundingBox外接矩形框的窗口回归两个向量;用Softmax代替SVM提出多任务损失函数思想,将深度网络和SVM分类两个阶段整合,即将分类问题和边框回归问题进行合并。

Fast R-CNN使用VGG-16代替AlexNet,将特征提取、目标分类和位置回归都整合到一个模型中,降低了复杂度,提高了检测精度和速度。在VOC2007数据集上mAP提高到70%,训练速度较R-CNN相比,提升了9倍。但该算法使用选择性搜索算法对候选区域进行特征提取会生成正负样本候选框,消耗大量时间,实时性问题没有得到解决。

2.1.5 Faster R-CNN算法

边界框、选择性搜索等使用候选区域生成方法大大阻碍了精确度的提高。针对该问题,2017年Ren等[32]提出了Faster R-CNN,引入了区域建议网络(Region Proposal Network,RPN)取代选择性搜索算法,RPN将区域建议提取集成到DCNN,与整个检测过程共享所有图像的卷积特征。RPN在每个位置同时预测目标边界框和类别置信度分数,实现了目标检测端到端的训练,加快了网络的计算速度。Faster R-CNN网络由卷积层、RPN网络、RoI Pooling层、分类和回归层等4部分组成。

Faster R-CNN使用VGG-16骨干网络,在PASCAL VOC 2007数据集上mAP为73.2%。但仍存在问题:(1)在特征图上采用anchor机制时候选框设置的尺度不适用于所有目标,尤其对小目标检测效果不好;(2)只采用VGG-16网络的最后一层卷积层所输出特征做预测,经过RoI Pooling层后导致网络特征丧失平移不变性,

降低准确性；(3)ROI池化层对高度重叠的区域存在重复计算，降低检测速度；(4)RPN生成包含背景的类似目标的区域，而不是目标实例，在处理非常大或者有形状的目标时效果不好。

2.1.6 R-FCN算法

R-CNN系列算法的思想和性能决定了目标检测的里程碑。该系列结构本质上由两个子网组成(Faster R-CNN添加了PRN，由三个子网组成)，前一个子网是提取特征的骨干网络，后一个子网用于完成目标检测的分类和定位。RoI池化层在两个子网之间，将多尺度特征映射转换成固定大小的特征映射，但该步骤破坏了网络的平移不变性，不利于目标分类。2016年，Dai等[25]在基于区域的全卷积网络(Region based Fully Convolutional Networks，R-FCN)算法中提出了包含目标位置信息的位置敏感分数图(Position-Sensitive Score Maps)，采用ResNet-101[33]骨干网络。R-FCN区域检测是基于整个图像的全卷积计算且共享全卷积网络，使用位置敏感映射图RoI池来协调整个卷积层的平移不变性和平移敏感性，提高目标定位。

R-FCN在PASCALVOC2007测试集上的mAP达83.6%，在PASCAL VOC 2012测试集上mAP为82.0%，每张图片检测时间缩短至170 ms，较Faster R-CNN提高了2.5~20倍。

2.1.7 Mask R-CNN算法

实例分割任务非常具有挑战性，由目标检测和实例分割两个独立功能组成。2017年，He等提出的Mask R-CNN[26]是Faster R-CNN的扩展，采用ResNet-101-FPN骨干网络。Mask R-CNN将多任务损失与分割分支损失、分类、边界框回归损失相结合，在目标分类和边框回归的基础上增加了一个用于RoI预测分割的Mask网络分支，实现实时的目标检测和实例分割。由于RoI池的整数量化，特征图区域和原始图像区域不对齐，因此在预测像素级掩码时会产生一个偏差。为解决在下采样和RoI池层等比例缩放对特征图尺度做取整操作引入误差的问题，He等提出RoIAlign[34]层代替RoI池层，使用双线性差值填补非整数位置的像素实现像素级对齐，提高目标检测分支的精度。在COCO数据集的mAP提升至39.8%，检测速度为5 frame/s。但是检测速度仍然难以满足实时要求，且实例分割标注代价过大。

2.1.8 Chained Cascade Network和Cascade RCNN算法

Cascade级联的本质是通过使用多阶段分类器学习更多的判别型分类器，即前期阶段丢弃大量简易负样本以便后期集中处理较困难样本。两阶段目标检测可以看作是一个级联，第一阶段检测器去除大量背景，第二阶段对剩余区域进行分类。最近，超过两级联分类器的端到端学习和通用目标检测的DCNNs提出了链接级联网络(Chained Cascade Network)[35]、级联RCNN的扩展[36]以及申请同步目标检测和实例分割[37]，在2018年COCO检测挑战获得成功。

2.1.9 Light Head RCNN算法

为了进一步提高R-FCN的检测速度，Li等[38]提出了Light Head RCNN，使检测网络的头部尽可能轻，以减少RoI计算。特别地，文献[38]应用了卷积来生成具有小信道数(例如，COCO有490个信道)和低成本RCNN子网的薄特征图，从而实现了速度和精度的良好平衡。

2.1.10 双阶段目标检测算法性能对比分析

双阶段目标检测实现了级联结构，且在COCO数据集上取得了实例分割的成功，检测精度一直在不断提高，但是检测速度普遍较慢。表1总结了双阶段目标检测算法的骨干网络以及在VOC2007测试集、VOC 2012测试集以及COCO测试集上的检测精度(mAP)和检测速度，“—”表示没有相关数据。表2总结了这些基于候选区域的目标检测算法的优缺点和适用场景。算法输入图像的大小、使用的的骨干网络和计算机配置会有差别，在一定程度上会影响算法结果，总体来说算法性能结果的对比还是客观公平的。

从表1可以看出，双阶段目标检测器引入了更深层的ResNet[39]、ResNeXt[40]等骨干网络，检测精度可以达到83.6%，但算法模型扩大引起计算量增加，检测速度只有11 frame/s，无法满足实时性要求。

从表2可以看出，双阶段目标检测算法一直在弥补之前算法的缺陷，但带来的模型规模大、检测速度慢等问题没有得到解决。对此，一些研究者提出一种将目标

表1 双阶段目标检测算法性能对比

算法	骨干网络	检测速度/(frame·s^{-1})	mAP/%		
			VOC2007	VOC2012	COCO
R-CNN[24]	AlexNet	0.03	58.5	—	—
	VGG-16	0.5	66.0		
SPP-Net[30]	ZF-5	2	59.2	—	—
Fast R-CNN[31]	VGG-16	7	70.0	68.4	19.7
Faster R-CNN[32]	VGG-16	7	73.2	70.4	21.9
	ResNet-101	5	76.4	73.8	34.9
R-FCN[25]	ResNet-101	9	83.6	82.0	29.9
Mask R-CNN[26]	ResNeXt-101	11	78.2	73.9	39.8

表2 双阶段目标检测算法的优缺点及适用场景

模型	优点	缺点	适用场景
OverFeat	使用CNN进行特征提取	使用滑窗,时间、空间开销大	目标检测
R-CNN	将CNN与候选框方法结合	特征提取复杂,耗时,固定图像输入大小	目标检测
SPP-Net	对整张图片做卷积运算,实现多尺度卷积计算	空间开销大	目标检测
Fast R-CNN	用ROI Pooling层提取特征,节省时间和特征存储空间	候选区域选取方式计算复杂	目标检测
Faster R-CNN	用RPN代替区域建议,加快训练速度和精度	模型复杂,空间量化粗糙	目标检测
R-FCN	定位精度提高	模型流程复杂,计算量大	目标检测
Mask R-CNN	解决特征图与原图不对准,结合检测与分割	实例分割代价大	目标检测、实例分割

检测转化到回归问题上的思路,简化算法模型,在提高检测精度的同时提高检测速度。

2.2 单阶段目标检测算法

单阶段的目标检测算法采用了回归分析的思想,也称为基于回归分析的目标检测算法。单阶段目标检测器省略了候选区域生成阶段,直接得到目标分类和位置信息,以YOLO系列和SSD系列为典型代表。

