



计算机科学

COMPUTER SCIENCE

计算机视觉下的旋转目标检测研究综述

王旭, 吴艳霞, 张雪, 洪瑞泽, 李广生

引用本文

王旭, 吴艳霞, 张雪, 洪瑞泽, 李广生. 计算机视觉下的旋转目标检测研究综述[J]. 计算机科学, 2023, 50(8): 79-92.

WANG Xu, WU Yanxia, ZHANG Xue, HONG Ruize, LI Guangsheng. [Survey of Rotating Object Detection Research in Computer Vision](#) [J]. Computer Science, 2023, 50(8): 79-92.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于字符特征的 DGA 域名检测方法研究综述](#)

Survey of DGA Domain Name Detection Based on Character Feature

计算机科学, 2023, 50(8): 251-259. <https://doi.org/10.11896/jsjcx.220700277>

[融合粗粒度代价体及双边网格的轻量级多视图三维重建](#)

Lightweight Multi-view Stereo Integrating Coarse Cost Volume and Bilateral Grid

计算机科学, 2023, 50(8): 125-132. <https://doi.org/10.11896/jsjcx.220600046>

[基于深度学习的图像描述优化策略](#)

Image Captioning Optimization Strategy Based on Deep Learning

计算机科学, 2023, 50(8): 99-110. <https://doi.org/10.11896/jsjcx.230200091>

[说话人生成研究现状与发展趋势](#)

Review of Talking Face Generation

计算机科学, 2023, 50(8): 68-78. <https://doi.org/10.11896/jsjcx.221000031>

[基于注意力机制的多模态在线评论有用性预测研究](#)

Study on Multimodal Online Reviews Helpfulness Prediction Based on Attention Mechanism

计算机科学, 2023, 50(8): 37-44. <https://doi.org/10.11896/jsjcx.220600204>

计算机视觉下的旋转目标检测研究综述

王旭 吴艳霞 张雪 洪瑞泽 李广生

哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001

(2750250486@hrbeu.edu.cn)

摘要 传统目标检测器通过水平边界框(Horizontal Bounding Box, HBB)定位目标,在检测方向角任意、分布密集、长宽比大、背景复杂的目标时,往往精度较低、泛化能力较差。在边界框中增加不同旋转角度的旋转目标框可有效解决上述问题,其被广泛应用在遥感图像、场景文本图像、货架商品图像等目标检测领域,具有重要研究价值。目前大多数工作旨在构建不同的旋转目标检测模型,对现有模型的归纳总结及深入分析的综述性工作较少。为此,对旋转目标检测现有研究成果进行了详细综述。首先根据当前流行的目标框表征方式,将目标框分为旋转矩形框(Oriented Bounding Box, OBB)、四边形边界框(Quadrilateral Bounding Box, QBB)和点集(Point set) 3种类型,并比较了不同旋转目标检测算法的优缺点、网络结构和性能;其次分析了目前常用的旋转目标检测数据集和性能评价指标;最后对目前研究中存在的问题进行简要总结和讨论,并对未来的发展趋势进行展望。

关键词: 计算机视觉;深度学习;目标检测;旋转目标;性能比较

中图法分类号 TP311

Survey of Rotating Object Detection Research in Computer Vision

WANG Xu, WU Yanxia, ZHANG Xue, HONG Ruize and LI Guangsheng

College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

Abstract Traditional object detector locates objects by horizontal bounding box(HBB), which often have low accuracy and poor generalization ability when detecting objects with arbitrary orientation angle, dense distribution, large aspect ratio and complex background. The above problems can be effectively solved by adding rotating target boxes with different rotation angles in the bounding boxes. This method is widely used in the fields of remote sensing images, scene text images, shelf goods images and other target detection, and has important research value. Most of the current works aim at constructing different models for rotating object detection, and there are fewer review works for summarizing and analyzing existing models in depth. Therefore, this paper provides a detailed review of existing research results on rotating object detection. Firstly, according to the current popular way of target box characterization, the target boxes are classified into three types of oriented bounding box(OBB), quadrilateral bounding boxes(QBB) and point set for generalized analysis, and simultaneously compare the advantages and disadvantages, network structures and performance of different rotating object detection algorithms. Secondly, the commonly used rotating object detection datasets and performance evaluation metrics are analyzed. Finally, the problems in the current study are briefly summarized and discussed, and the future development trend is prospected.

Keywords Computer vision, Deep learning, Object detection, Rotating object, Performance comparison

1 引言

目标检测是利用计算机在给定的图像或视频中找出所有感兴趣的目标,并确定其位置和类别,是机器视觉领域的核心任务之一。随着目标检测技术的快速发展,越来越多基于深度学习的水平框目标检测算法被提出,如双阶段通用网络模型 FastR-CNN^[1], Faster R-CNN^[2], SPP-Net^[3]等,单阶段通用检测网络模型 YOLO^[4], SSD^[5], Retina-Net^[6]等,大大提高了图像目标检测的性能。尽管上述算法在自然图像检测中取得了重大进展,但由于它们均使用水平矩形框勾勒目标轮廓,

在场景文本检测^[7]、遥感图像检测^[8]、零售场景检测^[9]等领域存在明显缺陷。具体表现为:

(1)上述场景中物体尺度往往变化较大,存在较多极端长宽比的目标(比如桥梁、舰船、文字等),使水平检测框包含许多背景区域,存在大量冗余像素,增加了分类任务的难度^[10],如图1所示。

(2)上述场景中往往存在小而密集的目标群,许多小物体实例挤在图像中的某块区域中,并呈现任意角度分布,使得该类算法使用非极大值抑制(Non-Maximum Suppression, NMS)作为后处理操作时,相邻目标的交并比较大,在执行

NMS 操作过程中可能抑制其中某个目标,导致目标漏检,降低检测精度^[11]。并且,对于这些密集目标,水平框检测方式难以进行紧凑、精准的定位,容易产生边界框混淆、目标展示不清晰、影响观察等问题。

(3)上述场景包含的飞机、船舶、车辆、文本等目标往往具有朝向和尺度信息,若使用水平边界框表征,无法获得目标的各种物理属性。



(a)大量冗余像素

(b)NMS 错误过滤

图1 水平边界框示例图

Fig.1 Diagram of horizontal bounding box

针对水平边界框在检测多方向目标时出现的问题,一些学者提出了 RRPN^[12] (Radar Region Proposal Network), R3Det^[13] (Refined Rotation RetinaNet), CAD-net^[14] (Context-Aware Detection Network), TIDE-Det^[15] (Task Interleaving and Orientation Estimation Detector)等深度学习算法,将目标角度信息融入区域生成网络或设计专门的角度预测网络,使用具有角度信息的旋转边界框来解决上述问题。相比水平边界框,使用旋转边界框检测多方向性目标具有以下优势:1)使用旋转边界框检测目标几乎不包含背景区域,减少了背景对目标分类的影响,如图2(a)所示;2)使用倾斜非极大值抑制方法(Skew Non-Maximum Suppression, Skew-NMS)作为后处理操作,避免了传统 NMS 过分抑制的问题;3)可以从旋转边界框中粗略获得物体的朝向信息及运动轨迹^[16],如图2(b)所示;4)边界框之间几乎没有重叠,可以更清楚地识别框中包含的目标。



(a)旋转边界框

(b)表征物体方向

图2 旋转边界框示例图

Fig.2 Diagram of rotating bounding box

旋转目标检测技术主要是针对大量杂散、背景复杂、任意方向的小物体的检测,通过重新定义目标框的表征形式,增加回归自由度的个数,使边界框按照物体朝向去勾勒目标,最大化地增加目标真实区域与边界框区域的重叠度,实现用更加紧凑的边界框检测多方向目标的目的。该技术在场景文本检测、遥感图像检测、人脸检测^[17]、顶拍鱼镜头下的人体或物体检测等领域有较好的检测效果,拥有广阔的应用价值。因此,基于深度学习的旋转目标检测逐渐成为计算机视觉领域的研究热点之一。

随着目标检测技术的发展,众多旋转目标检测算法被提出。虽然部分学者对个别场景下的旋转目标检测做了综述,例如文献^[18-20]针对遥感图像目标检测相关问题进行综述,Wang 等^[21]对场景文本检测与识别进行了研究,但是目前仍缺乏对旋转目标检测领域系统的总结归纳的文献。为了方便学者了解该领域的发展现状,本文对旋转目标检测领域中边界框表征方式、模型算法、存在的问题、优化策略等进行全面的梳理,并对各旋转目标检测算法的性能进行综合分析和总结;最后对该领域的现存问题进行总结,并对其发展趋势进行合理预测。

2 旋转目标检测方法归纳分析

对任意角度的目标具备识别能力,以实现旋转物体检测一直是目标检测领域的重要课题。由于其仍属于目标检测的范畴,研究者主要通过改造现有网络模型来实现旋转物体检测。一种解决思路是通过旋转变换扩充训练样本,来提高模型对旋转目标检测的稳健性^[22-23]。但由于仍使用水平边界框表征,不能很好地拟合物体,存在较大问题。

