

DOI:10.19651/j.cnki.cmt.2107833

基于深度学习的旋转目标检测方法研究进展*

安胜彪 姜慧儒 陈书旺 白宇

(河北科技大学 信息科学与工程学院 石家庄 050018)

摘要: 在遥感和场景文本图像中,目标具有方向多样性和尺度变化较大等特点,使得常见目标检测方法在这两场景中检测效果不佳,针对此问题,诞生了许多专门设计的检测方法。将朝向角度信息融入候选区域生成网络或设计专门的方向角度预测网络,是目前有朝向目标检测研究的主流方法,其对遥感和场景文本图像检测具有重要意义。本文综述了旋转目标检测在遥感和场景文本两场景中的研究现状,根据有无锚框将当前基于深度学习的旋转目标检测方法分为基于锚框的一阶段方法、基于锚框的二阶段方法和无锚框方法3类方法进行归纳分析,并从优缺点、骨干网络和适用场景等方面进行了对比。最后,对旋转目标检测方法的发展前景和研究方向进行了展望。

关键词: 卷积神经网络;目标检测;旋转目标;深度学习;计算机视觉

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.1050

Research progress of rotating target detection methods based on deep learning

An Shengbiao Lou Huiru Chen Shuwang Bai Yu

(School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China)

Abstract: In remote sensing and scene text images, the target has the characteristics of directional diversity and large scale change, which makes the common target detection methods have poor detection effect in these two scenes. Aiming at this problem, many specially designed detection methods have been born. Integrating the orientation angle information into the candidate area generation network or designing a special orientation angle prediction network is the mainstream method of orientation target detection, which is of great significance to remote sensing and scene text image detection. This paper summarizes the research status of rotating target detection in remote sensing and scene text. According to whether there is an anchor box or not, the current rotation detection methods based on deep learning are divided into three types: one-stage method based on anchor, two-stage method based on anchor and anchor free method, and compared from the aspects of advantages and disadvantages, backbone network and applicable scene. Finally, the development prospect and research direction of rotating target detection methods are prospected.

Keywords: convolution neural network; object detection; rotated object; deep learning; computer vision

0 引言

目标检测是计算机视觉领域的最基本和重要任务之一,其内涵是从图像中定位感兴趣的目标,准确地判断出每个目标的具体类别,并给出每个目标的边界框。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。传统检测方法大多是通过人工手动设计特征对图像进行处理,对特征的处理较为单一,适用性不强,并且存在很多弊端。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,外观形态的多样性、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,泛化性能差。随着深度学习技术的快速发展,运用卷积神经网络自动提取高层特征,提升了特征的表达能力,在准确率和效率上都大幅提升^[1]。目前主流的目标检测算法主要分为二阶段检测算法和一阶段检测算法^[2]。但这些目标检测方法都是通过水平框实现对目标轮廓的勾勒,如果应用到遥感和场景文本中,则会出现大量目标框重叠的现象,无法有效地将旋转目标与背景分离。因此,针对旋转目标的检测方法不断被提出。

遥感和场景文本是旋转检测器的主要应用场景。遥感

收稿日期:2021-09-10

* 基金项目:河北省自然科学基金(F2019208305)项目资助

图像目标检测广泛应用于灾害监测、资源调查、农业产值测算等领域,对于社会 and 经济发展具有重要的意义。另外,在海上救援、分析道路拥堵情况和分析战争形势等^[3]民用和军用领域也有着十分重要的用途。场景文本检测则有助于理解和应用自然场景图像的信息,如互联网图像存储检索、移动端 AR 文字翻译、自动驾驶等应用^[4]。但遥感图像分辨率高、背景复杂、目标分布密集、大宽高比、小尺寸,场景文本排列方式、字体大小呈现多样性以及光照条件的不确定性这些问题,仍是旋转目标检测面临的挑战。

旋转目标检测主要是针对带有朝向角度信息的物体的检测,主要思路是使边界框按照目标朝向去勾勒目标,最大化的减小目标真实区域与边界框区域的面积差,从而准确的将旋转目标与背景分离,解决目标密集、目标比例不统一、背景复杂等导致的旋转目标检测困难问题。唐建宇等^[5]提出了一种基于旋转目标框和注意力机制的遥感图像目标检测算法,在 YOLOv5 的基础上进行改进,并使用基于环形平滑标签的长边表示法来实现目标框的旋转,实现对密集目标的检测,但也存在计算复杂、分类繁多的缺点;姚树春等^[6]通过级联回归网络来实现旋转人脸的检测;张宏群等^[7]改进 YOLOv5,实现对被遮挡、排列密集的遥感舰船的监控,但精度仍有待提高;李明阳等^[8]针对现有目标检测方法在复杂场景中旋转不变性差的问题,提出了一种多尺度可变性卷积神经网络;文献[9-11]一直致力于旋转目标检测方法研究,并相继提出了 R3Det、SCRDet、SCRDet++ 等旋转目标检测方法,并不断刷新在遥感公开数据集上的榜单,但仍然存在网络复杂的问题。因此,基于深度学习的旋转目标检测渐渐成为计算机视觉领域的研究热点之一。

当前,随着目标检测技术的不断发展,部分学者对几种基于深度学习的目标检测方法进行了梳理总结,如王建新等^[12]致力于对自然场景文本检测与识别,刘天颖等^[13]专注于光学遥感图像的目标检测,史彩娟等^[14]着力于显著性目标检测,罗会兰等^[15]等分两类对目标检测技术进行综述,即一类是基于候选区域的目标检测算法,一类是基于回归的目标检测算法,并对每类方法分维度综述了研究者们在其基础算法上所做的改进研究。但目前还没有对旋转目标检测算法的综述,因此本文对现有旋转目标检测算法进行综述是必要的,本文对现有的基于深度学习的旋转目标检测模型进行了总结和分析。