2.2.1 YOLO及其改进目标检测算法

针对双阶段目标检测算法的低效问题,2016年,Redmon等提出了一种单阶段目标检测器YOLO(You Only Look Once)[41]。YOLO架构由24个卷积层和2个FC层组成,使用最顶层的特征图来预测边界框,直接评估每个类别的概率,使用P-Relu激活函数。YOLO将每个图像划分成$S\times S$的网格单元,每个网格单元只负责预测网格中心的目标,该算法舍去了候选区域生成阶段,将特征提取、回归和分类放在一个卷积网络中,简化了网络网络。在实时情况下,YOLO检测速度为45 frame/s,平均检测精度mAP为63.4%。但YOLO对小尺度目标的检测效果不佳,目标重叠遮挡环境下容易漏检。

针对YOLO定位不准,召回率和检测精度低等缺陷,2017年,Redmon等改进了YOLO的网络结构,提出了YOLOv2[42]算法。该模型以DarkNet-19[43]为骨干网络,增加了批量归一化预处理;采用更高分辨率的分类器;使用2个不同尺度的特征,以K-Means聚类的方式计算更好的anchor模板,即多尺度训练机制;采用binary cross-entropy损失函数替换Softmax损失函数,召回率和准确性得到了显著提升,检测速度达到67 frame/s,在VOC2007数据集上mAP提升到78.6%。但是对重叠度高和小尺度目标检测仍需改进。

2018年,Redmon等又在YOLOv2的基础上改进提出了YOLOv3[44]。该版本采用了全新设计的更深层次的DarkNet-53[45]残差网络来进行特征提取,并结合FPN网络;使用了3个尺度的特征图进行边界框的预测,同时增加了anchor数量。YOLOv3算法在COCO数据集上AP为33.0%,mAP为57.9%。YOLOv3模型结构更复杂,使检测速度有所降低,同时多尺度预测使对小目标的检测性能有明显提高,但是当IOU>0.5时,检测精度没有明显提升。

2020年4月,Bochkovskiy等在YOLOv3的基础上进行改进,提出了YOLOv4[46]。该模型主要做了以下改进:(1)采用CSPDarkNet53[47]主干网络代替DarkNet-53;(2)用SPP+PAN(Path Aggregation Network)代替FPN来融合不同尺寸特征图的特征信息,SPP模块增加了感受野,PAN进行多通道特征融合;(3)采用了CutMix数据增强和马赛克(Mosaic)数据增强;(4)DropBlock正则化。相较于YOLOv3,YOLOv4在保证速度的同时,大幅度提高了模型的检测精度。在目前所有实时性目标检测算法中精度最高,在COCO数据集上的mAP为43.5%。

2020年6月,Jocher提出了YOLOv5[48]。YOLOv5一共有YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x四个网络模型。YOLOv5的网络结构分为输入端、Backbone、Neck、Prediction四个部分:(1)输入端采用了Mosaic数据增强、自适应锚框计算、自适应图片缩放。(2)Backbone中加入了Focus结构和CSP结构,Focus结构用来实现切片操作,例如在YOLOv5s中,原始608×608×3的图像输入Focus结构后,采用切片操作,先变成304×304×12的特征图,再经过一次32个卷积核的卷积操作,最终变成304×304×32的特征图;YOLOv4中只有主干网络使用了CSP结构,而YOLOv5中设计了两种CSP结构,以YOLOv5s网络为例,CSP1_X结构应用于Backbone主干网络,另一种CSP2_X结构则应用于Neck中。(3)与YOLOv4相同,YOLOv5的Neck采用FPN+PAN的结构,另外Neck结构中的CSP2_X结构加强了网络特征融合。(4)Prediction部分采用GIOU_Loss做Bounding Box的损失函数。YOLOv5的检测速度很快,每幅图片的推理时间达到0.007 s,也就是140 frame/s。

YOLO系列在处理不常见的比例目标的泛化过程效果不好,需多次下采样得到标准特征。而且由于边界框预测时空间限制的影响,使其对小目标检测的检测效果不好。

2.2.2 SSD及其改进目标检测算法

RCNN系列和YOLO在速度和准确性上各有优势。RCNN系列具有较高的检测精度,但速度较慢。YOLO虽然检测速度快,但对大维度变化目标的泛化能力好,对小目标的检测效果较弱。结合Faster RCNN和

YOLO的优点,Liu等提出SSD(Single Shot multi-box Detector)算法[49]来平衡检测精度和检测速度,SSD使用VGG-16骨干网络进行特征提取,用第6、第7卷积层代替FC6和FC7,并添加了4个卷积层。SSD网络的设计思想是分层提取特征,将单级网络划分为6级,每个阶段提取不同语义层次的特征图进行目标分类和边界框回归,多尺度特征图与anchor机制相结合提升了算法对不同尺度目标的检测能力。另外,SSD还采用了目标预测机制,依据anchor在不同尺度上得到的候选框来判别目标种类和位置。这种机制存在以下优点:(1)通过卷积层预测目标位置和类别,减少了计算量;(2)没有对目标检测过程设置空间限制,可以有效检测成群的小目标物体。SSD在Nvida Titan X的运行速度提升至59 frame/s,明显优于YOLO;在VOC2007数据集上的mAP达到79.8%,是Faster R-CNN的3倍。可是,SSD对小目标的分类效果不好,由于不同尺度的特征图相互独立,造成不同尺寸的检测框对同一目标重复检测。

2017年,Jeong等提出了R-SSD(Rainbow SSD)[50]检测模型,增加了反卷积模块和预测模块。每个预测层包含预测模块中的残差单元,然后将残差块和预测层的输出因子相加;反卷积模块通过提高特征图的分辨率来增强特征,经过一个预测模块后,每一个反卷积层都可以预测不同大小的目标。提升了小目标物体检测效果。R-SSD算法在VOC2007数据集输入300×300尺寸时的mAP达到78.5%,检测速度为35 frame/s;在VOC2012数据集输入512×512尺寸时mAP达到80.8%,检测速度为16.6 frame/s。

为了提高SSD低层特征图的表达能力,同年Fu等提出了DSSD[51]检测模型。DSSD以ResNet-101为骨干网络,增强网络特征提取能力;添加了反卷积模块和跳跃连接,增强了低层特征图的表达,实现了一定程度的特征融合;设计预测模块,通过融合后的结构特征图进行目标检测。DSSD在VCO2007数据集的mAP达到81.5%,在VOC2012数据集上mAP为80.0%,在COCO数据集的mAP为33.2%。提升了小目标检测精度,但由于Resnet-101骨干网络太深,训练变慢,降低了检测速度。

2017年,Li等提出了F-SSD(Feature Fusion Single Shot Multibox Detector)[52]。该模型充分融合了不同层和不同尺度的特征,通过下采样块生成新的特征金字塔,将其反馈给多盒检测器来预测最终的检测结果,检测精度得到明显提升。在PASCAL VOC2007数据集上mAP达到82.7%,速度为65.8 frame/s。

Shen等提出了深度监督对象检测器(Deeply Learning Supervised Object Detectors,DSOD)[53]算法。该模型采用了DenseNet特征提取网络结构来避免梯度消失;利用Dense Prediction结构,大大减少模型参数量;设计stem结构来减少图片信息的丢失,以提升检测精度。DSOD检测框架不需要预训练模型从零开始训练数据,其目标检测效果相比SSD得到了提升。

单阶段检测器(YOLO和SSD)由于其轻量级骨架网络比双阶段框架需要更少的时间,在检测中避免了预处理算法,需要很少的候选区域。骨干网络的特征提取器在目标检测中耗时较长,使用性能更好的骨干网络加快了检测速度。

2.2.3 RetinaNet算法

2017年,Lin等[54]借鉴了Faster R-CNN和多尺度目标检测[55]的思想设计训练出RetinaNet目标检测器,该模型的主要思想是通过重塑Focal Loss损失函数来解决之前检测模型在训练过程中训练样本中出现的正负样本类不平衡问题。RetinaNet网络是由ResNet骨干网络和两个有特定任务的FCN子网络组成的单一网络,骨干网络负责在整个图像上计算卷积特征,Regression子网络在骨干网络的输出上执行图像分类任务,Classification子网络负责卷积边框回归。

