基于深度学习的小目标检测基准研究进展

童 康,吴一全*
(南京航空航天大学电子信息工程学院,江苏南京 211106)

摘 要: 小目标检测是计算机视觉中极具挑战性的任务. 它被广泛应用于遥感、交通、国防军事和日常生活等领域. 相比其他视觉任务,小目标检测的研究进展相对缓慢. 制约因素除了学习小目标特征的内在困难,还有小目标检测基准,即小目标检测数据集的稀缺以及建立小目标检测评估指标的挑战. 为了更深入地理解小目标检测,本文首次对基于深度学习的小目标检测基准进行了全新彻底的调查. 系统介绍了现存的35个小目标数据集,并从相对尺度和绝对尺度(目标边界框的宽度或高度、目标边界框宽高的乘积、目标边界框面积的平方根)对小目标的定义进行全面总结. 重点从基于交并比及其变体、基于平均精度及其变体以及其他评估指标这3方面详细探讨了小目标检测评估指标. 此外,从锚框机制、尺度感知与融合、上下文信息、超分辨率技术以及其他改进思路这5个角度对代表性小目标检测算法进行了全面阐述. 与此同时,在6个数据集上对典型评估指标(评估指标+目标定义、评估指标+单目标类别)下的代表性小目标检测算法进行性能的深入分析与比较,并从小目标检测新基准、小目标定义的统一、小目标检测新框架、多模态小目标检测算法、旋转小目标检测以及高精度且实时的小目标检测这6个方面指出未来可能的发展趋势. 希望该综述可以启发相关研究人员,进一步促进小目标检测的发展.

关键词: 小目标检测;深度学习;小目标评估指标;小目标数据集;小目标定义;小目标检测基准
基金项目: 国家自然科学基金(No.61573183)
中图分类号: TP389.1;TP391.41 **文献标识码:** A **文章编号:** 0372-2112(2024)03-1016-25
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20230624

Research Advances on Deep Learning Based Small Object Detection Benchmarks

TONG Kang, WU Yi-quan*
(*College of Electronic and Information Engineering*, *Nanjing University of Aeronautics and Astronautics*, *Nanjing*, *Jiangsu* 211106, *China*)

Abstract: Small object detection is an extremely challenging task in computer vision. It is widely used in remote sensing, intelligent transportation, national defense and military, daily life and other fields. Compared to other visual tasks such as image segmentation, action recognition, object tracking, generic object detection, image classification, video caption and human pose estimation, the research progress of small object detection is relatively slow. We believe that the constraints mainly include two aspects: the intrinsic difficulty of learning small object features and the scarcity of small object detection benchmarks. In particular, the scarcity of small object detection benchmarks can be considered from two aspects: the scarcity of small object detection datasets and the difficulty of establishing evaluation metrics for small object detection. To gain a deeper understanding of small object detection, this article conducts a brand-new and thorough investigation on small object detection benchmarks based on deep learning for the first time. The existing 35 small object detection datasets are introduced from 7 different application scenarios, such as remote sensing images, traffic sign and traffic light detection, pedestrian detection, face detection, synthetic aperture radar images and infrared images, daily life and others. Meanwhile, comprehensively summarize the definition of small objects from both relative scale and absolute scale. For the absolute scale, it mainly includes 3 categories: the width or height of the object bounding box, the product of the width and height of the object bounding box, and the square root of the area of the object bounding box. The focus is on exploring the evaluation metrics of small object detection in detail from 3 aspects: based on IoU (Intersection over Union) and its variants, based on average precision and its variants, and other evaluation metrics. In addition, in-depth analysis and comparison of the perfor-

收稿日期:2023-07-03;修回日期:2023-12-20;责任编辑:宋甲英
*通讯作者:吴一全

mance of some representative small object detection algorithms under typical evaluation metrics are conducted on 6 datasets. These categories of typical evaluation metrics can be further subdivided, including the evaluation metric plus the definition of objects, the evaluation metric plus single object category. More concretely, the evaluation metrics plus the definition of objects can be divided into 4 categories: average precision plus the definition of objects, miss rate plus the definition of objects, DoR-AP-SM (Degree of Reduction in Average Precision between Small objects and Medium objects) and DoR-AP-SL (Degree of Reduction in Average Precision between Small objects and Large objects). For the evaluation metrics plus single object category, it mainly includes 2 types: average precision plus single object category, OLRP (Optimal Localization Recall Precision) plus single object category. These representative small object detection methods mainly include anchor mechanism, scale-aware and fusion, context information, super-resolution technique and other improvement ideas. Finally, we point out the possible trends in the future from 6 aspects: a new benchmark for small object detection, a unified definition of small objects, a new framework for small object detection, multi-modal small object detection algorithms, rotating small object detection, and high precision and real time small object detection. We hope that this paper could provide a timely and comprehensive review of the research progress of small object detection benchmarks based on deep learning, and inspire relevant researchers to further promote the development of this field.

Key words:　small object detection; deep learning; evaluation metric of small objects; small object dataset; the definition of small objects; small object detection benchmark

Foundation Item(s):　National Natural Science Foundation of China (No.61573183)

1　引言

小目标检测是计算机视觉中的一项重要任务. 它指的是定位并识别图像或视频中尺寸较小的目标. 小目标检测在学术界和现实世界应用广泛,如无人机场景分析、智能交通、军事侦察监视与日常生活等. 本文主要关注基于深度学习的小目标检测基准(数据集和评估指标)的研究进展. 为了完整性和更好的可读性,本文也包括了一些其他相关工作.

与其他视觉任务相比,小目标检测发展相对缓慢. 我们认为制约因素有:(1)小目标数据集的稀缺;(2)建立小目标评估指标的挑战;(3)学习小目标特征的内在困难.

深度学习作为一种数据驱动技术,离不开各种数据集. 在深度学习时代的目标检测的整个发展过程中,数据集不仅在模型训练中发挥了关键作用,而且还是评估和验证检测器性能的通用标准. 因此,数据集在一定程度上推动了深度学习在小目标检测中取得成功. 然而,与通用目标检测相比,针对小目标检测的数据集仍然稀缺. 除了小目标数据集,小目标检测的评估指标也至关重要,图 1 为小目标检测算法在小目标数据集上评估示意图. 通过性能好坏与否的反馈,以指导小目标检测算法的进一步改进. 可以说,评估指标是连接小目标检测算法与小目标数据集的桥梁.