1 面向旋转目标的目标检测

目前面向旋转目标的检测模型大多都是针对遥感图像和自然场景文本图像提出的,其相对于普通目标检测是具有挑战性的。普通目标检测基本都是用水平矩形框来表示目标,而自然场景文本图像和遥感图像中的目标大多有方向性,如果依然使用水平矩形框,不但不能将目标与背景有效的分离,且不能反映目标真实的形状。使用水平区域建议框和旋转区域建议框对场景文本和遥感舰船的检测结果

如图 1(a)和(b)所示。可以明显看出,旋转矩形框将目标更准确的勾勒出来,有效地减小了目标真实区域与边界框区域的面积差。同时,计算目标的角度也是一项具有挑战性的任务,所以将目标检测模型应用到旋转目标检测时就需要对网络进行专门的设计,来更好地适应旋转目标检测。本文将现有的基于深度学习的旋转目标检测方法按照检测步骤从基于锚框的二阶段方法、基于锚框的一阶段方法和无锚框方法进行归纳综述。

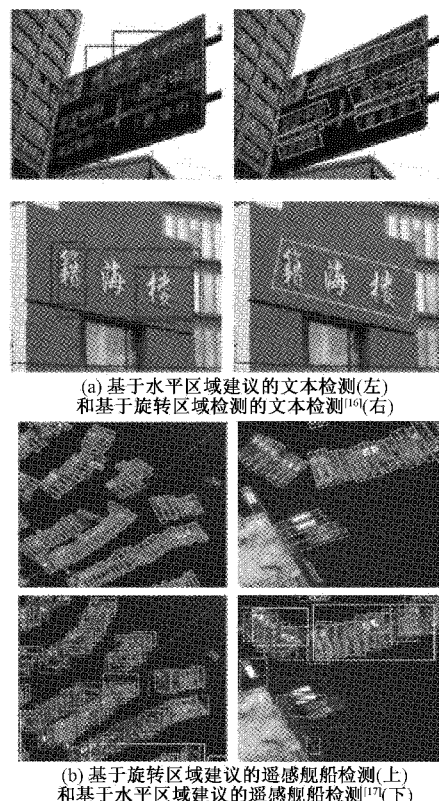


图 1 基于不同区域建议在不同场景中的检测效果

1.1 基于锚框的二阶段方法

基于锚框的二阶段旋转目标检测方法,首先由区域生成网络生成一系列作为样本的候选框,再通过卷积神经网络对样本进行分类。基于锚框的二阶段旋转目标检测方法有着准确度较高的优势,但同时存在速度相对较慢的劣势。常见的二阶段通用网络模型 R-CNN、Fast R-CNN、Faster R-CNN、SPP-Net^[18-21]等都是使用水平矩形框来勾勒目标轮廓,如果仍使用水平矩形框来勾勒旋转目标的轮廓,会包含大量的冗余信息,影响检测精度。因此,在自然场景文本图像检测和遥感图像检测领域诞生了许多新的检测算法。

利用区域生成网络(RPN)生成候选区域是 Faster R-CNN 算法的主要创新点。与启发式算法(selective search)相比,这种方法在速度和精度上都有所提升。RRPN 是第 1 个基于 RPN 架构引入旋转候选框实现任意方向场景文本检测的算法。它通过 RRPN 生成任意方向的区域建议,并进一步对区域建议边界框进行回归,以更好地适应文

本实例。然后把旋转感兴趣区域池化层(RROI Pooling)作为特征汇集层,将任意方向的区域建议映射到特征图,最后使用两个全连接层将具有旋转感兴趣特征的区域区分出文本和背景。RRCNN^[22]通过旋转感兴趣区域池化层来提取旋转目标特征,针对多类别使用多任务 NMS,最后回归出旋转目标边界框。R3-Net^[23]、R2PN^[24]和 RRPN 结构相似,都使用专门设计的旋转候选区域生成网络,从而生成任意方向的候选区域,这种方法可以有效检测出旋转目标,但计算量非常大。不同的是 R3-Net 提出了一种 BAR anchor,利用训练集中车辆的尺寸信息来初始化区域建议的可旋转锚框,并改进旋转区域池化为 R-PS 池化(rotatable position sensitive pooling),使得在对特征图进行下采样的同时保持目标的信息在特定的方向上。RoI Transformer^[25]针对 R2PN 网络计算量大的问题进行了优化,在使用 RPN 产生水平矩形框的候选区域的基础上,设计了一个旋转感兴趣区域学习机制可以将水平框转变成旋转框,这样不但减少了计算量还大大提高了检测精度。

CAD-Net^[26]是在网络 Faster-RCNN 和 FPN^[27]的基础

上,利用注意力机制获取全局和局部的上下文特征。另一方面,Gliding-Vertex^[28]提出了一种新的旋转框定义方式,它通过学习 4 个点在非旋转矩形上的偏移量,来更准确地定位出一个物体。APE^[29]还提出了一种新的旋转框表示方法,该角度由两个周期不同的二维周期向量表示,并该向量随形状变化而连续。另外,一些学者在模型训练过程中注意到旋转框回归会引起旋转敏感度误差^[30]的问题,从而导致模型训练的不稳定性,甚至影响到最终后的目标检测精度。SCRDet,SCRDet++设计了一种新的损失函数来缓解这一问题,但并没有从根本上解决。Chen 等^[31]提出了 OBB 多定义和选择策略(OBB-MDS-Strategy),通过改变 GT OBB (ground truth oriented bounding box)的定义,来改善目标角度变化问题。OBB 多定义和选择策略首先反转角度并改变原始 GT OBB 的高度和宽度,从而获得另外两个 OBB 定义,然后选择和预测的 OBB 具有相同定义的 OBB 作为真正的 GT OBB,并用于计算回归损失。各种二阶段的旋转目标检测方法的优缺点、主干结构、适用场景以及在 DOTA 遥感数据集上的 mAP 值对比如表 1 所示。