在单阶段检测器中,前景背景的类别不平衡是导致网络训练收敛的主要原因。在训练阶段,Focal Loss避免了大量简单负例,集中于难训练样本。通过训练不平衡正实例和负实例,继承了单阶段检测器的速度。实验结果表明,在MS COCO测试集上,RetinaNet使用ResNet-101-FPN骨干网络与DSSD513相比,AP提升了6%;使用ResNeXt-101-FPN,RetinaNet的AP提升了9%。

2.2.4 Tiny RetinaNet算法

2020年,Cheng等[56]提出Tiny RetinaNet,使用MobileNetV2-FPN作为骨干网络进行特征提取,主要由Stem块骨干网络和SEnet以及两个具有特定任务的子网组成,提高了准确性,减少了信息损失,使用RetinaNet Focal Loss作为分类损失。在PASCAL VOC2007和PASCAL VOC2012数据集的mAP分别为71.4%和73.8%。

2.2.5 M2Det算法

2019年,Zhao等提出了基于多级特征金字塔网络(Multi-Level Feature Pyramid Network,ML-FPN)的M2Det[57],解决了目标实例之间尺度变化的问题。该模型通过三个步骤来实现最终的增量特征金字塔:(1)从骨干网络的大量层中提取多层特征融合为基本特征;(2)将基层特征送入TUM(Thinned U-shape Modules)模块和FFM(Feature Fusion Modules)模块连接组成的一个块中,得到TUM解码层作为下一步骤的输入;(3)等效尺度的解码层集成构造出多层特征的特征金字塔。

M2Det采用VGG骨干网络,在MS COCO测试数据集上采用单尺度推理策略以1.8 frame/s速度获得41.0%AP,采用多尺度推理策略获得44.2%AP。

2.2.6 CornerNet和CenterNet算法

最近,Law和Deng[58]对锚框在SoA目标检测框架中

所起的主导作用提出质疑。文献[58]认为锚框的使用，特别是在单阶段检测器中有缺陷，例如在正例和负例之间造成巨大的不平衡，减慢训练速度和引入额外的超参数。借用了在多人姿态估计中关联嵌入的思想，提出了CornerNet，将边界框目标检测定义为检测左上角和右下角的一对关键点。在CornerNet的骨干网由两个stacked Hourglass Network组成[59]，采用简单的角池方法更好地定位角。CornerNet在COCO的AP达到42.1%，优于以往所有单阶段检测器；然而，在Titan X GPU上的平均推断时间约为4 frame/s，明显低于SSD和YOLO。因为CornerNet很难判断哪些对关键点应该被分组到相同的目标中，所以会产生错误的边界框。为了进一步改进CornerNet，Duan等[60]提出了CenterNet，通过在建议的中心额外引入一个关键点作为一个关键点的三元组来检测每个目标，将在COCO的AP提高到47.0%，但推理速度比CornerNet慢。

2.2.7 单阶段目标检测算法性能对比

单阶段目标检测算法起步晚于双阶段目标检测算法，但借助其结构更精简、计算高效的优势得到了很多研究者的关注，发展过程十分迅速。早期的单阶段目标检测算法往往检测速度快，但检测精度与两阶段检测算法差距大。随着计算机视觉的快速发展，目前单阶段目标检测框架的速度和精度性能都有很大的提升。表3总结了单阶段检测算法的骨干网络以及在PASCAL VOC2007测试集、PASCAL VOC2012测试集和COCO测试集上的检测精度（mAP）和检测速度，“—”表示无数据。表4总结了单阶段目标检测算法的优缺点及适用场景。

从表3看出，单阶段目标检测算法引入了新的骨干网络Darknet-53[45]和MobileNet[61]，检测速度在不断提高，且检测精度也在不断提升并超过双阶段目标检测算法。YOLOv4在速度和精度达到较高的平衡，YOLOv5的小尺寸模型使其可以快速部署、嵌入移动端。RetinaNet对中小型目标的检测精度有显著提高。

从表4看出，单阶段目标检测算法使用金字塔来处理姿态变化和小目标检测问题、新的训练策略、数据增加、不同骨干网络的结合、多种检测框架等来提升目标检测的性能。YOLO系列对小尺度和密集目标检测效

表3 单阶段目标检测算法性能对比

算法	骨干网络	检测速度/(frame·s^{-1})	mAP/%		
			VOC2007	VOC2012	COCO
YOLO[41]	VGG-16	45.0	63.4	57.9	—
YOLOv2[42]	Darknet-19	40.0	78.6	73.5	21.6
YOLOv3[44]	Darknet-53	51.0	—	—	57.9
YOLOv4[46]	CSPDarknet-53	23.0	—	—	43.5
SSD[49]	VGG-16	19.3	79.8	78.5	28.8
R-SSD[50]	ResNet	35.0	78.5	80.8	—
	VGG-16	16.6	80.8	—	
DSSD321[51]	ResNet-101	9.5	78.6	76.3	33.2
DSSD513[51]		5.5	81.5	80.8	
F-SSD[52]	VGGNet	65.8	82.7	—	—
DSOD300[53]	DS/64-192-48-1	17.4	77.7	76.3	—
RetinaNet[54]	ResNeXt-101+FPN	5.4	—	—	40.8
Tiny RetinaNet[56]	MobileNetV2-FPN	—	71.4	73.8	—

表4 单阶段目标检测算法优缺点及适用场景

模型	优点	缺点	适用场景
YOLO	将图像划分为网格单元，检测速度快	对密集和小目标检测效果不好	目标检测
YOLOv2	使用聚类产生锚框，提高分类精度	使用预训练，难迁移	目标检测
YOLOv3	借鉴残差学习思想，实现了多尺度检测	模型复杂，中、大尺度目标检测效果差	多尺度目标检测
YOLOv4	检测精度和检测速度的trade-off很优	检测精度有待提高	高精度实时目标检测
YOLOv5	模型尺寸小，降低部署成本，灵活性高，检测速度高	性能有待提高	目标检测
SSD	多尺度锚框对边界空间离散化	准确率低，模型难收敛，对小目标检测效果提升不大	多尺度目标检测
DSSD	将ResNet-101作为骨干网络，提升小目标检测效果	与SSD相比检测速度较慢	目标检测
R-SSD	改进特征融合方式，检测精度提高	模型计算复杂，检测速度一般	目标检测
F-SSD	重构金字塔特征图以融合不同尺度特征，有利于小目标检测	与SSD相比检测速度较慢	多尺度目标检测
DSOD	不需要预训练	检测速度一般	目标检测
RetinaNet	通过Focal Loss优化正负样本比例	在密集样本训练时会造成样本不平衡	轻量级、多尺度目标检测

果不佳,SSD系列对此进行了改进,实现了高精度、多尺度检测。

2.3 基于生成对抗网络的目标检测算法

Goodfellow等[62]于2014年提出了对抗生成网络(Generative Adversarial Networks,GANs),是一种无监督生成模型,基于最大似然原则工作,并使用对抗训练。对抗学习的思想是通过对抗网络生成遮挡和变形图像样本来训练检测网络,是生成数据分布的常用生成模型方法之一。GAN不仅仅是一个图像生成工具,它还从训练数据中检索有用的信息,以便在不同的领域中执行目标检测、分割和分类任务。

2.3.1 A-Fast-RCNN算法

2017年Wang等[63]引入了对抗网络的思想,提出了使用对抗网络生成困难正样本的A-Fast-RCNN算法。A-Fast-RCNN将对抗学习与Fast R-CNN结合,通过GANs来生成困难样本(Hard Example),以增加遮挡和姿态变化的目标的数量。区别于传统直接生成样本图像的方法,该方法在特征图上采用了一些变换:(1)在处理遮挡的对抗网络(Adversarial Spatial Dropout Network,ASDN)中,添加了一个Mask层来实现特征的部分遮挡,根据loss选择Mask;(2)在处理形变的对抗网络(Adversarial Spatial Transformer Network,ASTN)中,通过操纵特征相应来实现特征的部分变形。STN包括定位网络(localisation network)、网格生成器(grid generator)和采样器(sampler)三部分。定位网络 估计形变参数(如旋转、平移距离和缩放因子),网格生成器和采样器用这些参数来产生形变后的特征图。ASDN和ASTN提供了两种不同的变化,通过将这两种变形相结合(ASDN输出作为ASTN的输入),检测器可以训练得更加鲁棒。ASTN和随机抖动(random jittering)做了对比,使用AlexNet,mAP分别是58.1%和57.3%,使用VGG16,mAP分别是69.9%和68.6%,ASTN的表现都比随机抖动效果好。在与OHEM(Online Hard Example Mining)方法对比中,在VOC 2007数据集上,该方法略好(71.4% vs. 69.9%),而在VOC 2012数据集上,OHEM更好(69.0% vs. 69.8%)。将对抗网络引入目标检测,确实开了先河,从改善效果上来讲,并不比OHEM有多好的效果,某些遮挡样本可能会导致错分的情况。