随着技术的发展,在双阶段目标检测模型基础上,通过引入旋转角度 θ 这一回归参数定位旋转矩形框来实现旋转物体检测。后来,为了灵活表征物体,通过设计损失函数对4个顶点进行回归以定位多边形边界框来勾勒旋转目标轮廓^[24-25]。随着移动端对目标检测效率要求的提高,单阶段目标检测模型逐渐成为主流,并涌现出一些算法如 R3Det^[13], R4Det^[26]等,针对旋转目标检测的效率进行优化。然而无论是基于双阶段还是单阶段检测,上述两种边界框方法均存在与其表征相关的问题,如边界不连续^[27]、类方形问题^[28]、表征模糊^[29]等。为了解决上述问题,提高算法的检测性能,学者们提出了基于损失函数的改进、基于回归参数的改进、基于图像分割方式的改进等方案。随着 CornerNet^[30]算法的提出,兴起了对无锚框算法的研究,有些学者提出通过点集^[31]表征物体。点集表征目标更为灵活,可以适应任意形状,不受回归参数的限制,但也存在孤立点^[32]等问题。图3以时间为纵轴给出了深度学习旋转目标检测算法的发展路线。

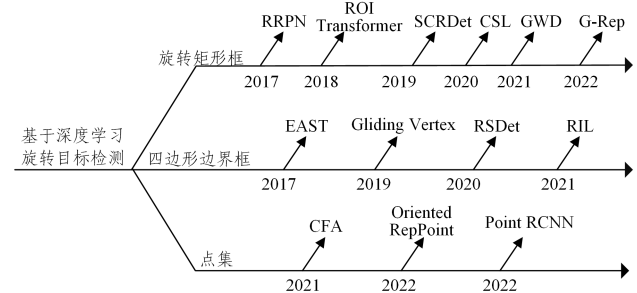


图3 旋转目标检测发展路线

Fig.3 Development route of rotating object detection

现阶段大多数研究工作都基于旋转边界框开展,其不同的表征形式会影响网络的设计以及目标的定位效果,恰当的旋转框表征形式不仅能使网络更加容易学习,还能够降低模型的复杂度。本文根据目前流行的旋转目标框表征方式将其分为 OBB, QBB 以及 Point set 这 3 种类型,并对不同类型的

表示方法、算法特点、性能、适用场景以及存在问题进行总结和分析。

2.1 基于旋转矩形边界框的目标检测算法

2.1.1 旋转矩形边界框定义

旋转矩形框是目前旋转目标检测最流行的表征方式,该方式可以简单直观地获得旋转角度。模型主要通过回归预测 (x, y, w, h, θ) 5 个参数,来获得旋转矩形框。具体的表征形式主要有 OpenCV 表示法和长边表示法^[33]。

OpenCV 表示法: (x, y) 为 OBB 中心点, θ 为水平轴在框最低点逆时针旋转与碰到矩形第一条边的夹角,角度的表示范围是 $[-90^\circ, 0^\circ]$ 。 w 与 h 不固定代表矩形的长短边,而是将与 x 轴的夹角小于或等于 90° 的边记为 w ,对应的另一条边记为 h 。具体形式如图 4(a) 所示。

长边表示法:长边为 w ,短边为 h ,旋转角 θ 为水平轴与宽所形成的夹角,向上或向下旋转 θ ,如图 4(b) 所示。这种表示方法与 OpenCV 定义方法相似,都使用中心点、宽高和旋转角表示 OBB,但宽高和旋转角的定义有所区别,长边定义法旋转度的范围为 $[-90^\circ, 90^\circ]$,宽高相对固定。两种表示方法可以进行相互转换,同时 OBB 也可以使用其他形式的参数进行表示,但都与上述两种方法思想类似。

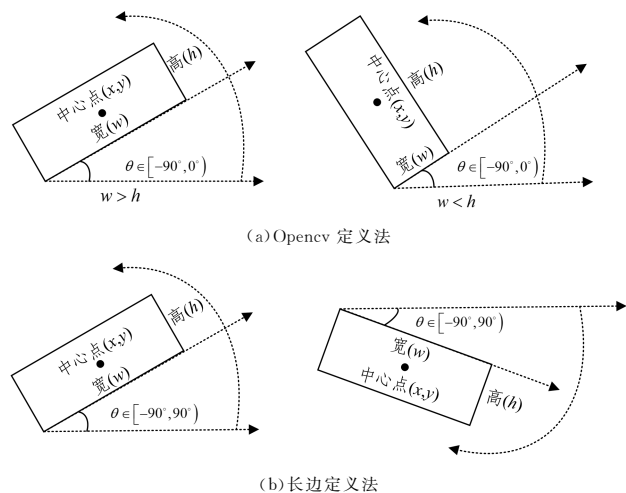


图 4 旋转矩形框定义方法

Fig. 4 Definition method of rotating rectangle box

2.1.2 基于 OBB 目标检测算法的发展

现有的基于角度回归的目标检测器大多使用上述参数表示方案。算法主要分为双阶段检测算法和单阶段检测算法。对于双阶段旋转目标检测算法^[34-38],一般使用 RPN 来生成旋转感兴趣区域 (Rotation Region-of-Interest, RROI), 然后使用 RROI Pooling 对 RRoI 进行特征提取,最后通过 RCNN Head 预测旋转边界框及相应的类别。对于单阶段检测器^[39]则是直接对参数进行回归,并根据稠密锚点对其进行分类以提高检测效率。

具体来说,为了实现任意方向场景的文本检测, Ma 等^[12]提出了 RRPN 模型,采用 (x, y, w, h, θ) 作为旋转边界框的位置参数,对于 θ 选取了 $-\frac{\pi}{6}, 0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}$ 这 6 个角度参数。该算法首先生成具有不同尺度、长宽比和旋转角度的

锚框,其次对参数进行微调得到带角度的候选区域,然后利用 RROI Pooling 提取建议区域内任意方向的有效特征,最后使用两层全连接层对候选区域进行分类。RRPN 是第一个基于 RPN 架构引入旋转候选框实现场景文本检测的算法^[40],该算法思想非常值得借鉴,但是 RRPN 网络用有限个数的角度参数去覆盖任意朝向的目标显然无法实现,存在一定缺陷。为此 Jiang 等提出了 R2-CNN^[41]模型,改进了旋转边界框表示方法,并在第二阶段检测中增加预测分支,用于预测旋转边界框的位置参数,还增加了 RoI Pooling 层的输出尺寸组合,用于捕捉更多的横向和纵向特征。为了解决 RRPN 算法中 RPN 网络产生大量组合锚框,导致检测速度低的问题, Ding 等^[34]提出了 RoI Transformer 算法,首先使 RPN 网络通过水平锚框的特征预测得到旋转锚框的回归参数,再将水平锚框转换成旋转锚框。此策略无需增加 RPN 中 Anchor 的数量即可获得较为精确的旋转区域候选框。图 5(a) 和图 5(b) 为 RRPN 和 RoI Transformer 生成 RRoI 方式的比较。

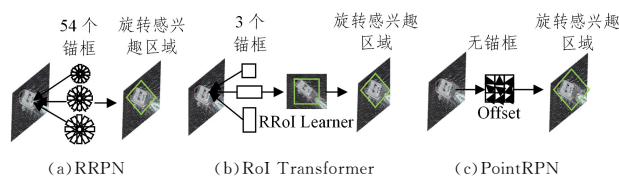


图 5 生成 RRoI 的不同方法的比较

Fig. 5 Comparison of different generation methods for RRoI

Yang 等^[16,42]对 R2-CNN 和 RRPN 算法思想进行迁移,提出了 AO-RCNN 模型,设计了自适应感兴趣区域对齐 (Adaptive ROI Align) 方法,减轻了候选框中冗余噪声的影响。APE^[43]提出了一种新的旋转框表示方法,用两个周期不同的二维周期向量表示角度参数,同时该向量可随形状变化而连续。R3Det^[13]提出了一个端到端精炼的单阶段旋转物体检测器,通过解决特征错位问题实现快速准确的物体定位。ICN^[44]使用图像级联、特征金字塔以及多尺寸卷积核提取特征,并使用旋转 RPN 以及多尺度预选框网络,得到最终的带角度边界框。Han 等提出了一种旋转等变检测器 ReDet^[35],在检测器中加入旋转等变网络来提取旋转等变特征,并提出了旋转不变 RoI 对齐 (RiRoI Align) 方法,根据 RoI 的方向自适应地从等变特征中提取旋转不变特征。

上述提到的旋转目标检测器均取得了一定效果,但都没有考虑参数表征的限制导致的旋转框边界突变问题,使模型训练出现不稳定的情况,甚至影响目标检测精度。接下来本文将详细介绍有关旋转矩形框表征存在的问题以及目前的改进策略。

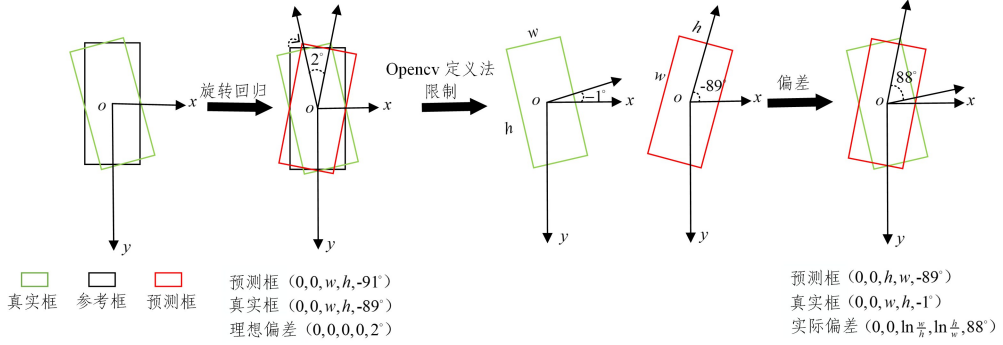
2.1.3 旋转矩形边界框局限性分析

OBB 一般使用五参数表示法,虽然参数简单,但是存在很大局限性。主要表现为以下 3 点:

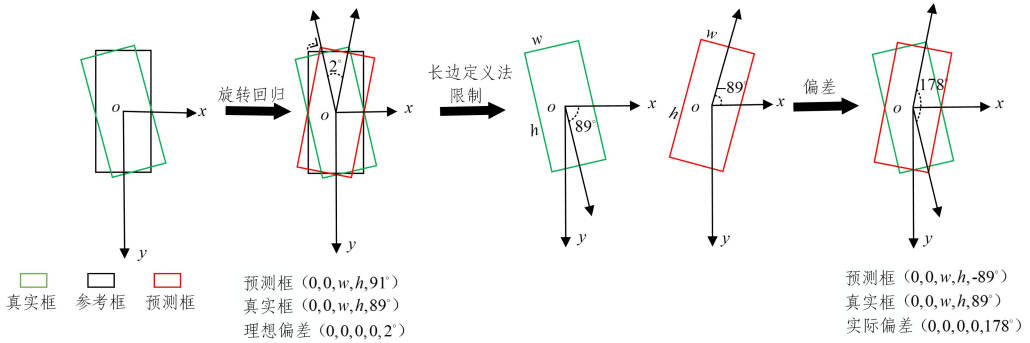
(1) 损失不连续问题^[27]:由于角和边的参数化,在旋转框处于水平或竖直方向这两个旋转角表示范围的边界时,会出现剧烈的 Loss 增长的现象。这主要是由旋转边界框的定义与损失函数不匹配造成的,具体表现如图 6(a) 和图 6(b) 所示。在长边表示法中,角度周期为 180° ,在边界值 0° 和 180° 时

边界框实际重合,但角度却在数值上有180°的差距,产生了损失突变,从而影响了角度的预测精度;Opencv 表示法中突变

问题更为复杂,在边界上除了角度产生突变以外,宽和高也会在边界发生交换,产生巨大损失。



(a)Opencv 表示法局限性分析



(b)长边表示法局限性分析

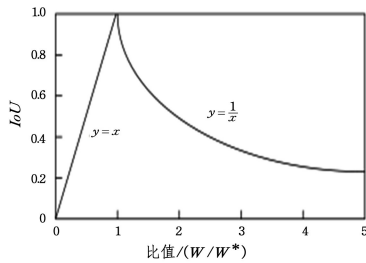
图 6 损失不连续问题示意说明

Fig. 6 Description of loss discontinuity problem

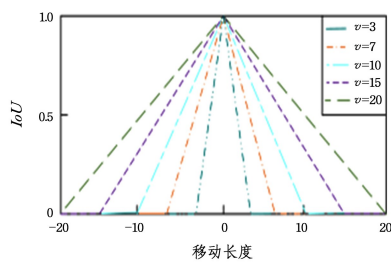
(2)类方形问题(SLP):对于类正方形目标,由于度量方式和损失的不一致问题,因此基于长边表示法的目标检测网络出现高IoU和高损失的情况,使参数回归困难[28]。

的关系是一个线性函数和反比例函数的组合。中心点和IoU之间的关系是一个对称的线性函数,如图7(b)所示。与其他参数完全不同的是,角度参数和IoU之间的关系是一个多重多项式函数,如图7(c)所示。图7(d)给出了3种关系的比较,这种回归的不一致性问题将导致训练收敛性和检测性能降低[27]。

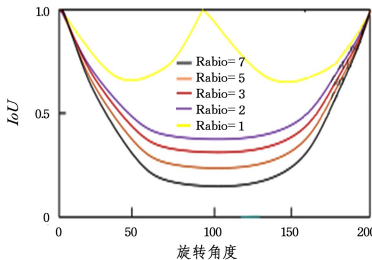
(3)回归不一致问题:在五参数表示法中,角度、宽度、高度和中心点坐标具有不同的测量单位,使它们与IoU表现出了不同的函数关系。如图7(a)所示,IoU和宽度(高度)之间



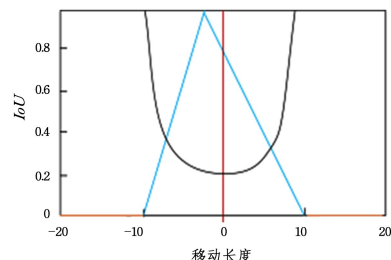
(a)宽度与IoU的关系



(b)中心点与IoU的关系



(c)角度和IoU的关系



(d)3种关系的比较

图 7 回归不一致问题

Fig. 7 Regression inconsistency problem

2.1.4 基于旋转矩形边界框算法的优化策略

为了解决上述问题,根据改进方式的不同将优化策略主要分为基于图像分割的改进、基于回归参数的改进,以及基于损失函数的改进。

(1) 基于图像分割改进的旋转目标检测

基于图像分割的改进指通过获得目标像素级分类,经过后处理操作后由掩膜(Mask)获得对应的旋转边界框。Mask OBB^[45]和 RADet^[46]基于分割产生的类别掩膜来生成最小外接矩形,避免了边界问题。RBMDeepNet^[47]提出了一种混合方法,将 CNN 和 RBM 相结合用于检测遥感图像中的汽车。Audebert^[48]提出了一种 Segment-Before-Detect 的框架来检测遥感图像中的汽车。这种基于分割的方法在文本检测中应用较多,在遥感目标检测领域应用较少。虽然基于分割的方法可有效解决边界问题,但其引入了额外的语义分割计算,计算复杂度较高,影响模型速度。

(2) 基于回归参数的改进

基于回归参数的改进主要是从回归参数入手,通过规避角度参数,避免出现与角度有关的边界突变问题。Yang 等提出的新的角度编码方式圆平滑标签(CSL)^[33],改变了神经网络的表示(Representation),将角度的计算从回归问题重新设计成了分类问题。该算法完全规避了角度边界问题,保证了预测的可接受度。但是由于 CSL 采用稀疏编码的方式设计标签,预测层参数数量巨大,影响了模型性能。同时,CSL 采用长边定义法并没有解决类方形问题。DCI^[49]算法提出了二进制编码和格雷码编码方式,相比 CSL,预测层的参数数量更少,训练速度更快,检测精度也更高。Wang 等^[50]提出了一种由粗粒度角度分类(CAC)和细粒度角度回归(FAR)组成的多粒度角度表示(MGAR)方法,提高了角度预测的精度,同时通过粗化离散角编码(DAE)降低了模型复杂度。

(3) 基于损失函数改进的旋转目标检测

旋转目标检测网络的回归损失大多沿用经典的范数式损失,如 Smooth-L1 损失,但在实际使用中,即使建议框与预测框拥有了相当的重合度,但由于损失函数与表示方法的局限,

损失值仍可能比较大,导致回归困难。因此,基于损失函数的改进成为了重要研究方向。

Yang 等提出了 SCRDet^[38]模型,结合 IoU 和 Smooth L1 损失提出了一个 IoU-Smooth L1 损失,通过在角度回归的边界处加入 IoU 常数因子,来缓解角度边界引起的损失急剧增加问题。后来 Yang 等^[51]结合 SCRDet 的去噪思想与 R3Det 的单阶段检测,提出了 SCRDet++,设计了改进的 Smooth L1 损失,通过常数因子缓解旋转矩形框回归的边界问题。Chen 等^[52]提出使用 PIoU(Pixels-IoU)损失来同时提高旋转角度和 IoU 的准确率。其本质是使用 IoU 的思想,但其以像素为单位进行计量。Constraint loss^[53]提出了一种解耦调制机制来解决损失突然变化的问题。S2A-NET^[54]提出了特征对齐模块(Feature Alignment Module,FAM)及方向检测模块(Oriented Detection Module,ODM)并采用 Focal Loss 损失函数,实现了全特征对齐,缓解了回归和分类之间不一致的问题。

Yang 等另辟蹊径提出 GWD^[28](Gaussian Wasserstein Distance)算法,该算法将旋转回归目标定义为高斯分布,使用 Wasserstein 距离度量高斯分布间的距离用于训练。无论边界框如何定义,GWD 均可有效解决边界不连续和类正方形问题。该算法作为启发式工作,设计了一种简单有效的度量方式,解决了旋转目标检测的基本问题,对旋转目标检测的未来发展起到了推进作用。但其本质上不是倾斜 IoU (SkewIoU)损失,Yang 等提出了一种基于卡尔曼滤波的倾斜 IoU,称为 KFIoU^[55],解决了 GWD 存在的问题。他们提出的方法可推广到三维空间检测,为该领域开辟了新的研究方向。未来通过将任意旋转矩形近似成一个二维的高斯分布,计算分布之间的 Wasserstein 距离,设计高斯回归损失,表征旋转框是非常值得研究的内容,具有广阔的应用前景。

各种基于 OBB 旋转目标检测方法的优缺点、主干结构、适用场景以及在典型数据集上的 mAP 值或 F 值对比如表 1 所列。

表 1 基于矩形边界框的旋转目标检测方法比较

Table 1 Comparison of rotating object detection methods based on rectangle bounding boxes