表 1 基于锚框的二阶段旋转目标检测方法比较

算法	优点	缺点	Backbone	适用场景	发表年份	mAP/% DOTA
RRPN ^[16]	1)提出了一种新的任意旋转文本区域优化的策略,从而优化了旋转文本的检测性能; 2)同时将 RRoI(rotation region-of-interest)和旋转感兴趣区域学习相结合,保证了文本检测过程中的效率。	生成 anchor 的角度不连续导致角度学习效果差;冗余旋转框多;速度慢	ResNet	1)场景文本; 2)任意旋转角度的场景文字。	2017	61.01
RRCNN ^[22]	1)增加旋转感兴趣区域池化层(RROD)、旋转边界框回归模型和多任务的不同类间的 NMS,提高检测精度。	与 RRPN 相比检测速度较慢	VGG-16	1)遥感卫星舰船图像。	2017	—
R2CNN ^[32]	1)增加了一些小尺寸的 anchor 提升对小文本的检测效果; 2)为了最大限度地利用文本特征,为每个 RPN 提议区域做了几个不同集合大小(7×7,11×3,3×11)的 ROI Poolings; 3)对倾斜框也做 NMS 操作,提高精度。	检测速度较慢	ResNet	1)文本。	2017	60.67
R3-Net ^[23]	1)提出了一种新的下采样方法用于 R-ROIS(rotatable region of interests)称为 R-PS 池化(rotatable position sensitive pooling),实现 R-ROI 特征映射的快速减少,并保持位置和方向的信息。	检测速度慢,精度有待提高	ResNet	1)遥感图像。	2018	—

续表 1

算法	优点	缺点	Backbone	适用场景	发表年份	mAP/% DOTA
R2PN ^[24]	1) 基于旋转 anchor, 有效提高预测效果; 2) 通过 RPN 得到旋转 ROI, 基于旋转 ROI 提取特征, 然后进行定位和分类。	Anchor 数量多, 计算量大, 检测速度慢	VGG-16	1) 遥感船舶图像。	2018	—
RoI Transformer ^[25]	1) 在 RPN 阶段通过全连接学习得到旋转 ROI, 减少了计算量, 然后基于旋转 ROI 提取特征, 进行定位和分类。	对大长宽比的目标检测不准确	ResNet	1) 遥感图像。	2018	69.56
CAD-Net ^[26]	1) 全局上下文网络(GCNet)和一个金字塔局部上下文网络(PLCNet), 有助于处理遥感图像纹理稀疏、对比度低、信息丢失严重的情况; 2) 空间感知和尺度感知的注意模块, 有助于处理纹理稀疏、背景对比度较低的目标, 并且尺度感知特征有助于处理不同尺度的目标。	检测速度慢, 精度有待提高	ResNet 和 FPN	1) 遥感图像。	2019	69.90
GlidingVertex ^[28]	1) 提出了一种新的旋转框表示方法。	边界框预测不准确, 检测速度慢	ResNet	1) 遥感图像; 2) 任意方向场景文本、任意长宽比目标。	2019	75.02
APE ^[29]	1) 提出了一种新的旋转框表示方法和新的 IoU 计算方式。	计算复杂, 检测速度慢	FPN	1) 遥感图像。	2019	75.75
Qian 等 ^[30]	1) 提出旋转灵敏度误差问题, 并给出一个新损失函数。	相比 Gliding Vertex 的精度低一些	ResNet	1) 遥感图像; 2) 任意方向场景文本。	2019	74.10
SCRDet ^[11]	1) 添加特征融合和空间、通道注意力机制提升对多尺度目标的检测能力; 2) 设计了一种改进的平滑 L1 损失, 专门用于解决旋转边界框回归的边界问题。	存在旋转角度的边界问题, 会产生巨大损失	ResNet	1) 遥感图像; 2) 小型、杂乱和旋转物体的航拍图像。	2019	72.61
SCRDet++ ^[12]	1) 在特征图上执行实例级去噪, 增强对小型杂乱物体的检测; 2) 改进 smooth L1 loss。	角的周期性影响检测精度	ResNet	1) 小型、杂乱和旋转物体的航拍图像。	2020	76.81
Chen 等 ^[31]	1) 提出级联网络(CA-Net), 用于增强感兴趣, 并抑制背景噪声; 2) 提出自适应特征串网络(AFC-NET), 自适应地堆叠从所有 FPN 级汇集的特征映射以及全局语义特征, 以处理对象的多尺度变化; 3) 提出 OBB 多定义和选择策略(OBB-MDS-策略), 更加平滑地回归出旋转边界框。	网络参数较多, 训练较慢	ResNet	1) 航空图像; 2) 大长宽比、密集分布和类别不平衡的旋转目标。	2021	76.27

1.2 基于锚框的一阶段方法

基于锚框的一阶段旋转目标检测方法是直接在主干网络中提取特征来预测物体类别和位置信息。这样的算法速度更快,但是精度相对二阶段(two-stage)目标检测网络略低。常用的一阶段通用检测网络有 YOLO^[33] 系列、SSD^[34] 系列、Retina-Net^[35] 等,旋转目标检测网络很多是在这些经典检测网络基础上改进的。