2.3.2 SOD-MTGAN算法

为了提升小目标检测精度,2018年Bai等[64]提出了一种端到端的多任务生成对抗网络(Small Object Detection via Multi-Task Generative Adversarial Network,SOD-MTGAN)算法。其生成器是一个超分辨率网络,可以将小的模糊图像上采样到精细图像,并恢复详细信息以进行更精确的检测。判别器是一个多任务网络,该网络用真/虚分数、目标类别分数和边界框回归量来描述每个超分辨图像块。此外,为了使生成器获得更多细节信息以便于检测,在训练过程中,将判别器中的分类和回归损失反向传播到生成器中。在COCO数据集上进行的大量实验证明了该方法从模糊的小图像中恢复清晰的超分辨图像的有效性,并表明检测性能(特别是对于小型物体)比最新技术有所提高。

2.3.3 SAGAN算法

传统的卷积生成对抗网络CGANs仅在低分辨率的特征地图上生成空间局部点的函数,从而生成高分辨率的细节。Zhang等[65]提出的自注意生成对抗网络(Self-Attention Generative Adversarial Network,SA-GAN),允许注意驱动和长期依赖建模的图像生成任务。其可以通过所有特征位置的线索生成细节,还应用了光谱归一化来增强训练的动态性,并取得了显著的效果。

2.3.4 Your Local GAN算法

Daras等[66]提出了一种二维局部注意力机制的生成模型(Two Dimensional Local Attention Mechanisms for Generative Models),引入了新的保留了二维几何和局部性的局部稀疏注意层。其替换了SAGAN(Self-Attention Generative Adversarial Networks)的密集注意力层,在ImageNet上,FID分数从18.65优化到15.94。该方法中提出的新层的稀疏注意模式使用信息流图的新信息理论准则来设计,同时还提出了一种新的逆转对抗生成网络注意的方法。

2.3.5 MSG-GAN稳定图像合成算法

生成对抗网络GAN虽然在图像合成任务中取得了一部分成就,但其无法适应不同的数据集,部分原因是训练期间的不稳定性和对超参数的敏感性。造成这种不稳定性的一个原因是当真实和虚拟分布的支持没有足够的重叠时,从鉴别器传递到生成器的梯度将变得信息不足。针对以上问题,Karnewar等[67]提出了多尺度梯度生成对抗网络(Multi-Scale Gradient Generative Adversarial Network ,MSG-GAN),其允许梯度在多个尺度上从鉴别器到流向产生器,为高分辨率图像合成提供了一种稳定的方法。MSG-GAN在不同大小、分辨率和域的数据集,以及不同的损失函数和架构上都可以稳定收敛。

研究者还派生出了大量的GAN变体,如CGAN、WGAN、Progressive GAN、图像到图像翻译GAN、循环GAN、SR GAN、文本到图像GAN、面部修复GAN、文本到语音GAN等,用于各种应用。图2按时间顺序展示了一些流行的GAN的演变过程。

2.4 其他改进算法

2.4.1 基于单阶段、双阶段目标检测算法的结合

目前,学者们对双阶段目标检测算法和单阶段目标检测算法都已经进行了大量的研究,使之都具备了一定

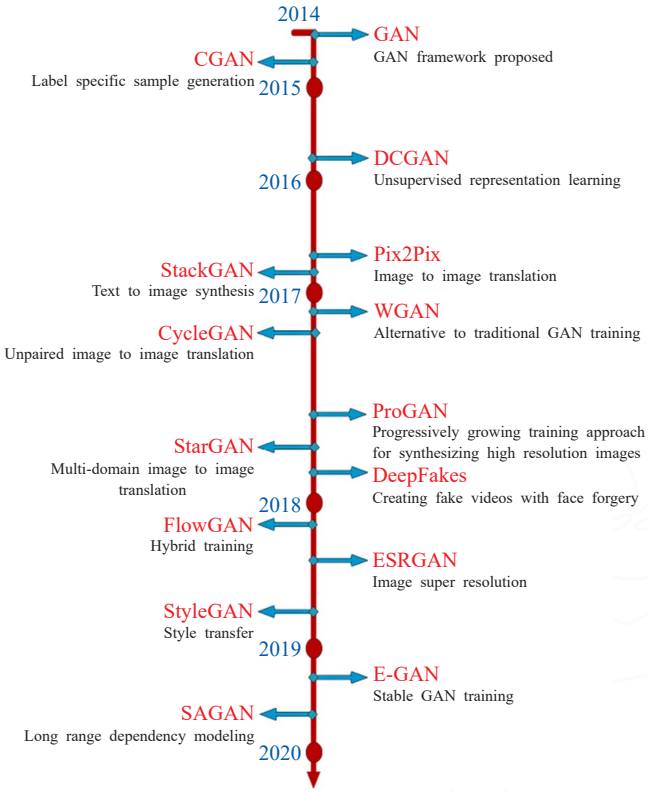


图2 主流GAN演变过程

的理论基础。双阶段目标检测算法在检测精度上占优势,需要不断改进以提高检测速度;单阶段目标检测算法在检测速度上占优势,需要不断提升模型以增加检测精度,所以有研究者将两类算法模型结合起来以平衡检测精度和检测速度。

(1)RON算法

2017年,RON(Reverse connection with Objectness prior Networks)[68]算法是在以Faster R-CNN为代表的双阶段检测框架和以YOLO、SSD为代表的单阶段检测框架的基础上提出的高效通用的目标检测算法框架。在全卷积网络下,与SSD相似,RON使用VGG-16作为骨干网络,不同之处在于RON将VGG-16网络的第14与第15全连接层变成了核大小为2×2、步长为2的卷积层以实现特征图降采样处理,同时采用反向连接方式将相邻位置的待检测特征图联系起来,实现多尺度目标定位并达到更好的效果;与Faster R-CNN相似,RON提出了与RPN网络思想类似的目标先验(Objectness Prior)策略,具体来说,在Softmax后添加一个3×3×2的卷积核来指示在每个参考框中是否存在物体,通过多任务损失函数优化反向连接、目标先验和目标检测器连接,从而使RON可以直接从各个特征图的各个位置预测最终的搜索结果,大大减少目标的搜索空间。在测试中,RON达到了先进的目标检测性能,输入384×384尺寸图片,在PASCAL VOC2007数据集的mAP达到81.3%,在PASCAL VOC 2012数据集上mAP提升至80.7%。

(2)RefineDet算法

2018年,Zhang等设计了RefineDet[69]算法,继承了单阶段检测器和双阶段检测器的优点。RefineDet用VGG-16或者ResNet-101作为骨干网络进行特征提取,将颈部结构(特征金字塔和特征融合)集成到头部结构中。RefineDet的头部结构由ARM(Anchor Refinement Module)模型、传递连接TCB(Transfer Connection)和目标检测模块(Object Detection Module,ODM)组成:

① ARM的四个特征图来自骨干网络的不同层,ARM旨从密集的候选框粗略地过滤负样本减少分类器的搜索空间,调整定义框的位置和大小,即分类和回归步骤。

② TCB旨在将Refined box转移到ODM,并将ARM的浅层和深层特征信息融合。

③ ODM类似于SSD结构,将refined box作为TCB的输入,输出预测多类别标签和refined box的定位,即分类和回归步骤。在测试阶段,通过NMS处理得到图像补丁级的预测结果。

