



基于无锚框的目标检测方法及其在复杂场景下的应用进展

刘小波 肖肖 王凌 蔡之华 龚鑫 郑可心

Anchor-free Based Object Detection Methods and Its Application Progress in Complex Scenes

LIU Xiao-Bo, XIAO Xiao, WANG Ling, CAI Zhi-Hua, GONG Xin, ZHENG Ke-Xin

在线阅读 View online: <https://doi.org/10.16383/j.aas.c220115>

您可能感兴趣的其他文章

基于中心点搜索的无锚框全卷积孪生跟踪器

AFST: Anchor-free Fully Convolutional Siamese Tracker With Searching Center Point

自动化学报. 2021, 47(4): 801–812 <https://doi.org/10.16383/j.aas.c200469>

目标检测模型及其优化方法综述

A Survey of Object Detection Models and Its Optimization Methods

自动化学报. 2021, 47(6): 1232–1255 <https://doi.org/10.16383/j.aas.c190756>

基于深度学习的光学遥感图像目标检测研究进展

Research Progress of Optical Remote Sensing Image Object Detection Based on Deep Learning

自动化学报. 2021, 47(9): 2078–2089 <https://doi.org/10.16383/j.aas.c190455>

基于显著图融合的无人机载热红外图像目标检测方法

Object Detection Method Based on Saliency Map Fusion for UAV-borne Thermal Images

自动化学报. 2021, 47(9): 2120–2131 <https://doi.org/10.16383/j.aas.c200021>

光学遥感图像目标检测算法综述

A Survey of Object Detection in Optical Remote Sensing Images

自动化学报. 2021, 47(8): 1749–1768 <https://doi.org/10.16383/j.aas.c200596>

结合目标检测的人体行为识别

Human Action Recognition Combined With Object Detection

自动化学报. 2020, 46(9): 1961–1970 <https://doi.org/10.16383/j.aas.c180848>

基于无锚框的目标检测方法及其在复杂场景下的应用进展

刘小波^{1,2,3,4} 肖肖^{1,2,3} 王凌⁵ 蔡之华⁶ 龚鑫^{1,2,3} 郑可心^{1,2,3}

摘要 基于深度学习的目标检测方法是当前计算机视觉领域的热点,在目标识别、跟踪等领域发挥了重要的作用.随着研究的深入开展,基于深度学习的目标检测方法主要分为有锚框的目标检测方法和无锚框的目标检测方法,其中无锚框的目标检测方法无需预定义大量锚框,具有更低的模型复杂度和更稳定的检测性能,是目前目标检测领域中较前沿的方法.在调研国内外相关文献的基础上,梳理基于无锚框的目标检测方法及各场景下的常用数据集,根据样本分配方式不同,分别从基于关键点组合、中心点回归、Transformer、锚框和无锚框融合等 4 个方面进行整体结构分析和总结,并结合 COCO (Common objects in context) 数据集上的性能指标进一步对比.在此基础上,介绍了无锚框目标检测方法在重叠目标、小目标和旋转目标等复杂场景情况下的应用,聚焦目标遮挡、尺寸过小和角度多等关键问题,综述现有方法的优缺点及难点.最后对无锚框目标检测方法中仍存在的问题进行总结并对未来发展的应用趋势进行展望.

关键词 无锚框, 关键点, 中心点, Transformer, 复杂场景, 目标检测

引用格式 刘小波, 肖肖, 王凌, 蔡之华, 龚鑫, 郑可心. 基于无锚框的目标检测方法及其在复杂场景下的应用进展. 自动化学报, 2023, 49(7): 1369–1392

DOI 10.16383/j.aas.c220115

Anchor-free Based Object Detection Methods and Its Application Progress in Complex Scenes

LIU Xiao-Bo^{1,2,3,4} XIAO Xiao^{1,2,3} WANG Ling⁵ CAI Zhi-Hua⁶ GONG Xin^{1,2,3} ZHENG Ke-Xin^{1,2,3}

Abstract The object detection method based on deep learning is a hot spot in the field of computer vision currently, which plays an important role in object recognition and tracking. With the in-depth development of research, object detection method are mainly divided into anchor-based object detection method and anchor-free object detection method, the anchor-free object detection method is state-of-the-art method currently, which can reduce model complexity and reach higher detection ability, owing to the fact that there is no need to predefine a large number of anchor boxes for it. Firstly, this paper summarizes the anchor-free object detection method and analyzes the performance on the common objects in context (COCO) dataset for further comparison in recent years, which are divided into keypoints combination, center point regression, Transformer, anchor-based fusing with anchor-free methods respectively, according to different sampling methods of anchor points. On the basis, in order to solve the above key problems, this paper focuses the applications of anchor-free object detection method in complex scenes. Finally, discusses the problems and the application trends of the anchor-free object detection method in the future development.

Key words Anchor-free, keypoints, center point, Transformer, complex scenes, object detection

Citation Liu Xiao-Bo, Xiao Xiao, Wang Ling, Cai Zhi-Hua, Gong Xin, Zheng Ke-Xin. Anchor-free based object detection methods and its application progress in complex scenes. *Acta Automatica Sinica*, 2023, 49(7): 1369–1392

收稿日期 2022-02-24 录用日期 2022-11-03

Manuscript received February 24, 2022; accepted November 3, 2022

国家自然科学基金(61973285, 62076226, 61873249, 61773355), 地质探测与评估教育部重点实验室主任基金 (GLAB2023ZR08) 资助
Supported by National Natural Science Foundation of China (61973285, 62076226, 61873249, 61773355) and Opening Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education (GLAB2023ZR08)

本文责任编辑 刘青山

Recommended by Associate Editor LIU Qing-Shan

1. 中国地质大学(武汉)自动化学院 武汉 430074 2. 复杂系统先进控制与智能自动化湖北省重点实验室 武汉 430074 3. 地球探测智能化技术教育部工程研究中心 武汉 430074 4. 地质探测与评估教育部重点实验室, 中国地质大学(武汉) 武汉 430074 5. 清华大学自动化系 北京 100084 6. 中国地质大学(武汉)计算机学院 武汉 430074

深度学习方法在目标检测领域具有优秀的性能,在视频监控、智能交通、国防安全等领域得到了广泛应用^[1]. 基于深度学习的目标检测方法在海量数据的驱动下,能通过特征提取网络学习到具有更强语义表征能力的特征,同时在神经网络前向传播

1. School of Automation, China University of Geosciences, Wuhan 430074 2. Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074 3. Engineering Research Center of Intelligent Technology for Geo-Exploration, Ministry of Education, Wuhan 430074 4. Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan 430074 5. Department of Automation, Tsinghua University, Beijing 100084 6. School of Computer Science, China University of Geosciences, Wuhan 430074

过程中避免了大量窗口的冗余计算,提升整体的检测速度的同时,检测精度也得到大幅度的提升.其中,基于深度学习的目标检测方法根据样本选择方式可分为:基于锚框的目标检测方法和基于无锚框的目标检测方法.

基于锚框的目标检测方法首先对每个像素点预设不同尺度和宽高比的锚框;其次使用特征提取网络来获取特征并生成特征图;然后利用检测头(Head)模块选择包含待测目标的锚框并计算目标类别置信度,利用位置损失和类别损失分别修正锚框的具体位置和类别;最后利用非极大值抑制(Non-maximum suppression, NMS)^[2]筛选同一目标的冗余锚框,实现最终的检测任务.其中,在训练阶段,常设置 1:1、1:2 和 2:1 等比例的锚框宽高比,并且常通过实验经验、K-means 聚类等方法选择锚框尺寸,用以适配不同目标.但由于预设锚框无法有效匹配小目标、宽高比例较大的目标,因此常根据特征金字塔来合理分配预设锚框,以提高检测的有效性和灵活性.此外,通过计算锚框和真实框的交并比(Intersection-over-union, IoU)来选择正负样本,即将大于 IoU 阈值的锚框作为正样本,剩余为负样本.在损失函数设计部分,常使用 L_1 损失、Smooth L_1 损失作为回归损失,用于回归锚框中心点坐标以及宽高参数,并且利用交叉熵损失或者焦点损失(Focal loss)来预测锚框类别的分类得分.此外,在预测阶段,计算每层特征金字塔的分类得分和位置回归偏置值,在对所有层的输出分类按类别进行排序后,将得分最高的部分锚框进行位置回归,最后利用非极大值抑制处理来得到最后的检测结果.其中,具有代表性的有锚框检测算法是 SSD (Single shot MultiBox detector)^[3]、R-CNN (Region based convolutional neural network)^[4-5]、YOLO (You only look once) v2^[6] 和 YOLO v3^[7] 等算法,这类算法在标准目标检测数据集上取得了较好的效果,推动目标检测的新发展,广泛应用于人脸识别、车牌识别、交通路况检测、农业病虫害监测、海洋救援等领域^[8].

但由于此类检测方法对锚框依赖性较强,从而导致以下问题:

- 1) 锚框的设计需要尺度、特征和类别等先验信息,而固定锚框尺寸、锚框预测宽高比和锚框数量在一定程度上影响检测网络的普适性;
- 2) 锚框的设置方法需引入大量冗余超参数,网络调节过程更加困难,且训练时需使用 IoU 计算锚框与真实框匹配程度,整体计算资源消耗较大;
- 3) 为了尽可能地精准匹配目标,需要预设数量繁多的锚框,而其中大部分锚框为仅含背景信息的负样本,只有少量为包含目标信息的正样本,因此会加剧目标检测存在的正负样本失衡问题.

针对有锚框检测方法的缺陷,Huang 等^[9]率先提出一种无锚框目标检测方法 DenseBox.该方法无需锚框预设过程,直接预测目标的边界框,不仅减少了大量锚框相关超参数,而且提升了整体模型的检测效率、降低模型复杂度,在无人驾驶、交通指挥以及国土安全等领域得到广泛应用.图 1 和图 2 分别表示有锚框和无锚框目标检测方法的整体框架,两类方法的不同之处在于是否定义先验锚框来提取候选预测框.

除了目标检测算法外,数据集也是目标检测发展的重要基石.公共数据集能有效地对不同算法进行横向比较,极大地促进了目标检测算法的发展,近年涌现一批目标检测公共数据集,具体统计特征如表 1 所示.其中 Pascal VOC (Pattern analysis statistical modeling and computational learning visual object classes)^[10]、COCO^[11] 数据集专用于评价通用水平框检测模型的性能;DOTA (Dataset for object detection in aerial images)^[12] 有针对性地检测高分辨率复杂图像中飞机、舰船等分布不均、角度多变的的多类目标,且同时使用水平框和旋转框,其中目标方向多变、尺度变化范围大,是目前使用最广、挑战性最强的数据集之一;此外,UCAS-AOD^[13] 包含汽车和飞机两类目标;ICDAR2015^[14] 针对场景文本数据进行归纳;CUHK-SYSU^[15]、PRW^[16]、

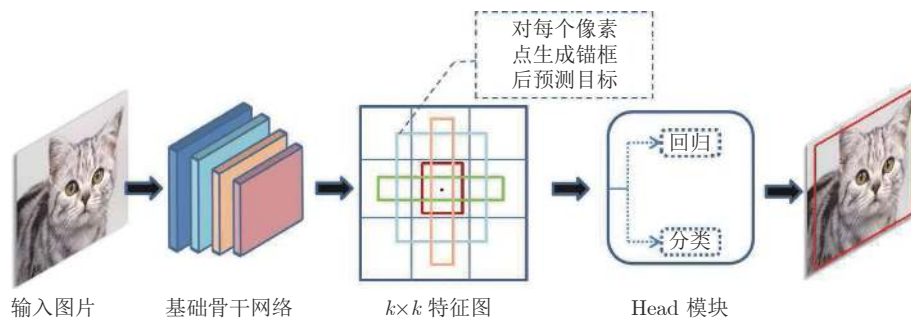


图 1 基于锚框的目标检测方法整体框架

Fig. 1 The overall framework of anchor-based object detection method

CrowdHuman^[17] 等数据集专用于行人检测, 其中 CrowdHuman 数据集拥有规模庞大的行人数据, 是评价行人检测效果的代表性数据集之一; 在遥感船舰目标检测方面, HRSC2016^[18]、SSDD^[19]、HRSID^[20] 对多源船舰数据进行整理, 对复杂遥感船舰目标检测性能的提升发挥了至关重要的作用。

本文首先围绕样本标签分配方式来展开无锚框目标检测算法的分类; 其次关注近期无锚框目标检测算法的发展, 探索无锚框算法和有锚框算法的本质区别, 在主流数据集上对比典型算法的性能表现; 然后讨论和研究基于无锚框的目标检测方法在目标重叠排列、目标尺寸小和目标角度多等情形下的算法设计与应用; 最后对无锚框目标检测方法进行总

结并展望未来的研究方向。

1 基于无锚框的目标检测方法

无锚框是指无需对每个像素点预设宽高不同的锚框, 而直接对像素点进行类别判断或匹配多个同类像素点的操作, 避免了锚框相关超参数的设计步骤, 提高了模型对多类目标的普适性并且降低了模型复杂度。根据样本分配方式, 无锚框的目标检测方法分为基于关键点组合、中心点回归、Transformer、锚框和无锚框融合 4 大类, 各类方法的对比情况如表 2 所示。本节分别阐述各类无锚框目标检测方法的主要思想、网络架构和检测效果, 分析归纳无锚框目标检测方法的创新点及优缺点。

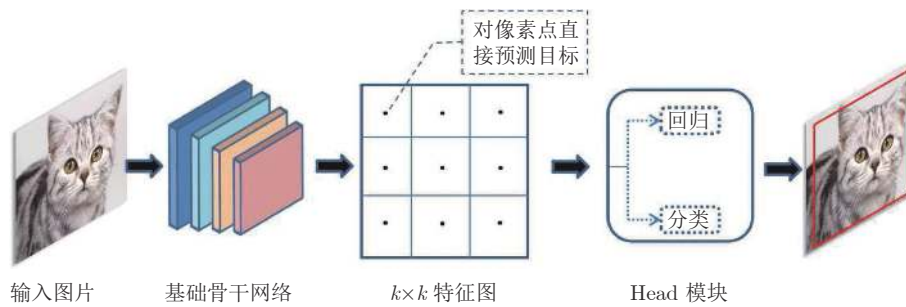


图 2 基于无锚框的目标检测方法整体框架

Fig. 2 The overall framework of anchor-free object detection method

表 1 目标检测公共数据集对比

Table 1 Comparison of public datasets for object detection

数据集	类别数	图片数量	实例数量	图片尺寸 (像素)	标注方式	使用场景	发表年份
Pascal VOC ^[10]	20	~23 k	~55 k	800 × 800	水平框	综合	2010
COCO ^[11]	80	~123 k	~896 k	—	水平框	综合	2014
DOTA ^[12]	15	~2.8 k	~188 k	800 ~ 4000	水平框/旋转框	综合	2018
UCAS-AOD ^[13]	2	~1 k	~6 k	1280 × 1280	旋转框	汽车、飞机	2015
ICDAR2015 ^[14]	1	1.5 k	—	720 × 1280	旋转框	文本	2015
CUHK-SYSU ^[15]	1	~18 k	~96 k	50 ~ 4000	水平框	行人	2017
PRW ^[16]	1	~12 k	~43 k	—	水平框	行人	2017
CrowdHuman ^[17]	1	~24 k	~470 k	608 × 608	水平框	行人	2018
HRSC2016 ^[18]	1	~1.1 k	~3 k	~1000 × 1000	旋转框	船舰	2017
SSDD ^[19]	1	1.16 k	~2.5 k	500 × 500	水平框	船舰	2017
HRSID ^[20]	1	~5.6 k	~17 k	800 × 800	水平框	船舰	2020

表 2 基于无锚框的目标检测方法对比

Table 2 Comparison of anchor-free object detection method

方法类型	基于关键点组合	基于中心点回归	基于 Transformer	基于锚框和无锚框融合
方法动机		无需设计锚框, 减少锚框带来的超参数, 简化模型		
方法思想	组合关键点并检测	中心点回归预测框位置	Transformer 的编码和解码直接预测	优化样本标签分配策略
方法优点	充分利用边界和内部信息	减少回归超参数数量	实现端到端, 简化流程	缓解正负样本不平衡
方法难点	不同类别关键点的误配对	中心点重叠目标的漏检	小目标检测性能较差	自适应标签分配不连续
计算速度	检测速度相对较慢	检测速度相对较快	收敛速度相对较慢	检测速度相对较慢

1.1 基于关键点组合的无锚框目标检测方法

关键点是指预测框中有代表性目标特征的像素点. 基于关键点组合的目标检测方法首先利用特征提取网络获得信息丰富的特征图, 然后使用多个关键点预测模块定位特征提取后的预测框关键点, 通过组合关键点来回归目标预测框, 即计算两条边界框长度(宽、高)来生成目标的检测框, 实现目标检测任务. 该类方法通过关键点匹配操作, 寻找图像中特征关键点的对应关系, 从而确定预测框的位置和类别, 用于解决不同类别目标间的像素点误匹配和“背景-目标”像素点误匹配的问题.