小目标数据集和小目标检测评估指标对基于深度学习的小目标检测器至关重要. 此外,小目标检测算法的研究对推动小目标检测的发展也大有裨益. 与大目标、中等目标相比,小目标更难被准确检测. 这是由于小目标检测的几大困难. 首先,小目标分辨率低,特征不足;其次,目标尺度跨度大,多尺度并存;再次,小

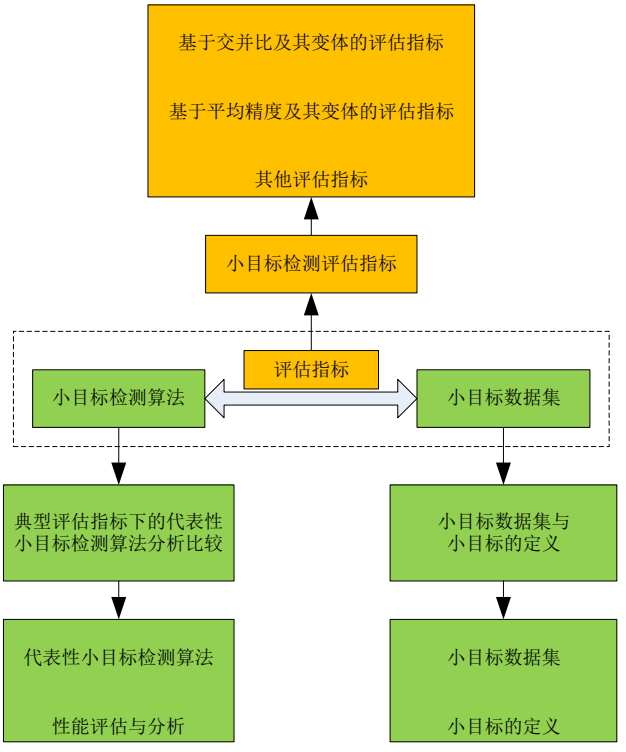


图 1　小目标检测算法在小目标数据集上的评估示意图

目标样本很少,即生成的小尺寸锚框数量不足以匹配小目标以及成功匹配真值的实例数量不足;最后,小目标的类别不平衡. 对于小目标,大多数锚框与真值的重叠度较低或没有重叠,这将带来大量的负例,而锚框与真值匹配达到预定阈值的正例则很少,这导致了两类的极度不平衡,进一步加大了检测小目标的难度. 为了解决这些困难,研究人员提出了一系列的工

作. 文献[1]通过一阶段、二阶段与多阶段的分类方法来阐述小目标检测. 文献[2]从多尺度特征学习、数据增强、训练策略、基于上下文的检测和基于生成对抗网络的检测5个方面综述现有的基于深度学习的小目标检测方法. 在此基础上,其他文献添加了新的分类方法来阐述小目标检测,如文献[3]中的无锚机制、文献[4]中的改善定位精度和感兴趣区域池化层设计、文献[5]中的锚框机制以及基于损失函数的策略. 文献[6]从光学遥感图像、合成孔径雷达图像以及红外图像小目标检测这3个方面系统总结了领域内的代表性小目标检测方法. 文献[7]针对小目标检测的难点给出了相应解决方案:改进小目标特征图、结合小目标上下文信息、增加小目标训练样本、改善小目标前景和背景类不平衡. 此外,文献[8]基于深度学习框架,从多尺度表示、上下文信息、超分辨率、区域建议这4个方面对深度学习时代小目标检测方法进行阐述. 表1给出了国内外学者对基于深度学习的小目标检测算法的总结与对比.

表1 国内外小目标检测综述对比

分类	文献	主要涉及的算法或技术
国内	文献[1]	一阶段算法、二阶段算法、多阶段算法
	文献[3]	数据扩充、多尺度学习、上下文学习、生成对抗学习、无锚机制
	文献[4]	特征增强、上下文信息、锚框设计、数据集差异性处理、感兴趣区域池化层设计、数据扩充、改善定位精度、尺度自适应检测
	文献[6]	数据扩充、特征融合、超分辨率、背景建模、优化训练策略、注意力特征感知、半监督学习、背景噪声抑制
国外	文献[2]	多尺度特征学习、数据增强、训练策略、基于上下文的检测、基于生成对抗网络的检测
	文献[7]	改进小目标特征图、结合小目标上下文信息、改善小目标前景和背景类不平衡、增加小目标训练样本
	文献[8]	多尺度表示、上下文信息、超分辨率、区域建议
	文献[5]	超分辨率技术、基于上下文信息、多尺度表示学习、锚框机制、训练策略、数据增强、基于损失函数的策略
国内外综述总评		优点:这些综述从方法分类的角度对小目标检测展开调查,以推动小目标检测领域发展 缺点:不同分类方法之间存在一定重叠,对小目标数据集和小目标定义探讨不深,缺乏对小目标检测评估指标的深入分析
本文		不同于传统的技术方法分类视角,本文首次从基准(数据集和评估指标)的角度对基于深度学习的小目标检测进行全面综述

上述综述从算法分类的角度对小目标检测展开调查,以推动深度学习技术引领的小目标检测发展. 然而,它们存在一定的局限性. 首先,不同分类方法之间存在一定的重叠;其次,它们对小目标数据集和小目标定义探讨不够深入;最后,它们都缺乏对小目标检测评估指标的具体阐述. 不同于上述综述采用的方法分类视角,本文首次从基准(数据集和评估指标)的角度对基于深度学习的小目标检测展开全面调查. 我们希望该综述可以为视觉社区提供及时的回顾,并启发相关研究者,以进一步促进小目标检测的发展. 本文贡献总结如下:(1)全面调查小目标数据集与小目标的定义;(2)深入阐述小目标检测评估指标;(3)剖析代表性小目标检测算法;(4)评估与分析典型评估指标下算法的性能;(5)指出未来小目标检测的潜在发展方向.

2 小目标数据集与小目标的定义

2.1 小目标数据集

小目标数据集不仅能为数据驱动的深度学习算法提供足够的数据,还可以建立不同检测算法性能比较的平台. 因此,它在小目标检测中起着不可或缺的作用. 本节基于不同的应用场景(如遥感图像、交通标志与交通灯检测、行人检测、人脸检测、合成孔径雷达(Synthetic Aperture Radar,SAR)图像和红外图像、日常生活、其他等)总结了35个小目标数据集. 图2给出了不同应用场景中一些小目标的示例图. 对于每一个应用场景中的小目标数据集,我们按照时间顺序来介绍.

(1)遥感图像

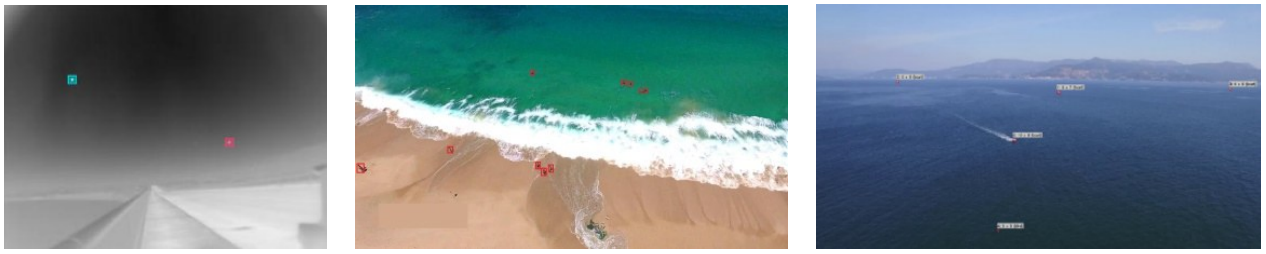
SODA-A[9]. 为了促进航空场景中小目标检测的发展,文献[9]构建了一个大规模的小目标检测数据集SODA-A. 该数据集包含2 510幅高分辨率航空图像以及9个类别的800 203个目标实例. 此外,该数据集中的目标具有不同的位置和方向,并且目标较为密集.