作为一种单阶段目标检测模型,RefineDet具有两阶段模型(即两步级联回归和两级分类)的特点,可以更好地预测难检测的目标,尤其对小目标的准确定位。RefineDet在VOC2007和VOC2012数据集的mAP分别为85.8%和86.8%,输入320×320尺寸时运行速度为40.2 frame/s。

2.4.2 Relation Networkfor Detecting Objects算法

图像中呈现的不同目标具有不同的外观特征和几何信息,需要相互作用。考虑这一需求,2018年,Hu等[70]提出了Relation network目标检测算法,增加一种改进的注意力网络,称为对象关系模块,该模块插入到网络的两个完全连接层之前,提取优越的特征,以便对目标对象进行适当的分类和定位。这种增加的关联单元为回归器和分类器提供了更好的特征,但代替了NMS后处理步骤,从而提高了整体检测精度。

2.4.3 DCN算法

常规卷积神经网络考虑了预先定义的正方形大小核的特征,这使得可服从区域字段不能完全覆盖目标对象的整个像素区域。Zhu等[71]提出了可变形卷积网络(Deformable Convolutional Network v2,DCNv2)算法,该模型反映了目标空间支持区域的几何不变性。可变形卷积网络产生可变形核,而且从网络中学习固定尺寸初始卷积核的偏移量。但可变形RoI可以提高不同形状的目标定位。DCN的检测精度比传统convents提高了近4%,并在MSCOCO上实现了37.5%的mAP。DCNv2比DCNv1使用更多的层,可学习标量对所有可变形层进行调制,提高了精度和可变形效果。

2.4.4 NAS-FPN算法

2019年,NAS-FPN[72](Neural Architecture Search Feature Pyramid Network)是从谷歌大脑适应性神经结构搜索中发现的一种新特征金字塔结构,为不同尺度的特征融合提供了自顶向下和自底向上的连接。在搜索阶段,它重复 N 次FPN结构并以纪念性结构的形式串联起来,通过高级特征层模拟任意级别的特征。大多数高效结构通过利用高分辨率输入特征图和输出层之间的连接产生高分辨特征来识别小目标。其采用大容量架构、叠加金字塔网络和增加特征维数显著提高了检测精度。在COCO测试集上,NAS-FPN算法的平均检测精度mAP比原来采用ResNet-50骨干网络的FPN提升了2.9%,NAS-FPN使用 AmoebaNet[73]骨干网络在COCO测试集达到48.3%mAP。

2.4.5 其他改进算法的性能对比

除了单、双阶段目标检测算法,研究者还探究了以上其他改进算法,总体而言,这些算法的检测精度和速度都有提高,实现了多尺度目标检测。表5总结了上述改进算法的骨干网络以及在PASCAL VOC2007测试集、PASCAL VOC2012测试集和COCO测试集上的检测精度(mAP)和检测速度,"—"表示无数据。表6总结了这些检测算法的优缺点及适用场景。

从表5可以看出,结合单、双阶段的RON和RefineDet算法的检测精度很高,NAS-FPN算法的检测速度为92.1 frame/s,达到最快。

从表6可以看出,研究者们试图从多方面来改进算法的性能,增加小目标和遮挡目标的检测效果,DCNv2可以对几何变换建模学习,使其可以应用到更多现实场景。

3 目标检测算法的应用

目标检测作为计算机视觉的三大基本任务之一,在现实场景中得到了广泛的应用。在实际应用场景中根据不同的任务需求实现了不同的技术,主要包括人脸检测、显著目标检测、行人检测、遥感图像检测、医学图像检测和粮虫检测等重要领域。

3.1 人脸检测

人脸检测是目标检测中最重要的应用领域,是人脸识别、人脸对齐、性别识别、情绪分析的基础。在实际场景中,由于面部特征、光照、亮度、遮挡、分辨率、噪声和相机畸变等多方面因素影响,使目标检测更具挑战性。

人脸检测的目的是确定图像中是否有人脸并找出其位置。随着深度学习时代的到来,基于深度学习的人脸检测取得了很大的成功。Li等提出了一种Cascade CNN[74],其包含大量的级联DCNN分类器在一定程度上解决了实际场景中光照和角度的敏感性问题。Zhang等[75]提出了一种使用CascasdeRCNN级联架构的多任务人脸检测算法,即MTCNN(Multi-Task Cascaded Convolutional Neural Networks)。为提升分类器的性能,Jiang等[76]提出了一种基于Faster RCNN的Face RCNN,添加了基于softmax的中心损失。Najibi等[77]提出SSH(Single Stage Headless)通过检测不同尺度特征图来实现多尺度人脸检测。

3.2 行人检测

行人检测在智能监控、自动驾驶和机器人导航中得到了广泛的应用。行人检测面临的问题比目标检测要复杂得多,行人目标同时具有静态和动态的特点,更容易受到姿态、抓捕的角度、模糊程度、背景、光照的影响,使行人检测的难度更高。

表5 其他改进算法性能对比

算法	骨干网络	检测速度/(frame·s^{-1})	mAP/%		
			VOC2007	VOC2012	COCO
RON[68]	VGG-16	15.0	81.3	80.7	—
RefineDet320[69]	VGG-16	40.3	80.0	78.1	44.4
RefineDet512[69]		24.1	81.8	80.1	48.3
Relation Network[70]	ResNet 50	—	—	—	35.2
DCNv2[71]	ResNet-101	—	—	—	37.5
NAS-FPN[72]	ResNet-50	92.1	—	—	48.3

表6 其他改进算法优缺点及适用场景

模型	优点	缺点	适用场景
RON	提升小目标检测性能	检测速度不高	目标检测
RefineDet	提升小目标和遮挡检测效果	没有解决特征校准问题	目标检测、实例分割、全景分割
Relation Network	提高目标识别准确率,NMS被轻量的关系网络替换	多个目标重叠时不确定具体学习到什么内容	目标检测、场景分割
DCNv2	引入可变形卷积结构,能对几何变换建模学习	增加对无用信息检测	目标检测
NAS-FPN	更好的准确性和延迟权衡,生成多尺度特征表示,随时检测	模型难训练,耗时多	目标检测

Faster RCNN在行人检测中的效果并不好。Zhang等[78]分析后提出利用基于Faster RCNN的RPN处理小目标和负样本,然后使用随机森林对候选区域分类。对于行人的多尺度问题,Li等[79]设计了两个子网络同时检测大尺度和小尺度目标,然后利用scale-aware合并两个子网络。为了解决遮挡问题,Tian等[80]提出了DeepParts,将人体分割成多个部分进行检测,然后再合并。

3.3 遥感影像检测

遥感影像检测主要用于军事侦察、国土资源调查、城市规划和交通导航等领域。这些目标包括飞机、船舶、车辆、道路、港口和各种建筑物。遥感图像主要存在以下困难:(1)遥感图像视距大导致图像分辨率大,对于图像中较小尺寸的目标在大视点角度下,导致遥感图像小目标检测比较困难;(2)常见图像中的目标大多都是水平的,在拍摄遥感图像时,目标的旋转不变性也是一个重要的问题;(3)遥感图像的背景相当复杂。

目前,基于深度学习的遥感图像检测正致力于解决这些问题。Etten[81]基于高速YOLOv2提出了YOLT(You Only Look Twice),通过两次检测提高了高分辨率遥感图像的检测速度。同时,为了提升小目标检测效率,提高了特征图的分辨率,Long等[82]提出了一种结合非最大抑制的无监督评分边界框回归(Unsupervised Score-Based Bounding Box Regression,USB-BBR)方法来优化边界框,增强对小目标物体的定位能力。Wang等[83]提出一种端到端多尺度视觉注意力网络(Multi-Scale Visual Attention Network,MS-VANs),核心思想是为每个尺度的特征图学习一个注意网络,以突出目标和抑制背景。

3.4 医学图像检测

医学图像检测可以帮助医生快速找出病变区域,提高医学诊断的准确性,减少医生的人工工作量。

Li等[84]将基于DCNN的注意机制应用于青光眼的检测。Kawahara等[85]提出了Multi-stream CNN对皮肤损伤进行分类。Kong等[86]提出了一种结合LSTM-RNN和CNN检测心脏(Magnetic Resonance Imaging,MRI)图像中的舒张末期和收缩末期框架的方法。Hwang等[87]提出了弱监督深度学习的方法来检测结节中的病变和胸部X-rays。