Wang 等^[21]首次提出一种基于点连接网络(Point linking network, PLN)的目标检测方法, 将检测问题视作点检测和点连接问题, 充分利用局部特征以提升重叠目标及极端宽高比目标的检测精度, 但由于感受野的范围限制, 导致角点难以定位目标位置. 针对该问题, Law 等^[22]提出 CornerNet 算法, 整体网络结构如图 3 所示. 该网络使用角点来作为关键点, 并利用修正后的残差模块、角池化模块和卷积模块来分别预测出左上角点和右下角点(图 4(a))的热力图(Heatmaps)、嵌入向量(Embeddings)和偏移量(Offsets), 充分利用了目标的边界信息. 此外, CornerNet 通过角点对来调整预测框的位置, 以完成目标检测任务. CornerNet 首次使用角池化以获取预测框角点, 通过提高 x, y 方向像素点向量的正交程度来加强边缘信息的获取, 达到修正角点位置的效果. 虽然 CornerNet 不使用预训练模型也能获得较理想的检测效果, 但该算法因角点匹配方法的类别误判, 导致不同类别角点误匹配和误检等问题. 针对 CornerNet 中存在的角点误匹配问题, Law 等^[23]提出 CornerNet-Lite 算法, 其中包括 CornerNet-Saccade 和 CornerNet-Squeeze, 该类算法通过减少像素点的数量来改进角点匹配, 在不牺牲精度的情况下同时提高检测效率和准确性, 且能

应用于离线检测和实时检测任务.

由于 CornerNet 及 CornerNet-Lite 算法过度关注目标的边缘信息, 而忽略目标内部的特征信息. 因此, Zhou 等^[24]在沿用 CornerNet 整体框架和损失函数的基础上, 提出一种自下而上的无锚框目标检测算法 ExtremeNet, 将关键点分组方式从卷积映射转换成几何结构分组, 该方法通过判断目标的 4 个极值点和 1 个中心点(如图 4(b)), 增加关键点的数量来充分学习目标的表征信息. ExtremeNet 不足之处是使用极值点聚类而产生过多重叠预测框, 且在关键点匹配时容易产生假阳性样本, 影响整体检测效果. Duan 等^[25]在 CornerNet 的基础上, 提出一种旨在解决角点误匹配问题的 CenterNet-Triplets 算法, 其关键思想在于将左上角点、右下角点和中心点替代两个角点, 使用级联角池化来获取丰富的角点信息, 并且利用中心池化来获取目标的内部信息、判断检测框的筛选、矫正角点的预测位置, 有效减少误检并提高召回率, 不足之处在于中心点漏检会导致预测框位置偏移量过大. 为了进一步增强目标内部特征的表达能力, 并且优化角点匹配导致的误检, Dong 等^[26]提出一种向心偏移量来匹配同类目标角点对的检测算法 CentripetalNet, 该算法使用向心偏移模块预测角点的向心偏移量(如图 4(c)), 利用位置解码操作对齐角点的向心偏移量, 使用十字星可变形卷积来丰富角点的表观特征, 并且添加一个实例掩码模块, 将向心位移预测区域作为候选区域, 此外利用轻量卷积网络对分割掩码进行目标检测, 大幅提升了模型对不同类别目标的适应能力. 在此基础上, 为了进一步加强局部和全局特征信息的表达能力, SaccadeNet^[27]用 5 个关键点(4 个角点和 1 个中心点)来获取目标信息. 首先使用中心注意力模块预测目标中心的类别和位置, 其次使用注意力传递模块预测相应边界框角点的粗略位置, 然后使用角点注意力模块来提取目标信息的角点特征, 从而使回归边界框更加准确, 最

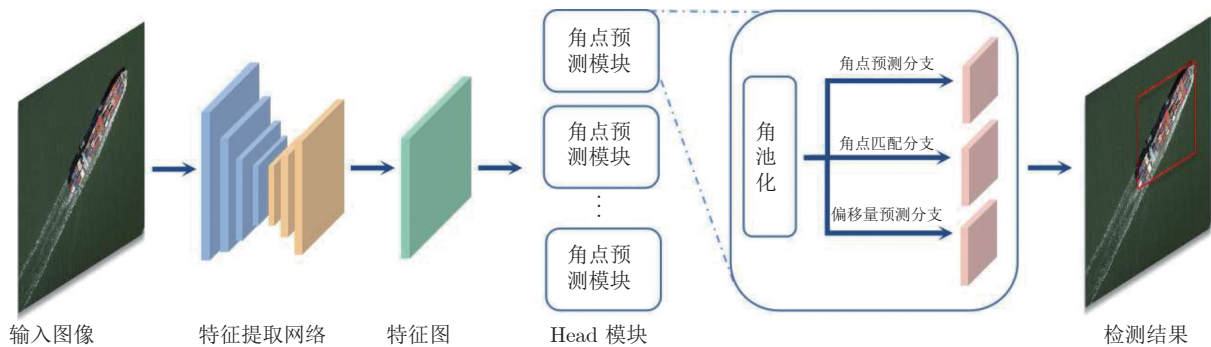


图 3 基于角点组合的 CornerNet 目标检测方法

Fig.3 CornerNet framework of object detection method based on corner points combination

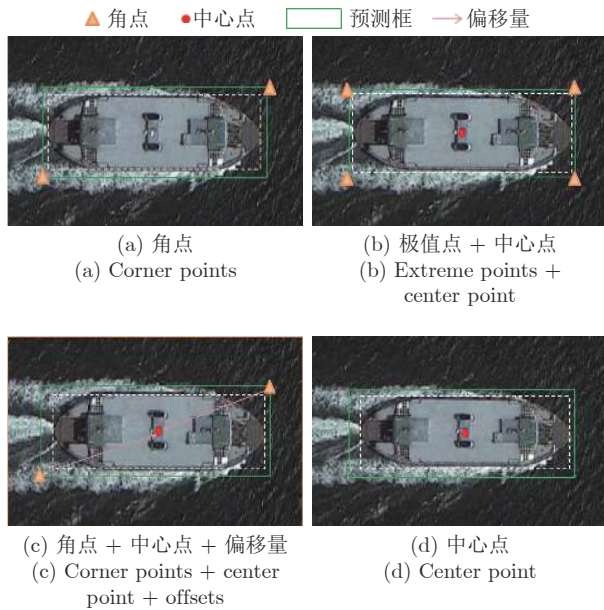


图 4 预测框采样方法

Fig. 4 The sampling methods of prediction box

后使用聚合注意力模块来融合中心和角点信息, 不断调整预测目标的精确位置. 该算法在提取目标多尺度特征的基础上, 同步提升检测速度和检测精度.

在充分利用特征表达能力的基础上, 为了减少不同几何形状目标的误检, Duan 等^[28] 提出分类和回归相独立的角点提议算法 CPNDet (Corner proposal network for object detection), 并建立角点和目标内部像素点间的联系来细调预测框的位置和类

别. 该算法第 1 阶段利用 CornerNet 的角点检测方法来遍历角点, 并且组合成角点对构成候选预测框; 第 2 阶段利用两类分类器来滤除大部分误检的预测框, 然后使用更精细的多类别分类器来确定预测框的类别以及相关置信度. 该方法能根据对应的区域特征来区分目标和错误匹配预测框之间的关键点, 在不降低检测速度的情况下, 整体网络更灵活地定位任意几何形状的目标. 上述算法均使用多角点和中心点来充分获取目标的边缘和内部信息, 尽可能地利用丰富的目标信息, 并且通过角点和中心点预测模块不断细化预测框的关键点调整过程, 降低预测框位置的偏移量, 实现更精准的目标定位.

综上所述, 基于关键点组合的目标检测方法不再列举大量可能存在目标的锚框, 而是从边缘关键点能够获取目标的代表性边界信息这一思想出发, 不断改进边缘关键点的选取方式、分组方式, 并且逐渐聚焦中心点对检测效果的影响, 通过中心点来同步加强内部信息和全局信息的利用, 从而利用中心点联合边缘关键点来设计偏移量预测模块, 通过修正关键点的位置来生成高质量预测框, 进一步减少误检和漏检的问题^[29]. 表 3 详细分析了各算法的性能和优缺点, 其中包括每秒传输帧数 (Frames per second, FPS) 和平均检测精度 (Mean average precision, mAP). 从表 3 可知, 大部分基于关键点组合的算法都在 CornerNet 的基础上进行改进, 如 CornerNet-Saccade 设计轻量级的网络用于实时检测, ExtremeNet 等通过结合中心点来减少预测框

表 3 基于关键点组合的无锚框目标检测算法在 COCO 数据集上的性能及优缺点对比

Table 3 Comparison of the keypoints combination based anchor-free object detection methods on the COCO dataset

算法	特征提取网络	输入尺寸 (像素)	处理器配置及检测 速度(帧/s)	mAP (%)	优点	缺点	收录来源	发表年份
PLN ^[21]	Inception-V2	512 × 512	GTX 1080 —	28.9	重叠及特殊形状目标的检测效果好	感受野范围较小	arXiv	2017
CornerNet ^[22]	Hourglass-104	511 × 511	TitanX × 10 4.1	42.1	使用角池化来精确定位目标	同类别的角点匹配易出错	ECCV	2018
CornerNet-Saccade ^[23]	Hourglass-54	255 × 255	GTX 1080Ti × 4 5.2	42.6	无需对每个像素点进行类别检测	小目标的误检率较高	BMVC	2020
CornerNet-Squeeze ^[23]	Hourglass-54	255 × 255	GTX 1080Ti × 4 33	34.4	大幅提升检测速度	角点类别的判断较易出错	BMVC	2020
ExtremeNet ^[24]	Hourglass-104	511 × 511	TitanX × 10 3.1	43.7	极值点和中心点充分获取目标信息	容易产生假阳性样本	CVPR	2019
CenterNet-Triplets ^[25]	Hourglass-104	511 × 511	Tesla V100 × 8 2.94	47.0	用角点和中心点获取充分目标信息	中心点遗漏时位置偏移量大	ICCV	2019
CentripetalNet ^[26]	Hourglass-104	511 × 511	Tesla V100 × 16 —	48.0	改进 CornerNet 的角点误匹配问题	中心区域的缩放依赖超参数	CVPR	2020
SaccadeNet ^[27]	DLA-34-DCN	512 × 512	RTX 2080Ti 28	40.4	获取局部和整体特征, 提高特征利用率	需要平衡检测精度与速度	CVPR	2020
CPNDet ^[28]	Hourglass-104	511 × 511	Tesla V100 × 8 —	49.2	多种分类器提升角点类别判断准确率	检测头计算效率较低	ECCV	2020

误检问题, CentripetalNet 等改进关键点选取和组合方式以提升检测精度, 但大多数关键点组合的方法仍有较多后处理操作, 因此检测速度有待进一步提升.

1.2 基于中心点回归的无锚框目标检测方法

中心点能反映目标的内部特征信息. 基于中心点回归的无锚框目标检测方法首先利用特征提取网络获取目标的特征图; 其次用检测头模块的热力图峰值表示目标中心点的位置, 宽高预测分支用于回归预测框的尺度信息, 中心点预测分支用于修正中心点偏移量; 最后利用后处理操作筛选出最优预测框, 整体框架如图 5 所示. 基于中心点回归的无锚框目标检测方法相比于基于关键点组合的无锚框目标检测方法, 减少了关键点匹配操作, 缓解了因关键点类别误判导致的误检问题, 提高了模型的整体性能.

在无锚框目标检测算法研究初期, UnitBox 检测网络^[30]首次使用预测框和真实框 4 个顶点的 IoU 损失 L_{IoU} (如式 (1) 所示) 来替代 DenseBox^[9] 的 L_2 损失 L_{L_2} (如式 (2) 所示), 以考虑预测框坐标间的相关性, 实现更精准的预测位置修正, 更好地适应目标的尺寸和外观变化, 在人脸检测上获得良好的检测效果, 为通用无锚框检测算法的精准定位奠定了坚实的基础. Redmon 等^[31]受此启发, 利用回归的思想提出 YOLO v1, 将输入图像划分为子网格单元并对每个网格内进行目标中心点的检测, 即获取包含目标中心点的网格并预测网格内相应目标的边界框以及类别概率向量. 该检测网络能够通过划分输入图像区域来实现快速目标检测, 但由于每个网格只能检测一个目标中心点, 因此当多个目标中心落入同一网格内, 易出现漏检问题. 为了分别提升 Head 模块中的分类、回归分支的检测性能并且解决同网格多目标的漏检问题, Ge 等^[32]在 YOLO v3^[7]基础上提出 YOLOX. 该算法把正样本限定在目标中心的一定网格区域内, 通过增大正样

本数量来减少重叠漏检问题, 并且通过解耦检测头来拆分回归和分类分支结构, 以此来加快模型收敛速度、提高检测精度. 此外, YOLOX 使用 SimOTA 样本匹配方法, 缓解极端宽高比的物体的检测效果差、极端大小目标正负样本不均衡的问题, 整体 YOLOX 检测模型还提供多平台的资源部署版本, 扩大使用平台和范围.

IoU 损失计算如下:

$$L_{IoU} = 1 - IoU = 1 - \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (1)$$

其中, B 为预测框, B^{gt} 为真实框.

L_2 损失计算如下:

$$L_{L_2} = \frac{1}{n} \sum_{i=1}^n (y_i^p - y_i)^2 \quad (2)$$

其中, y_i^p 为预测值, y_i 为真实值.

除了可以划分输入图像来筛选目标中心点位置的范围外, Tian 等^[33]使用全卷积网络逐像素预测的全卷积单阶段目标检测算法 (Fully convolutional one-stage object detector, FCOS), 该算法将落入真实框边界范围内的中心点视为正样本, 相比有锚框检测方法增大了正样本的数量, 缓解正负样本不均衡的问题, 并且使用多尺度特征预测方法来预测不同尺度的目标边界框, 解决 YOLO v1 中的中心点重叠的漏检问题, 加入中心置信度 (Centerness) 子分支来筛选远离目标中心点的低置信度预测框, 并采用 NMS 后处理筛选冗余框, 得到最终与目标适配度高的检测结果. 其中, 中心置信度计算如下:

$$centerness = \sqrt{\frac{\min(l, r)}{\max(l, r)} \times \frac{\min(t, b)}{\max(t, b)}} \quad (3)$$

其中, $l = x - x_0^{(i)}$, $t = y - y_0^{(i)}$, $r = x_1^{(i)} - x$, $b = y_1^{(i)} - y$, l, t, r, b 分别表示该中心点到预测框四边的距离, $(x_0^{(i)}, y_0^{(i)})$, $(x_1^{(i)}, y_1^{(i)})$ 为第 i 个预测框的左上和右下顶点坐标, (x, y) 为中心点坐标.