SDOTA[10]. SDOTA是一个自建的数据集. 它包含4个类别的1 508幅图像和227 656个目标实例. 数据集中大多数是小于50个像素的小目标. 此外,对于小型车辆类别,很多实例都小于10个像素.

SDD[10]. SDD来自DOTA数据集和DIOR数据集. 它包括5个类别的12 628幅航空图像和343 961个标注实例. 涵盖的5个类别分别是车辆、飞机、船舶、风车和游泳池.

DIOR[10,11]. DIOR是一个用于评估光学遥感目标检测的大规模数据集. 它包括20个类别的23 463幅图像和190 288个目标实例. 数据集中每个实例都由专家使用水平边界框标注.

(a) 遥感图像检测　(b) 交通场景检测　(c) 人脸检测　(d) 日常生活中的检测



(e) 红外图像检测　(f) 行人检测　(g) 其他场景检测

图2　不同应用场景中一些小目标的示例

AI-TOD[12]. 该数据集专注检测航空图像中的微小目标. 它共有8个类别的700 621个目标实例,涵盖28 036幅尺寸为800×800像素的图像. 该数据集中目标的平均大小约为12.8个像素,远小于现有的航空图像目标检测数据集. 数据集AI-TOD-v2[13]对AI-TOD的标注进行了精确修复,重标注后的实例数为752 745.

UAVDT[14]. UAVDT是一个具有挑战性的大规模无人机检测和跟踪数据集. 由于无人机拍摄视角的高度较高,因此该数据集中的目标通常很小. 特别地,数据集中27.5%的目标像素小于400,相当于一帧图像的0.07%. 它包含80 000幅图像和841 500个目标实例.

DOTA[15]. 它是一个大规模航空图像目标检测数据集. 它包含18个类别和11 268幅图像中的1 793 658个实例. 该数据集中包含不同形状、方向和尺寸的目标. 训练、测试和验证集的比例分别为1/2,1/3,1/6.

Stanford Drone Dataset[16]. 该数据集是一个大型数据集,包含在真实的大学校园中移动和交互的各类目标的图像和视频. 训练和验证集包含69 673幅图像,测试集包含53 224幅图像.

DLR 3k Munich Dataset[17]. 这是一个用于车辆检测的数据集. 它包含20幅超大图像. 10幅用于训练,10幅用于测试. 有14 235辆车通过边界框手动标注.

(2)交通标志与交通灯检测

SODA-D[9]. 文献[9]为了促进自动驾驶场景中小目标检测的发展,构建了一个大规模小目标检测数据集SODA-D. 该数据集包含24 704幅高质量交通图像以及9个类别的277 596个目标实例. 此外,该数据集在位置、天气、时间、拍摄视角和场景方面具有丰富的多样性.

Bosch Small Traffic Lights[18]. 该数据集包含13 427幅分辨率为1 280×720像素的图像,其中大约有24 000个带标注的交通灯. 每个交通灯实例都使用红绿灯的边界框以及当前灯的状态来标注.

Tsinghua-Tencent 100K(TT100K)[19]. TT100K是一个真实的交通标志检测数据集,包含100 000幅图像中的30 000个交通标志实例,涵盖45个常见的中文交通标志. 它的每个实例都使用边界框和实例级掩码进行注释.

German Traffic Sign Detection Benchmark(GTSDB)[20]. GTSDB是一个包含农村、城市和高速公路驾驶等场景的图像数据集,其中大多数交通标志只出现一次. 该数据集中交通标志尺寸的最长边在16到128像素之间.

LISA Traffic Sign Dataset[21]. 它包含6 610个视频帧中的49个US交通标志和7 855个标注. 每个交通标志都标注了大小、类型、位置等. 其尺寸介于6×6与167×168像素.

(3)行人检测

TinyPerson[22]. 文献[22]提出的TinyPerson数据集促进了远距离、大背景下的微小目标检测的发展. 它包含5个类别的72 651个标注实例和1 610幅图像(训练和验证图像分别为794幅和816幅).

TinyCityPersons[22]. 为量化绝对尺寸减小对性能的影响,文献[22]对CityPersons进行4×4下采样,以构建数据集TinyCityPersons.

EuroCity Persons[23]. 文献[23]从12个欧洲国家的31个城市的移动车辆上收集了该数据集的图像. 它含有47 300幅图像,并提供了大量准确和详细的城市场景中行人以及其他骑手的注释(约238 200个实例).

CityPersons[24]. 文献[24]在Cityscapes[25]基础之上构建了CityPersons数据集. 它是在27个城市、3个季节和多种天气条件下,由常见人群组成的行人检测数据集. 在含有30个类别的总共5 000幅图像中,有大约35 000个行人实例标注和大约13 000个忽略区域的标注.

Caltech[26]. Caltech是流行的行人检测数据集,包含250 000帧,共有约350 000个标注的边界框. 训练集和测试集包含的实例数分别为19 200和155 000.

(4)人脸检测

WIDER FACE[27]. WIDER FACE是一个面向精确人脸检测的大规模数据集,其中人脸在尺度、姿态、遮挡、表情、外观和光照等方面存在显著差异. 它由32 203幅图像和393 703个标记的人脸组成. 该数据集基于60个事件类别,每个事件类别随机选择40%,10%和50%的数据分别作为训练、验证和测试集.

PASCAL FACE[28]. PASCAL FACE数据集是PASCAL-VOC的一个子集,在851幅图像中有1 335幅被标记的人脸,这些人脸的外观和姿势变换很大.

(5)SAR图像和红外图像

SAR-ACD[29]. 该数据集由11幅高分3号下1 m分辨率的SAR图像构建而成. 它包含6个民用飞机类别和14个其他飞机类别的4 322个目标实例. SAR-ACD包含不同机场(如上海虹桥机场和北京首都国际机场等)的复杂场景,目标种类丰富且实例大小差异大,这给检测分类带来较大难度.

HRSID[30]. 该数据集中的SAR图像由Sentinel-1,TerraSAR-X和TanDEMX卫星拍摄. HRSID拥有5 604幅800×800像素的SAR图像,涵盖不同海域和港口场景以便于舰船的检测与分割. 在16 951个目标实例中,像素占比小于3.29%的舰船目标达到99%以上.

SIRST[31]. 该数据集是红外小目标数据集,由短波长、中波长以及950 nm波长的红外图像组成. SIRST数据集极具挑战,它含有许多暗淡的目标,并且这些目标大多隐匿在复杂背景中. 每个红外序列中只选择1幅具有代表性的图像以防止训练验证和测试集之间的重叠. SIRST共有427幅图像和480个目标实例.