3.5 粮虫检测

粮虫检测是目标检测在农业研究领域的重要应用之一。储粮过程中,谷虫侵虐会造成粮食品质下降,高效检测储粮害虫可以减少经济损失,提升储粮环境质量。在实际粮虫检测中,粮虫图像检测存在诸多问题,如谷虫与粮粒之间的混淆、谷粒对谷虫的遮挡、谷虫对谷虫的遮挡等。

为解决真实场景中虫害多的问题,研究者们做了大量研究工作。Shen等[88]采用Faster R-CNN模型对6类储粮害虫的检测准确率达88.02%。Liu等[89]提出了PestNet网络结构,并使用RPN和位置评分敏感性等技术对田间害虫进行了检测,取得了不错的效果。Xia等[90]提出基于神经网络的区域建议网络来检测农作物中的昆虫,取得了较高的准确率。Shi等[91]提出改进的R-FCN网络结构解决了8种常见储粮害虫的检测和分类问题。

3.6 显著目标检测

显著目标检测的作用是突出图像中的主要目标区域,也称作显著区域,是视频中目标检测的一个重要而活跃的领域,越来越受到研究者的关注。

DCNN由于其较强的特征表达能力被引入显著目标检测。Zhao等[92]提出了一种多情景深度学习框架(Multi-Context Deep Learning,MCDL),基于多层感知器(Multi-Layer Perceptron,MLP)提取局部和全局上下文,对前景和背景进行分类。虽然基于多层感知器的方法在性能上有所提高,但它对空间信息不敏感,且耗时较长。目标最先进的显著目标检测是基于全卷积网络的。Hu等[93]提出了一种深度网络集来生成一种紧凑而统一的显著性映射从物体边界中区分像素。Kousik等[94]结合RCNN建立了一个利用时间、空间和局部约束线索来实现全局优化的时空模型,实现在基准动态视频数据集中寻找显著目标的任务。

3.7 异常检测

异常检测技术在欺诈检测、医疗保健监测和天气监视中发挥了重要作用。目前的异常检测技术使用逐点标准[95]来研究数据。为了分析连续时间和空间间隔,Barz等[96]提出了一种称为最大发散间隔的无监督异常检测技术。

3.8 3D Pointclouds目标检测

3D点云目标检测提供了深度信息,可以进一步精确定位目标和表征形状。基于激光雷达点云的3D目标检测在自动驾驶、机器人和虚拟现实应用中发挥着重要作用。但该技术还面临着激光雷达点云稀疏、3D空间采样不一致、遮挡和相对位姿变化等挑战。Qi等[97]开发了一种称为PointNet的端到端深度神经网络,可以直接从激光雷达云学习特征。为了有效地映射和处理海量3D数据,Engelcke等[98]提出了稀疏卷积层和L1正则化。Zhou等[99]开发了一种名为VoxelNet的通用端到端3D目标检测框架,它可以通过学习点云的鉴别特征表示来预测正确的3D边界框。

3.9 细粒度视觉识别

细粒度视觉识别的目的是在每个基本层次类别中识别一个确切的目标类别,比如识别汽车的模型或识别

哺乳动物的种类。不同类别间的视觉差异极小,而且易受姿势、目标的位置和给定图像中的视点等因素影响。Krause等[100]利用3D目标表达在位置和局部特征的层次上跨多个视点进行泛化。Lin等[101]提出了由两个CNN流组成的双线性模型,其中两个CNN流的输出在每个图像位置使用外积相乘,然后将外积合并在一起得到一个图像描述符。He等[102]提出一种基于显著性引导的faster R-CNN的细粒度判别定位方法,后来又提出了一种用于快速细粒度图像分类的弱监督版本。

3.10 小目标检测

小目标普遍存在像素低、被遮挡现象,造成误检、漏检等检测精度不高的问题,很多学者对小目标检测展开了很多研究。冯小雨等[103]对Faster RCNN进行改进优化使其在对空中目标检测时弥补了对弱小目标和被遮挡目标不敏感的缺陷,提高了检测速度和精度。梁延禹等[104]提出了一种多尺度非局部注意力网络的小目标检测方法,在浅层利用非局部通道注意力模块整合特征的全局空间信息,进而对通道间的信息进行校准,提升了小目标的检测精度。奚琦等[105]提出一种改进的MDSSD小目标实时检测算法,在VOC2007数据集的mAP达到81.7%。岳晓新等[106]通过对YOLOv3进行改进,然后对道路小目标检测,使检测精度提升了2.36%。

4 未来研究展望

基于深度学习的目标检测技术由于其强大的学习能力和在处理遮挡、尺度变化和背景交换方面的优越性,使其成为一个高度热门的研究领域。本文介绍了基于深度卷积神经网络的目标检测算法的发展情况,在网络架构层面总结了单阶段、双阶段两大类目标检测器,即基于区域的目标检测网络和基于回归分析的目标检测网络以及其他改进算法;深入分析了各类算法的网络结构、优势缺点和适用场景,并对比分析常用数据集以及各类相关算法在主流数据集上的实验结果,最后总结了目标检测的一些应用领域以全面了解目标检测,并分析了其未来发展趋势。随着视觉目标检测器在安全、交通、军事等领域的应用日益强大,目标检测的应用也随之急剧增加。尽管取得了这些进步,但仍有许多进一步发展的空间。下面提供了这一领域的一些最新趋势,以促进使用深度学习进行视觉目标检测的未来研究。

(1)视频目标检测:视频目标检测存在运动目标不均匀、目标非常微小、截断和遮挡等问题,很难实现高精度和高效率。因此,研究基于运动的目标和视频序列等多方面数据源将是最有前景的未来研究领域之一。

(2)弱监督目标检测:弱监督目标检测模型旨在于利用一小组完全注释的图像来检测大量的非注释的对应目标。因此,使用大量带有目标物体和边界框的标注和标记图像来高效训练网络以达到较高的有效性,是未来研究的一个重要问题。

(3)多领域目标检测:针对区域的检测器往往性能更好,在预定义的数据集上实现了较高的检测精度。因此,开发一种通用的目标检测器使其能够在没有任何先验知识的情况下检测出多领域的目标是未来的基本研究方向。

(4)显著目标检测:指该区域用于强调图像中的显著目标区域。显著性目标检测应用范围广泛,适用于不同领域的目标检测应用。在每一帧中突出的目标重要区域有助于准确检测连续场景或视频序列中的目标。因此,在重要的识别和检测任务中,显著性引导的目标检测可以看作是一个初步的过程。

(5)多任务学习:为了提高检测性能,积累主干架构的多层特征是一个重要的步骤。同时执行目标检测、语义和实例分割等多种计算机视觉任务可以利用更丰富的信息在很大程度上提高性能。采用这种方法可以有效地将多个任务组合在一个模型中,如何在不影响处理速度的情况下提高检测精度是研究人员面临的一系列挑战。

(6)无监督目标检测:研究自动标注技术以摆脱人工标注是无监督目标检测的一个令人渴望和有前景的趋势。对于尖锐的检测任务,无监督目标检测是未来的研究方向。

(7)遥感实时检测:遥感图像在军事和农业领域都有广泛的应用。自动检测模型和集成硬件单元将促进这些领域的快速发展。

(8)基于GAN的目标检测器:基于深度学习的目标检测器通常需要大量的数据来进行训练,而基于GAN的目标检测器是一种影响较大的结构,它会产生虚拟的图像。结合真实场景和GAN生成的模拟数据,有助于检测器获得更好的鲁棒性,并获得更强的泛化能力。

基于深度学习的目标检测技术的研究还需要进一步深入。希望基于深度学习的目标检测器未来能够为便利人们生活做出更多贡献。

参考文献:

[1] 李航.基于深度学习目标检测的算法研究[D].长春:中国科学院大学(中国科学院长春光学精密机械与物理研究所),2020.

[2] DOLLAR P,WOJEK C,SCHIELE B,et al.Pedestrian detection:a benchmark[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,2009:304-311.