FCOS 的中心置信度分支能降低远离目标中心

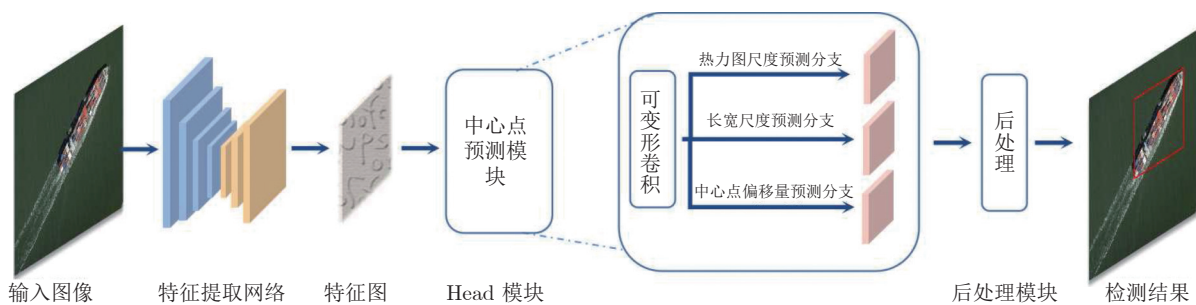


图 5 基于中心点回归的无锚框目标检测方法整体框架

Fig. 5 The overall framework of anchor-free object detection method based on center point regression

的预测框得分, 此外针对同一像素点同时包含两个目标的情况, FCOS 直接回归出当前像素点所指向目标的类别和最小预测框, 极大缓解了重叠目标的类别模糊问题, 相比有锚框检测算法减少了正样本数量和训练内存占用并提高检测精度, 然而仍会出现同尺度层的误检问题. 由于 FCOS 仍使用后处理操作来筛除多余的预测框, 增加了训练和预测的计算复杂度. 因此, Sun 等^[34] 设计一个端到端的目标检测算法 OneNet. 该方法同样使用 FCOS 的中心点来定位预测框, 额外提出最小化损失分配策略, 仅对每个目标分配一个预测框, 并将损失最小的预测框视为唯一正样本, 去除计算量大的后处理操作, 缓解正负样本不均衡的问题并提高检测速度, 但检测精度有待进一步提升.

Zhou 等^[35] 提出一种无需划分输入图像的目标检测算法 CenterNet, 该算法将多点检测转换成单中心点检测问题 (如图 4(d)), 利用中心点感知目标内部信息的思想, 解决了多点匹配导致的类别判断错误、定位不准确的问题. 此外, CenterNet 无需 NMS 后处理操作来去除冗余预测框, 减少超参数数量并降低模型复杂度, 实现检测速度和精度的平衡, 但缺点是仅利用中心点回归目标, 导致无法充分利用全局信息, 出现因中心点重叠所产生的漏检问题. 为了解决 CenterNet 的漏检问题, Zhou 等^[36] 进一步提出 CenterNet2 算法. 该算法分为两个阶段, 首先在第 1 阶段使用 CenterNet 提取区域级别的特征并对其分类, 判断校准后目标和背景的相似性; 然后在第 2 阶段使用级联分类器, 最大化真实目标边界框的对数似然, 并在预测时使用对数似然计算最终的检测分数. CenterNet2 与两阶段有锚框目标检测算法相比, 更加精准有效地选择感兴趣区域. 此外, CenterNet2 与 CenterNet 相比, 其 Head 模块更加精简, 减少类别预测数量并加速检测, 显著提升检测准确率.

由于 CenterNet 系列方法过于依赖中心点局部信息, 因此 HoughNet^[37] 受广义 Hough 变换的启发, 利用对数极坐标投票机制从近距离和远距离位置收集选票, 增加局部和全局信息都有效正样本的比例. 为了进一步体现模型的有效性, 提取 HoughNet 的投票模块来筛选正样本并嵌入到两个生成对抗网络 (Generative adversarial networks, GANs) 模型 (CycleGAN^[38] 和 Pix2Pix^[39]) 中, 提高生成对抗网络模型的样本生成效果. 除此之外, Gird R-CNN^[40] 在 Faster R-CNN^[5] 的基础上, 选择 4 个角点、预测框的边界中点和目标中心点作为网格点, 利用网格定位机制选择候选区域, 在回归分支利用空洞卷积和反卷积网络替代全连接层, 实现候选区

域的修正, 加入特征融合模块以聚合深层语义信息和浅层位置信息, 从而增强有效特征的表达能力, 然而特征点的采样区域范围较广, 加重了正负样本不均衡的问题. 在 Gird R-CNN 的基础上, Grid R-CNN Plus^[41] 利用阈值去除无关键点的候选区域, 减少特征相关的冗余计算, 缩小目标的检测范围、减少计算量的同时聚焦区域性特征, 提升整体检测速度.

综上所述, 表 4 总结了基于中心点回归的无锚框目标检测方法, 能够利用 IoU 来加强目标类别和位置的联系、使用子网格划分图像的方法减少中心点采样数量、使用热力图等方法获取目标的有效特征中心点. 在直接使用中心点来回归目标的基础上, 添加网格定位点提高定位精度、并利用中心置信度过滤低质量预测框等方法, 在无需进行计算复杂并且误差较高的角点匹配思想下, 进一步简化模型复杂度和提升检测性能, 提供了无锚框目标检测方法的新思路. 然而基于中心点回归的方法仍使用多尺度检测的方法, 计算效率有待提升, 因此针对特征金字塔的耗时问题需进一步优化.

1.3 基于 Transformer 的无锚框目标检测方法

近年来, Transformer 利用注意力机制有效捕捉目标的感受野, 广泛用于机器翻译、语音识别等结构化任务. 基于 Transformer 的目标检测方法主要由堆叠的编码器和解码器组成, 而编码器由交替的自注意力机制模块和多层前馈神经网络组成. 自注意力机制能为输入的目标向量学习得到一个权重, 寻找当前目标和预测目标之间的关系, 解码器则在编码器的基础上增加“编码-解码”注意力模块, 用于寻找编码前后特征向量间的关系. 整体网络无需手工设计参数, 无需由于后处理操作, 实现了端到端的目标检测结构, 简化了目标检测流程.

基于 Transformer 的端到端目标检测器 (End-to-end object detection with Transformers, DETR)^[42] 首次将 Transformer 成功地整合到目标检测中, 整个 DETR 架构如图 6 所示.

在图 6 中, DETR 首先使用骨干网络提取、合并目标特征集并添加位置编码, 然后输入到 Transformer 组成的“编码器-解码器”结构, 使用自注意力机制式地对序列中所有特征组合进行交互式建模, 最后将解码器的输出传递给固定数量的预测头来获取预测框的类别和位置信息. DETR 能够结合二分匹配损失和 Transformer, 同步完成并行解码, 预测目标和全局图像上下文之间的关系, 仅用单次传递就能预测、删除无目标锚框和完成非极大值抑制操作, 删除重复预测并简化检测流程, 建立了一个新的目标检测范式. 然而, 二分匹配损失在训练

初期无法稳定地匹配预测框和真实框, 导致小目标的检测性能较差且模型收敛速度较慢. 针对 Transformer 解码器交叉注意力机制的训练收敛速度极慢、二分匹配损失稳定性较差的问题, Sun 等^[43] 结合 FCOS 采样方式和 DETR 的编码器来构成新的检测网络 TSP-FCOS (Transformer-based set prediction of FCOS), 通过在骨干网络和编码器之间加入并行的辅助子网和分类子网模块, 并且在分类子网上引出一个感兴趣特征选择器来筛选多尺度特征. 此外, 还设计一个新的二分图匹配损失来精确定位目标, 在检测速度均衡的情况下有效提高 DETR 和 FCOS 的检测精度. 同时为了提高检测效率, Deformable DETR^[44] 在 DETR 的基础上, 使用多尺度可变形注意力机制替代 DETR 的注意力机制模块, 融合可变形卷积的稀疏空间采样与 Transformer 的特征关联性建模能力, 有效关注稀疏空间的目标位置, 且能在有效信息中区分各特征的重要

程度, 选择并生成目标特征信息丰富的特征图. 此外, 该算法优化高分辨率图像的特征图选择和样本选择, 通过改进标签分配策略来增大正样本比例, 解决 DETR 训练周期长、收敛速度慢的问题, 同时提升小目标的检测性能, 然而整体模型的计算量较大. 同时, Dai 等^[45] 提出基于动态注意力机制的算法 Dynamic DETR, 针对 DETR 编码器的自注意力模块计算复杂度较高的问题, 使用动态编码器以兼顾空间、尺度和代表性特征的动态调整, 并且将解码部分的交叉注意力机制替换为基于感兴趣区域的动态注意力机制, 使 Transformer 有效区分粗粒度和细粒度感兴趣区域的关注程度, 加速模型的收敛速度. 另一方面, DETR 的收敛速度慢的原因是目标查询与目标特征的匹配过程比较复杂, 因此, Zhang 等^[46] 提出一种语义对齐匹配的目标检测算法 (Semantic aligned matching DETR, SAM-DETR). 该网络利用对齐语义将目标查询投射到与编

表 4 基于中心点回归的无锚框目标检测算法在 COCO 数据集上的性能及优缺点对比

Table 4 Comparison of the center point regression based anchor-free object detection methods on the COCO dataset

算法	特征提取网络	输入尺寸 (像素)	处理器配置及检测速度 (帧/s)	mAP (%)	优点	缺点	收录来源	发表年份
YOLO v1 ^[31]	—	—	—	—	用网格划分法提高中心点搜寻效率	目标中心点在同网格内的漏检	CVPR	2016
FCOS ^[33]	ResNet-101	$800 \times \leq 1333$	— 9.3	41.5	用中心度降低远离中心点的预测框得分	同尺度特征层中出现目标误检	ICCV	2019
CenterNet ^[35]	Hourglass-104	511×511	Titan X 7.8	45.1	用中心点定位目标减少角点匹配操作	目标中心点重合, 产生漏检	arXiv	2019
Grid R-CNN ^[40]	ResNet-101	800×800	Titan Xp $\times 32$ 3.45	41.5	用网格定位机制精准定位边界框	特征采样区域范围过于广泛	CVPR	2019
Grid R-CNN Plus ^[41]	ResNet-101	800×800	Titan Xp $\times 32$ 7.69	42.0	缩小小特征表达区域尺寸, 减少计算量	非代表性特征区域存在遗漏	arXiv	2019
HoughNet ^[37]	Hourglass-104	512×512	Tesla V100 $\times 4$ —	46.4	用投票机制改进全局信息缺失的问题	投票机制使计算量增大	ECCV	2020
YOLOX ^[32]	Darknet53	640×640	Tesla V100 $\times 8$ 90.1	47.4	解耦分类和回归分支, 提升收敛速度	难分类样本的检测精度较低	arXiv	2021
OneNet ^[34]	ResNet-101	$512 \times \leq 853$	Tesla V100 $\times 8$ 50	37.7	用最小匹配损失提升预测框和标签的匹配	单像素点检测单目标, 产生漏检	ICML	2021
CenterNet2 ^[36]	Res2Net-101-DCN-BiFPN	1280×1280	Titan Xp —	56.4	清晰区分目标特征和背景区域的特征	分步分类、回归的效率较低	arXiv	2021

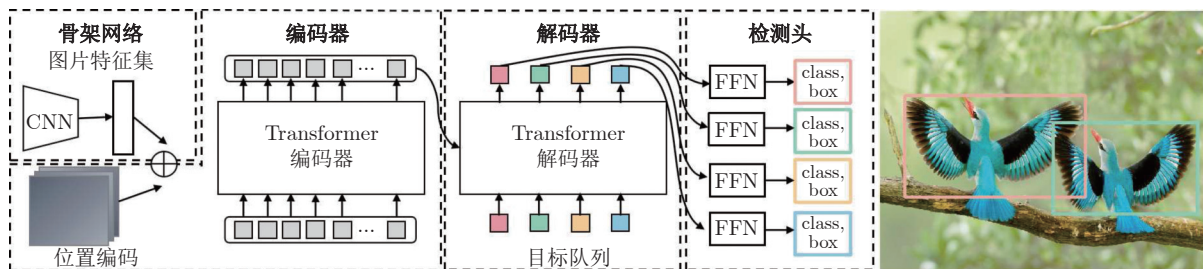


图 6 DETR 整体框架

Fig. 6 The overall architecture of DETR

码图像特征相同的嵌入空间里, 且使用具有代表性的目标特征来搜索目标关键点, 高效地实现语义对齐特征匹配, 较好地缓解现有模型收敛速度慢的问题。

上述方法在利用骨干网络提取特征后, 通过修正 Transformer 编码器、解码器的注意力机制来提高 DETR 的收敛速度, 提升整体检测精度。同时, 仍有其他改进策略来提高 DETR 的检测性能, 如 YOLOS^[47] 算法不使用卷积模块组成的骨干网络来提取特征, 转而从视觉 Transformer (Vision Transformer, ViT)^[48] 的角度出发, 参照 DETR 的“编码器-解码器”架构设计并进行适当调整。该网络丢弃 ViT 用于图像分类的分支并添加随机初始化的标签检测分支, 避免图像与标签先验信息的偏置计算, 而且使用 DETR 的二分匹配损失替换分类损失, 不断调整整体网络权重。此外, YOLOS 仅需对 ViT 结构进行微调, 即可成功迁移到极具挑战性的 COCO 目标检测任务中, 进一步揭示 Transformer 在目标检测的有效性, 然而检测性能有待进一步提升。在 YOLOS 的基础上, ViDT (Vision and detection Transformer)^[49] 引入了一个重配置的注意力模块 RAM, 将基于层级和滑窗的 Transformer 骨干网络扩展为一个独立的目标检测器, 其次采用轻量级的无编码器检测颈 (Neck) 架构来减少计算开销, 在不增加计算负载的情况下轻量化模型并提高检测性能。除上述方法外, Li 等^[50] 首次提出全新的去噪训练方法 (Denoising DETR, DN-DETR) 来解决 DETR 二分图匹配不稳定的问题。该方法给解码器的嵌入向量加入噪声并作为标签, 并在训练时通过去噪模块来训练含噪声的预测数据, 在预测

阶段丢弃去噪模块, 在不改变模型结构的前提下, 模型收敛速度翻倍并显著提升检测精度。

从表 5 中可以看出, 模型浮点计算量 (Floating point operations, FLOPs) 等参数能反映模型整体复杂度, 基于 Transformer 的目标检测网络能够将“编码器-解码器”结构作为部分网络结构或整体检测网络来实现目标检测, 并且能够通过注意力机制模块来降低编码器和解码器结构复杂度, 通过修改二分匹配损失的结构来解决损失稳定性较差的问题, 此外还能够提高特征匹配程度来解决模型收敛速度慢的问题。然而现在常用多尺度特征、高分辨率特征图来提高小目标检测精度, 而基于 Transformer 的检测方法大幅增加计算量和空间复杂度, 难以处理高分辨率图像的特征, 在小目标上的检测精度需进一步提高。

1.4 基于锚框和无锚框融合的目标检测方法

基于锚框的目标检测方法能够通过锚框枚举得到定位较精准的预测框, 而基于无锚框的方法能利用目标特征点来直接对目标位置进行预测, 减少相关超参数调节过程, 降低模型参数量和计算量。因此合理融合有锚框和无锚框的目标检测方法有助于提高检测精度和检测效率, 有效提高目标检测的性能。

针对有锚框方法因预设锚框宽高比、尺寸无法精准匹配宽高比较大、尺度变化较大的目标的问题, Wang 等^[51] 受启发于 UnitBox^[30] 检测大尺度范围目标的思想, 集成有锚框和无锚框检测方法并构成检测各类人脸的 SFace 网络。该网络使用有锚框预测

表 5 基于 Transformer 的无锚框目标检测算法在 COCO 数据集上的性能及优缺点对比
Table 5 Comparison of the Transformer based anchor-free object detection methods on the COCO dataset