(6)日常生活

Mini6KClean[32]. 该数据集是MS-COCO train2017的子集,它修复了原始MS-COCO数据集中的标签错误. Mini6KClean包含6 000幅图像、80个目标类别以及55 111个目标实例.

SDOD-MT[33]. 小而密集的商品检测在仓库管理和实体零售中具有很高的应用价值. 文献[33]构建了一个新数据集SDOD-MT以促进特定商品的检测. 该数据集包含13个类别的16 919幅图像. 实例数量高达392 969. 数据集中训练和测试集的划分比约为4∶1.

Small Object Dataset[34]. 文献[34]从MS-COCO数据集中挑选了10个较小的目标类别(如刀、叉等)以构建该小目标数据集. 这些类别来自日常生活的复杂场景,大约包含74 531个目标实例的标注.

SOD[35]. 文献[35]通过SUN[36]和MS-COCO数据集的子集构建SOD. SOD包括来自10个类别的4 925幅图像中的大约8 393个目标实例. 该数据集中目标实例大都小于30 cm.

MS-COCO[37]. MS-COCO数据集中的图片源于自然背景下复杂的日常生活场景. 该数据集是全场景理解(包括目标检测、语义分割、关键点检测等)的流行数据集,已被研究者们广泛使用. 它包含32.8万幅图片和大约250万个标注实例.

PASCAL-VOC[38]. PASCAL-VOC是视觉目标识别和检测的经典数据集. 它主要包含VOC2007和VOC2012这两个常用版本. 这两个版本都包含20个目标类别. VOC2007包含9 963幅图像和24 640个目标实例. 而VOC2012包含33 260幅图像和超过27 450个目标实例.

(7)其他

Small Object Dataset[39]. 该数据集中的图像源于大学不同教室的视频记录. 它包含约2 200幅手动标注的人员头部图像. 训练,验证和测试集图像数分别为550,550和1 100.

USC-GRAD-STDdb[40,41]. 该数据集是从YouTube上检索的一个视频数据集,包含115个视频片段、5个目标类别、超过25 000个带注释的高质量图像. 数据集中小目标的像素范围为16~256像素,标注的小目标实例数超过56 000.

DeepScores[42]. 这是一个大型公共数据集,拥有大约1亿小目标. 它包含30万幅高质量的乐谱图像,涵盖123个不同的符号类别.

Lost and Found[43]. 该数据集专注检测道路上的小危险物和丢失货物. 它一共包含37种不同的障碍物类别和2 104个标注实例.

表2从遥感图像、交通标志与交通灯检测、行人检测、人脸检测、SAR图像和红外图像、日常生活、其他等场景汇总了现存的流行小目标数据集,并给出相应数据集获取的链接,以方便研究人员下载. 遥感图像具有幅面广大、成像高度多变以及场景多样等特点,这无疑会对小尺寸目标的检测带来极大挑战. 交通标志与交通灯检测通常涉及的类别较杂,而且尺寸很小,这些常规交通标志和交通灯的准确检测对智能交通的发展至关重要. 行人检测和人脸检测虽然涉及的类别较为单一,但却容易出现目标遮挡、目标密集等情况,这将进一步加大对小尺寸行人和小尺度人脸检测的难度. SAR图像和红外图像中的目标不仅尺寸很小,而且目

标暗淡,此外更容易受到杂波和背景噪声的影响. 日常生活场景的图像虽然分辨率不是特别高,但目标类别和实例数量较多,这也在一定程度上阻碍了对小目标的准确检测. 鉴于以上分析以及数据集的可获取性,我们认为 SODA-A, AI-TOD-v2, SODA-D, TinyPerson, SIRST, MS-COCO 和 Mini6KClean 这些数据集可以为后续小目标检测算法的公平比较提供帮助. 具体而言,SODA-A 和 AI-TOD-v2 作为遥感领域最新且极具挑战性的两个数据集,可以作为评估遥感小目标检测的基准. SODA-D 作为最新的大规模交通场景数据集,能够为该领域的小目标检测算法提供公平比较的平台. TinyPerson 和 SIRST 可以分别作为微小行人检测以及红外小目标检测的基准. MS-COCO 作为经典的大规模全场景数据集,能引导日常生活中的小目标检测算法进行公平的性能比较. 与数据集 MS-COCO 不同,Mini6KClean 作为一个全新修复的小规模数据集,能够为小目标检测算法的快速验证提供便利.

表 2　现有的流行小目标数据集

应用场景	数据集	发表刊物	年份	类别数	图片数	实例数	公开链接与否	链接
遥感图像	SODA-A[9]	ECCV	2022	9	2 510	800 203	是	https://shaunyuan22.github.io/SODA/
	AI-TOD-v2[13]	ISPRS-JPRS	2022	8	28 036	752 745	是	https://chasel-tsui.github.io/AI-TOD-v2/
	SDOTA[10]	J-STARS	2021	4	1 508	227 656	否	—
	SDD[10]	J-STARS	2021	5	12 628	343 961	否	—
	DIOR[10,11]	ISPRS-JPRS	2020	20	23 463	190 288	是	http://www.escience.cn/people/gongcheng/DIOR.html
	AI-TOD[12]	ICPR	2020	8	28 036	700 621	是	—
	UAVDT[14]	ECCV	2018	3	80 000	841 500	是	https://opendatalab.com/UAVDT
	DOTA[15]	CVPR	2018	15	2 806	188 282	是	https://captain-whu.github.io/DOTA
	SDD[16]	ECCV	2016	6	122 897	20 000	是	https://cvgl.stanford.edu/projects/uav_data/
	DLR[17]	GRSL	2015	7	20	14 235	是	https://pan.baidu.com/s/1xH12NLMZtxPTlyVvzSb_Xg#list/path=%2F
交通标志与交通灯检测	SODA-D[9]	ECCV	2022	9	24 704	277 596	是	https://shaunyuan22.github.io/SODA/
	Bosch[18]	ICRA	2017	19	13 427	24 000	是	https://hci.iwr.uni-heidelberg.de/content/bosch-small-traffic-lights-dataset
	TT100K[19]	CVPR	2016	45	100 000	30 000	是	http://cg.cs.tsinghua.edu.cn/traffic%2Dsign/
	GTSDB[20]	IJCNN	2013	4	900	1 206	是	https://www.kaggle.com/datasets/safabouguezzi/german-traffic-sign-detection-benchmark-gtsdb
	LISA[21]	TITS	2012	49	6 610	7 855	是	http://cvrr.ucsd.edu/lisa/lisa-traffic-sign-dataset.html
行人检测	TinyPerson[22]	WACV	2020	5	1 610	72 651	是	https://github.com/ucas-vg/TinyBenchmark
	EuroCity Persons[23]	TPAMI	2019	7	47 300	238 200	是	https://eurocity-dataset.tudelft.nl/eval/overview/home
	CityPersons[24]	CVPR	2017	30	5 000	48 188	是	https://pan.baidu.com/s/1OOcNvhB6FBolD6YKh74Q4w?pwd=z1zs
	Caltech[26]	TPAMI	2012	1	250 000	350 000	是	http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/