算法	骨干网络	主要贡献	局限性分析	适用场景	数据集	mAP	F 值
RRPN ^[12]	ResNet	1)提出了一种基于区域提取的方法,并将 RRoI 和旋转兴趣区域学习相结合,提高了文本检测过程中的效率	冗余旋转框多,计算量大;检测速度慢;旋转角度有限	场景文本检测 建筑物检测	DOTA ICDAR	61.01	— 77.0
		2)提出了一种新的任意旋转文本区域优化策略,优化了旋转文本检测器的性能			MSRA-TD500	—	74.0
R2-CNN ^[41]	ResNet101	1)将任意方向的文本检测转化为一个多任务问题 2)对每个 RPN 候选框进行了多个尺寸的 ROI 池化,充分利用了文本特征,提高了精度 3)增加了更小尺寸的 anchor,提升了小文本检测效果	检测速度慢	场景文本检测	ICDAR	—	82.5
RoI Transformer ^[34]	ResNet101	1)提出了一种旋转 RoI 的监督学习器,可以将 HRoIs 转化到 RRoIs,提高了检测精度	角度回归边界问题;结构复杂,检测速度慢	遥感图像检测	DOTA	69.56	—
		2)设计了一个旋转位置敏感的 RoI Align 模块用于空间不变性特征提取,具有高效、易嵌入的特点			HRSC	86.20	

(续表)

算法	骨干网络	主要贡献	局限性分析	适用场景	数据集	mAP	F 值
APE ^[43]	ResNet101	1) 提出了一种新边界框表征形式, 规避了回归角度出现的问题 2) 提出了新的 IOU 计算方式 (LIoU)	结构复杂, 检测速度慢	遥感图像检测	DOTA	75.75	—
R3Det ^[13]	ResNet101	1) 提出了一种近似 SkewIoU loss, 解决了 SkewIoU 的计算不导问题 2) 提出了单阶段端到端检测器并设计了特征精炼模块减少了检测边界框的数量, 提高了模型检测速度	模型训练速度较慢	场景文本检测 遥感图像检测	DOTA UCAS-AOD HRSC	73.79 96.17 89.26	—
ICN ^[44]	ResNet101	1) 提出了一种新的图像级联和特征金字塔网络 2) 从主干到检测头, 改善了整个网络的特征表示, 提高了检测精度 3) 可在更大图像尺度上提取信息	边界突变问题	遥感图像检测	DOTA UCAS-AOD	68.16 95.67	—
RADet ^[46]	ResNeXt101	1) 提出了旋转不变的 RoI Align, 可根据 RoI 的方向, 从等值特征中自适应地提取旋转不变的特征 2) 解决了图像大尺度问题	边界突变问题	航空图像检测	DOTA	69.09	—
Mask OBB ^[45]	ResNeXt101	1) 将角度预测转化为像素级分类问题, 规避了边界问题 2) 提出了 ILC-FPN (Inception Lateral Connection Network) 网络, 能够处理目标尺度变化大的问题 3) 提出了 SAN (Semantic Attention Network) 网络, 提供语义特征区分目标和背景	狭长目标检测效果较差	航拍图像	DOTA HRSC	75.33 96.7	—
CSL ^[33]	ResNet152	1) 提出了基于圆形平滑标 CSL 2) 将角度预测转为分类任务规避边界问题	存在分支不平衡问题	遥感图像检测 人脸检测	DOTA HRSC ICDAR	76.17 89.62 —	— — 83.6
SCRDet ^[38]	ResNet101	1) 设计了特征融合结构并提出了一种基于监督的多维注意力网络, 提高了密集小目标的检测精度 2) 增加 IoU 常数因子, 设计了一种改进的 Smooth L1 loss 解决旋转边界框回归的边界问题	存在边界问题 S	遥感图像检测	DOTA ICDAR	73.89 —	— 80.1
SCRDet++ ^[51]	ResNet101	1) 设计了新的实例级去噪 (InLD) 模块, 提高了对小物体的检测精度 2) 改进了 Smooth L1 loss, 改善了回归边界问题	角度周期问题	遥感图像检测	DOTA S2TLD DIOR	76.81 95.94 77.80	—
PIoU Loss ^[52]	DLA-34	1) 提出了新的损失函数 PIoU 损失, 提高了倾斜目标的检测效果 2) 提出了新的零售场景数据集 Retail50K, 可以更好地进行 OBB-based 算法的评估	存在边界问题	遥图像检测 零售场景检测	DOTA HRSC	60.5 89.2	—
GWD ^[28]	ResNet152	1) 提出 GWD (Gaussian Wasserstein Distance) 描述旋转 bbox 间的距离 2) 提出 GWD-based 损失解决了旋转角度范围不连续问题和方形问题	不具备尺度不变性, 对中心点优化时出现检测结果偏移; 边界框水平情况下退化到常用的损失函数时, 仍存在损失不一致问题	遥感图像检测 场景文本检测	DOTA HRSC	80.23 89.85	—
KFIOU ^[55]	Swin-Tiny	1) 将旋转框转换到高斯分布, 解决了度量方式和损失函数不一致的问题 2) 解决了边界不连续和类正方形问题 3) 将旋转目标检测从二维扩展到了三维空间	结构复杂, 检测速度慢	3D 目标检测 人脸检测 场景文本检测	DOTA KITTI(3D)	80.93 70.91	—

2.2 基于四边形边界框的目标检测算法

2.2.1 四边形边界框的定义

一些研究者提出使用八参数的四边形边界框来表征目标。这种方式规避了上文提到的非坐标参数导致的单位不一致问题, 具有参数自然一致性, 一定程度上避免了回归角度的问题。同时, 进行任意四边形的回归, 更贴合非矩形目标, 相较于矩形边界框表征物体更加准确, 适用于更复杂的应用场景。目前八参数表示法主要有有序四边形定义法和 Xu 等^[25]提出的顶点偏移定义法。

有序四边形定义法: 定义方式包含 $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$ 这 8 个自由度, 即目标 4 个顶点的横、纵坐标, 通过 4 个顶点坐标对水平框与衍生旋转框做逆时针排列来表征物体, 如图 8(a) 所示。

顶点偏移定义法: 借助与水平矩形框 4 个角点的相对偏移距离定义四边形, 水平矩形表示为 (x, y, w, h) , 四边形表示为 $(x, y, \alpha_1 = \frac{s_1}{w}, \alpha_2 = \frac{s_2}{h}, \alpha_3 = \frac{s_3}{w}, \alpha_4 = \frac{s_4}{h})$, (x, y) 为矩形边界框中点。具体形式如图 8(b) 所示。

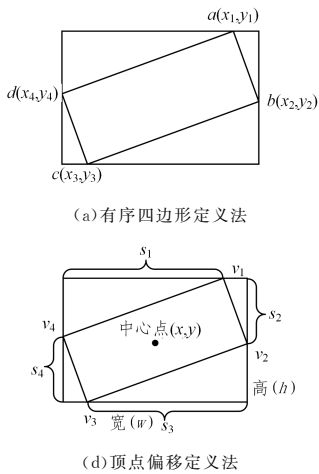


图8 八参数定义法

Fig. 8 Definition method of eight parameters

2.2.2 基于QBB的目标检测算法发展

He等在DirectRegression方法^[56]中首次提出“直接回归”的概念,不使用锚框作为桥梁,直接学习目标参数和GT参数的偏移量,得到用于表示任意方向目标的四边形来完成有向目标的检测。Zhou等^[57]提出EAST模型,解决了水平框表示倾斜文本的缺陷问题,该模型通过回归任意四边形检测框的位置坐标 $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$ 8个参数,来表征一些扭曲变形文本。Zhang等^[58]提出LOMO算法,算法由直接回归单元(DR)、迭代优化模块(IRM)和任意形状表达模块(SEM)3部分组成;通过SEM模块在文本区域的几何特征和IRM的基础上重建更加精确的多边形任意方向目标区域边界表示。TOSO^[59]提出学生T-分布,将分类任务与旋转角度回归相关联,通过预测目标的水平边界框和相对滑动偏移来表示四边形边界框。

尽管四边形边界框不存在与角度回归有关的问题,但是也存在与其表征形式相关的问题。

2.2.3 四边形边界框局限性分析

四边形边界框局限性的主要表现为以下两点。

(1) 表征模糊问题

基于四边形回归的方法通常在定义Ground-Truth(GT) 4个顶点顺序时含糊不清,导致对某些方向的目标产生意外的检测结果。具体表现为:四边形边界框主要是通过回归4个点,再连接成4条边。对于一个凸的四边形而言,4个顶点有 $P_4^4 = 24$ 种组合方式,均能表示唯一的凸四边形,这24种表示方式是等价的局部最优解。但是实际回归时,使用Smooth L1损失,只能指定一种情况学习,损失向唯一的全局最小进行优化,导致出现次优回归过程,使损失震荡,收敛速度变慢。

(2) 边界不连续问题

使用八参数表示,损失不连续的现象仍然存在,具体情况如图9所示。灰框为锚框,红框和绿框均为GT,顶点均已排序。当绿框为真实框时,理想和实际的回归方式都是: $a \rightarrow a_1, b \rightarrow b_1, c \rightarrow c_1, d \rightarrow d_1$;当红框为GT时,实际回归方式如图9中实线箭头所示,为: $a \rightarrow a_2, b \rightarrow b_2, c \rightarrow c_2, d \rightarrow d_2$;而理想回归方式如图9中虚线箭头所示,为: $a \rightarrow d_2, b \rightarrow a_2, c \rightarrow b_2, d \rightarrow c_2$,