基于锚框的一阶段旋转目标检测方法将目标检测问题转化成了回归问题,虽然精度有所下降,但是更加高效。EAST^[36] 使用全卷积网络(FCN^[37])和非极大值抑制(NMS^[38]),大大提高了检测效率,同时为了适应各个方向文本区域目标时,采用了旋转框(RBOX)、水平框(QUAD)两种几何形状,并且这两种几何形状的预测框分别使用了不同的损失函数。TextBoxes++^[39] 和 DRBox 都是在 SSD 的基础上加入了适应角度变化而采用的旋转候选框检测方式。TextBoxes++ 可以产生有旋转角度的矩形文本检测框或是 4 边形检测框去适应带有旋转角度的文本。DRBox 还通过实验分析了水平矩形框相较于旋转矩形框检测的一些不足之处。R3Det 在特征精细化阶段使用旋转锚框以适应密集场景,并且为了解决基于锚框的一阶段方法特征未对齐的问题,设计了一个特征细化模块,通过

特征插值实现特征重构和对齐,精度有了很大地提升^[41]。R4Det^[40] 与 R3Det 使用相同的旋转目标表示方法,但不同的是 R4Det 使用了 RFP(recursive feature pyramid)模块,在 FPN 的基础上增加了反馈机制,使特征提取性能优于 FPN;另外在骨干网络之后添加 RCB(recursive connection block),来整合高低维特征图丰富 FRM 的输入特征。焦军锋等^[41] 采用旋转矩形框和注意力模块 CBAM 对 Retina-Net 进行改进,并在位置回归网络中采用旋转角度参数,从而提升近岸舰船检测精度。Yang 等^[42] 考虑到角的周期性和边的交换性问题,设计了 one-hot 编码和 CSL(circular smooth label),并增加相邻角度之间的容忍误差,通过将角度的回归问题转换成了一个分类问题,以适应角度的周期性。这种方法虽然效果不错,但还会存在一些问题,例如角度类别太多,角度类别不平衡等,导致网络头部过厚计算复杂、损失较大。最近 Yang 等^[43] 针对预测层过于厚重和对类正方形目标检测不友好这两问题对 CSL 进行了优化,提出 DCL(densely coded labels),在精度和速度上有了提升,但代价是存在理论误差以及模型参数量和计算量的增加。各种一阶段旋转目标检测方法的优缺点、主干结构、适用场景以及在不同数据集上的 mAP 值和 F 值对比如表 2 所示。

表 2 基于锚框的一阶段旋转目标检测方法比较

算法	优点	缺点	Backbone	适用场景	发表年份	F 值	
						mAP/%	F 值
						DOTA	ICDAR2015
						HRSC2016	COCO-Text
DRBox ^[47]	1)能有效定位任意方向旋转目标; 2)DRBox 的训练迫使检测网络去学习目标的正确角度,从而实现旋转不变性。	结构复杂,检测速度慢	CNN	1)遥感卫星图像上的车辆、船舶和飞机。	2017	—	—
EAST ^[36]	1)没有冗余的检测步骤; 2)能检测任意角度的文字,速度快。	对一些弯曲的文本框检测效果不好	FCN	1)文本; 2)单词级别或者文本行级别。	2017	—	0.807 2 0.394 5
TextBoxes++ ^[39]	1)能够准确高效的实现对任意方向的场景文本的检测,并且不需要除了非极大值抑制之外的后处理。	冗余框较多,实时性差	VGG-16	1)文本; 2)自然场景中多方向、小尺度、极端宽高比的文本。	2018	—	0.829 0 0.587 2
R3Det ^[49]	1)在精细化阶段使用旋转 anchors 以适应密集场景; 2)设计了一个特征细化模块,通过特征插值实现特征重构和对齐。	训练速度较慢	FPN	1)遥感图像或文本; 2)大长宽比、密集分布和类别不平衡的旋转目标。	2019	73.74 89.33	0.849 6 —
R4Det ^[40]	1)检测多尺度目标; 2)改进 Focal Loss 函数优化类别不平衡和尺度不平衡问题。	检测速度较慢	FPN	1)航空图像; 2)大长宽比、密集分布和类别不平衡的旋转目标。	2020	75.84 89.56	— —

续表 2

算法	优点	缺点	Backbone	适用场景	发表年份	F 值	
						mAP/% DOTA HRSC2016	ICDAR2015 COCO-Text
Yang 等 ^[42]	1) 针对基于回归的角度预测方法都有或多或少的边界问题,将角度回归问题转换成了分类问题。	角度类别太多,会导致 RetinaNet 的 head 部分过于厚	FPN	1) 航空图像。	2020	76.24 96.10	— —
Yang 等 ^[43]	1) 针对过于厚重的预测层以及对类正方形目标检测的不友好进行了优化。	存在理论误差以及模型参数量和计算量的增加	FPN	1) 航空图像。	2021	77.37 96.41	— —

1.3 无锚框方法

anchor-free 方法不需要设定锚框来产生候选区域,而是直接通过关键点来预测出目标所在的位置坐标。因为 anchor-free 方法在通用目标检测上表现出来的优秀检测效果,以及 anchor-free 方法快捷高效的优点,一些学者开始尝试将 anchor-free 的思想应用到旋转目标的检测上。He 等^[44]通过直接回归的方式,学习目标与真实框参数的偏移量,不利用 anchor 作为桥梁,而是直接进行旋转矩形框回归或顶点回归,得到最终的倾斜矩形或者四边形检测框。Lyu 等^[45]通过角点检测和文本区域分割来实现带方向的场景文本检测。首先通过 DSSD 网络进行角点检测,同时另一分支生成位置敏感得分图,把检测到的所有角点进行组合得到候选框,然后根据得分图滤除掉得分低的候选框。