续表

应用场景	数据集	发表刊物	年份	类别数	图片数	实例数	公开链接与否	链接
人脸检测	WIDER FACE[27]	CVPR	2016	60	32 203	393 703	是	http://shuoyang1213.me/WIDERFACE/
	PASCAL FACE[28]	IVC	2014	1	851	1 341	是	https://paperswithcode.com/dataset/pascal-face
SAR图像和红外图像	SAR-ACD[29]	TGRS	2022	20	11	4 322	否	—
	HRSID[30]	ACCESS	2021	1	5 604	16 951	否	—
	SIRST[31]	WACV	2021	1	427	480	是	https://github.com/YimianDai/sirst
日常生活	Mini6KClean[32]	JVCIR	2023	80	6 000	55 111	是	https://github.com/graceveryear/dataquality
	SDOD-MT[33]	ACM MM	2019	13	16 919	392 969	是	https://pan.baidu.com/s/1Yape24qyL8o4E2DOraT1UA
	Small Object Dataset[34]	ICCC	2018	10	—	74 531	否	—
	SOD[35]	ACCV	2016	10	4 925	8 393	是	http://www.merl.com
	MS-COCO[37]	ECCV	2014	91	328 000	2 500 000	是	http://cocodataset.org
	PASCAL-VOC[38]	IJCV	2010	20	9 963	24 640	是	http://host.robots.ox.ac.uk/pascal/VOC/
其他	Small Object Dataset[39]	Access	2021	1	2 200	—	否	—
	USC-GRAD-STDdb[40,41]	EAAI	2020	5	>25 000	56 000	否(需求获取)	—
	DeepScores[42]	ICPR	2018	123	300 000	80 M	是	https://tuggeluk.github.io/deepscores/
	Lost and Found (LF)[43]	IROS	2016	37	2 104	—	是	http://shujujishi.com/dataset/6ed3302d-415f-4c2f-adc1-30c00862b78b.html

2.2 小目标的定义

小目标是指明确目标的尺寸到底有多小,或它们在图像中所占的像素数. 目前,定义小目标主要有两种方法:绝对尺度和相对尺度.

国际光学工程学会对小目标的定义是目标尺寸小于原始图像的0.12%. 文献[44]表明,小目标的边界框应覆盖小于原始图像的1%. 表3总结了相对尺度下小目标的定义. 除了相对尺度,表4总结了绝对尺度下小目标的定义.

小结:本节主要从遥感图像、交通标志与交通灯检测、行人检测、人脸检测、SAR图像和红外图像、日常生活、其他等场景全面汇总了现存的流行小目标数据集. 为方便相关研究人员获取数据集,本节给出了相应数据集的下载链接. 此外,通过对数据集的介绍与分析,本节给出了不同情况下选择特定数据集的一些建议. 具体来说,SODA-A和AI-TOD-v2作为遥感领域极具挑战性的两个全新构建的数据集,可以作为评估遥感小目标检测的基准. 全新的大规模数据集SODA-D,能够为交通场景下的小目标检测算法提供公平比较的平台. 此外,TinyPerson和SIRST可以分别作为微小行人检测以及红外小目标检测的基准. MS-COCO作为经典的大规模全场景数据集,能引

导日常生活中的小目标检测算法进行公平的性能比较. 而 Mini6KClean 作为一个新修复的小规模数据集,能够为小目标检测算法的快速验证提供便利. 除了对小目标检测数据集的阐述,本节还从绝对尺度和相对尺度对小目标的定义进行全面梳理,以帮助研究者更好地理解小目标. 我们期望对小目标检测数据集以及小目标定义的深入探讨,能够助力小目标检测的发展.

表3　相对尺度下小目标的定义

数据集	相对尺度下小目标的具体定义
UAVDT[14]	大多数微小目标仅包含帧的0.005%像素
SDD[16]	所有目标的尺寸都不超过图像尺寸的0.2%,其中大量实例的尺寸介于图像尺寸的0.1%和0.15%之间
TT100K[19]	目标尺寸占图像尺寸20%的目标被视为小目标. 若为方形形状的交通标志,那么当其边界框的宽度小于图像的20%且边界框的高度小于图像的高度时,它就被认为是一个小目标
Small Object Dataset[34]	相比于大目标,该数据集包含厨房场景中的小目标,如刀、叉、瓶子、酒杯、杯子、勺子、碗等
SOD[35]	同一类别中所有目标实例的相对面积中值介于0.08%和0.58%之间的目标被当作小目标
PASCAL-VOC[5,38]	通过计算数据集中同一类别所有目标实例的相对面积中值来定义小目标,即相对面积的中值小于5%的目标(如瓶子、汽车、盆栽、羊和船)被视为小目标

表4　绝对尺度下小目标的定义

定义方式	数据集	绝对尺度下小目标的具体定义
目标边界框的宽度或高度	SDD[10],DIOR[10,11]	将宽度或高度小于50像素的目标视为小目标
	SDOTA[10]	大多是小于50像素的小目标,小型车辆类别含有许多10像素以下的小目标
	DOTA[15]	水平边界框的高度介于10到50像素的目标视为小目标
	EuroCity Persons[23]	高度介于30至60像素之间,且遮挡或截断小于40%的目标视为小目标
	Caltech [26]	高度小于30像素的目标
	GTSDB[20]	交通标志的最长边为16~128像素的视为小尺寸标志
	Bosch[18]	交通灯的宽度为1~85像素的视为小目标
	SDOD-MT[33]	水平边界框的高度为10~50像素的定义为小目标
	WIDER FACE[27] PASCAL FACE[28]	人脸高度为10~50像素的视为小尺度人脸
	Lost and Found[43]	高度低至5 cm的小障碍物视为小目标
目标边界框宽高的乘积(即面积)	SODA-A[9] SODA-D[9]	面积小于2 000像素的实例被定义为小目标,具体为:极其微小目标面积小于256像素;相对微小目标面积为256~576像素;一般微小目标面积为576~1 024像素;小目标面积为1 024~2 000像素
	Mini6KClean[32],MS-COCO[37]	面积小于或等于32×32像素的目标属于小目标
	Small Object Dataset[39]	不超过32×32像素的目标被视为小目标
	USC-GRAD-STDdb[40,41]	占据16×16像素以下区域的目标被定义为小目标
	DeepScores[42]	包含许多非常小的目标,像素面积低至几个像素
	DLR[17]	小尺寸车辆定义为小于30×12像素
	LISA[21]	小交通标志介于6×6像素和167×168像素之间
目标边界框面积的平方根	AI-TOD[12] AI-TOD-v2[13]	2~8像素范围内的目标视为非常微小的目标;8~16像素范围内的目标视为微小目标;16~32像素范围内的目标视为小目标
	TinyPerson[22] TinyCityPersons[22]	粗分2组:微小目标像素为2~20;小目标像素为20~32 微小目标进一步细分为3组:微小目标1的像素为2~8;微小目标2的像素为8~12;微小目标3的像素为12~20
	CityPersons[22,24]	微小目标1的像素范围为8~32;微小目标2的像素范围为32~48;微小目标3的像素范围为48~80;小目标像素范围为80~128