由此产生较大的损失值。四边形定义法等基于点的表示方法均存在这类边界问题。

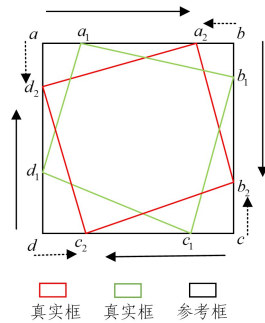


图9 八参数表示法边界问题示意图(电子版为彩图)

Fig. 9 Boundary problem of eight parameters

2.2.4 基于四边形边界框算法的优化策略

为了解决任意四边形表征存在的问题,他们改进方式的不同主要分为基于回归参数的改进和基于损失函数的改进。

(1) 基于回归参数改进

Liu等^[60]首次指出基于四角点回归方法会遇到顺序标注敏感的问题,为了解决这个问题,作者提出了SBD算法,利用与标签序列无关的不变表示形式反推边界框坐标,然后学习序列标记匹配类型以找出最佳拟合的四边形,摆脱了训练目标的模糊性。该方法是第一个根据四边形边界框的顺序解决文本检测歧义的方法,对提高检测精度起到了重要作用。Xu等^[25]提出了Gliding Vertex算法,使用顶点偏移定义法,通过改变框的表示方式避免了排序的问题。算法首先获得物体的水平框候选区域,然后通过回归水平框4个顶点的相对偏移得到物体的倾斜边界框。Wei等^[61]利用中心点和两条与中心点相互垂直的线段表示物体边界框的范围,共需要10个参数进行表示。除了以上介绍的在直角坐标系下的算法,也有研究人员通过PolarDet^[62]极坐标系方法获得物体坐标。

(2) 基于损失函数改进

RSDet^[27]采用基于叉积的算法来获得4个顶点的序列,同时提出了一种调制的旋转损失以消除损失的不连续,又将新损失与八参数回归相结合,解决了回归不一致问题。此算法取得了一定的改善效果,但算法的损失函数设计较为复杂,进一步增加了双阶段检测算法的计算量,影响了检测效率。

Ming等^[29]提出了一种表征不变性损失RIL来优化回归旋转物体的边界框。具体来说,RIL将一个定向物体的多种表征视为多个等价的局部最小值,将边界框回归转变为与这些局部最小值的自适应匹配过程。采用匈牙利匹配算法来获得最佳回归策略,有效解决了四边形边界框表征方式导致的次优学习问题。

四边形相比旋转矩形的定位更加精确,可以更加拟合物体的外形,并且可以很大程度缓解边界问题。但是当前基于四边形检测器的运算较为复杂,性能相较于旋转矩形检测器较差。

各种基于QBB旋转目标检测方法的优缺点、主干结构、适用场景以及在典型数据集上的mAP值或F值对比如表2所列。

表 2 基于四边形边界框的旋转目标检测方法比较

Table 2 Comparison of rotating object detection methods based on quadrangle bounding boxes

算法	骨干网络	主要贡献	局限性分析	适用场景	数据集	mAP/%	F 值
EAST ^[57]	FCN	提出了一种由两阶段组成的场景文本检测方法,精简了检测步骤,检测速度较快	对弯曲文字识别能力较差	场景文本检测	ICDAR MSRA- TD500	—	80.72 76.08
LOMO ^[58]	ResNet101	1)提出 TFAM,动态调整感受野 2)提出 PA-NMS,通过合并可靠的预测结果来提升性能 3)提出 Instance-aware IoU 损失,平衡不同尺度的训练	存在边界突变问题	场景文本检测	ICDAR	—	87.20
TOSO ^[59]	ResNet101	1)提出使用水平边界框和相对滑动偏移来表示四边形边界框 2)提出学生 T-分布,将分类任务与旋转角度回归相关联	存在边界突变问题	遥感图像检测	DOTA HRSC2016	57.92 79.29	—
SBD ^[60]	ResNet101	提出了一种简单而有效的匹配策略重构四边形边界框,摆脱了训练目标的模糊性,提高了检测精度	检测速度较慢	场景文本检测	ICDAR	—	90.10
Gliding Vertex ^[25]	ResNet101	1)提出了一种新的边界框表征方法,通过四边形边界框精确描述目标 2)提出了一种倾斜因子,解决了近水平目标的混淆问题	存在边界突变问题; 结构复杂,检测速度较慢	航空图像中的目标检测、场景文本检测、鱼眼图像中的行人检测	DOTA MSRA- TD500	75.02 —	— 86.50
PolarDet ^[62]	ResNet101	提出了极坐标表示方法,避免了角度边界问题	检测精度较低	航空图像检测	DOTA UCAS-AOD HRSC2016	76.64 97.01 90.46	—
RSDet ^[27]	ResNet152	提出了调制旋转损失,消除损失的不连续性	存在边界问题	遥感图像检测	DOTA UCAS-AOD	74.10 96.50	—
RIDet-O ^[29]	ResNet101	1)提出了一种表示不变性损失(RIL)算法,解决了次优回归优化及损失度量与预测的定位精度之间的不一致问题 2)提出了匈牙利匹配算法,找到了最佳回归策略 3)提出了归一化的旋转损失,减轻了不同变量之间的弱相关性	检测速度较慢	遥感图像检测	DOTA HRSC2016 UCAS-AOD	77.6 89.63 89.62	—

2.3 基于点集的目标检测算法

2.3.1 点集定义

上述两种表征方式无论是基于角度还是无角度的方法,对于 bounding box 的依赖都非常严重,这些 bounding box 都是规则且相对固定的候选框,只能对目标提供一个较为粗略的定位,导致提取出来的特征相对粗糙,检测精度较低。同时,上述两种表征方法都存在边界突变等问题。为此,有学者提出通过几个独立的点(Point Set)来表示目标。具体表征形式如下。

点集:通过一组 $\{(x_1, y_1), \dots, (x_i, y_i)\}$ 点集提供更细粒度的位置表示和便于分类的信息。如图 10 所示,这些点分布在目标的空间范围和具有重要语义信息的位置。

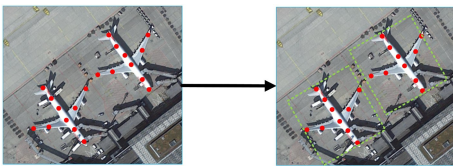


图 10 点集定义

Fig. 10 Definition of point set

2.3.2 点集局限性分析及优化策略

虽然通过点集表征具有较好的效果,但也存在有关孤立点的问题,具体如下。

孤立点问题:由于各点之间的独立优化,使得高精度检测器对孤立点非常敏感,尤其是对于大长宽比的物体,轻微的

偏差就会导致 IoU 急剧下降,虽然大部分的点都预测正确,但是出现一个孤立的点就会导致最终预测失败,情况如图 11 所示。



图 11 点集孤立点问题

Fig. 11 Isolated point problem of point set

这种基于点集的算法,需要设计辅助方法来获取最终的目标检测框。CFA^[63]提出了凸包特征适应方法(Convex-hull Feature Adaptation, CFA),通过可变性卷积生成 9 个特征点,并利用 Jarvis March 算法生成最小凸包。Li 等提出了 Oriented RepPoints^[64]算法,这是一种有效的自适应点学习方法。将自适应点集^[31]作为细粒度表示,同时引入定向转换函数,将代表点驱动在定向对象上实现自适应地向适当的位置移动,能够表征杂乱环境中方向变化剧烈的物体的几何形状。

Zhou 等提出了 PointRCNN^[65]算法,该算法主要由 Point-RPN 和 PointReg 组成,对于给定的输入特征图,PointRPN 以粗到细的方式为每个特征点学习一组代表点,并

通过 OpenCV^[66] 的 MinAreaRect 函数生成 RRoI,如图 5(c)所示。最后,PointReg 应用 RoI Align 算子来提取 RRoI 特征,并对角点的 8 个参数进行细化和分类,实现任意物体检测。

通过从代表性的点中提取特征能大大降低背景或包含少量语义信息的前景区域的影响,并且点集可以表征任意形状

的物体,很大程度上缓解了边界问题。但是,目前其检测性能还相对较低。

各种基于 Pointset 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 3 所列。

表 3 基于点集的旋转目标检测方法比较

Table 3 Comparison of rotational object detection methods based on point set

算法	骨干网络	主要贡献	局限性分析	适用场景	数据集	mAP%
CFA ^[63]	ResNet101	提出凸壳特征自适应(CFA)方法,有效检测密集方向任意物体	存在语义模糊问题	遥感图像检测	DOTA	75.05
Oriented RepPoint ^[64]	ResNet101	1)在 RepPoints 的基础上改进了 3 种定向转换函数	存在语义模糊问题	遥感图像检测	DOTA	76.52
		2)提出了一种有效针对自适应点的质量评估和样本分配方案(APAA)			UCAS-AOD	90.11
		3)检测精度高,解决了边界突变问题			DIOR-R	66.71
Point RCNN ^[65]	Swin-Tiny	1)提出了一种无角度旋转目标检测框架	存在语义模糊问题	遥感图像检测	HRSC2016	90.38
		2)提出了一种平衡数据集策略,解决了严重的不平衡数据集问题			DOTA	81.32
					HRSC2016	90.53