算法	特征提取网络	输入尺寸 (像素)	处理器配置及检测速度 (帧/s)	mAP (%)	浮点计算量 (FLOPs/G)	优点	缺点	收录来源	发表年份
DETR ^[42]	ResNet-50	(480, 800) × (800, 1333)	Tesla V100 × 16 28	42.0	86	用 Transformer 减少手工设计参数数量	收敛速度慢, 小目标检测性能较差	ECCV	2020
TSP-FCOS ^[43]	ResNet-50	(640, 800) × (800, 1333)	Tesla V100 × 8 15	43.1	189	添加辅助子网来提高多尺度特征的提取	模型计算量、复杂度较高	ICCV	2021
Deformable DETR ^[44]	ResNet-50	(480, 800) × (800, 1333)	Tesla V100 19	43.8	173	有效关注稀疏空间的目标位置	模型计算量、复杂度较高	ICLR	2021
Dynamic DETR ^[45]	ResNet-50	—	Tesla V100 × 8 —	47.2	—	用动态注意力机制加速收敛	未说明模型的计算量、复杂度	ICCV	2021
YOLOS ^[47]	DeiT-base	(480, 800) × (800, 1333)	— 2.7	42.0	538	不依赖卷积骨干网络, 性能良好	检测速度较低, 计算量较高	NeurIPS	2021
SAM-DETR ^[46]	ResNet-50	(480, 800) × (800, 1333)	Tesla V100 × 8 —	41.8	100	利用语义对齐加速模型收敛速度	检测精度有待进一步提升	CVPR	2022
ViDT ^[49]	Swin-base	(480, 800) × (800, 1333)	Tesla V100 × 8 11.6	49.2	—	用新的骨干网络和检测颈减少计算开销	浅层难以直接获取目标的有用信息	ICLR	2022
DN-DETR ^[50]	ResNet-50	—	Tesla A100 × 8 —	44.1	94	利用去噪训练法大幅提升检测性能	仅使用均匀分布的噪声	CVPR	2022

分支来充分利用目标的预定义信息,精准定位常规尺度目标,并在金字塔底端引出一个无锚框检测分支,用于处理大尺度变化的人脸.此外提出基于IoU回归的重打分机制来标识分类结果,以更好地联合两个分支的分类和回归策略,实现两个分支的有效互补.然而SFace方法仅用单特征层来检测多尺度目标,无法针对目标尺寸选择最合适的特征层.因此Zhu等^[52]在RetinaNet^[53]网络的基础上加入无锚框特征选择模块(Feature selective anchor-free, FSAF),该模块在每一层特征金字塔上引出一个无锚框分支对实例进行编码,以获取用于分类和回归的参数,在训练过程中动态调整金字塔网络的不同感受野和目标空间信息,并为目标选择合适的特征层尺寸来提取特征,改善特征选择不准确、像素点采样重叠的问题.为了进一步优化特征选择,Zhu等^[54]又提出加权锚点目标检测算法(Soft anchor-point object detection, SAPD),在FSAF基础上加入与FCOS的中心置信度相似的锚点权重注意力机制,对每个锚点赋予不同的损失权重因子,降低目标边缘点的权重值,使网络更加关注代表性强的目标中心区域.此外,SAPD利用特征金字塔选择模块和联合训练策略,对不同级别特征层的每个特征进行预测并分配不同权重,寻找目标的最优特征层并选择出最具代表性的目标特征,在不影响预测速度的情况下提高检测精度.然而SAPD的有锚框和无锚框分支仍相互独立,未能实现真正融合.

由于有锚框的目标检测方法需预设置锚框、手动调整样本超参数等策略来进行标签分配,为实现标签自动化分配、探索标签分配问题对检测网络的影响,涌现了大量相关新方法,图7表示基于优化标签分配各算法的关系.

FreeAnchor^[55]在RetinaNet的基础上,将检测器的训练过程定义成一个极大似然估计问题,通过

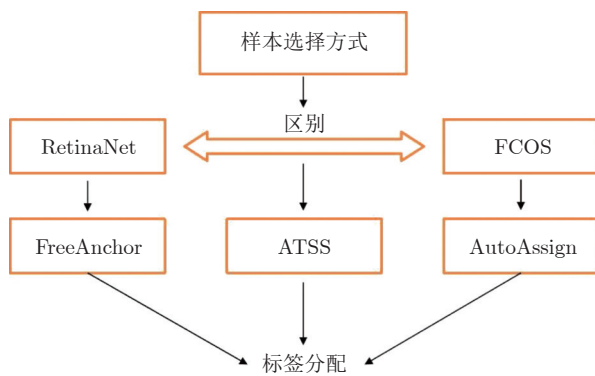


图7 基于优化标签分配算法的关系

Fig.7 The relationship between label assignment optimization algorithms

构建损失函数以实现自适应锚框设置过程,使网络能够自主学习与目标相匹配的锚框.Zhang等^[56]通过对比有锚框检测算法RetinaNet^[54]和无锚框检测算法FCOS^[57]发现,RetinaNet借助IoU阈值来划分正负样本,FCOS则利用目标中心区域和目标尺度选择正负样本,它们的本质区别在于正负训练样本的选择方式不同.因此Zhang等^[56]提出一种自适应的训练样本选择方法(Adaptive training sample selection, ATSS),根据目标的统计特性自动选择正负训练样本,缩小两类检测算法之间的性能差异,在不增加额外开销的情况下,大幅提高检测性能.但是FreeAnchor和ATSS本质上仍是通过中心先验、IoU、空间和尺度约束来进行标签分配,无法避免大量超参数的手动调整,未完全实现样本的自适应选择.为了实现无需参数调节的自适应标签分配,Zhu等^[57]在FCOS的基础上首次提出了一种基于可微样本分配策略的目标检测算法AutoAssign.该网络使用一个中心加权模块来调整特定类别的先验分布,适应不同类别目标的外观和尺度分布,利用一个置信度加权模块,动态地对目标在空间和尺度上的位置进行权重映射,改进样本分配来优化重叠目标共用边界类别模糊的情况,完成端到端的动态样本分配,能够在无任何参数修改的情况下扩展到其他数据集和检测任务上,但样本分配策略的权重机制较复杂.

除上述方法外,Duan等^[58]又发现目标检测、实例分割和姿态估计都能视为位置敏感的计算机视觉任务,并研究出统一的位置敏感网络(Location sensitive network, LSNet)来实现3类视觉识别任务.该网络使用ATSS的正负样本采样策略,额外将可变形卷积的偏移量等比例映射至相邻的特征金字塔网络(Feature pyramid network, FPN),利用多层金字塔特征融合方法充分提取预测点的标志性特征信息,并且在CenterNet的中心点样本选择策略基础上,额外添加指向极值点的4个向量来确定预测框的位置,有效获取极值点包含的目标语义信息并提升获取显著性特征的能力.此外,LSNet使用新的Cross-IoU损失 L_{CIoU} (如式(4)所示)来替换CenterNet的损失函数,将二维IoU损失压缩到一维,从垂直和水平方向分解预测框与真实框计算向量的重合程度,细化预测框定位步骤,整体网络实现了精度和速度的同步提升.Li等^[59]针对上述FCOS、AutoAssign等标签分配方法的正负样本权重相关性高、无法有效赋予样本权重等问题,提出基于双权重加权的标签分配方法(Dual weighting, DW).该网络使用分类和回归置信度的一致性来动态赋予正样本权重,使用负样本的概率和重要程度来调整

负样本权重, 有效过滤位置偏移量大的预测框, 筛选出分类置信度高和位置精准的预测框, 此外还提出预测框细化模块, 通过聚合预测框中心以及 4 个边界点来精确预测框位置, 整体性能得到大幅提升。

Cross-IoU 损失计算如下:

$$L_{\text{CIoU}} = 1 - \frac{\min(vp_x, vg_x) + \min(vp_y, vg_y)}{\max(vp_x, vg_x) + \max(vp_y, vg_y)} \quad (4)$$

其中, vp_x, vp_y 为预测框向量的 x 分向量和 y 分向量, vg_x, vg_y 为真实框向量的 x 分向量和 y 分向量。

表 6 根据分析对比有锚框和无锚框相结合的各种检测方法, 说明现有的标签分配策略使用锚框或者中心点预测等方式, 为了让检测方法对多类型数据分布的适配性更好, 标签分配策略使用多维度标签分配等策略, 缓解正负样本不均衡的问题。ATSS 也说明无锚框和有锚框方法的不同之处在于正负样本的选择方式不同, 有锚框分支能够在无锚框检测方法中提高预测框定位精度, 无锚框方法能够用中心点替代锚框预设过程, 增加正样本所占比例, 两类方法能够相互促进, 共同实现检测精度和速度的同步提升。然而由于理想的标签分配策略需为连续的计算过程, 上述大部分模型使用离散化的 IoU 阈值来划分正负样本, 因此如何设计连续的自适应标签分配方案是待解决的研究问题。

1.5 分析与总结

根据上述对比和分析, 基于关键点组合的方法利用角点、极值点和中心点等来作为关键点, 以获得充足的目标内部信息和边缘信息, 不断修正目标定位机制以校准最终预测框, 提高整体检测精度和检测效率。基于中心点回归的方法不使用多关键点组合, 直接获取目标的特征中心点并以此判断中心点类别, 减少了因类别误判产生的预测框数量和多

关键点匹配产生的计算复杂度。此外近期发展迅速的基于 Transformer 目标检测方法, 将文本检测的结构扩展到目标检测领域, 通过优化样本标签方法, 修正 Transformer 的“编码器-解码器”结构, 聚焦于目标的局部信息, 极大程度上简化模型设计复杂度, 且有望在未来进一步提出目标检测新范式。除了上述方法外, 融合无锚框和有锚框的目标检测方法充分利用有锚框的定位分支, 并引申出一系列改进标签分配的方法, 不仅缓解正负样本不均衡的问题, 而且有效提高目标检测的性能。

总而言之, 基于无锚框的目标检测方法在不断发展过程中, 去除锚框的预设以及相关计算过程, 降低了模型复杂度, 逐步优化了模型误检、漏检等问题, 使得目标检测更加简单、快速。此外, 基于无锚框的目标检测方法在目标密集、目标尺寸小和目标角度多变等复杂场景下也得到了广泛应用, 并逐渐在自动驾驶、工业自动化和卫星遥感图像目标检测等领域占据重要地位^[60]。

2 无锚框目标检测方法在复杂场景下的应用

目前基于深度学习的目标检测算法在普通场景下已经达到较为理想的检测效果, 但在目标重叠、目标过小、目标旋转等复杂场景下的检测任务仍有较大的挑战性^[61-63]。本节针对复杂场景下存在的各种问题, 从重叠目标、小目标、旋转目标等 3 个方面介绍无锚框目标检测方法在不同问题上的模型设计, 并分析各方法之间的优势与不足。

2.1 基于无锚框的重叠目标检测方法

重叠目标是指包含共用的像素点表示目标信息的相邻目标, 重叠目标检测在复杂场景下的应用十

表 6 基于锚框和无锚框融合的目标检测算法在 COCO 数据集上的性能及优缺点对比

Table 6 Comparison of the anchor-based and anchor-free fusion object detection methods on the COCO dataset

算法	特征提取网络	输入尺寸 (像素)	处理器配置及检测 速度 (帧/s)	mAP (%)	优点	缺点	收录来源	发表年份
FSAF ^[52]	ResNeXt-101	800 × 800	Tesla V100 × 8 2.76	44.6	动态选择最适合目标的特征层	未区分不同特征的关注程度	CVPR	2019
SAPD ^[54]	ResNeXt-101	800 × 800	GTX 1080Ti 4.5	47.4	能筛选出有代表性的目标特征	未能真正将有锚框和无锚框分支融合	ECCV	2020
ATSS ^[56]	ResNeXt-101	800 × (800, 1333)	Tesla V100 —	50.7	能根据统计特性自动训练样本	未完全实现无需参数调节的样本分配	CVPR	2020
AutoAssign ^[57]	ResNeXt-101	800 × 800	—	52.1	无需手动调节的动态样本分配	样本的的权重分配机制相对复杂	arXiv	2020
LSNet ^[58]	ResNeXt-101	800 × (800, 1333)	Tesla V100 × 8 5.1	50.4	用位置敏感网络大幅提高定位精度	小目标的定位和分类精度较低	arXiv	2021
DW ^[59]	ResNeXt-101	800 × 800	GPU × 8 —	49.8	有效获取分类和回归置信度高的框	小目标的检测性能仍需进一步提升	CVPR	2022

分广泛,然而因目标间的相互遮挡,导致目标部分信息的缺失,因此出现目标重复检测、漏检等问题(如图8(a)和图8(b)所示).

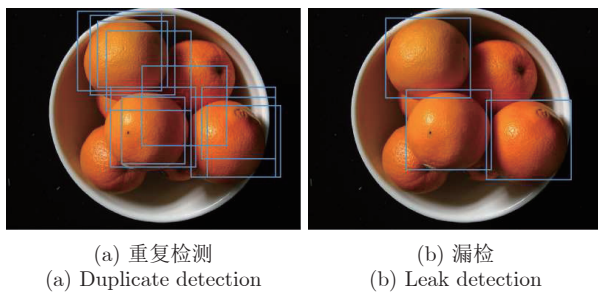


图8 重叠目标检测问题

Fig.8 The detection problems of overlapping object

现有的目标检测算法为了更有效地筛选冗余预测框,常用 Soft-NMS^[64]、Adaptive-NMS^[65]和 R²NMS^[66]的柔和筛选与密集程度判断等方法来筛选密集排列预测框,然而计算成本有所增加.此外,Aggregation Loss^[67]、Rep-GIoU Loss^[68]和 NMS Loss^[69]等算法通过修正损失,拟合密集排列场景的目标形状和位置.但此类方法对高置信度预测框的作用不明显,无法充分利用低质量预测框的特征信息.本节针对重叠目标检测的问题,阐述各算法在不同场景下目标检测方法的模型架构、检测性能和优缺点,最后对算法进行纵向对比分析与总结.

密集行人是重叠目标检测场景的典型应用,中心和尺度预测算法(Center and scale prediction, CSP)^[70]作为重叠目标检测的先驱,对行人的位置和尺寸进行分解,将目标检测简化为预测目标的中心点和尺度等高级语义特征信息的任务.通过堆叠的卷积模块预测两个映射图,其中一个映射图利用热力图获得预测框的中心位置和对应预测框的置信度,另一个映射图利用尺度变化预测对应预测框的尺寸,两个映射图共同作用得到最终预测框的位置及类别,在密集人群检测数据集上获得良好的检测效果.缺点是中心点无法充分利用特征与预测框之间的关联性,从而导致多个重叠目标的漏检问题.针对行人检测任务中的重叠目标漏检问题,特征对齐行人搜索网络(Feature aligned person search network, AlignPS)^[71]考虑行人重识别特征优先的原则,利用可变形卷积将采样点聚焦于行人内部,重点关注 ReID 特征的学习,使用一个特征对齐和融合模块,解决特征学习中区域和尺度不对齐的问题,在重叠排放目标中获得良好的检测效果.除此之外, Cao 等^[72]提出一个新型 Transformer 的行人搜索网络(Transformer-based person search,

PSTR),通过引入特定行人搜索模块(Person search-specialized, PSS),并使用具有多级监督方案的结构共享解码器,提高重识别特征的可判别性,并额外引入包含显著性特征重识别解码器的注意力模块,寻找各特征间的关联性.综上,PSTR 作为首个端到端的单步 Transformer 行人搜索网络,使用多尺度解码器来精准实现行人多尺度特征匹配,大幅提高行人检测精度.