3 小目标检测评估指标

除了第2.2节提到的小目标数据集和小目标的定义之外，小目标检测的评估指标也同样重要．本节从3个方面进行阐述：基于交并比（Intersection over Union，IoU）及其变体的评估指标、基于平均精度（Average Precision，AP）及其变体的评估指标以及其他评估指标．

3.1 基于IoU及其变体的评估指标

交并比IoU反映两个边界框之间的重叠率，计算方式如下：

$$\mathrm{IoU}=\frac{\mathrm{A}\cap\mathrm{B}}{\mathrm{A}\cup\mathrm{B}} \tag{1}$$

其中，A和B分别表示真值边界框和预测边界框，∩和∪分别表示交集和并集操作．

在TinyPerson[22]中，大多数被忽略的区域要比一个人的区域大得多．因此对忽略区域，使用交检比（Intersection over Detection，IoD）替换IoU．IoD指标略微不同于IoU，形式如下：

$$\mathrm{IoD}=\frac{\mathrm{A}\cap\mathrm{B}}{\mathrm{B}} \tag{2}$$

其中，A和B分别表示真值边界框和预测边界框；∩表示交集操作．

广义交并比（Generalized IoU，GIoU），利用两个边界框的最小闭包来解决IoU可能造成的梯度消失问题．计算方式如下：

$$\mathrm{GIoU}=\mathrm{IoU}-\frac{\mathrm{C}\backslash\mathrm{A}\cup\mathrm{B}}{\mathrm{C}} \tag{3}$$

其中，A和B分别表示真值边界框和预测边界框；C为包含边界框A和边界框B的最小框；\表示从左边集合中排除右边集合．GIoU取值范围为$[-1,1]$，不具有IoU的归一化形式．

距离交并比（Distance IoU，DIoU），考虑了目标与锚框之间的距离、重叠率和尺度．计算方式如下：

$$\mathrm{DIoU}=\mathrm{IoU}-\frac{l_1^2}{l_2^2} \tag{4}$$

其中，l_1是A和B中心点之间的欧几里得距离；l_2是覆盖这两个框的最小外框的对角线长度．

完整交并比（Complete IoU，CIoU），通过考虑重叠面积、距离和纵横比这几个因素来提高性能．计算方式如下：

$$\mathrm{CIoU}=\mathrm{IoU}-\frac{l_1^2}{l_2^2}-\alpha V \tag{5}$$

$$V=\frac{4}{\pi^2}\left(\arctan\frac{w_{\mathrm{A}}}{h_{\mathrm{A}}}-\arctan\frac{w_{\mathrm{B}}}{h_{\mathrm{B}}}\right)^2 \tag{6}$$

其中，α是权衡参数；V表示纵横比的一致性；w_{A}和w_{B}分别表示边界框A和边界框B的宽度；h_{A}和h_{B}分别表示边界框A和边界框B的高度．

IoU是评估两个边界框之间位置关系最广泛使用的度量标准．然而只有当两个边界框重叠时，IoU才有效．为了解决该问题，学者们提出了一些评估标准．

GIoU[45]通过添加两个边界框的最小闭包的权重改善了IoU的计算过程，在两个边界框不相交时仍能学习训练．文献[46]引入两个边界框之间的中心点距离以及基于IoU的最小闭包对角线的长度构造了DIoU．DIoU可以直接最小化两个目标边界框的距离，收敛速度更快．基于DIoU，CIoU进一步考虑添加了边界框长宽比这一惩罚项，使对目标检测的评估更加准确．

图3显示了IoU及其变体指标的比较．图3中的A和B分别表示真值边界框和预测边界框，C为包含边界框A和B的最小框．

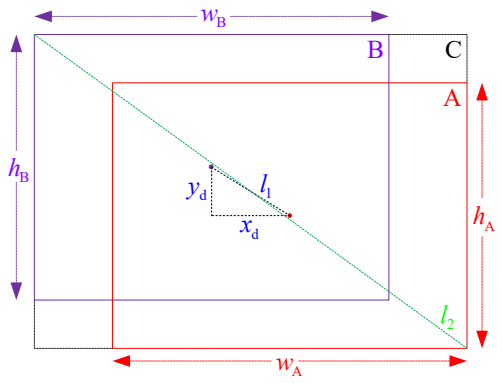


图3 IoU及其变体评估指标的简要比较．

然而，受限于非归一化取值范围的数学性质，GIoU，DIoU和CIoU最初被设计用于非极大值抑制（Non-Maximum Suppression，NMS）和损失函数，很难在无需修改的情况下作为阈值．此外，这些改进仅仅是对IoU施加权重的微调，并没有从本质上解决小目标对IoU极度敏感的问题．图4为IoU对小尺寸目标（网球）和正常尺寸目标（人）的敏感性分析，每个网格表示一个像素，红色边界框表示真值，绿色和蓝色边界框分别表示右向下（对角线）偏差为1个像素和3个像素的预测边界框．从图4可以看出，IoU对不同尺寸目标的敏感性差异很大．对于5×5像素的小目标（网球），微小像素的位置偏差将导致IoU显著下降（从0.47下降到0.09），这将造成不准确的标签分配．而对于30×90像素的正常尺寸目标（人），在相同的位置偏差下，IoU略有变化（从0.92降至0.77）．

鉴于上述分析可知，IoU对小目标来说并不是一个很好的度量标准．小目标边界框中往往会存在一些背景像素，因为大多数真实目标都不是严格的矩形．在这些边界框中，前景像素和背景像素分别集中在边界框的中心和边界上．为了更好地描述边界框中不同像素的权重，文献[13]将边界框建模为二维高斯分布，并通

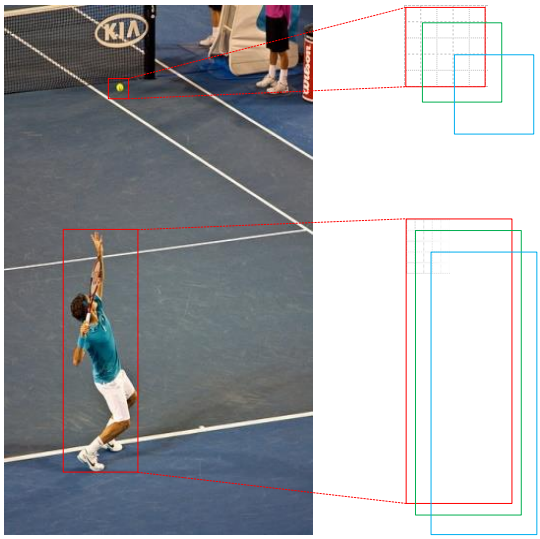


图4　IoU对小尺寸目标(网球)和正常尺寸目标(人)的敏感性分析

过最优传输理论的Wasserstein距离的指数函数来计算相似性. 该评估指标称为归一化Wasserstein距离(Normalized Wasserstein Distance,NWD),计算方式如下:

$$\mathrm{NWD}\left(N_{\mathrm{a}},N_{\mathrm{b}}\right)=\exp\left(-\frac{\sqrt{W_2^2\left(N_{\mathrm{a}},N_{\mathrm{b}}\right)}}{\mathrm{c}}\right) \tag{7}$$

$$W_2^2\left(N_{\mathrm{a}},N_{\mathrm{b}}\right)=\left\|m_1-m_2\right\|_2^2+\left\|\Sigma_1^{1/2}-\Sigma_2^{1/2}\right\|_F^2 \tag{8}$$

其中,c是一个可学习常数;$W_2^2\left(N_{\mathrm{a}},N_{\mathrm{b}}\right)$表示真值框和预测框之间的Wasserstein距离;$\|\cdot\|_F$是Frobenius范数;m是框的中心;Σ_1和Σ_2是它们的协方差.