2.4 小结

每种边界框表征方法都有不同特点,均存在一些缺陷,目前有学者提出了 G-Rep^[67] 统一高斯表示法,以构建 OBB, QBB 和 Point set 的高斯分布,有效解决了上述 3 种表征方法存在的问题,实现了对各种表示法和边界问题的统一解决。

为了对不同表征算法进行充分评估,验证其对旋转目标的检测能力,本文以 DOTA-v1.0 数据集为基准,在显卡为 NVIDIA RTX 2080Ti 上进行了对比实验分析,并选取平均精度均值、检测速度、模型参数量、模型计算量作为不同检测算法的评估指标。各主要算法的结果如表 4 所列,加粗字体代表该类别的最优效果。

表 4 不同算法在 DOTA 数据集上的结果对比

Table 4 Comparison of results of different algorithms on DOTA dataset

数据集	边界框	模型	Backbone	Image Size	mAP/%	Speed/fps	Params/M	FLOPs/G	对比分析	
DOTA-v1.0	OBB	RoI-Trans-KFIoU	Swin-T-FPN	1024×1024	80.93	14.6	60.52	229.99	准确率较高,但参数量、计算量较大	
		Oriented R-CNN	ResNet50-FPN	1024×1024	80.87	16.2	41.13	211.42		
		KLD+R3Det	ResNet50-FPN	1024×1024	80.17	12.4	41.90	335.74		
		S2A-Net	ResNet50-FPN	1024×1024	79.42	16.0	38.57	196.91		
	R3Det	ResNet152-FPN	1024×1024	76.47	12.4	41.90	350.74			
	PolarDet	ResNet50	1024×1024	75.02	17.1	—	—			
	QBB	RIDet-O	ResNet101	1024×1024	74.70	15.8	—	—		准确率较低
		GlidingVertex	ResNet50-FPN	1024×1024	73.23	16.4	41.14	211.30		
	Point Set	Oriented RepPoints	Swin-T-FPN	1024×1024	81.32	16.7	—	—		准确率较低,但参数量较小
			ResNet50-FPN	1024×1024	75.97	16.1	36.61	194.32		
		CFA	ResNet50-FPN	1024×1024	73.45	12.5	36.62	243.00		

旋转目标检测除了边界问题外,还存在目标模糊、小目标规模大、尺度大等特点,诸多学者根据自然目标检测技术的思想,做了很多研究^[68-74],这对于提高检测精度、探索高性能的

旋转框框检测算法十分必要,有利于各个领域的发展。表 5 列举了目前常见表征方式的优缺点及适用场景,可供相关人员选择一款适合场景的目标检测器。

表 5 不同边界框表征方式比较

Table 5 Comparison of different bounding boxes representation methods

边界框	参数形式	自由度	是否含有角度	边界问题	通用性	适用场景
OBB	OpenCV 定义 (x, y, w, h, θ)	5	有	2 种	一般	目标方向任意 需要角度信息 目标形状偏矩形
OBB	长边定义 (x, y, w, h, θ)	5	有	3 种	一般	目标方向任意 需要角度信息 目标形状偏矩形
QBB	有序四边形定义 ($x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4$)	8	无	1 种	好	目标形状各异、方向任意
QBB	顶点偏移定义 ($x, y, \alpha_1 = \frac{s_1}{w}, \alpha_2 = \frac{s_2}{h}, \alpha_3 = \frac{s_3}{w}, \alpha_4 = \frac{s_4}{h}$)	6	无	1 种	好	目标形状各异、方向任意
Point Set	点集 ($(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$)	—	无	1 种	好	目标形状各异、方向任意

3 数据集和算法评价指标

本节介绍旋转目标检测常用的数据集和评估指标。

3.1 数据集

目前旋转目标检测的主要应用场景是遥感图像、文本图像及零售场景图像。下面介绍上述领域常见的公共数据集。

3.1.1 遥感数据集

DOTA^[75]数据集由武汉大学和华中科技大学联合制作和发布,是目前最大的定向遥感图像目标检测数据集之一。在DOTA-v1.0版本中包含2806张来自不同传感器和平台收集的航空遥感图像,共15个类别,包含188282个实例,每个实例使用4个顶点构成的边界框顺时针标注其位置。该数据集对旋转目标检测算法的发展起到了巨大的推动作用。

UCAS AOD^[76]数据集是由国科大模式识别实验室标注的遥感图像数据集,仅包含汽车、飞机两类目标。其中飞机数据集包括600张图像和3210架飞机,车辆数据集包括310张图像和2819辆车辆。后来团队对数据集进行补充,最终得到1000张飞机图像,共7482个实例;510张汽车图像,共7114个实例;910张反例图像。

HRSC2016^[77]数据集是由中科院自动化研究所Liu等从谷歌地球收集并标注的高分辨率光学遥感图像舰船检测数据集,其中包含1061张图片和2976个实例。图像尺寸范围在300×300到1500×900像素之间,大部分图像尺寸大于1000×600像素。

以上为光学遥感图像,下面介绍几个合成孔径雷达(Synthetic Aperture Radar, SAR)图像数据集。

RSDD-SAR^[78]数据集是舰船目标斜框检测数据集,是以国产高分3号卫星数据和TerraSAR-X卫星数据为基础,构建的一套包含多种成像模式、多种极化方式、多种分辨率的SAR舰船斜框检测数据集。该数据集包含多种成像模式、

多种极化方式、多种分辨率切片7000张,舰船实例10263个,具有旋转方向任意、长宽比大、小目标占比高和场景丰富的特点。

SSDD+^[79]数据集是在SSDD舰船数据集基础上构建的,将垂直边框变成了旋转边框,总共包含1160张图片,2456个舰船,平均每张图片的舰船数量为2.12。

3.1.2 文本数据集

ICDAR^[80]是进行场景文本检测和识别(Scene Text Detection and Recognition)任务最知名和常用的数据集。该数据集是由Google Glass在未聚焦的情况下随机拍摄的街头或者商场图片,包括1000张训练图片、500张测试图片。所有图片的尺寸均是1280×720像素点。

MSRA-TD500^[81]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,共500张自然场景图片。

COCO-Text^[82]数据集是在微软提供的目标识别数据集MSCOCO的基础上完善成型的,包含背景复杂的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的,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。

3.1.3 零售场景数据集

Retail50K^[52]数据集是来自不同国家和地区的志愿拍摄的超市图像,共47000张图片,唯一标注类型为货架层,在标注之前,图像尺寸被统一调整为600×800。该数据集可用于货架零售标签检测、自动货架分层、货架层和图像偏角估计等。

表6列出了每个数据集样例的统计信息,包括发布者和内容描述、目标类别的数量以及数据集中包含的图像。每种数据集的样例如图12所示。正是这些公开的图像数据集使得基于深度学习的旋转目标检测技术迅速发展。

表6 旋转目标检测数据集对比

Table 6 Comparison of rotating object detection datasets

数据集	类别总数	图像总数	实例总数	边界框	应用领域	公开时间
DOTA-v1.0 ^[75]	15	2806	188282	旋转框	遥感图像物体检测	2018
UCAS-AOD ^[76]	2	2420	14596	旋转框	遥感图像飞机及车辆检测	2015
HRSC2016 ^[77]	27	1061	2976	旋转框	遥感图像舰船检测	2016
RSDD-SAR ^[78]	1	7000	10263	旋转框	SAR舰船斜框检测	2022
SSDD+ ^[79]	1	1160	2456	旋转框	SAR舰船斜框检测	2017
ICDAR2015 ^[80]	—	1500	—	旋转框	场景文本检测	2015
MSRA-TD500 ^[81]	—	500	—	旋转框	场景文本检测	2012
COCO-Text ^[82]	—	63686	173589	旋转框	场景文本检测	2016
Retail50K ^[52]	—	—	—	旋转框	零售场景检测	2020



图12 数据集样例图

Fig. 12 Sample diagram of datasets

3.2 评价指标

目前旋转目标检测领域的主要评价内容包括精度和速度两个方面。

在速度方面,通常采用每秒检测帧数(Frames Per Second, FPS),即算法平均每秒能够检测的图像数量来评价其执行速度。FPS值越大,说明检测器的实时性越好。

在精度方面,评价指标包括召回率(Recall)、精确率(Precision)和平均精度(Average Precision, AP)、平均精度均值(Mean Average Precision, mAP)。

精确率指正确检测的目标占检测出总目标个数的比重,召回率指正确检测的目标占真实标注目标个数的比重。在旋转目标检测中,设 P_T 为正确检测的目标个数, P_F 为误检的目标个数, N_F 为漏检的目标个数。精确率、召回率的计算公式如下:

$$P = \frac{P_T}{P_T + P_F} \quad (1)$$

$$R = \frac{P_T}{P_T + N_F} \quad (2)$$

为确定一个检测框是否正确检测,还需要计算预测框与真实框之间的交集面积和预测框与真实框之间的并集面积之比,即 IOU(Intersection Over Union)。一般来说,IOU 会设置阈值,通常为 0.5,如果 IOU 大于阈值,那么判定预测结果正确;如果 IOU 小于阈值,那么判定预测结果错误。