除了密集行人的检测方法得到不断发展之外,解决目标重叠排列的通用方法也得到飞速发展.如第 1.2 节中所提及的 FCOS^[33]能直接回归出当前像素点所指向目标的类别和最小预测框,优化目标中心区域和目标样本的标签分配方法,极大缓解了重叠目标的类别模糊问题.在此基础上,一系列优化特征提取和表达能力的方法获得启发.其中,FCOS v2^[73]替换 FCOS 的部分网络结构.该网络将正样本点的选择范围从目标内部缩小成位于目标的中心区域,使用加权双向特征金字塔网络(Bidirectional feature pyramid network, BiFPN)替代特征金字塔网络,获得信息更丰富的多尺度特征,加入可变形卷积来改变感受野范围,增强目标特征的获取能力,在分类分支修改中心度的计算方式,更精确回归出预测框的位置,修改原有的 IoU 损失为广义 IoU (Generalized IoU, GIoU)^[74]损失 L_{GIoU} (如式(5)所示),用于准确计算重叠预测框的位置损失.通过对 FCOS 的改进,FCOS v2 能增强多尺度目标特征的选择和利用,减少同类目标的误检问题,在密集目标检测中取得了较好的效果.与此同时,为了提升 FCOS 的特征表示能力, Qin 等^[75]在 FCOS^[33]和 ATSS^[56]架构上去除中心度分支,构建了一个新的密集目标检测器 VarifocalNet,将位置预测合并到分类损失计算中,通过预测定位相关的感知 IoU 分类得分(IoU-aware classification score, IACS),优化检测结果的排序过程.此外设计一种高效星型预测框的特征表示方法,获取预测框的几何信息和上下文信息,并且充分利用焦点损失的加权思想,提出变焦损失(Varifocal Loss),通过调制因子来回归每张图上连续的 IoU 关联分类分数 IACS,用于均衡正负样本权重和聚焦难分类样本,从大量预测框中选择位置和类别平衡的最优预测框,显著提升检测精度.针对 FCOS 使用单特征点无法完整表达重叠目标边界信息的问题, BorderDet^[76]通过分析重叠目标检测的特征表达能力,使用一种边界对齐(Border align, BA)的特征提取操作,以实现自适应的边界特征提取.此外,将边界对齐操作封装成边界对齐模块(Border align module, BAM)并集成到 FCOS 网络中,以高效的边界特征提取策略来

提高重叠目标检测精度.

GIoU 损失计算如下:

$$L_{\text{GIoU}} = 1 - \frac{|B \cap B^{\text{gt}}|}{|B \cup B^{\text{gt}}|} + \frac{|C|}{|B^{\text{gt}} \cup B|} \quad (5)$$

其中, B 为预测框, B^{gt} 为真实框, C 为包含预测框和真实框的最小方框, $|C|/(B^{\text{gt}} \cup B)$ 为最小方框 C 的面积减真实框 B^{gt} 和预测框 B 相与的面积.

在样本标签分配方面, 第 1.4 节所提及的 ATSS^[56]、AutoAssign^[57] 等方法虽然能够动态地选择正负样本, 调整目标预测框的相对位置, 但未考虑目标尺寸、形状或边界遮挡的情况. 因此, Cheng 等^[77] 提出一种弱监督多模式注释分割模型 (Weakly supervised multimodal annotation segmentation, WSMA-Seg), 去除 NMS 操作以及相关超参数, 利用多模式分段注释的拓扑结构提取每个目标的轮廓, 使用多尺度池化分割 MSP-Seg 提高多尺度目标的边缘提取, 用静态样本分配策略来有效提高严重遮挡目标的检测精度. 然而上述方法需手动设置层和层内部空间的各种参数和结构, 无法动态地选择正负样本, 因此会导致正负样本不均衡的问题. 在静态和动态样本分配方法的基础上, 为了寻找全局最优的样本分配方法, Ge 等^[78] 丢弃了为每一个目标对象单独寻求最优分配的方法, 提出基于最优传输理论的目标检测样本匹配策略 (Optimal transport assignment, OTA), 利用全局信息来寻找全局最佳样本分配结果, 为图像中的所有目标寻找全局置信度高的样本分配方式, 应用于 FCOS 方法且取得了良好的检测效果. 在此基础上, Ge 等^[79] 又提出基于损失感知的样本分配策略 (Loss-aware label assignment, LLA), 通过联合计算分类和回归损失, 解决锚点和对应真实框样本分配不一致的问题, 提高人群场景中行人检测的性能, 并且通过实验证明了这种简单的样本分配策略能大幅提升 FCOS 的检测精度. 除此之外, Zhang 等^[80] 将 FreeAnchor 扩展为 LTM (Learning-to-match) 检测算法, 该网络将目标与特征的匹配定义为极大似然估计问题, 并将极大似然概率转换为损失函数, 使手动设计的锚点转变为自动特征选择, 解决目标遮挡或特征中心不对齐时, 预测框和真实框间的 IoU 难以度量目标与特征匹配的问题. LTM 不仅在 COCO 数据集上获得良好的检测效果, 而且大幅提升密集、重叠目标的检测性能.

除了上述方法外, 前文所描述的 Transformer 检测算法 DETR^[41]、Deformable DETR^[44] 等也能够通过层叠的解码器级联结构在一定程度上提高重叠目标的特征匹配程度. 为了进一步提高密集场景下

各类目标的检测性能、简化整体结构, Yao 等^[81] 在 DETR 基础上提出了一种简单有效的端到端检测算法 Efficient DETR. 该算法利用密集先验知识初始化检测网络, 降低解码器层数较少造成的性能抑制, 同时共享一个检测头来简化结构, 仅用 3 个编码器和 1 个解码器就达到较高的检测精度并提高收敛速度, 优于 Deformable DETR 等算法的密集目标检测效果. 此外, Yu 等^[82] 提出了用于端到端行人搜索的级联阻塞注意变换检测算法 (Co-scale convolutional image Transformers, COAT), 使用三段级联设计来检测和提升目标的检测性能, 实现每个阶段注意力结构的紧密特征交叉效果, 使网络从粗到细地学习目标特征, 更清晰地区分目标和背景特征, 并通过实验模拟目标遮挡的效果, 验证该方法的先进性. 在同一时期, Zheng 等^[83] 在 Deformable DETR 基础上, 提出一种适用于密集场景的渐进式端到端目标检测器. 通过选择低置信度检测点进行后续的渐进式优化, 设计关联信息提取模块, 寻找低质量预测点与附近高置信度像素点的相关性, 此外通过局部自注意力机制构建队列更新模块, 寻找相邻像素点的空间关联性, 并使用一对一的标签分配策略, 解决 Transformer 在重叠目标场景下的重复检测问题.

表 7 和表 8 说明各通用密集目标检测方法分别从加强边缘特征信息、对齐区域和特征、优化正负样本分配和优化 Transformer 等方面, 改变目标位置的获取方式, 提取并强化目标的语义信息, 优化共用像素点的分类方法, 充分聚焦局部特征信息与整体信息的联系, 精准定位各类别密集排列的目标, 有效提高重叠目标的检测性能. 综上所述, 重叠目标因共用边界导致的类别模糊、漏检和误检等, 可以从寻找多尺度特征间关联、加强特征重识别等角度解决密集行人的重复检测等问题. 现已有大量重叠目标检测方法并取得了丰硕的成果, 但是在训练数据有限的情况下, 考虑到遮挡问题的复杂性, 未来可以从半监督、无监督的学习方式, 利用对抗神经网络扩充数据集, 优化 Transformer 的注意力机制等方面着手, 进一步优化通用场景下的重叠目标检测性能.

2.2 基于无锚框的小目标检测方法

小目标是图像中尺寸小于 32×32 像素或尺寸占比小于 0.12% 的目标 (如图 9 所示). 小目标像素点占比少、覆盖面积小、可视化特征信息极少, 因此难以将小目标与背景、相似目标区域进行区分. 虽然已有 CoupleNet^[84]、NETnet^[85] 和 AF R-CNN^[86] 等算法聚焦目标的主要特征、不断增强并充分利用

表 7 解决目标重叠排列问题的不同检测方法的性能对比

Table 7 Performance comparison of detection methods to solve the problem that objects are densely arranged

问题	算法	数据集	输入尺寸 (像素)	骨干网络	处理器配置	检测速度 (帧/s)	mAP (%)	收录来源	发表年份
目标重叠排列	VarifocalNet ^[75]	COCO	$(480, 960) \times 1333$	ResNeXt-101	Tesla V100 $\times 8$	6.7	50.8	TMI	2019
	WSMA-Seg ^[77]	COCO	—	MSP-Seg	—	—	38.1	arXiv	2019
	FCOS v2 ^[73]	COCO CrowdHuman	$800 \times \leq 1333$	ResNeXt-101 ResNet-50	GTX 1080Ti	—	50.4 87.3	TPAMI	2022
	BorderDet ^[76]	COCO	$800 \times \leq 1333$	ResNeXt-101	GPU $\times 8$	—	50.3	ECCV	2020
	AlignPS ^[71]	CUHK-SYSU PRW	900×1500	ResNet-50	Tesla V100	16.4	94.0 46.1	CVPR	2021
	OTA-FCOS ^[78]	COCO CrowdHuman	$(640, 800) \times \leq 1333$	ResNeXt-101 ResNet-50	GPU $\times 8$	—	51.5 88.4	CVPR	2021
	LLA-FCOS ^[79]	CrowdHuman	$800 \times \leq 1400$	ResNet-50	GPU $\times 8$	—	88.1	Neuro-computing	2021
	LTM ^[80]	COCO	$800 \times \leq 1333$	ResNeXt-101	Tesla V100 $\times 8$	1.7	46.3	TPAMI	2022
	Efficient DETR ^[81]	COCO CrowdHuman	—	ResNet-101 ResNet-50	—	—	45.7 90.8	arXiv	2021
	PSTR ^[72]	CUHK-SYSU PRW	900×1500	ResNet-50	Tesla V100	—	94.2 50.1	CVPR	2022
	COAT ^[82]	CUHK-SYSU PRW	900×1500	ResNet-50	Tesla A100	11.1	94.2 53.3	CVPR	2022
	Progressive DETR ^[83]	COCO CrowdHuman	$(480, 800) \times \leq 1333$	ResNet-50	GPU $\times 8$	—	46.7 92.1	CVPR	2022

表 8 解决目标重叠排列问题的不同检测方法优缺点对比

Table 8 Feature comparison of detection methods to solve the problem that objects are densely arranged

问题	算法	方法	优点	缺点/难点
目标重叠排列	CSP ^[70]	增加中心点偏移量预测分支和尺度预测分支	解决行人检测任务中漏检问题	特征与框间的关联度较低
	VarifocalNet ^[75]	预测 IACS 分类得分、提出 Varifocal Loss 损失函数	有效抑制同目标重叠框	小目标检测效果需提升
	WSMA-Seg ^[77]	利用分割模型构建无需 NMS 后处理的目标检测模型	准确利用重叠目标边缘特征	分割算法的模型复杂度较高
	FCOS v2 ^[73]	将中心度子分支加入回归分支, 并修正中心度计算方式	减少类别判断错误数量	针对不同尺度特征仅使用相同检测头, 限制模型性能
	BorderDet ^[76]	用边界对齐的特征提取操作自适应地提取边界特征	高效获取预测框的位置	边界点选取数量较多
	AlignPS ^[71]	使用特征对齐和聚合模块	解决区域、尺度不对齐的问题	未扩展到通用目标检测任务
	OTA-FCOS ^[78]	用最优传输理论寻找全局高置信度样本分配方式	有助于选择信息丰富区域	模型的计算复杂度较高
	LLA-FCOS ^[79]	使用基于损失感知的样本分配策略	锚点和真实框对应性更好	仅在密集人群中的效果较好
	LTM ^[80]	目标与特征的匹配定义为极大似然估计问题	提高目标遮挡和不对齐的精度	检测速度有待进一步提高
	Efficient DETR ^[81]	用密集先验知识初始化来简化模型结构	减少编码器和解码器数量	检测精度有待进一步提升
	PSTR ^[72]	使用 Transformer 构成首个人搜索网络	提高特征的可判别性和关联性	未扩展到通用目标检测任务
	COAT ^[82]	用三段级联设计来检测和完善的检测和重识别	更清晰地区分目标和背景特征	部分阶段过度关注 ReID 特征, 牺牲部分检测性能
	Progressive DETR ^[83]	设计关联信息提取模块和队列更新模块	加强低置信点的复用	检测精度有待进一步提升

目标特征信息, 以此来提高小目标检测的性能. 然而, 真实场景通常会存在光照剧烈变化、目标遮挡、目标稠密相连和目标尺度不一等错综复杂的变化情况, 这些因素进一步增大了小目标检测的难度, 因此检测任务在小目标场景下仍然充满挑战. 此外在实际应用中, 精确地从人群中检测出犯罪嫌疑人,

能够有助于案件的侦破; 在自动驾驶中准确检测出行人或障碍物, 能够减少交通事故的发生; 在海洋安全中精确检测出尺寸小的船舶, 有助于寻找非法渔船、禁止非法入侵, 因此小目标检测具有广泛的实际应用价值和重要的研究意义. 本节主要聚焦无锚框的小目标检测方法, 从增强特征选择和利用样



图9 小目标示例

Fig.9 The object example of too few pixels

本标签分配来进一步提高复杂背景下小目标检测的性能。

为了给不同尺度的特征图选择最丰富的小目标特征, Samet 等^[87]提出一种基于预测池化的检测算法 (Prediction pooling detection network, PP-Det). 该网络使用一种新的样本分配策略来汇总各个特征的预测值, 减少训练过程中非判别性特征的影响, 并且遵循尺度分配策略来预测类别概率向量和边界框坐标, 自动降低训练过程中非目标区域特征和非判别性目标特征的贡献, 减少目标上非判别性特征、预测框内背景特征、遮挡物的标签噪音. 在样本标签分配方面, 第 1.3 节的 Deformable DETR^[44] 也使用可变形注意力机制替代基于 Transformer 的注意力机制, 进一步关注稀疏空间的小目标位置, 提升小目标检测精度. He 等^[88] 为了提高船舰小目标的检测性能, 提出了自适应特征编码策略 (Adaptive feature encoding strategy, AFE), 逐步将深层语义特征融合到浅特征层, 实现自适应的空间权重信息融合, 以此来自适应地增强小目标信息的表达能力. 此外, 针对样本分配不均衡的问题, 根据软采样的思想和 FCOS 整体架构来构造高斯引导检测头, 进一步优化目标定位准确度, 在小目标船舰数据集中获得良好的检测性能。

在增强小目标有效特征信息获取能力的基础上, 为了加强小目标特征的表达能力, Yang 等^[89] 提出一种新的用于表征目标特征的代表点集 (Representative points, RepPoints). 该方法通过自上而下的方式自适应学习特征图的极值点和关键语义信息, 并构建基于 RepPoints 的目标检测算法 RPDet, 减少背景内容或前景区域无效信息, 使特征包含更丰富的形状和姿态等目标语义信息, 提高小目标的定位精准度, 但仍过度依赖回归分支. 为了更进一步减少误检的概率, RepPoints v2^[90] 在 RepPoints 的基础上融合回归和辅助分支来对验证任务建模, 增加角点验证分支来判断当前像素点为角点的概率, 同时借助高斯分布平滑正负样本的分布曲线; 利用多任务学习、验证模块判断特征映射点是否位于目标对象框内, 进一步提升小目标检测性能. RepPoints

及其改进方法, 利用点集替代边界框 (如图 10 所示), 改变特征的采样和选择、重分配分类分支的像素点权重, 能获得更具目标内部和边缘信息的特征, 此类方法适合特征不充足的小目标检测, 还能应用于其他计算机视觉任务. 针对空中目标环境复杂且目标摆放角度多样的问题, Li 等^[91] 提出一种基于自适应学习样本点的方法 Oriented RepPoints. 该算法在 RepPoints 的基础上, 提出极值函数、局部极值函数和基于矩的函数来替换原有损失, 以用于捕获任意方向目标的几何信息、方便准确地定向定位和分类小目标. 针对 RepPoints 的关键点定位不准确的问题, 提出一种质量评估和样本分配方法 (Adaptive points assessment and assignment, APAA), 用于在训练期间选择具有代表性的定向正样本点. 该方法引入空间约束来惩罚非特征点, 提升相邻对象或背景噪声中非轴对齐目标特征的捕获能力, 在空域小目标、旋转目标等场景下有较好的检测效果。