对于微小目标来说,其边界框的绝对尺寸和相对尺寸都远小于中等目标和大目标. 微小目标的边界框的宽度和高度的重要性远低于中心点,因此一些研究将微小目标看作点. 文献[47]提出了一种名为偏差率(Deviation Rate,DR)的评价指标来衡量预测边界框和真值边界框的位置接近程度,计算方式如下:

$$\mathrm{DR}=\sqrt{{D_x}^2+{D_y}^2} \tag{9}$$

$$D_x=\frac{x_{\mathrm{d}}}{w_{\mathrm{A}}} \tag{10}$$

$$D_y=\frac{y_{\mathrm{d}}}{h_{\mathrm{A}}} \tag{11}$$

其中,x_{d}和y_{d}分别为沿x轴和y轴的偏移距离;D_x和D_y表示偏差程度.

与DR相似,文献[48]提出了另一个评估指标,称为点距离(Dot Distance,DotD). 它被定义为两个边界框中心点之间的归一化欧氏距离,计算方式如下:

$$\mathrm{DotD}=\mathrm{e}^{-\frac{l_1}{Q}} \tag{12}$$

$$Q=\sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N_i}w_{ij}\times h_{ij}}{\sum_{i=1}^{M}N_i}} \tag{13}$$

其中,Q表示数据集中所有目标的平均大小;M表示数据集中的图像数量;N_i表示第i幅图像中标注的边界框数量;w_{ij}和h_{ij}分别表示第i幅图像中第j个边界框的宽和高.

3.2　基于AP及其变体的评估指标

目标检测中,计算一个类别的平均精度(Average Precision,AP)涉及一组具有置信度分数的检测结果和一组边界框. 基于预定义的IoU阈值(如大于0.5),将检测与真值框相匹配. 每个真值框只能匹配一个检测,如果有多个检测满足IoU标准,则匹配具有最高置信度得分的检测. 与真值匹配的检测被计为TP(True Positive),不匹配的检测为FP(False Positive),不匹配的真值视为FN(False Negative). 给定置信度阈值s,具有比s更高的置信度分数的检测被保留,其余检测被丢弃. 通过系统地改变s,计算精度和召回对,以获得精度召回(Precision Recall,PR)曲线. PR曲线下的面积决定了一个类别的AP,并且检测器在所有类别(如类别数为n)上的性能mAP(mean Average Precision)仅通过平均每个类别的AP值来获得. 上述指标的具体公式定义如下:

$$\mathrm{Precision}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}} \tag{14}$$

$$\mathrm{Recall}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}} \tag{15}$$

$$F_1=\frac{2\times\mathrm{Recall}\times\mathrm{Precision}}{\mathrm{Recall}+\mathrm{Precision}} \tag{16}$$

$$\mathrm{AP}=\int_0^1 P(R)\,\mathrm{d}R \tag{17}$$

$$\mathrm{mAP}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{AP}_i \tag{18}$$

其中,N_{TP},N_{FP}和N_{FN}分别表示TP、FP和FN的数量;F_1指标为精度和召回的调和均值.

AP已成为目标检测性能评估的业界标准,很少受到挑战. 但是其也存在一定的缺陷. 首先,AP无法区分非常不同的PR曲线. 召回率低且精度高的检测器、召回率高且精度低的检测器,这两类检测器虽然特性不同,但却存在AP值相同的情况. 其次,AP无法确切衡量定位精度,即不能从它推断边界框检测的紧密程度. 也就是说,AP旨在比较检测器的整体性能,而不比较检测器的最大值.

为了解决AP的局限性,文献[49,50]提出了一个称为定位召回精度(Localization Recall Precision,LRP)

的新评估指标. LRP误差可以等效地表示为TP的平均定位质量、精度误差和召回误差的加权组合,这三个分量作为LRP误差的分量. 计算方式简写为

$$\mathrm{LRP}=\frac{1}{N_{\mathrm{TP}}+N_{\mathrm{FP}}+N_{\mathrm{FN}}}\left(\sum_{i=1}^{N_{\mathrm{TP}}}\varepsilon_{\mathrm{Loc}}(i)+N_{\mathrm{FP}}+N_{\mathrm{FN}}\right) \tag{19}$$

其中,$\varepsilon_{\mathrm{Loc}}(i)$表示第$i$个TP的归一化(即在0和1之间)定位误差.

在LRP基础之上,文献[49,50]又提出了一个称为最优定位召回精度(Optimal Localization Recall Precision, OLRP)的指标来获得更可靠的定位性能评估. OLRP误差定义为

$$\mathrm{OLRP}\left(X,Y_{\mathrm{s}}\right)=\min_{s}\frac{1}{Z}\left(\frac{N_{\mathrm{TP}}}{1-\tau}\mathrm{LRP}_{\mathrm{IoU}}\left(X,Y_{\mathrm{s}}\right)+\left|Y_{\mathrm{s}}\right|\mathrm{LRP}_{\mathrm{FP}}\left(X,Y_{\mathrm{s}}\right)+\left|X\right|\mathrm{LRP}_{\mathrm{FN}}\left(X,Y_{\mathrm{s}}\right)\right) \tag{20}$$

$$Z=N_{\mathrm{TP}}+N_{\mathrm{FP}}+N_{\mathrm{FN}} \tag{21}$$

$$\mathrm{LRP}_{\mathrm{IoU}}\left(X,Y_{\mathrm{s}}\right)=\frac{1}{N_{\mathrm{TP}}}\sum_{i=1}^{N_{\mathrm{TP}}}\left(1-\mathrm{IoU}\left(x_i,y_{x_i}\right)\right) \tag{22}$$

$$\mathrm{LRP}_{\mathrm{FP}}\left(X,Y_{\mathrm{s}}\right)=1-\mathrm{Precision}=\frac{N_{\mathrm{FP}}}{\left|Y_{\mathrm{s}}\right|} \tag{23}$$

$$\mathrm{LRP}_{\mathrm{FN}}\left(X,Y_{\mathrm{s}}\right)=1-\mathrm{Recall}=\frac{N_{\mathrm{FN}}}{\left|X\right|} \tag{24}$$

其中,X和Y_{s}分别表示真值框和检测框;$s\in[0,1]$为置信度得分阈值;$\tau\in[0,1)$为IoU阈值;$\mathrm{IoU}\left(x_i,y_{x_i}\right)$表示$x_i\in X$与其指定的检测$y_{x_i}\in Y_{\mathrm{s}}$之间的IoU. 上述分析可知,LRP和OLRP能够更加精确地预测边界框包围目标的紧密程度,进一步缓解原始评估指标AP对目标定位不准的缺陷.