平均精度(Average Precision, AP)是 Precision-Recall 曲线下与坐标轴包围的面积,计算式如下:

$$AP = \sum_{n=0} (r_{n+1} - r_n) \rho_{\text{interp}}(r_{n+1}) \text{ with } \rho_{\text{interp}}(r_{n+1}) = \max_{r: r \geq r_n} \rho(r) \quad (3)$$

其中, $\rho(r)$ 指在召回率为 r 时的准确率,它被定义为正确识别的物体数占总识别的物体个数的百分数。而评估所有类别的检测准确度的指标为平均精度均值(Mean Average Precision, mAP),其被定义为所有类别检测的平均准确度,计算方式如式(4)所示:

$$mAP = \frac{1}{N} \sum_{i=1,2,\dots,N} AP_i \quad (4)$$

本文将 mAP 作为检测算法性能评估的最终指标。

4 研究展望

旋转目标检测算法主要研究如何从复杂背景中定位相对较小的目标,如何获得具有丰富信息的特征层,如何在只添加少量参数的情况下有效获得精确旋转框,以及如何解决各种表征方法的边界问题。结合现有的检测算法和当前的主要应用领域,本小节主要分析计算机视觉下旋转目标检测技术存在的问题,并对未来的发展趋势进行合理展望。

目前现存的问题主要包括:

(1)数据规模不足问题。数据在深度学习中占据着非常重要的地位,没有数据的支撑,再好的深度学习算法只能是一纸空谈。目前基于旋转框的数据集数量与水平标注的数据集数量相差较大,前者数据规模较小,并且仅局限于遥感和文本等部分场景中,限制了模型的发展。

(2)超大图像尺寸问题。对于遥感图像领域,遥感图像具有超大的图像尺寸和覆盖面积。现有方法尚不能直接对图像

全局进行检测,典型的分块处理方式效率较低,引入了过多冗余的计算,同时容易割裂目标,造成精度上的损失。

(3)目标大小各异、密集分布等问题。由于旋转目标检测主要应用在文本和遥感图像领域,而在这些领域的目标往往存在密集的小目标,这些小目标在特征提取过程中提供的细节信息有限,且随着网络的加深,小目标的细节信息会逐渐丢失,同时还存在目标分布不均和尺度变化剧烈等问题。

(4)背景复杂问题。旋转目标检测图像背景往往极为丰富,存在大量形似目标,并且背景包含大量噪声信息,弱化了目标特征,极易导致漏检错检。

(5)旋转表征问题。现阶段普遍使用的旋转表征方法主要有3种:1)带有角度的旋转矩形框;2)由4个顶点表征的四边形边界框;3)通过几个单独的点来表征任意方向的物体。3种方法均存在一定的不足,比如第一种和第二种存在边界突变、模糊表征、类正方形等问题,第三种存在孤立点敏感问题。

针对上述问题,结合自然图像中深度学习目标检测方法的发展趋势,本文展望了该领域未来的发展趋势。

(1)丰富数据集规模,提高数据集质量。目前数据集标注大都由人工标注,费时费力且效率较低。由于生成式对抗网络模型具有很强的数据扩充能力,研究基于生成式对抗网络的数据增强方法,能够提高数据集的完备性、泛化性以及表征图像特点的充分性,使得数据足以支撑网络学习,这也是所有目标检测算法共同努力的方向。

(2)超大尺寸图像目标检测。研究采用基于图像整体进行感兴趣区域提取的方式来替代分块处理方式,快速滤除大部分背景区域,从而避免计算冗余,提高算法效率。

(3)注意力机制区分图像前后背景。由于注意力机制可选择聚焦位置,产生更具分辨性的特征,因此对于旋转目标检测,研究结合注意力机制的旋转目标检测方法,可快速提取注意力感知(Attention-aware)特征,找到感兴趣区域,忽略其他干扰因素,实现对图像前后背景更准确的检测。

(4)骨干网络设计。底层骨干网络性能好坏直接影响旋转目标检测算法的性能,因此获取高性能的骨干网络将对检测精度与后续研究产生重要影响。未来可以从以下几点进行研究:研究针对不同目标检测领域的专用骨干网络,提高骨干网络针对不同任务的专一性;研究提高骨干网络针对小目标的敏感性,从根本上提高旋转目标检测算法的性能;研究提高骨干网络的处理效率,在尽量减少精度损失的前提下减少计算量,提高检测效率。对骨干网络的设计也将是旋转目标检测的重点研究方向。

(5)解决旋转边界突变问题。对于边界突变问题有以下几种解决方式:1)寻找一种新的旋转目标定义方式,该方式中不含周期变化性的参数,但又能表示周期旋转的目标物体,从根本上杜绝边界问题的产生,比如使用点集(Point Set)方法;2)从损失函数入手,使用合理的函数,赋予损失函数和角度同样的周期性,使得边界处角度差值可以很大,但损失值变化实际很小,或综合考虑所有回归参数的影响,使用旋转 IoU 损失函数规避边界问题;3)将旋转角由回归问题转为分类问题。目前来看,使用基于点集检测算法可较好地解决旋转框边界

突变问题,值得学者深入探索。

关于旋转目标检测存在的这些问题,目前还是开放问题,没有统一的解决方案。旋转目标检测也会随着自然目标检测技术的发展以及算法的进步,进一步提高检测精度。

结束语 本文以旋转目标检测为主题,着重对近年来出现的遥感图像、文本图像等领域的旋转目标检测方法进行了综述。文章首先分析了旋转目标检测的背景和意义;接着按照旋转边界框不同表征方法,详细阐述了现有旋转目标检测算法的结构及优化策略;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。

参 考 文 献

- [1] GIRSHICK R. Fast R-CNN [C]//GIRSHICK R. GIRSHICK R [C]//IEEE International Conference on Computer Vision. Piscataway:IEEE Press,2015:1440-1448.
- [2] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.
- [3] 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.
- [4] 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. Las Vegas, NV, USA: IEEE Comp. Soc. ,2016:779-788.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [C]//European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.
- [6] 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. Venice, Italy: ICCV, 2017: 2980-2988.
- [7] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist text proposal network [C]//European Conference on Computer Vision. Berlin: Springer, 2016: 56-72.
- [8] LIAO M, SHI B, BAI X, et al. TextBoxes: A Fast Text Detector with a Single Deep Neural Network [C]//Proceedings of the AAAI Conference on Artificial Intelligence. San Francisco, USA: ACM, 2017, 31(1): 4161-4167.
- [9] SHEN Y, LIU D, ZHANG F, et al. Fast and accurate multi-class geospatial object detection with large-size remote sensing imagery using CNN and Truncated NMS [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 191: 235-249.
- [10] SHI P, ZHAO Z, FAN X, et al. Remote Sensing Image Object Detection Based on Angle Classification [J]. IEEE Access, 2021, 9: 118696-118707.
- [11] ZHANG L, ZHANG Y S, YU Y, et al. Survey on object detection in tilting box for remote sensing images [J]. National Remote Sensing Bulletin, 2022, 26(9): 1723-1743.
- [12] MA J, SHAO W, YE H, et al. Arbitrary-oriented scene text detection via rotation proposals [J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.
- [13] YANG X, YAN J, FENG Z, et al. R3det: Refined single-stage detector with feature refinement for rotating object [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Online: AAAI, 2021, 35(4): 3163-3171.
- [14] 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.
- [15] MING Q, MIAO L, ZHOU Z, et al. Task interleaving and orientation estimation for high-precision oriented object detection in aerial images [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 196: 241-255.
- [16] YANG X, SUN H, SUN X, et al. Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network [J]. IEEE Access, 2018, 6: 50839-50849.
- [17] SHI X, SHAN S, KAN M, et al. Real-time rotation-invariant face detection with progressive calibration networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 2295-2303.
- [18] LIU T, LI W G, GUAN J H. A Review of Object Detection Methods in Optical Remote Sensing Image Based on Deep Learning [J]. Radio Communication Technology, 2020, 46(6): 624-634.
- [19] NIE G T, HUANG H. A survey of object detection in optical remote sensing images [J]. Acta Automatica Sinica, 2021, 47(8): 1749-1768.
- [20] SONG Z N, SUI H G, LI Y C. A survey on ship detection technology in high-resolution optical remote sensing images [J]. Geomatics and Information Science of Wuhan University, 2021, 46(11): 1703-1715.
- [21] WANG J X, WANG Z Y, TIAN X. A Review of Text Detection and Recognition in Natural Scenes Based on Deep Learning [J]. Journal of Software, 2020, 31(5): 1465-1496.
- [22] ZHAO T Z, YANG C W, LIU W. Remote sensing image target detection method based on non-local feature enhancement [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2021, 49(9): 47-51.
- [23] SCHILLING H, BULATOV D, NIESSNER R, et al. Detection of vehicles in multisensor data via multibranch convolutional neural networks [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(11): 4299-4316.
- [24] LIAO M, SHI B, BAI X. Textboxes++: A single-shot oriented scene text detector [J]. IEEE Transactions on Image Processing, 2018, 27(8): 3676-3690.
- [25] XU Y, FU M, WANG Q, 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.
- [26] SUN P, ZHENG Y, ZHOU Z, et al. R4 Det: Refined single-stage detector with feature recursion and refinement for rotating ob-