[3] UIJLINGS J R,VAN DE SANDE K E A,GEVERS T,et al.Selective search for object recognition[J].International Journal of Computer Vision,2013,104(2):154-171.

[4] VEDALD A,GULSHAN V,VARMA M,et al.Multiple kernels for object detection[C]//IEEE 12th International Conference on Computer Vision,2009:606-613.

[5] YU Y,ZHANG J,HUANG Y,et al.Object detection by context and boosted HOG-LBP[C]//ECCV Workshop on PASCAL VOC,2010.

[6] GUO B T,WANG X R,CHEN Y J,et al.High-accuracy infrared simulation model based on establishing the linear relationship between the outputs of different infrared imaging systems[J].Infrared Physics & Technology,2015,69:155-163.

[7] LIENHART R,MAYDT J.An extended set of haar-like features for rapid object detection[C]//Proceedings of International Conference on Image Processing,2002.

[8] 刘威,靳宝,周璇,等.基于特征融合及自适应模型更新的相关滤波目标跟踪算法[J/OL].智能系统学报:1-8[2020-10-10].http://kns.cnki.net/kcms/detail/23.1538.TP.20200827.1329.012.html.

[9] VIOLA P,JONES M.Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,2001.

[10] 张夏豪,庄毅.基于改进一类支持向量机的云计算故障检测策略[J/OL].计算机科学:1-8[2020-09-10].http://kns.cnki.net/kcms/detail/50.1075.TP.20200414.1535.038.html.

[11] 马原东,罗子江,倪照风,等.改进SSD算法的多目标检测[J].计算机工程与应用,2020,56(23):23-30.

[12] 凌晨,张鑫彤,马雷.基于Mask R-CNN算法的遥感图像处理技术及其应用[J].计算机科学,2020,47(10):151-160.

[13] KRIZHEVSKY A,SUTSKEVER I,HINTON G.ImageNet classification with deep convolutional neural networks [C]//Proceedings of NIPS,2012.

[14] DING L,LI H,HU C,et al.Alexnet feature extraction and multi-kernel learning for object-oriented classification[J].Int Arch Photogramm Remote Sens Spatial Inf Sci,2018,42:277-281.

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

[16] ITO S,CHEN P,COMTE P,et al.Fabrication of screen printing pastes from TiO2 powders for dye sensitised solar cells[J].Progress in Photovoltaics:Research and Applications,2007,15(7):603-612.

[17] MARRIS H,DEBOUDT K,AUGUSTIM P,et al.Fast changes in chemical composition and size distribution of fine particles during the near-field transport of industrial plumes[J].Science of the Total Environment,2012,427:126-138.

[18] LIN T Y,MAIRE M,BELONGIE S,et al.Microsoft coco:common objects in context[C]//European Conference on Computer Vision.Cham:Springer,2014:740-755.

[19] KRASIN I,DUERIG T,ALLDRIN N,et al.OpenImages:a public dataset for large-scale multi-label and multi-class image classification[EB/OL].[2020-08-10].https://storage.googleapis.com/openimages/web/index.html.

[20] 祝文韬,谢宝蓉,王琰,等.光学遥感图像中的飞机目标检测技术研究综述[J/OL].计算机科学:1-8[2020-11-04].http://kns.cnki.net/kcms/detail/50.1075.TP.20200701.0840.002.html.

[21] 刘颖,刘红燕,范九伦.基于深度学习的小目标检测研究与应用综述[J].电子学报,2020,48(3):590-601.

[22] ZITNICK C L,DOLLAR P.Edge boxes:locating object proposals from edges[C]//European Conference on Computer Vision.Cham:Springer,2014:391-405.

[23] HU Q,ZHAI L.RGB-D image multi-target detection method based on 3D DSF R-CNN[J].International Journal of Pattern Recognition and Artificial Intelligence,2019,33(8):1954026.

[24] GIRSHICK R,DONAHUE J,DARRELL T,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.

[25] DAI J,LI Y,HE K,et al.R-FCN:object detection via region-based fully convolutional networks[C]//Advances in Neural Information Processing Systems,2016:379-387.

[26] HE K,GKIOXARI G,PIOTR D,et al.Mask R-CNN[C]//IEEE International Conference on Computer Vision,2017.

[27] SERMANET P,EIGEN D,ZHANG X,et al.Overfeat:integrated recognition,localization and detection using convolutional networks[J].arXiv:1312.6229,2013.

[28] NAQVI S F,ALI S S A,YAHYA N,et al.Real-time stress assessment using sliding window based convolutional neural network[J].Sensors,2020,20(16):4400.

[29] LIN S,JI R,CHEN C,et al.Holistic CNN compression via low-rank decomposition with knowledge transfer[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,41(12):2889-2905.

[30] HE K,ZHANG X,REN S,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.

[31] GIRSHICK R.Fast R-CNN[C]//IEEE International Conference on Computer Vision,2016.

[32] REN S,HE K,GIRSHICK R,et al.Faster R-CNN:towards real-time object detection with region proposal net-

works[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2017,39(6):1137-1149.

[33] 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.

[34] 林凯瀚,赵慧民,吕巨建,等.基于Mask R-CNN的人脸检测与分割方法[J].计算机工程,2020,46(6):274-280.

[35] OUYANG W,WANG K,ZHU X,et al.Chained cascade network for object detection[C]//IEEE International Conference on Computer Vision,2017:1938-1946.

[36] CAI Z,VASCONCELOS N.Cascade R-CNN:delving into high quality object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:6154-6162.

[37] CHEN K,PANG J,WANG J,et al.Hybrid task cascade for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019:4974-4983.

[38] LI Z,PENG C,YU G,et al.Light-head R-CNN:in defense of two-stage object detector[J].arXiv:1711.07264,2017.

[39] ALLEN-ZHU Z,LI Y.What can ResNet learn efficiently,going beyond kernels?[C]//Advances in Neural Information Processing Systems,2019:9017-9028.

[40] HITAWALA S.Evaluating ResNeXt model architecture for image classification[J].arXiv:1805.08700,2018.

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

[42] REDMON J,FARHADI A.YOLO9000:better,faster,stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:6517-6525.

[43] WANG S,LI A,CHEN J,et al.RSnet:an improvement for Darknet[J].arXiv:2002.03729,2020.

[44] REDMON J,FARHADI A.Yolov3:an incremental improvement[J].arXiv:1804.02767,2018.

[45] KIM K J,KIM P K,CHUNG Y S,et al.Performance enhancement of YOLOv3 by adding prediction layers with spatial pyramid pooling for vehicle detection[C]//2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance(AVSS),2018:1-6.

[46] BOCHKOVSKIY A,WANG C Y,LIAO H Y M.YOLOv4:optimal speed and accuracy of object detection[J].arXiv:2004.10934,2020.

[47] MAHTO P,GARG P,SETH P,et al.Refining Yolov4 for vehicle detection[J].International Journal of Advanced Research in Engineering and Technology(IJARET),2020,11(5):409-419.

[48] JOCHER G.Yolov5[EB/OL].[2020-08-10].https://github.com/ultralytics/yolov5.

[49] LIU W,ANGUELOV D,ERHAN D,et al.SSD:single shot multibox detector[C]//European Conference on Computer Vision.Cham:Springer,2016:21-37.

[50] JEONG J,PARK H,KWAK N.Enhancement of SSD by concatenating feature maps for object detection[J].arXiv:1705.09587,2017.

[51] FU C Y,LIU W,RANGA A,et al.DSSD:deconvolutional single shot detector[J].arXiv:1701.06659,2017.

[52] LI Z,ZHOU F.FSSD:feature fusion single shot multibox detector[J].arXiv:1712.00960,2017.

[53] SHEN Z,LIU Z,LI J,et al.DSOD:learning deeply supervised object detectors from scratch[C]//Proceedings of the IEEE International Conference on Computer Vision,2017:1919-1927.

[54] LIN T Y,GOYAL P,GIRSHICK R,et al.Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision,2017:2980-2988.

[55] ERHAN D,SZEGEDY C,TOSHEV A,et al.Scalable object detection using deep neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2014:2147-2154.

[56] CHENG M,BAI J,LI L,et al.Tiny-RetinaNet:a one-stage detector for real-time object detection[C]//Eleventh International Conference on Graphics and Image Processing(ICGIP 2019),2020.