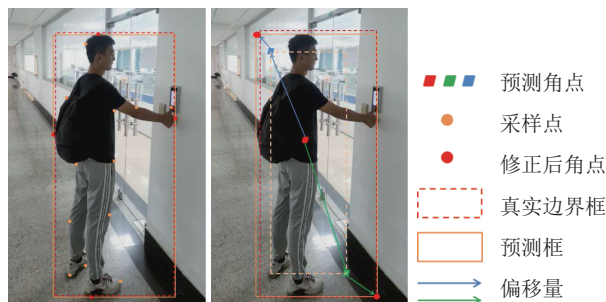


图 10 RepPoints 系列点集表示示例

Fig.10 The example of RepPoints series point set

为了充分利用和融合小目标的浅层位置信息和深层语义信息, DuBox^[92] 使用具有多尺度特性的双尺度残差单元, 加入先验框来提高模型处理尺度不变性特征的能力, 优化小目标预测框的定位准确度, 扩大边界框中心点正样本范围, 减少小目标边缘信息的漏检. 此外, FoveaBox 算法^[93] 联合预测目标中心区域的位置以及预测框的有效位置, 在多层特征图上检测不同尺度目标的特征, 直接学习分类目标类别并回归出目标的位置, 模型具备良好的鲁棒性和泛化能力, 大幅提高小目标检测精度. 为了拓宽尺度多样性、降低背景与目标的强关联性, Fu 等^[94] 提出了一种新的特征均衡与精炼的检测方法 (Feature balancing and refinement network, FBR-Net), 直接学习编码后的边界框, 解决船舶排列稀疏的漏检问题, 并使用基于多尺度的注意力机制来平衡不同层次的小目标语义信息, 实现特征均衡与网络细化. FBR-Net 整体网络优化小目标尺度多样性、减少近岸背景与船舶的强关联性, 显著提升小

目标的检测精度. 为了无需提高图像分辨率且实现快速有效的小目标检测, Yang 等^[95]提出一种基于级联稀疏查询机制 (Cascading sparse query mechanism, CSQ) 的检测网络 QueryDet. 该网络利用特征金字塔的特征查询来加快目标检测网络的预测速度, 该网络在低分辨率特征图上预测小目标的粗略位置, 利用稀疏卷积来获取高分辨率特征, 用粗调联合精调的方式进行动态预测, 减少浅层检测头的计算开销并提高小目标的位置精确度.

综上所述, 从样本标签分配、特征注意力机制和特征重要程度等方法中能充分利用目标相关的特征信息, 解决小目标因各类正负样本比例极不平衡、类间不平衡导致的特征学习困难等问题. 通过表 9 和表 10 分析各算法的检测性能可知, 多尺度特征

融合方法同时考虑了浅层的表征信息和深层的语义信息, 提高小目标的特征提取能力, 利用稀疏查询机制进一步缓解因特征图分辨率下降导致的小目标采样丢失、定位不准确和复杂背景误报等问题, 减少计算开销, 大幅提升小目标的检测效果. 虽然已有的无锚框小目标检测方法已获得了良好的检测性能, 但检测精度和速度与大、中目标检测还有一定差距, 并且随着小目标应用场景的迅速扩张, 小目标检测仍值得进一步研究.

2.3 基于无锚框的旋转目标检测方法

由于实际检测场景的高复杂性, 目标呈现排列密集、排放方向任意的特点, 仅用水平框检测目标会使每个预测框内包含大量非待测目标的像素点和

表 9 解决目标尺寸过小问题的不同检测方法性能对比

Table 9 Performance comparison of detection methods to solve the problem that object pixels are too few

问题	算法	数据集	输入尺寸 (像素)	骨干网络	处理器配置	检测速度 (帧/s)	mAP (%)	收录来源	发表年份
目标尺寸过小	RepPoints ^[89]	COCO	$(480, 960) \times \leq 960$	ResNet-101	GPU $\times 4$	—	46.5	ICCV	2019
	DuBox ^[92]	COCO VOC 2012	800×800 500×500	ResNet-101 VGG-16	NVIDIA P40 $\times 8$	—	39.5 82.0	arXiv	2019
	PPDet ^[87]	COCO	800×1300	ResNet-101	Tesla V100 $\times 4$	—	45.2	BMVC	2020
	RepPoints v2 ^[90]	COCO	$(800, 1333) \times \leq 1333$	ResNet-101	GPU $\times 8$	—	48.1	NeurIPS	2020
	FoveaBox ^[93]	COCO VOC 2012	800×800	ResNet-101 ResNet-50	GPU $\times 4$	— 16.4	42.1 76.6	TIP	2020
	FBR-Net ^[94]	SSDD	448×448	ResNet-50	RTX 2080Ti	25.0	92.8	TGRS	2021
	FCOS (AFE-GDH) ^[88]	HRSID SSDD	800×800	ResNet-50	NVIDIA Titan Xp	15.2	67.4 28.5	Remote Sensing	2022
	Oriented RepPoints ^[91]	DOTA HRSC2016	1024×1024 $(300, 900) \times (300, 1500)$	ResNet-101 ResNet-50	RTX 2080Ti $\times 4$	—	76.5 97.3	CVPR	2022
	QueryDet ^[95]	COCO	—	ResNet-50	RTX 2080Ti $\times 8$	14.4	39.5	CVPR	2022

表 10 解决目标尺寸过小问题的不同检测方法优缺点对比

Table 10 Feature comparison of detection methods to solve the problem that object pixels are too few

问题	算法	方法	优点	缺点/难点
目标尺寸过小	RepPoints ^[89]	使用点集形式表征目标的特征	自适应地学习极值点和语义信息	过度依赖回归分支
	DuBox ^[92]	使用有多尺度特性的双尺度残差单元	减少小目标边缘和内部信息的漏检	分割模型的复杂度较高
	PPDet ^[87]	使用框内部为正样本点的新标记策略	提高判别性目标特征的贡献程度	小目标特征信息不足
	RepPoints v2 ^[90]	增加角点验证分支来判断特征映射点	获得更具目标内部和边缘信息的特征	预测框定位准确度低
	FoveaBox ^[93]	在多层特征图上检测多尺度目标特征	对目标形状和分布有很强的适应能力	难以区分目标和背景区域
	FBR-Net ^[94]	用多尺度注意力机制选择特征重要性	减少背景区域与小目标间的强关联性	检测精度仍需进一步提升
	FCOS (AFE-GDH) ^[88]	使用自适应特征编码策略 (AFE) 和构造高斯引导检测头	有效增强小目标表达能力	仅说明船舰目标有效性
	Oriented RepPoints ^[91]	提出质量评估、样本分配方案和空间约束	提升非轴对齐小目标特征的捕获能力	仅涉及空域小目标检测
	QueryDet ^[95]	使用基于级联稀疏查询机制进行动态预测	减少检测头计算开销、提高小目标的位置精确度	提高分辨率导致误判概率提高

无关背景信息, 导致边界框的重叠度较高、边界不连续和预测框位置偏移等问题. 如舰船目标密集排列且旋转角度多样, 使用水平矩形框表示舰船目标会产生大量重合, 具体情况如图 11(a) 所示. 因此引入旋转矩形框来检测目标 (如图 11(b) 所示), 获取充足的目标特征信息, 如目标的姿态等, 减少冗余背景信息, 且能更准确地反映目标的位置信息, 减少相邻预测框的重叠. 除此之外, 在农业病虫害检测、交通车辆检测等任务中有更强的实用性.

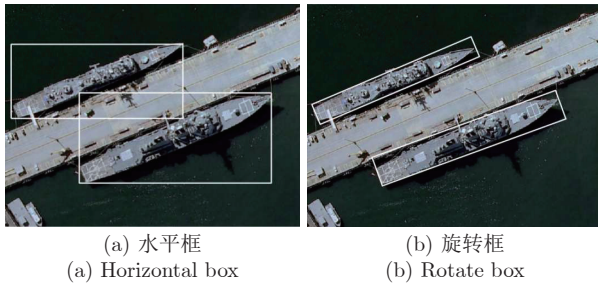


图 11 多角度目标检测结果示例

Fig. 11 The detection result of arbitrary rotation objects

旋转目标的预测框回归依赖于不同坐标系, 为了减小目标在不同坐标系中的参数量和回归难度, Yi 等^[96]将水平框的目标检测算法扩展到旋转目标检测任务中, 提出一个基于边缘感知向量的目标检测算法 (Box boundary-aware vectors, BBAVectors). 该方法回归一个边缘感知向量来替代原有参数 (宽度 w 、高度 h 和旋转角度 θ), 使所有的回归参数均位于同一坐标系内, 简化旋转预测框的计算复杂度, 且进一步将有方向的预测框分成了水平框和旋转框两大类, 解决目标旋转角度较小时预测框类型难以区分的问题. 另一方面, Zhou 等^[97]首次将极坐标系应用于目标检测任务中, 提出一个基于极坐标的检测算法 (Polar remote sensing object detector, P-RSDet). 整体网络通过回归一个极半径和两个极角来实现任意方向物体的检测, 此外还构建了极坐标环形面积损失函数, 避免角度回归的周期性及多边形预测框的顶点排序问题, 减少回归参数的数量, 简化检测模型结构.

在第 2.2 节提出的 Oriented RepPoints 算法^[91]不仅能精确检测小目标, 该算法的标签分配方案 APAA 也能有效提取具有显著性特征的非轴对齐目标信息, 提高空域旋转目标的检测精度. 在优化正负样本分配策略方面, Li 等^[98]将 FCOS 扩展为带有角度旋转的检测算法 (Rotated FCOS, FCOSR). 其中使用基于高斯分布的椭圆中心采样策略, 用于定义合适的旋转框正样本选择区域, 并设计模

样本分配策略, 解决 FCOS 样本分配方法在航空场景下忽略模糊样本的问题, 解决相邻水平框的目标区域重叠问题, 额外加入多尺度采样模块来解决像素点采样不充分的问题. 在优化正样本选择的基础上, 增大正样本比例, 有效提高空域目标的检测效果. 除此之外, 为了解决水平框检测目标所占预测框整体比例较少的问题, Cheng 等^[99]提出基于选择提议生成器的检测算法 (Anchor-free oriented proposal generator, AOPG). 该算法将特征图映射到图像上, 并将位于真实框中心区域的顶点视为正样本, 以此来构建新的区域标签分配模块, 缓解正样本所占比例少的问题. 此外将特征金字塔的每一层特征输入到粗定位模块 (Coarse location module, CLM) 来生成粗定位的旋转预测框, 然后使用对齐卷积来消除特征和旋转框间的不对齐, 采用 Faster R-CNN 的检测头来生成精确定位和分类的高质量预测框, 并在多个旋转目标数据集上进行实验来验证其有效性. 同时, Huang 等^[100]提出一种无锚框的目标自适应标签分配策略 (General Gaussian heatmap label assignment, GGHL), 用于定义基于二定向高斯热力图的正样本点, 从热力图中获取任意方向目标的形状和方向特征. 此外, 在标签分配策略的基础上开发用于定向预测框的标签分配组件 (Oriented bounding box representation component, ORC), 通过卷积层自适应地调整高斯中心先验权重以适应不同目标的特征, 并且设计具有面积归一化和动态置信度加权的联合优化损失 (Joint optimization loss, JOL), 以完善不同分类和回归子任务的非对齐优化任务, 除了提升检测精度外, 检测速度也得到大幅提升.

为了充分利用旋转目标的特征, Wei 等^[101]提出逐像素点预测的旋转目标检测算法 (Oriented objects detection network, O²-DNet). 该算法将二值化热力图中连通区域中线的交点作为关键点, 通过回归映射来预测两条相应的中线, 解决旋转目标角度边界不连续的问题, 且使用感知方向中心度的方法, 优化正样本点的加权方式, 引导网络从复杂背景中学习可判别性特征. 此外, 在损失函数内加入中心度损失, 更精准定位目标预测框的位置. 在此基础上, 为了获得更具判别性的方向信息, Wang 等^[102]提出一种基于尺度感知的旋转目标检测算法 (Scale-aware rotated object detection, SARD). 该算法首次使用一种新的特征融合模块 FPSN (Feature pyramid stack network), 充分获取高级语义信息和低级位置信息, 以适应各类大尺度变化的目标, 同时联合 IoU 损失 L_{IoU} (如式 (1) 所示) 和 Smooth L₁ 损失 $L_{Smooth L_1}$ (如式 (6) 所示) 修正回归分支的权重

参数, 提高预测框的位置判断准确度. 此外, 使用归一化策略, 更精准地检测任意方向和密集排列的目标. 在另一方面, Zhang 等^[103] 提出一个直接预测像素点边界框全部参数的网络 DARDet (Dense anchor-free rotated object detector), 该网络设计了一个高效对齐卷积模块来提取对齐特征, 此外还引入 Pixels-IoU (PIoU) 损失^[104] L_{PIoU} (如式 (7) 所示) 来有效处理边界不连续性问题. 为了进一步改进旋转目标损失不连续和小目标检测性能较差的问题, Lang 等^[105] 提出一种面向对象的无锚框检测网络 DAFNe (Dense anchor-free network). 该网络结合中心点与角点间的间距来联合调整边界框预测策略, 以提高检测器的定位精度, 并引入一种基于方向感知的边界框中心度函数, 首次将中心度函数推广到任意四边形, 以此来充分利用目标特征并降低低质量预测框的权重, 进一步提高旋转小目标的定位精度.

Smooth L_1 损失^[5] 的计算如下:

$$L_{\text{smooth } L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{否则} \end{cases} \quad (6)$$

其中, x 表示预测值.

PIoU 损失^[104] 的计算如下:

$$L_{PIoU} = -\frac{1}{|M|} \sum_{(\mathbf{b}, \mathbf{b}') \in M} \ln PIoU(\mathbf{b}, \mathbf{b}') \quad (7)$$

$$PIoU(\mathbf{b}, \mathbf{b}') = \frac{S_{\mathbf{b} \cap \mathbf{b}'}}{S_{\mathbf{b} \cup \mathbf{b}'}} \quad (8)$$

其中, \mathbf{b}' 为旋转预测框, \mathbf{b} 为旋转真实框, $S_{\mathbf{b} \cap \mathbf{b}'}$ 、 $S_{\mathbf{b} \cup \mathbf{b}'}$ 分别表示定向预测框和真实框的交集和并集, M 为正样本集.

在提高特征的获取和融合能力后, 提高旋转目标检测模型的自适应能力成为焦点. Pan 等^[106] 提出一种适应定向和密集目标任务的动态优化检测算法 (Dynamic refinement network, DRN), 通过自适应的特征选择模块来自动提取、融合各形状和方向的有效特征, 缓解各类目标与轴对齐感受野之间的不匹配问题, 同时提出两种检测头来分别动态优化分类和回归任务, 完成对样本唯一性和特殊性的建模, 提高检测性能的同时增强模型的灵活性. 为了进一步充分利用目标的尺寸和方向等先验信息, Zhang 等^[107] 提出一种基于中心点提取的检测算法 (Center-head point extraction detector, CHPDet). 该算法通过将角度回归任务转化为中心点估计任务, 解决角度周期性导致的边界不连续问题, 不仅实现任意方向的目标检测, 而且可判断目标的朝向 (如船首和船尾), 适应多类复杂场景下的检测任务.

由表 11 和表 12 可知, 上述方法从改变坐标表示、修正样本标签分配方法、充分利用特征信息、提高特征融合效率、充分映射方向不变性特征和修正损失函数等策略, 利用旋转框代替水平框检测目标, 解决旋转目标的特征利用不充分、预测框和目标间的特征不对齐、预测框位置偏移等问题, 在各旋转目标检测数据集上定位准确度高、误判概率低, 检测效率高. 虽然无锚框旋转目标检测方法已经获得良好的检测性能, 然而旋转目标损失函数仍存在不连续和小尺寸目标占比过小的问题. 目前现存基于锚框的目标检测方法 GWD^[108]、KLD^[109]、RSDet++^[110] 等所提出的损失函数能在保证连续的基础上, 进一步拟合旋转目标形状, 并在复杂目标数据集中获得良好的检测效果, 未来可以将此类损失函数应用于基于无锚框的目标检测任务中.