IENet^[46]借鉴了 FCOS^[47]在其结构上添加了一条分支用于预测角度,并使用交并比引导模型学习。O²-DNet^[48]和 P-RSDet^[49]则是基于关键点的策略,前者用中心点加两条中线表示旋转框,后者则将坐标系转换到极坐标系,在此基础上可通过中心点和旋转角及半径来构成一个旋转框^[16]。Pan 等^[50]设计了自适应感受野调整模块(feature selection module),使模型能够根据目标形状、旋转角度,自适应的调整感受野,缓解单一的感受野与多变的目标之间的矛盾。王明阳等^[51]提出一种 R-CenterNet 方法,在原网络结构中加入旋转因子,为检测框提供角度信息。各种无锚框的旋转目标检测方法的优缺点、主干结构、适用场景以及在不同开源数据集上的 mAP 值比较如表 3 所示。

1.4 不同检测方法分析比较

基于锚框的二阶段方法和一阶段方法是现在旋转目标检测领域的主流方法,随着深度学习研究的不断深入,无锚框方法也渐渐成为主流方法,在精度和速度方面,这 3 种方法都优于传统的目标检测方法。

1) 基于锚框的二阶段方法多针对复杂场景中旋转目标密集排列和尺度过小的情况,设计专门的旋转区域生成网络来刻画目标轮廓,减少冗余检测区域,提高检测精度。

但此方法往往结构设计复杂,导致检测速度慢;基于锚框的一阶段方法则为了减少冗余步骤和加快检测速度,摒弃了生成候选区域的过程,直接对原图进行特征提取,同时加入旋转角度预测,最后回归得到旋转目标框。这两种算法设计时既要控制模型推理速度又要保证模型检测效果,很难平衡预选框数量和检测性能。

2) 无锚框方法摆脱了使用锚框而带来的复杂计算,通过卷积神经网络直接回归得到目标边界框坐标,其结构比基于锚框的一阶段、二阶段方法简单,且模型参数少。这种方法简单、快速、高效而且没有任何的 NMS 的后处理操作,可以直接端到端地进行训练。但是这种基于特征图中特征点的算法,需要设计辅助方法来获取最终的目标检测框,容易出现角度偏差,还有可能出现目标语义模糊,从而影响检测精度。

2 旋转目标检测性能评价

2.1 数据集

随着目标检测技术的不断发展,不断涌现出一些旋转目标检测的数据集。下面介绍几种常用的公开数据集,每种数据集的样例图如图 2 所示。

DOTA^[52]广泛应用于遥感航空图像检测领域,DOTA 中的图像大小从 800×800 像素到 4 000×4 000 像素不等,其中包含 188 282 个不同尺寸方向和形状的实例,小型实例(10~50 像素)和中型实例(50~300 像素)居多,有大量具有大宽高比的实例。在进行注释时共分为 15 个目标类别,主类有 14 个,其中 small vehicle 和 large vehicle 都是 vehicle 的子类。其标注方式为 4 点确定的任意形状和方向的四边形。数据集划分为 1/6 验证集、1/3 测试集、1/2 训练集。此数据集对检测密集小型实例、极大实例以及具有任意取向目标具有重要意义和挑战性。

NWPU VHR-10^[53]是由西北工业大学发布的航天遥感目标检测数据集,总共包含 800 张图像,其中包含目标的图像有六百余张,背景图像 150 张。目标包括飞机、港口、桥梁、舰船、油罐、棒球场、网球场、篮球场、田径场、车辆 10 种。

表 3 无锚框旋转目标检测方法比较

算法	优点	缺点	Backbone	适用场景	发表年份	mAP/%		F 值/%	
						HRSC2016	DOTA	ICDAR2015	COCO-Text
He 等 ^[44]	1)直接进行旋转框回归。	精度有待提高	VGG16	场景文本 图像	2018	—	—	82.00	—
Lyu 等 ^[45]	1)每个角点都是独立的; 2)角点检测不用考虑方向性和文本框的长宽比大小。	角点较多,组合出 很多无关框,增加 网络计算时间	VGG16+CNN	场景文本 图像	2018	—	—	84.30	42.50
O ² -DNet ^[48]	1)基于关键点,无需 NMS 操作。	精度有待提高	Hourglass	场景文本和 遥感图像	2019	—	71.04	82.97	—
IENET ^[46]	1)无锚式单级定向检测器; 2)利用基于自注意机制的 IE 模块作为特征交互模块,结合特征进行方向预测。	精度有待提高	CNN	遥感图像	2020	75.01	61.24	—	—
P-RSDet ^[49]	1)模型简单,回归参数少。	精度有待提高	CNN	遥感检测	2020	—	72.30	—	—
Pan 等 ^[50]	1)FSM 能够根据目标物体的形状和方向位置调整神经元的感受野; 2)DRH 能使模块以目标感知的方式动态地优化预测。	检测速度较慢、参 数量多	FPN	自然场景 和 遥 感 图 像	2020	92.70	73.23	—	—
王明阳等 ^[51]	1)在网络结构中加入旋转因子为检测框提供角度信息; 2)将目标特征注意力信息与多尺度池化信息相融合,得到目标的多尺度信息。	精度有待提高	CNN	遥感图像	2021	—	—	—	—