- ject detection in aerial images[J]. *Image and Vision Computing*, 2020,103:104036.
- [27] QIAN W, YANG X, PENG S, et al. Learning modulated loss for rotated object detection[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Online: AAAI, 2021, 35 (3): 2458-2466.
- [28] YANG X, YAN J, MING Q, et al. Rethinking rotated object detection with gaussian wasserstein distance loss[C]// *International Conference on Machine Learning*. PMLR, 2021: 11830-11841.
- [29] MING Q, MIAO L, ZHOU Z, et al. Optimization for arbitrary-oriented object detection via representation invariance loss[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 19: 1-5.
- [30] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints[C]// *Proceedings of the European Conference on Computer Vision (ECCV)*. Munich: ECCV, 2018: 734-750.
- [31] YANG Z, LIU S, HU H, et al. Reppoints: Point set representation for object detection[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE, 2019: 9657-9666.
- [32] HOU L, LU K, YANG X, et al. G-Rep: Gaussian Representation for Arbitrary-Oriented Object Detection[J]. *arXiv*: 2205. 11796, 2022.
- [33] YANG X, YAN J. Arbitrary-oriented object detection with circular smooth label[C]// *European Conference on Computer Vision*. Berlin: Springer, 2020: 677-694.
- [34] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seoul, Korea (South): ICCV, 2019: 2849-2858.
- [35] HAN J, DING J, XUE N, et al. Redet: A rotation-equivariant detector for aerial object detection[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Montreal Canada: ICCV, 2021: 2786-2795.
- [36] MA J, SHAO W, YE H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. *IEEE Transactions on Multimedia*, 2018, 20(11): 3111-3122.
- [37] XIE X, CHENG G, WANG J, et al. Oriented R-CNN for object detection[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal Canada: ICCV, 2021: 3520-3529.
- [38] YANG X, YANG J, YAN J, et al. Scrdet: Towards more robust detection for small, cluttered and rotated objects[C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): ICCV, 2019: 8232-8241.
- [39] LYU C, ZHANG W, HUANG H, et al. Rtmtdet: An empirical study of designing real-time object detectors[J]. *arXiv*: 2212. 07784, 2022.
- [40] AN S B, LOU H R, CHEN S W, et al. Research progress of rotating target detection method based on deep learning [J]. *Electronic Measurement Technology*, 2021, 44(21): 168-178.
- [41] JIANG Y, ZHU X, WANG X, et al. R2CNN: Rotational region CNN for orientation robust scene text detection [J]. *arXiv*: 1706. 09579, 2017.
- [42] YANG X, SUN H, FU K, et al. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks[J]. *Remote Sensing*, 2018, 10(1): 132.
- [43] ZHU Y, DU J, WU X. Adaptive period embedding for representing oriented objects in aerial images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(10): 7247-7257.
- [44] AZIMI S M, VIG E, BAHMANYAR R, et al. Towards multi-class object detection in unconstrained remote sensing imagery [C]// *Asian Conference on Computer Vision*. Berlin: Springer, 2018: 150-165.
- [45] WANG J, DING J, GUO H, et al. Mask OBB: A semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images[J]. *Remote Sensing*, 2019, 11(24): 2930.
- [46] LI Y, HUANG Q, PEI X, et al. RADet: Refine feature pyramid network and multi-layer attention network for arbitrary-oriented object detection of remote sensing images[J]. *Remote Sensing*, 2020, 12(3): 389.
- [47] KHOSHBORESH M M, SHAH-HOSSEINI R. A hybrid deep learning-based model for automatic car extraction from high-resolution airborne imagery[J]. *Applied Geomatics*, 2020, 12(2): 107-119.
- [48] AUDEBERT N, LE SAUX B, LEFÈVRE S. Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images[J]. *Remote Sensing*, 2017, 9(4): 368.
- [49] YANG X, HOU L, ZHOU Y, et al. Dense label encoding for boundary discontinuity free rotation detection[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Montreal, Canada: ICCV, 2021: 15819-15829.
- [50] WANG H, HUANG Z, CHEN Z, et al. Multi-Grained Angle Representation for Remote Sensing Object Detection[J]. *arXiv*: 2209. 02884, 2022.
- [51] YANG X, YAN J, YANG X, et al. SCRDet++: Detecting Small, Cluttered and Rotated Objects via Instance-Level Feature Denoising and Rotation Loss Smoothing[J]. *arXiv*: 2004. 13316, 2020.
- [52] CHEN Z, CHEN K, LIN W, et al. PIoU Loss: Towards Accurate Oriented Object Detection in Complex Environments[C]// *European Conference on Computer Vision*. Berlin: Springer, 2020: 195-211.
- [53] ZHANG L, WANG H, WANG L, et al. Constraint Loss for Rotated Object Detection in Remote Sensing Images[J]. *Remote Sensing*, 2021, 13(21): 4291.
- [54] HAN J, DING J, LI J, et al. Align deep features for oriented object detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-11.
- [55] YANG X, ZHOU Y, ZHANG G, et al. The KFIoU Loss for Rotated Object Detection[J]. *arXiv*: 2201. 12558, 2022.
- [56] HE W, ZHANG X Y, YIN F, et al. Deep direct regression for multi-oriented scene text detection [C]// *Proceedings of the IEEE International Conference on Computer Vision*. Venice-Italy: ICCV, 2017: 745-753.
- [57] ZHOU X, YAO C, WEN H, et al. East: an efficient and accurate scene text detector[C]// *Proceedings of the IEEE Conference on*

- Computer Vision and Pattern Recognition. Venice, Italy; ICCV, 2017; 5551-5560.
- [58] ZHANG C, LIANG B, HUANG Z, et al. Look more than once: An accurate detector for text of arbitrary shapes[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seoul, Korea(south); ICCV, 2019; 10552-10561.
- [59] FENG P, LIN Y, GUAN J, et al. Toso: Student's distribution aided one-stage orientation target detection in remote sensing images[C]// 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020). Barcelona, Spain; IEEE, 2020; 4057-4061.
- [60] LIU Y, HE T, CHEN H, et al. Exploring the capacity of an orderless box discretization network for multi-orientation scene text detection[J]. International Journal of Computer Vision, 2021, 129(6): 1972-1992.
- [61] WEI H, ZHANG Y, CHANG Z, et al. Oriented objects as pairs of middle lines[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 169: 268-279.
- [62] ZHOU L, WEI H, LI H, et al. Arbitrary-oriented object detection in remote sensing images based on polar coordinates[J]. IEEE Access, 2020, 8: 223373-223384.
- [63] GUO Z, ZHANG X, LIU C, et al. Convex-hull Feature Adaptation for Oriented and Densely Packed Object Detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(8): 5252-5265.
- [64] LI W, CHEN Y, HU K, et al. Oriented reppoints for aerial object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, Louisiana; CVPR, 2022; 1829-1838.
- [65] ZHOU Q, YU C. Point RCNN: An Angle-Free Framework for Rotated Object Detection[J]. Remote Sensing, 2022, 14(11): 2605.
- [66] BRADSKI G. The OpenCV library[J]. Dr. Dobb's Journal: Software Tools for the Professional Programmer, 2000, 25(11): 120-123.
- [67] HOU L, LU K, YANG X, et al. G-Rep: Gaussian Representation for Arbitrary-Oriented Object Detection[J]. arXiv:2205.11796, 2022.
- [68] CHEN H B, JIANG S, HE G, et al. TEANS: A target enhancement and attenuated nonmaximum suppression object detector for remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(4): 632-636.
- [69] HOU L P, LU K, XUE J, et al. Cascade detector with feature fusion for arbitrary-oriented objects in remote sensing images[C]// Proceedings of 2020 IEEE International Conference on Multimedia and Expo. Piscataway; IEEE Press, 2020; 1-6.
- [70] LU X, JI J, XING Z, et al. Attention and feature fusion SSD for remote sensing object detection[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-9.
- [71] FU K, CHANG Z, ZHANG Y, et al. Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 161: 294-308.
- [72] DENG Z, SUN H, ZHOU S, et al. Multi-scale object detection in remote sensing imagery with convolutional neural networks[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 145: 3-22.
- [73] DONG Z, WANG M, WANG Y, et al. Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable object scale features[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 58(3): 2104-2114.
- [74] 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.
- [75] 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, UT, USA; CVPR, 2018; 3974-3983.
- [76] ZHU H, CHEN X, DAI W, et al. Orientation robust object detection in aerial images using deep convolutional neural network[C]// 2015 IEEE International Conference on Image Processing (ICIP). Quebec City, QC, Canada; IEEE, 2015; 3735-3739.
- [77] 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.
- [78] XU C A, SU H, LI J W, et al. RSDD-SAR; SAR ship oblique frame detection dataset[J]. Journal of Radar, 2022, 11(4): 581-599.
- [79] ZHANG T, ZHANG X, LI J, et al. Sar ship detection dataset (ssdd): Official release and comprehensive data analysis[J]. Remote Sensing, 2021, 13(18): 3690.
- [80] 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). Tunis, Tunisia; IEEE, 2015; 1156-1160.
- [81] 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. Providence, RI, USA; IEEE, 2012; 1083-1090.
- [82] VEIT A, MATERA T, NEUMANN L, et al. Coco-text: Dataset and benchmark for text detection and recognition in natural images[J]. arXiv:1601.07140, 2016.



WANG Xu, born in 1996, Ph.D candidate. His main research interests include rotating target detection and image processing.



WU Yanxia, born in 1979, Ph.D, professor, is a member of China Computer Federation. Her main research interests include computer architecture and compiler technology.