[57] ZHAO Q,SHENG T,WANG Y,et al.M2Det:a single-shot object detector based on multi-level feature pyramid network[C]//Proceedings of the AAAI Conference on Artificial Intelligence,2019:9259-9266.

[58] LAW H,DENG J.CornerNet:detecting objects as paired keypoints[C]//Proceedings of the European Conference on Computer Vision,2018:734-750.

[59] NEWELL A,YANG K,DENG J.Stacked Hourglass networks for human pose estimation[C]//European Conference on Computer Vision.Cham:Springer,2016:483-499.

[60] DUAN K,BAI S,XIE L,et al.CenterNet:keypoint triplets for object detection[C]//Proceedings of the IEEE International Conference on Computer Vision,2019:6569-6578.

[61] Chen H Y,SU C Y.An enhanced hybrid MobileNet[C]//2018 9th International Conference on Awareness Science and Technology(iCAST),2018:308-312.

[62] GOODFELLOW I,POUGET-ABADIE J,MIRZA M,et al.Generative adversarial nets[C]//Advances in Neural Information Processing Systems,2014:2672-2680.

[63] WANG X,SHRIVASTAVA A,GUPTA A.A-Fast-RCNN:hard positive generation via adversary for object detec-

tion[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:2606-2615.
[64] BAI Y,ZHANG Y,DING M,et al.SOD-MTGAN:small object detection via multi-task generative adversarial network[C]//Proceedings of the European Conference on Computer Vision,2018:206-221.
[65] ZHANG H,GOODFELLOW I,METAXAS D,et al.Self-attention generative adversarial networks[C]//International Conference on Machine Learning,2019:7354-7363.
[66] DARAS G,ODENA A,ZHANG H,et al.Your local GAN:designing two dimensional local attention mechanisms for generative models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020:14531-14539.
[67] KARNEWAR A,WANG O.MSG-GAN:multi-scale gradient GAN for stable image synthesis[J].arXiv:1903.06048,2019.
[68] KONG T,SUN F,YAO A,et al.RON:reverse connection with objectness prior networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:5936-5944.
[69] ZHANG S,WEN L,BIAN X,et al.Single-shot refinement neural network for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:4203-4212.
[70] HU H,GU J,ZHANG Z,et al.Relation networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:3588-3597.
[71] ZHU X,HU H,LIN S,et al.Deformable ConvNets v2:more deformable,better results[J].arXiv:1811.11168,2018.
[72] GHIASI G,LIN T Y,LE Q V.NAS-FPN:learning scalable feature pyramid architecture for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019:7036-7045.
[73] SHAH S A R,WU W,LU Q,et al.AmoebaNet:an SDN-enabled network service for big data science[J].Journal of Network and Computer Applications,2018,119:70-82.
[74] LI H,LIN Z,SHEN X,et al.A convolutional neural network cascade for face detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015:5325-5334.
[75] ZHANG K,ZHANG Z,LI Z,et al.Joint face detection and alignment using multitask cascaded convolutional networks[J].IEEE Signal Processing Letters,2016,23(10):1499-1503.
[76] JIANG H,LEARNED-MILLER E.Face detection with the faster R-CNN[C]//IEEE International Conference on Automatic Face & Gesture Recognition,2017:650-657.
[77] NAJIBI M,SAMANGOUEI P,CHELLAPPA R,et al.SSH:single stage headless face detector[C]//Proceedings of the IEEE International Conference on Computer Vision,2017:4875-4884.
[78] ZHANG L,LIN L,LIANG X,et al.Is faster R-CNN doing well for pedestrian detection?[C]//European Conference on Computer Vision.Cham:Springer,2016:443-457.
[79] LI J,LIANG X,SHEN S M,et al.Scale-aware fast R-CNN for pedestrian detection[J].IEEE Transactions on Multimedia,2017,20(4):985-996.
[80] TIAN Y,LUO P,WANG X,et al.Deep learning strong parts for pedestrian detection[C]//Proceedings of the IEEE International Conference on Computer Vision,2015:1904-1912.
[81] VAN ETTEN A.You only look twice:rapid multi-scale object detection in satellite imagery[J].arXiv:1805.09512,2018.
[82] LONG Y,GONG Y,XIAO Z,et al.Accurate object localization in remote sensing images based on convolutional neural networks[J].IEEE Transactions on Geoscience and Remote Sensing,2017,55(5):2486-2498.
[83] WANG C,BAI X,WANG S,et al.Multiscale visual attention networks for object detection in VHR remote sensing images[J].IEEE Geoscience and Remote Sensing Letters,2018,16(2):310-314.
[84] LI L,XU M,WANG X,et al.Attention based glaucoma detection:a large-scale database and CNN model[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019:10571-10580.
[85] KAWAHARA J,HAMARNEH G.Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers[C]//International Workshop on Machine Learning in Medical Imaging.Cham:Springer,2016:164-171.
[86] KONG B,ZHAN Y,SHIN M,et al.Recognizing end-diastole and end-systole frames via deep temporal regression network[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention.Cham:Springer,2016:264-272.
[87] HWANG S,KIM H E.Self-transfer learning for weakly supervised lesion localization[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention.Cham:Springer,2016:239-246.
[88] SHEN Y,ZHOU H,LI J,et al.Detection of stored-grain insects using deep learning[J].Computers and Electronics in Agriculture,2018,145:319-325.
[89] LIU L,WANG R,XIE C,et al.PestNet:an end-to-end deep learning approach for large-scale multi-class pest detection and classification[J].IEEE Access,2019,7:45301-45312.

[90] XIA D,CHEN P,WANG B,et al.Insect detection and classification based on an improved convolutional neural network[J].Sensors,2018,18(12):4169.

[91] SHI Z,DANG H,LIU Z,et al.Detection and identification of stored-grain insects using deep learning:a more effective neural network[J].IEEE Access,2020.

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

[93] HU P,SHUAI B,LIU J,et al.Deep level sets for salient object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:2300-2309.

[94] KOUSIK N V,NATARAJAN Y,RAJA R A,et al.Improved salient object detection using hybrid convolution recurrent neural network[J].Expert Systems with Applications,2020:114064.

[95] SENIN P,LIN J,WANG X,et al.Grammarviz3.0:interactive discovery of variable-length time series patterns[J].ACM Transactions on Knowledge Discovery from Data,2018,12(1):1-28.

[96] BARZ B,RODNER E,GARCIA Y G,et al.Detecting regions of maximal divergence for spatio-temporal anomaly detection[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,41(5):1088-1101.

[97] QI C R,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,2017:652-660.

[98] ENGELCKE M,RAO D,WANG D Z,et al.Vote3deep:fast object detection in 3d point clouds using efficient convolutional neural networks[C]//2017 IEEE International Conference on Robotics and Automation(ICRA),2017:1355-1361.

[99] ZHOU Y,TUZEL O.Voxelnet:end-to-end learning for point cloud based 3d object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:4490-4499.

[100] KRAUSE J,STARK M,DENG J,et al.3d object representations for fine-grained categorization[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops,2013:554-561.

[101] LIN T Y,ROYCHOWDHURY A,MAJI S.Bilinear CNN models for fine-grained visual recognition[C]//Proceedings of the IEEE International Conference on Computer Vision,2015:1449-1457.

[102] HE X,PENG Y,ZHAO J.Fine-grained discriminative localization via saliency-guided faster R-CNN[C]//Proceedings of the 25th ACM International Conference on Multimedia,2017:627-635.

[103] 冯小雨,梅卫,胡大帅.基于改进Faster R-CNN的空中目标检测[J].光学学报,2018,38(6):250-258.

[104] 梁延禹,李金宝.多尺度非局部注意力网络的小目标检测算法[J].计算机科学与探索,2020,14(10):1744-1753.

[105] 奚琦,张正道,彭力.基于改进MDSSD的小目标实时检测算法[J].激光与光电子学进展,2020,57(20):97-105.

[106] 岳晓新,贾君霞,陈喜东,等.改进YOLO V3的道路小目标检测[J].计算机工程与应用,2020,56(21):218-223.