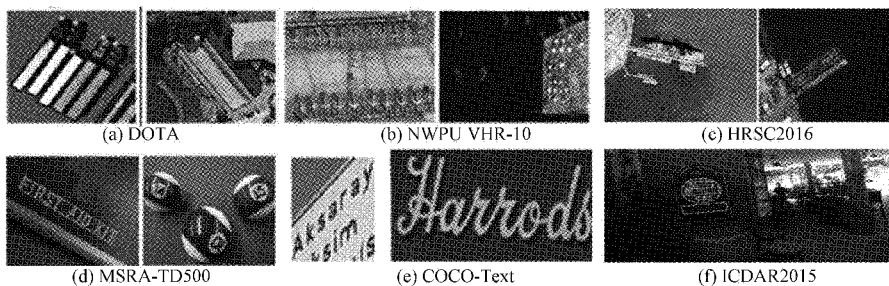


图 2 数据集样例图

HRSC2016^[54] 遥感舰船公开数据集,分辨率从 2~0.4 m 不等,图像大小从 300×300 到 1 500×900 不等,大部分比 1 000×600 大。含有 3 个级别任务(分别为单类 class,4 类 category 和 19 类 type 舰船检测识别)。

MSRA-TD500^[55]是由华中科技大学于 2012 年提供的文本检测数据集。数据集的图像分别是在办公室、商场和街道等多种场景中通过相机拍摄的图片,图片的文本由不同方向的中文和英文组成。

COCO-Text^[56]数据集是在微软提供的目标识别数据

集 MS COCO 的基础上完善成型的。COCO-Text 数据集的图片包含背景复杂自然图片和生活场景图片,由于图片是在不关注文本的情况下收集,因此大部分图片中的文本目标尺度小甚至不清晰,图片中也可能不包含文本内容。

ICDAR2015^[57]该数据集是 2015 年 ICDAR 在 RRC 中增加的偶然场景文本(incidental scene text)阅读竞赛提供的公共数据集。数据集是由 Google Glass 在未聚焦的情况下随机拍摄的街头或者商场图片,旨在帮助文本检测和识别

模型提高泛化性能。各种开源数据集的应用领域、大小、标注方式、数量、类别以及最佳检测结果对比如表 4 所示。

表 4 旋转目标检测数据集对比

数据集	应用领域	数据集大小	标注方式	数量	类别	最佳检测结果	
						mAP/%	检测方法
DOTA ^[32]	遥感航空图像检测	—	旋转框	包含 2 806 张航空图像	包含 15 个类别共计 188 282 个实例	76.81 —	SCRDet++ ^[25]
NWPU VHR-10 ^{L53}	空间物体检测	约 73M	水平框	包含 650 张包含目标的图像和 150 张背景图像, 共计 800 张	包含 10 类, 包括飞机、舰船、油罐、棒球场、网球场、篮球场、田径场、港口、桥梁和汽车	91.75 —	SCRDet ^{L21}
HRSC2016 ^[54]	遥感舰船	—	旋转框	包含 1 070 张图片 2 976 个实例	包含 1 个类别	96.41 —	DCL ^[48]
MSRA-TD500 ^{L55}	任意方向文本检测, 自然场景图像检测, 文本行	—	旋转框	包含 500 张自然图像	—	— 0.760 8	EAST ³⁰
COCO-Text ^[36]	单词级	12.58 GB	旋转框	包含 63 686 幅图像, 173 589 个文本实例	—	— 0.587 2	TextBoxes++ ^[49]
ICDAR2015 ^[57]	单词级	—	旋转框	包含 1 500 张图片	—	— 0.849 6	R3Det ^[35]

2.2 旋转目标检测方法性能指标

旋转目标检测方法的分类和定位能力是其最主要的性能体现。遥感图像检测中 mAP (mean average precision) 值是其检测效果最直观的表达方式, 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 (frame per second) 值来体现, FPS 值越大, 说明算法模型实时性越好。场景文本检测中, 现有的性能评估方法主要以 3 个表征检测结果的参数即准确率 (Precision, P 值)、召回率 (Recall, R 值)、综合评价指标 (F-measure, F 值) 和表征检测速度的参数即每秒检测帧数 (frames per second, FPS) 作为参考标准, 本文则选择参考文献中所列 F 值作为参考。

3 展 望

随着目标检测需求不断提升, 旋转目标检测研究已经成为当前的一个研究热点。旋转目标检测主要应用于遥感目标检测和文本检测。本文将近几年涌现出来的在文本检测领域和遥感图像检测领域的旋转目标检测方法分类进行了综述, 并介绍了在文本检测领域和遥感图像检测领域中常用的几种数据集。这些算法虽然取得了不错的

效果, 但依然存在一些挑战, 未来旋转目标检测在如下几个方面值得进一步研究。

1) 探索更优的旋转区域生成网络。目前大多数目标检测方法都已实现了按照目标朝向去勾勒目标轮廓, 但这个过程会生成很多无用框, 增加模型计算量, 所以需要探索如何在模型计算量更少的情况下准确的勾勒出旋转目标的轮廓, 提升对背景复杂、分布密集、小尺度和形态多样目标的检测能力。

2) 平衡检测精度和速度。目前大多数应用在遥感图像和场景文本图像中的旋转目标检测方法, 都是基于 anchor 的方法, 加上辅助角度回归的网络, 造成网络参数多、检测速度慢, 所以需要平衡检测精度和速度的问题, 在不降低精度的情况下, 快速实现对旋转目标的检测。例如, 移动端部署的 YOLO-Fastest 大小只有 1.3 MByte, 而 mAP 只下降了大约 5%。