3 总结与展望

本文简要回顾和介绍了基于锚框的目标检测方法整体流程和问题, 系统地对比了现有基于无锚框的目标检测方法, 并针对密集目标、小目标、旋转目标的检测任务, 总结目前的研究现状. 虽然硬件条件不断完善, 但是随着图像分辨率的提高, 数据量逐渐增大, 对目标检测也提出新的挑战. 例如在实际应用中, 目标检测面临的样本数量少、背景复杂等检测问题. 因此, 在后续的研究中应考虑以下几个研究方向:

1) 由于目标检测算法通常依赖于大量高质量标注的图像数据集, 而标注过程复杂且效率较低, 现常用样本生成方法来扩充样本数量. 然而该方法本质上是数据增强操作, 无法从根本上解决数据匮乏的问题. 因此使用弱监督目标检测技术, 通过对部分标注的数据集进行小样本训练, 使用生成对抗网络来扩充数据集是后续的重要研究方向.

2) 图像数据在获取过程中会受到气候、光照等因素的影响, 现已有图像去雾去噪、多尺度感知等方法来生成清晰且无颜色失真的图像, 但实际场景下仍会造成图像细节丢失、目标边缘模糊等问题. 如何更好地实现复杂多变场景下的目标检测, 是后续研究的重要发展方向.

3) 由于图像中存在多类不同尺寸的目标, 且提升多种尺度目标的检测精确度有一定挑战, 现有方法利用多尺度信息提取多样的目标特征, 但是多尺度检测方法仅对目标的尺寸进行区分, 并未关注目标类别和尺寸间的相关性. 现已将文本检测的 Transformer 结构引入至目标检测任务, 即通过对输入图像的切片化处理, 不断聚焦目标局部特征信

表 11 解决目标方向变化问题的不同检测方法性能对比

Table 11 Performance comparison of detection methods to solve the problem that object direction changeable

问题	算法	数据集	输入尺寸 (像素)	骨干网络	处理器配置	检测速度 (帧/s)	mAP (%)	收录来源	发表年份
目标方向 变化	SARD ^[102]	DOTA HRSC2016	800 × 800	ResNet-101	Tesla P100	— 1.5	72.9 85.4	IEEE Access	2019
	P-RSDet ^[97]	DOTA UCAS-AOD	512 × 512	ResNet-101	Tesla V100 × 2	—	72.3 90.0	IEEE Access	2020
	O ² -DNet ^[101]	DOTA ICDAR2015	800 × 800	ResNet-101	Tesla V100 × 2	—	71.0 85.6	P&RS	2020
	DRN ^[106]	DOTA HRSC2016	1024 × 1024 768 × 768	Hourglass-104	Tesla V100	—	73.2 92.7	CVPR	2020
	BBAVectors ^[96]	DOTA HRSC2016	608 × 608	ResNet-101	GTX 1080Ti × 4	— 11.7	75.4 88.6	WACV	2021
	FCOSR ^[98]	DOTA HRSC2016	1024 × 1024 800 × 800	ResNeXt-101	Tesla V100 × 4	7.9 —	77.4 95.7	arXiv	2021
	DARDet ^[103]	DOTA HRSC2016	1024 × 1024	ResNet-50	RTX 2080Ti	12.6 —	71.2 78.9	GRSL	2021
	DAFNe ^[105]	DOTA HRSC2016	1024 × 1024	ResNet-101	Tesla V100 × 4	—	76.9 89.5	arXiv	2021
	CHPDet ^[107]	UCAS-AOD HRSC2016	1024 × 1024	DLA-34	RTX 2080Ti	—	89.6 88.8	TGRS	2021
	AOPG ^[99]	DOTA HRSC2016	1024 × 1024 (800, 1333) × (800, 1333)	ResNet-101 ResNet-50	RTX 2080Ti	10.8 —	80.2 96.2	TGRS	2022
GGHL ^[100]	DOTA SSDD+	800 × 800 —	Darknet53	RTX 3090 × 2	42.3 44.1	76.9 90.2	TIP	2022	

表 12 解决目标方向变化问题的不同检测方法优缺点对比

Table 12 Feature comparison of detection methods to solve the problem that object direction changeable

问题	算法	方法	优点	缺点/难点
目标方 向变化	SARD ^[102]	用尺度感知方法融合深层和浅层特征信息	对大尺度变化和多角度变化目标适应度好	整体检测效率较低
	P-RSDet ^[97]	回归一个极半径和两个极角, 实现多角度物体的检测	避免角度周期性及预测框的顶点排序问题	极坐标的后处理操作相关复杂度较高
	O ² -DNet ^[101]	用纵横比感知方向中心度的方法, 学习判别性特征	网络从复杂背景中学习更具判别性的特征	特征融合方法的实际融合效果较差
	DRN ^[106]	使用自适应的特征选择模块和动态优化的检测头	缓解目标特征和坐标轴之间的不对齐问题	检测精度有待进一步提升
	BBAVectors ^[96]	使用边缘感知向量来替代原回归参数	在同坐标系中回归所有参数, 减少计算量	向量的类型转化过程处理较复杂
	FCOSR ^[98]	使用基于高斯分布的椭圆中心采样策略	修正样本分配方法在航空场景下漏检问题	未实现标签分配方案的自适应
	DARDet ^[103]	设计高效对齐卷积模块来提取对齐特征	可一次性预测出所有的预测框相关参数	损失函数的角度预测偏移量较大
	DAFNe ^[105]	使用基于方向感知的边界框中心度函数	降低低质量框的权重并且提高定位精度	损失函数仍存在部分旋转敏感度误差
	CHPDet ^[107]	使用方位不变模块 OIM 生成方位不变特征映射	确定旋转目标的朝向 (如车头、船头等)	存在目标预测框的位置偏移量
	AOPG ^[99]	使用区域标签分配模块和粗定位模块	缓解标签分配不均衡、目标特征不对齐的问题	未实现标签分配方案的自适应
GGHL ^[100]	使用二维定向高斯热力图进行定向框标签分配	实现动态标签分配对齐回归和分类任务	检测精度有待进一步提升	

息的提取并加强特征的联系. 在后续研究中应考虑将其其他视觉任务扩展至目标检测领域, 以自适应的方式提高不同输入图像尺寸、目标尺寸的检测性能, 并获得泛化能力更强的模型.

4) 针对图像中目标密集排列、重叠摆放所导致预测框定位困难的问题, 现存的解决方案 SCRDet^[111]等算法利用倾斜框替代水平框, 使预测框包含更多目标像素点, 但是未从根本上解决密集目标边界特

征提取难的问题. 未来可以将目标检测与图像分割任务相结合, 先对密集目标进行分割, 再利用分割后的目标完成特征提取和检测.

5) 现有的目标检测算法通常针对单张图像进行目标检测, 忽略了视频帧之间的相关性. 而高清视频中的实时目标检测/跟踪对于视频监控和自动驾驶具有重要意义, 现已有 MOT^[112]、FairMOT^[113] 等算法来实现多目标跟踪任务, 在目标检测基础上实现应用更广泛的视频跟踪任务. 因此未来应继续探索视频帧序列之间的空间、时间相关性, 以进一步改善检测性能并扩大应用范围.

6) 现存目标检测方法检测速度较慢, 为了能够在移动设备上获得流畅的运行结果, 通常会投入大量资金来购买算力更高的设备去加速算法检测过程, 这在一定程度上阻碍了各类检测算法的落地进程. MobileNet^[114]、ShuffleNet^[115] 和 Pelee^[116] 等轻量化网络简化网络结构并降低算力要求, 提供了移动端的目标检测的坚实基础, 为在轻量设备上检测目标提供可能性. 因此, 研究轻量快速且高精度的无锚框目标检测方法是未来必不可少的一个发展趋势.

References

- Nie Guang-Tao, Huang Hua. A survey of object detection in optical remote sensing images. *Acta Automatica Sinica*, 2021, 47(8): 1749–1768
(聂光涛, 黄华. 光学遥感图像目标检测算法综述. 自动化学报, 2021, 47(8): 1749–1768)
- Neubeck A, Van Gool L. Efficient non-maximum suppression. In: Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06). Hong Kong, China: IEEE, 2006. 850–855
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, et al. SSD: Single shot multibox detector. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 21–37
- Girshick R. Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 1440–1448
- Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Proceedings of the 29th International Conference on Neural Information Processing Systems. Montreal, Canada: 2015. 91–99
- Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 6517–6525
- Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv preprint arXiv: 1804.02767, 2018.
- Xiao Yu-Qing, Yang Hui-Min. Research on application of object detection algorithm in traffic scene. *Computer Engineering and Applications*, 2021, 57(6): 30–41
(肖雨晴, 杨慧敏. 目标检测算法在交通场景中应用综述. 计算机工程与应用, 2021, 57(6): 30–41)
- Huang L C, Yang Y, Deng Y F, Yu Y N. DenseBox: Unifying landmark localization with end to end object detection. arXiv preprint arXiv: 1509.04874, 2015.
- Everingham M, Van Gool L, Williams C K I, Winn J, Zisserman A. The PASCAL visual object classes (VOC) challenge. *International Journal of Computer Vision*, 2010, 88(2): 303–338
- Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: Common objects in context. In: Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 740–755
- Xia G S, Bai X, Ding J, Zhu Z, Belongie S, Luo J B, et al. DOTA: A large-scale dataset for object detection in aerial images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 3974–3983
- Zhu H G, Chen X G, Dai W Q, Fu K, Ye Q X, Jiao J B. Orientation robust object detection in aerial images using deep convolutional neural network. In: Proceedings of the IEEE International Conference on Image Processing (ICIP). Quebec City, Canada: IEEE, 2015. 3735–3739
- Karatzas D, Gomez-Bigorda L, Nicolaou A, Ghosh S, Bagdanov A, Iwamura M, et al. ICDAR 2015 competition on robust reading. In: Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR). Tunis, Tunisia: IEEE, 2015. 1156–1160
- Xiao T, Li S, Wang B C, Lin L, Wang X G. Joint detection and identification feature learning for person search. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 3376–3385
- Zheng L, Zhang H H, Sun S Y, Chandraker M, Yang Y, Tian Q. Person re-identification in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 3346–3355
- Shao S, Zhao Z J, Li B X, Xiao T T, Yu G, Zhang X Y, et al. CrowdHuman: A benchmark for detecting human in a crowd. arXiv preprint arXiv: 1805.00123, 2018.
- Liu Z K, Yuan L, Weng L B, Yang Y P. A high resolution optical satellite image dataset for ship recognition and some new baselines. In: Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods. Porto, Portugal: SciTePress, 2017. 324–331
- Li J W, Qu C W, Shao J Q. Ship detection in SAR images based on an improved faster R-CNN. In: Proceedings of the SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA). Beijing, China: IEEE, 2017. 1–6
- Wei S J, Zeng X F, Qu Q Z, Wang M, Su H, Shi J. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation. *IEEE Access*, 2020, 8: 120234–120254
- Wang X G, Chen K B, Huang Z L, Yao C, Liu W Y. Point linking network for object detection. arXiv preprint arXiv: 1706.03646, 2017.
- Law H, Deng J. CornerNet: Detecting objects as paired keypoints. In: Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 765–781
- Law H, Teng Y, Russakovsky O, Deng J. Cornernet-lite: Efficient keypoint based object detection. In: Proceedings of the 31st British Machine Vision Conference. BMVC, 2020.
- Zhou X Y, Zhuo J C, Krähenbühl P. Bottom-up object detection by grouping extreme and center points. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 850–859
- Duan K W, Bai S, Xie L X, Qi H G, Huang Q M, Tian Q. CenterNet: Keypoint triplets for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 6568–6577
- Dong Z W, Li G X, Liao Y, Wang F, Ren P J, Qian C. CentripetalNet: Pursuing high-quality keypoint pairs for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE,

2020. 10516–10525
- 27 Lan S Y, Ren Z, Wu Y, Davis L S, Hua G. SaccadeNet: A fast and accurate object detector. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020. 10394–10403
- 28 Duan K W, Xie L X, Qi H G, Bai S, Huang Q M, Tian Q. Corner proposal network for anchor-free, two-stage object detection. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 399–416
- 29 Wang Yan-Qing, Ma Lei, Tian Yuan. State-of-the-art of ship detection and recognition in optical remotely sensed imagery. *Acta Automatica Sinica*, 2011, **37**(9): 1029–1039 (王彦清, 马雷, 田原. 光学遥感图像舰船目标检测与识别综述. 自动化学报, 2011, **37**(9): 1029–1039)
- 30 Yu J H, Jiang Y N, Wang Z Y, Cao Z M, Huang T. UnitBox: An advanced object detection network. In: Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, The Netherlands: ACM, 2016. 516–520
- 31 Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 779–788
- 32 Ge Z, Liu S T, Wang F, Li Z M, Sun J. YOLOX: Exceeding YOLO series in 2021. arXiv preprint arXiv: 2107.08430, 2021.
- 33 Tian Z, Shen C H, Chen H, He T. FCOS: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 9626–9635
- 34 Sun P Z, Jiang Y, Xie E Z, Shao W Q, Yuan Z H, Wang C H, et al. What makes for end-to-end object detection? In: Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021. 9934–9944
- 35 Zhou X Y, Wang D Q, Krahenbuhl P. Objects as points. arXiv preprint arXiv: 1904.07850, 2019.
- 36 Zhou X Y, Koltun V, Krähenbühl P. Probabilistic two-stage detection. arXiv preprint arXiv: 2103.07461, 2021.
- 37 Samet N, Hicsonmez S, Akbas E. HoughNet: Integrating near and long-range evidence for bottom-up object detection. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 406–423
- 38 Chu C, Zhmoginov A, Sandler M. CycleGAN, a master of steganography. arXiv preprint arXiv: 1712.02950, 2017.
- 39 Isola P, Zhu J Y, Zhou T H, Efros A A. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 5967–5976
- 40 Lu X, Li B Y, Yue Y X, Li Q Q, Yan J J. Grid R-CNN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 7355–7364
- 41 Lu X, Li B Y, Yue Y X, Li Q Q, Yan J J. Grid R-CNN plus: Faster and better. arXiv preprint arXiv: 1906.05688, 2019.
- 42 Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 213–229
- 43 Sun Z Q, Cao S C, Yang Y M, Kitani K. Rethinking transformer-based set prediction for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021. 3591–3600
- 44 Zhu X Z, Su W J, Lu L W, Li B, Wang X G, Dai J F. Deformable DETR: Deformable transformers for end-to-end object detection. In: Proceedings of the 9th International Conference on Learning Representations. ICLR, 2021.
- 45 Dai X Y, Chen Y P, Yang J W, Zhang P C, Yuan L, Zhang L. Dynamic DETR: End-to-end object detection with dynamic attention. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021. 2968–2977
- 46 Zhang G J, Luo Z P, Yu Y C, Cui K W, Lu S J. Accelerating DETR convergence via semantic-aligned matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022. 939–948
- 47 Fang Y X, Liao B C, Wang X G, Fang J M, Qi J Y, Wu R, et al. You only look at one sequence: Rethinking transformer in vision through object detection. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021. 26183–26197
- 48 Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. An image is worth 16×16 words: Transformers for image recognition at scale. In: Proceedings of the 9th International Conference on Learning Representations. ICLR, 2021.
- 49 Song H, Sun D Q, Chun S, Jampani V, Han D, Heo B, et al. ViDT: An efficient and effective fully transformer-based object detector. In: Proceedings of the 10th International Conference on Learning Representations. ICLR, 2022.
- 50 Li F, Zhang H, Liu S L, Guo J, Ni L M, Zhang L. DN-DETR: Accelerate DETR training by introducing query DeNoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022. 13609–13617
- 51 Wang J F, Yuan Y, Li B X, Yu G, Jian S. SFace: An efficient network for face detection in large scale variations. arXiv preprint arXiv: 1804.06559, 2018.
- 52 Zhu C C, He Y H, Savvides M. Feature selective anchor-free module for single-shot object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 840–849
- 53 Lin T Y, Goyal P, Girshick R, He K M, Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 2999–3001
- 54 Zhu C C, Chen F Y, Shen Z Q, Savvides M. Soft anchor-point object detection. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 91–107
- 55 Zhang X S, Wan F, Liu C, Ji R R, Ye Q X. FreeAnchor: Learning to match anchors for visual object detection. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2019. Article No. 14
- 56 Zhang S F, Chi C, Yao Y Q, Lei Z, Li S Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020. 9756–9765
- 57 Zhu B J, Wang J F, Jiang Z K, Zong F H, Liu S T, Li Z M, et al. AutoAssign: Differentiable label assignment for dense object detection. arXiv preprint arXiv: 2007.03496, 2020.
- 58 Duan K W, Xie L X, Qi H G, Bai S, Huang Q M, Tian Q. Location-sensitive visual recognition with cross-IOU loss. arXiv preprint arXiv: 2104.04899, 2021.
- 59 Li S, He C H, Li R H, Zhang L. A dual weighting label assignment scheme for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022. 9377–9386
- 60 Liu Xiao-Bo, Liu Peng, Cai Zhi-Hua, Qiao Yu-Lin, Wang Ling, Wang Min. Research progress of optical remote sensing image object detection based on deep learning. *Acta Automatica Sinica*, 2021, **47**(9): 2078–2089 (刘小波, 刘鹏, 蔡之华, 乔禹霖, 王凌, 汪敏. 基于深度学习的光学遥感图像目标检测研究进展. 自动化学报, 2021, **47**(9): 2078–