3) 尝试用 Transformer 实现旋转目标检测。Transformer 根据 CNN 中的卷积思想, 结合了多头注意力, 实现并行计算, 解决了传统神经网络算法训练较慢的缺陷。Transformer 与 CNN 相比, 所需的参数更少, 并且在训练效率上具有显著优势。在架构上, 可以采用纯

Transformer 的方式使用,也可以与 CNN 结合的方式使用。将 Transformer 应用到旋转目标检测上指日可待。

4 结 论

本文将近几年涌现出来的遥感图像检测领域和文本图像检测领域的旋转目标检测方法进行了综述,以便学者了解旋转目标方向的发展现状。首先分析了旋转目标检测的背景,接着对面向旋转目标的检测方法进行分析和比较,分别为基于锚框的一阶段方法、基于锚框的二阶段方法和无锚框方法进行了归纳总结;然后介绍了旋转目标领域常用的公开数据集和性能评价指标;最后总结了旋转目标的未来发展趋势和可能的研究重点。虽然目前的旋转目标检测方法已经有一定成效,但检测性能仍需进一步提升,如检测精度、速度、复杂背景、小目标等。整体来说,研究者们仍然可以跟进通用目标检测领域的最新进展,并在其基础上针对旋转目标的特点进行改进。总而言之,旋转目标检测模型会朝着实时、高效的方向不断发展。

参考文献

- [1] 许德刚,王露,李凡. 深度学习的典型目标检测算法研究综述[J]. 计算机工程与应用, 2021, 57(8): 10-25.
- [2] 化嫣然,张卓,龙赛,等. 基于改进 YOLO 算法的遥感图像目标检测[J]. 电子测量技术, 2020, 43(24): 87-92.
- [3] 朱煜,方观寿,郑兵兵,等. 基于旋转框精细定位的遥感目标检测方法研究[J/OL]. 自动化学报, 2021: 1-11 [2021-05-30].
- [4] 易尧华,杨德齐,王新宇,等. 自然场景文本检测关键技术及应用[J]. 数字印刷, 2020(4): 1-11, 22.
- [5] 唐建宇,唐春晖. 基于旋转框和注意力机制的遥感图像目标检测算法[J]. 电子测量技术, 2021, 44(13): 114-120.
- [6] 姚树春,蔡黎亚,刘正. 基于级联回归网络的多尺度旋转人脸检测方法[J]. 电子测量与仪器学报, 2019, 33(5): 32-38.
- [7] 张宏群,班勇苗,郭玲玲,等. 基于 YOLOv5 的遥感图像舰船的检测方法[J]. 电子测量技术, 2021, 44(8): 87-92.
- [8] 李明阳,胡显,雷宏. 基于可变形卷积神经网络的遥感图像飞机目标检测[J]. 国外电子测量技术, 2020, 39(7): 121-126.
- [9] YANG X, LIU Q, YAN J, et al. R3Det: Refined single-stage detector with feature refinement for rotating object [J/OL]. (2020-2-21) [2021-10-09]. <https://arxiv.org/abs/1908.05612v5>.
- [10] YANG X, YANG J, YAN J, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects [C]. Proceedings of the IEEE International Conference on Computer Vision, Seoul: IEEE, 2019: 8232-8241.
- [11] YANG X, YANG J C, YAN X K, et al. SCRDet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing [J/OL]. (2020-4-28) [2021-10-09]. <https://arxiv.org/pdf/2004.13316v1.pdf>.
- [12] 王建新,王子亚,田莹. 基于深度学习的自然场景文本检测与识别综述[J]. 软件学报, 2020, 31(5): 1465-1496.
- [13] 刘天颖,李文根,关倩红. 基于深度学习的光学遥感图像目标检测方法综述[J]. 无线电通信技术, 2020, 46(6): 624-634.
- [14] 史彩娟,张卫明,陈厚儒,等. 基于深度学习的显著性目标检测综述[J]. 计算机科学与探索, 2021, 15(2): 219-232.
- [15] 罗会兰,陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报, 2020, 48(6): 1230-1239.
- [16] MA J Q, SHAO W Y, YE H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.
- [17] LIU L, PAN Z, LEI B. Learning a rotation invariant detector with rotatable bounding box[J/OL]. (2017-11-26) [2021-10-09]. <https://arxiv.org/pdf/1711.09405v1.pdf>.
- [18] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, USA: IEEE, 2014: 580-587.
- [19] GIRSHICK R. Fast R-CNN[C]. Proceedings of IEEE International Conference on Computer Vision, USA: IEEE, 2015: 1440-1448.
- [20] 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 & Machine Intelligence, 2017, 39(6): 1137-1149.
- [21] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-16.
- [22] LIU Z, HU J, WENG L, et al. Rotated region based cnn for ship detection [C]. 2017 IEEE International Conference on Image Processing, Beijing: IEEE, 2017: 900-904.
- [23] LI Q P, MOU L C, XU Q Z, et al. R3-Net: A deep network for multi-oriented vehicle detection in aerial