- 2089)
- 61 Gong Hao-Tian, Zhang Meng. Lightweight anchor-free object detection algorithm based on KeyPoint detection. *Computer Science*, 2021, **48**(8): 106–110
(龚浩田, 张萌. 基于关键点检测的无锚框轻量级目标检测算法. 计算机科学, 2021, **48**(8): 106–110)
- 62 Shao Xiao-Wen, Shuai Hui, Liu Qing-Shan. Person re-identification based on fused attribute features. *Acta Automatica Sinica*, 2022, **48**(2): 564–571
(邵晓雯, 帅惠, 刘青山. 融合属性特征的行人重识别方法. 自动化学报, 2022, **48**(2): 564–571)
- 63 Liu Yang, Zhan Yin-Wei. Survey of small object detection algorithms based on deep learning. *Computer Engineering and Applications*, 2021, **57**(2): 37–48
(刘洋, 战荫伟. 基于深度学习的小目标检测算法综述. 计算机工程与应用, 2021, **57**(2): 37–48)
- 64 Bodla N, Singh B, Chellappa R, Davis L S. Soft-NMS: Improving object detection with one line of code. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 5562–5570
- 65 Liu S T, Huang D, Wang Y H. Adaptive NMS: Refining pedestrian detection in a crowd. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 6452–6461
- 66 Huang X, Ge Z, Jie Z Q, Yoshie O. NMS by representative region: Towards crowded pedestrian detection by proposal pairing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020. 10747–10756
- 67 Zhang S F, Wen L Y, Bian X, Lei Z, Li S Z. Occlusion-aware R-CNN: Detecting pedestrians in a crowd. In: Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 657–674
- 68 Yang Shan, Wang Jian, Hu Li, Liu Bo, Zhao Hao. Research on occluded object detection by improved RetinaNet. *Computer Engineering and Applications*, 2022, **58**(11): 209–214
(阳珊, 王建, 胡莉, 刘波, 赵皓. 改进 RetinaNet 的遮挡目标检测算法研究. 计算机工程与应用, 2022, **58**(11): 209–214)
- 69 Luo Z K, Fang Z, Zheng S X, Wang Y B, Fu Y W. NMS-Loss: Learning with non-maximum suppression for crowded pedestrian detection. In: Proceedings of the International Conference on Multimedia Retrieval. Taipei, China: ACM, 2021. 481–485
- 70 Liu W, Hasan I, Liao S C. Center and scale prediction: A box-free approach for pedestrian and face detection. arXiv preprint arXiv: 1904.02948, 2019.
- 71 Yan Y C, Li J P, Qin J, Bai S, Liao S C, Liu L, et al. Anchor-free person search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE, 2021. 7686–7695
- 72 Cao J L, Pang Y W, Anwer R M, Cholakkal H, Xie J, Shah M, et al. PSTR: End-to-end one-step person search with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022. 9448–9457
- 73 Tian Z, Shen C H, Chen H, He T. FCOS: A simple and strong anchor-free object detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, **44**(4): 1922–1933
- 74 Rezatofighi H, Tsoi N, Gwak J Y, Sadeghian A, Reid I, Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 658–666
- 75 Qin Y L, Wen J, Zheng H, Huang X L, Yang J, Song N, et al. Varifocal-Net: A chromosome classification approach using deep convolutional networks. *IEEE Transactions on Medical Imaging*, 2019, **38**(11): 2569–2581
- 76 Qiu H, Ma Y C, Li Z M, Liu S T, Sun J. BorderDet: Border feature for dense object detection. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 549–564
- 77 Cheng Z H, Wu Y X, Xu Z H, Lukasiewicz T, Wang W Y. Segmentation is all you need. arXiv preprint arXiv: 1904.13300, 2019.
- 78 Ge Z, Liu S T, Li Z M, Yoshie O, Sun J. OTA: Optimal transport assignment for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE, 2021. 303–312
- 79 Ge Z, Wang J F, Huang X, Liu S T, Yoshie O. LLA: Loss-aware label assignment for dense pedestrian detection. *Neurocomputing*, 2021, **462**: 272–281
- 80 Zhang X S, Wan F, Liu C, Ji X Y, Ye Q X. Learning to match anchors for visual object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, **44**(6): 3096–3109
- 81 Yao Z Y, Ai J B, Li B X, Zhang C. Efficient DETR: Improving end-to-end object detector with dense prior. arXiv preprint arXiv: 2104.01318, 2021.
- 82 Yu R, Du D W, LaLonde R, Davila D, Funk C, Hoogs A, et al. Cascade transformers for end-to-end person search. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022. 7257–7266
- 83 Zheng A L, Zhang Y, Zhang X Y, Qi X J, Sun J. Progressive end-to-end object detection in crowded scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022. 847–856
- 84 Zhu Y S, Zhao C Y, Wang J Q, Zhao X, Wu Y, Lu H Q. CoupleNet: Coupling global structure with local parts for object detection. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 4146–4154
- 85 Li Y Z, Pang Y W, Shen J B, Cao J L, Shao L. NETNet: Neighbor erasing and transferring network for better single shot object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020. 13346–13355
- 86 Zhong Z Y, Sun L, Huo Q. An anchor-free region proposal network for Faster R-CNN-based text detection approaches. *International Journal on Document Analysis and Recognition (IJ DAR)*, 2019, **22**(3): 315–327
- 87 Samet N, Hicsonmez S, Akbas E. Reducing label noise in anchor-free object detection. In: Proceedings of the 31st British Machine Vision Conference. BMVC, 2020.
- 88 He B K, Zhang Q Y, Tong M, He C. An anchor-free method based on adaptive feature encoding and Gaussian-guided sampling optimization for ship detection in SAR imagery. *Remote Sensing*, 2022, **14**(7): Article No. 1738
- 89 Yang Z, Liu S H, Hu H, Wang L W, Lin S. RepPoints: Point set representation for object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 9656–9665
- 90 Chen Y H, Zhang Z, Cao Y, Wang L W, Lin S, Hu H. RepPoints v2: Verification meets regression for object detection. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020. Article No. 33
- 91 Li W T, Chen Y J, Hu K X, Zhu J K. Oriented RepPoints for aerial object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022. 1819–1828
- 92 Chen S, Li J P, Yao C Q, Hou W B, Qin S, Jin W Y, et al. DuBox: No-prior box objection detection via residual dual scale detectors. arXiv preprint arXiv: 1904.06883, 2019.
- 93 Kong T, Sun F C, Liu H P, Jiang Y N, Li L, Shi J B. FoveaBox: Beyond anchor-based object detection. *IEEE Transactions on Image Processing*, 2020, **29**: 7389–7398

- 94 Fu J M, Sun X, Wang Z R, Fu K. An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, **59**(2): 1331–1344
- 95 Yang C, Huang Z H, Wang N Y. QueryDet: Cascaded sparse query for accelerating high-resolution small object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022. 13658–13667
- 96 Yi J R, Wu P X, Liu B, Huang Q Y, Qu H, Metaxas D. Oriented object detection in aerial images with box boundary-aware vectors. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE, 2021. 2149–2158
- 97 Zhou L, Wei H R, Li H, Zhao W Z, Zhang Y, Zhang Y. Arbitrary-oriented object detection in remote sensing images based on polar coordinates. *IEEE Access*, 2020, **8**: 223373–223384
- 98 Li Z H, Hou B, Wu Z T, Jiao L C, Ren B, Yang C. FCOSR: A simple anchor-free rotated detector for aerial object detection. arXiv preprint arXiv: 2111.10780, 2021.
- 99 Cheng G, Wang J B, Li K, Xie X X, Lang C B, Yao Y Q, et al. Anchor-free oriented proposal generator for object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, **60**: Article No. 5625411
- 100 Huang Z C, Li W, Xia X G, Tao R. A general Gaussian heatmap label assignment for arbitrary-oriented object detection. *IEEE Transactions on Image Processing*, 2022, **31**: 1895–1910
- 101 Wei H R, Zhang Y, Chang Z H, Li H, Wang H Q, Sun X. Oriented objects as pairs of middle lines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, **169**: 268–279
- 102 Wang Y S, Zhang Y, Zhang Y, Zhao L J, Sun X, Guo Z. SARD: Towards scale-aware rotated object detection in aerial imagery. *IEEE Access*, 2019, **7**: 173855–173865
- 103 Zhang F, Wang X Y, Zhou S L, Wang Y Q. DARDet: A dense anchor-free rotated object detector in aerial images. *IEEE Geoscience and Remote Sensing Letters*, 2021, **19**: Article No. 8024305
- 104 Chen Z M, Chen K, Lin W Y, See J, Yu H, Ke Y, et al. PIoU loss: Towards accurate oriented object detection in complex environments. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer, 2020. 195–211
- 105 Lang S, Ventola F, Kersting K. DAFNe: A one-stage anchor-free approach for oriented object detection. arXiv preprint arXiv: 2109.06148, 2021.
- 106 Pan X J, Ren Y Q, Sheng K K, Dong W M, Yuan H L, Guo X W, et al. Dynamic refinement network for oriented and densely packed object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020. 11204–11213
- 107 Zhang F, Wang X Y, Zhou S L, Wang Y Q, Hou Y. Arbitrary-oriented ship detection through center-head point extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, **60**: Article No. 5612414
- 108 Yang X, Yan J C, Ming Q, Wang W T, Zhang X P, Tian Q. Rethinking rotated object detection with Gaussian wasserstein distance loss. In: Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021. 11830–11841
- 109 Yang X, Yang X J, Yang J R, Ming Q, Wang W T, Tian Q, et al. Learning high-precision bounding box for rotated object detection via Kullback-Leibler divergence. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021. 18381–18394
- 110 Qian W, Yang X, Peng S L, Zhang X J, Yan J C. RSDet++: Point-based modulated loss for more accurate rotated object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, **32**(11): 7869–7879
- 111 Yang X, Yang J R, Yan J C, Zhang Y, Zhang T F, Guo Z, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 8231–8240
- 112 Milan A, Leal-Taixé L, Reid I, Roth S, Schindler K. MOT16: A benchmark for multi-object tracking. arXiv preprint arXiv: 1603.00831, 2016.
- 113 Zhang Y F, Wang C Y, Wang X G, Zeng W J, Liu W Y. FairMOT: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, 2021, **129**(11): 3069–3087
- 114 Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv: 1704.04861, 2017.
- 115 Zhang X Y, Zhou X Y, Lin M X, Sun J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 6848–6856
- 116 Wang R J, Li X, Ling C X. Pelee: A real-time object detection system on mobile devices. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: Curran Associates Inc., 2018. 1967–1976



刘小波 中国地质大学(武汉)自动化学院副教授。2008年获得中国地质大学(武汉)计算机学院计算机科学与技术硕士学位。2012年获得中国地质大学(武汉)计算机学院地学信息工程博士学位。主要研究方向为机器学习,演化计算和高光谱遥感图像处理。本文通信作者。

E-mail: xbliu@cug.edu.cn

(LIU Xiao-Bo Associate professor at the School of Automation, China University of Geosciences. He received his master degree in computer software and theory from the School of Computer Science, China University of Geosciences in 2008. He received his Ph.D. degree in geoinformation engineering from the School of Computer Science, China University of Geosciences in 2012. His research interest covers machine learning, evolutionary computation, and hyperspectral remote sensing image processes. Corresponding author of this paper.)



肖肖 中国地质大学(武汉)自动化学院硕士研究生。2020年获得江汉大学物理与信息工程学院学士学位。主要研究方向为遥感图像处理,目标检测。E-mail: xxiao@cug.edu.cn

(XIAO Xiao Master student at the School of Automation, China University of Geosciences. She received her bachelor degree from the School of Physics and Information Engineering, Jiangnan University in 2020. Her research

interest covers remote sensing image processing and object detection.)



王 凌 清华大学自动化系教授. 1995 年获得清华大学自动化系学士学位. 1999 年获得清华大学自动化系控制理论与控制工程专业博士学位. 主要研究方向为智能优化理论、方法与应用, 复杂生产过程建模、优化与调度. E-mail: wangling@tsinghua.edu.cn

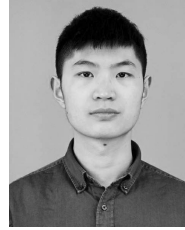
(WANG Ling Professor in the Department of Automation, Tsinghua University. He received his bachelor degree from the Department of Automation, Tsinghua University in 1995. He received his Ph.D. degree in control theory and control engineering from the Department of Automation, Tsinghua University in 1999. His research interest covers intelligent optimization theory, method and application, and complex production process modeling, optimization and scheduling.)



蔡之华 中国地质大学(武汉)计算机学院教授. 1986 年获得武汉大学学士学位. 1992 年获得北京工业大学硕士学位. 2003 年获得中国地质大学(武汉)博士学位. 主要研究方向为数据挖掘, 机器学习和演化计算. E-mail: zhcai@cug.edu.cn

(CAI Zhi-Hua Professor at the School of Computer Science, China University of Geosciences. He received his bachelor degree from Wuhan University in 1986. He

received his master degree from Beijing University of Technology in 1992. He received his Ph.D. degree from China University of Geosciences, in 2003. His research interest covers data mining, machine learning, and evolutionary computation.)



龚 鑫 中国地质大学(武汉)自动化学院硕士研究生. 2020 年获得江汉大学物理与信息工程学院学士学位. 主要研究方向为遥感图像处理, 架构搜索. E-mail: xgong@cug.edu.cn

(GONG Xin Master student at the School of Automation, China University of Geosciences. He received his bachelor degree from the School of Physics and Information Engineering, Jiangnan University in 2020. His research interest covers remote sensing image processing and neural architecture search.)



郑可心 中国地质大学(武汉)自动化学院硕士研究生. 2019 年获得长江大学物理与光电工程学院学士学位. 主要研究方向为遥感图像处理.

E-mail: zhengkexin@cug.edu.cn

(ZHENG Ke-Xin Master student at the School of Automation, China University of Geosciences. He received his bachelor degree from the School of Physics and Optoelectronic Engineering, Yangtze University in 2019. His main research interest is remote sensing image processing.)