- images and videos [J]. IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society, 2019, DOI:2019,10.1109/TGRS.2019.2895362.
- [24] ZHANG Z, GUO W, ZHU S, et al. Toward arbitrary-oriented ship detection with rotated region proposal and discrimination networks[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(11):1745-1749.
- [25] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach; IEEE, 2019:2849-2858.
- [26] ZHANG G, LU S, ZHANG W. CAD-Net: A context-aware detection network for objects in remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(12):10015-10024.
- [27] 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; IEEE, 2017:936-944.
- [28] XU Y C, FU M T, WANG Q M, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(4):1452-1459.
- [29] ZHU Y X, WU X Q, DU J. Adaptive period embedding for representing oriented objects in aerial images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(10):7247-7257.
- [30] QIAN W, YANG X, PENG S L, et al. Learning modulated loss for rotated object detection[J/OL]. (2019-12-20) [2021-10-09]. <https://arxiv.org/pdf/1911.08299v3.pdf>.
- [31] CHEN L, LIU C, CHANG F, et al. Adaptive multi-level feature fusion and attention-based network for arbitrary-oriented object detection in remote sensing imagery[J]. Neurocomputing, 2021, 451:67-80.
- [32] JIANG Y Y, WANG X B, ZHU X Y, et al. R2CNN: Rotational region cnn for orientation robust scene text detection[J/OL]. (2017-6-30) [2021-10-09]. <https://arxiv.org/ftp/arxiv/papers/1706/1706.09579.pdf>.
- [33] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, USA; IEEE, 2016:779-778.
- [34] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, Amsterdam, Netherlands; IEEE, 2016, 21-37.
- [35] 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, Piscataway; IEEE, 2017:2980-2988.
- [36] ZHOU X Y, YAO C, WANG Y Z, et al. EAST: An efficient and accurate scene text detector [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA; IEEE, 2017:2642-2651.
- [37] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]. Computer Vision and Pattern Recognition, USA; IEEE, 2015:3431-3440.
- [38] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression[C]. International Conference on Pattern Recognition, China; IEEE, 2006:850-855.
- [39] LIAO M H, SHI B G, BAI X. TextBoxes++: A single-shot oriented scene text detector [J]. IEEE Transactions on Image Processing, 2018, 27(8):3676-3690.
- [40] SUN P, ZHENG Y, ZHOU Z, et al. R¹ Det: Refined single-stage detector with feature recursion and refinement for rotating object detection in aerial images[J]. Image and Vision Computing, 2020, DOI: 10.1016/j.imavis.2020.104036.
- [41] 焦军峰, 靳国旺, 熊新, 等. 旋转矩形框与 CBAM 改进 RetinaNet 的 SAR 图像近岸舰船检测[J]. 测绘科学技术学报, 2020, 37(6):603-609.
- [42] YANG X, YAN J C, HE T, et al. On the arbitrary-oriented object detection: Classification based approaches revisited [J/OL]. (2021-3-27) [2021-10-09]. <https://arxiv.org/pdf/2003.05597v3.pdf>.
- [43] YANG X, HOU L P, ZHOU Y, et al. Dense label encoding for boundary discontinuity free rotation detection[J/OL]. (2021-5-25) [2021-10-09]. <https://arxiv.org/pdf/2011.09670.pdf>.
- [44] HE W H, ZHANG X Y, YIN F, et al. Multi-oriented and multi-lingual scene text detection with direct regression [J]. IEEE Transactions on Image Processing, 2018, 27(11):5406-5419.
- [45] LYU P Y, YAO C, WU W H, et al. Multi-oriented scene text detection via corner localization and region segmentation [C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2018:7553-7563.
- [46] LIN Y, FENG P, GUAN J. IENet: Interacting embranchment one stage anchor free detector for

- orientation aerial object detection[J/OL]. (2021-3-21) [2021-10-09]. <https://arxiv.org/pdf/1912.00969.pdf>.
- [47] TIAN Z, SHEN C, CHEN H, et al. FCOS: Fully convolutional one-stage object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, Beijing: IEEE, 2019: 9627-9636.
- [48] WEI H, ZHOU L, ZHANG Y, et al. Oriented objects as pairs of middle lines [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 169: 268-279.
- [49] ZHOU L, WEI H, LI H, et al. Objects detection for remote sensing images based on polar coordinates[J/OL]. (2020-9-21) [2021-10-09]. <https://arxiv.org/pdf/2001.02988v3.pdf>.
- [50] PAN X J, REN Y Q, SHENG K K, et al. Dynamic refinement network for oriented and densely packed object detection[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA: IEEE, 2020: 11204-11213.
- [51] 王明阳, 王江涛, 刘琛. 基于关键点的遥感图像旋转目标检测[J]. 电子测量与仪器学报, 2021, 35(6): 102-108.
- [52] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City: IEEE, 2018: 3974-3983.
- [53] CHENG G, HAN J, ZHOU P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98: 119-132.
- [54] LIU Z, WANG H, WENG L, et al. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(8): 1074-1078.
- [55] YAO C, BAI X, LIU W, et al. Detecting texts of arbitrary orientations in natural images[C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, USA: IEEE, 2012: 1083-1090.
- [56] VEIT A, MATERA T, NEUMANN L, et al. Coco-text: Dataset and benchmark for text detection and recognition in natural images [J/OL]. (2016-6-19) [2021-10-09]. <https://arxiv.org/pdf/1601.07140.pdf>.
- [57] KARATZAS D, GOMEZ-BIGORDA L, NICOLAOU A, et al. ICDAR 2015 competition on robust reading[C]. 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunisia: IEEE, 2015: 1156-1160.

作者简介

安胜彪, 硕士, 副教授, 主要研究方向为集成电子系统和集成电路的研究。

E-mail: 33588253@qq.com

娄慧儒, 硕士生, 主要研究方向为深度学习、计算机视觉。

E-mail: 925350832@qq.com

陈书旺, 博士后, 教授, 主要研究方向为现代传感技术、工业自动检测技术、红外人体信息检测及图像处理技术。

E-mail: 27984960@qq.com

白宇, 博士, 讲师, 主要研究方向为信息物理系统(CPS)、同步系统、深度学习、基于模型的系统设计、形式化方法。

E-mail: baiyu@hebust.edu