3.3 其他评估指标

与mAP相似,召回率也有相同的概念,等效的度量指标为mAR(mean Average Recall).

漏检率(Miss Rate, MR). 在行人和肿瘤检测的现实应用中,MR是主要目标. 为了避免产生重大后果(如事故或癌症),则不应该漏检任何小目标. 因此较低的MR总是可取的.

错误率(Error Rate, ER). 它可以被定义为未分类像素的总数除以像素总数. 在该情形下,深度神经网络的训练可以通过最小化误差来进行优化.

降低程度(Degree of Reduction, DoR). DoR表示大或中等目标与小目标间性能差距. 具体来说,DoR-AP-SM (Degree of Reduction in Average Precision between Small objects and Medium objects)表示小目标和中等目标间AP性能差距,而DoR-AP-SL(Degree of Reduction in Average Precision between Small objects and Large objects)表示小目标和大目标间AP的性能差距. DoR-AP-SM和DoR-AP-SL值较大时,则小目标性能越差. 帧每秒(Frame Per Second, FPS)主要衡量算法的运行时间,以评估其对实时检测的适用性. 因此具有更高FPS的算法可以实现对小目标的实时检测.

小结:本节从基于IoU及其变体的评估指标、基于AP及其变体的评估指标以及其他评估指标这3个方面详细阐述了小目标检测的评估指标. 具体总结如表5所示. 在基于IoU及其变体的评估指标中,我们认为IoU作为最广泛的目标检测评估指标,它更适用于中大目标的检测. 对于小目标和微小目标来说,我们建议使用NWD和DotD作为评估指标,因为它们能够更好地缓解IoU对小目标微小移动极度敏感的问题. 此外,在基

表5 小目标检测评估指标总结

评估指标分类	名称	相关描述
基于IoU及其变体	IoU	评估两个边界框之间位置关系最广泛使用的指标;当两个边界框无重叠或相互包容时,IoU会失效;对微小目标位置偏差很敏感
	IoD	仅适用于TinyPerson中的忽略区域
	GIoU, DIoU, CIoU	它们最初被设计用于NMS和损失函数,很难在无需修改的情况下作为阈值;仅仅是对IoU施加权重的微调,未从本质上解决小目标对IoU敏感的问题
	NWD	对位置偏差的平滑性;尺度平衡;能够测量不重叠或相互包容的框之间的相似度
	DR	适合衡量小目标边界框之间的接近程度
	DotD	具有规范化的形式;关注两个边界框中心点之间的位置关系,更适合微小目标
基于AP及其变体	Precision, Recall, F_1	在为分类输出定义的分类任务中,这些是众所周知的度量标准
	AP, mAP	目标检测评估中的业界标准;无法区分非常不同的PR曲线;无法推断边界框检测的紧密程度
	LRP, oLRP	解决了AP的局限性,能获得更可靠的定位性能评估
其他评估指标	mAR, MR, ER	这些指标使用相对较少;MR更多会应用于行人和肿瘤检测中;ER可以帮助优化深度神经网络的训练
	DoR	主要用于衡量小目标与中等或大目标之间的性能差距
	FPS	评估小目标检测算法的实时性

于AP及其变体的评估指标中,mAP作为典型的检测评估指标,它更适用于对数据集中目标类别进行整体的性能评估. 而在其他评估指标中,如FPS,它更多适用于一些实时应用场景的目标检测. 我们期望对小目标检测评估指标的探讨能够启发相关研究者设计出全新的评估指标以更好地评估小目标的检测性能.

4 典型评估指标下的代表性小目标检测算法分析比较

本节在6个数据集(AI-TOD-v2, AI-TOD, TinyPerson, MS-COCO, SODA-D和SODA-A)上对典型评估指标下的一些代表性小目标检测算法进行性能分析比较. 数据集AI-TOD, AI-TOD-v2和SODA-A用于遥感小目标检测. 数据集MS-COCO用于日常生活中的小目标检测. 数据集TinyPerson和SODA-D分别用于微小行人检测和交通场景下的小目标检测. 本文主要从两大分类(评估指标+目标定义、评估指标+单目标类别)来评估一些代表性小目标检测算法. 具体而言,评估指标+目标定义包括:AP+目标定义、MR+目标定义、DoR-AP-SM和DoR-AP-SL. 评估指标+单目标类别包括:AP+单目标类别,OLRP+单目标类别. 表6概述了6个数据集上的典型评估指标与一些代表性小目标检测算法.

表6　不同数据集上典型评估指标与代表性小目标检测算法

数据集	评估指标具体描述	代表性小目标检测算法
AI-TOD	AP+目标定义:AP, AP_{50}, AP_{75}, AP_{vt}, AP_{t}, AP_{s}, AP_{m} AP:在0.5和0.95之间的10个IoU阈值(间隔为0.05)上对AP取平均值 AP_{50}和AP_{75}分别以单个IoU阈值0.5和0.75计算 AP_{vt}:2~8像素的非常微小目标的AP AP_{t}:8~16像素的微小目标的AP AP_{s}:16~32像素的小目标的AP AP_{m}:32~64像素的中等目标的AP	M-CenterNet[12], NWD-RKA[13], DotD[48], RFLA[57], RetinaNet[69], ATSS[56], TridentNet[60]
	AP+单目标类别:飞机、桥、储罐、船舶、游泳池、车辆、人、风力发电机	M-CenterNet[12], TridentNet[60], RepPoints[54], Grid R-CNN[53]
	OLRP+单目标类别:飞机、桥、储罐、船舶、游泳池、车辆、人、风力发电机	
AI-TOD-v2	AP+目标定义:AP, AP_{50}, AP_{75}, AP_{vt}, AP_{t}, AP_{s}, AP_{m} (具体说明同上述AI-TOD数据集)	ADAS-GPM[58], NWD-RKA[13], DotD[48], ATSS[56], RetinaNet[69], SPNet[66], RFLA[57], TridentNet[60], DyHead[70], SM[22], RepPoints[54], Cascade RCNN[51], Grid R-CNN[53]
	AP+单目标类别:飞机、桥、储罐、船舶、游泳池、车辆、人、风力发电机	
TinyPerson	AP+目标定义:AP_{50}(Tiny, Tiny1, Tiny2, Tiny3, Small), AP_{25}(Tiny), AP_{75}(Tiny) AP_{50}(Tiny):2~20像素的微小目标的AP_{50} AP_{50}(Tiny1):2~8像素的微小目标1的AP_{50} AP_{50}(Tiny2):8~12像素的微小目标2的AP_{50} AP_{50}(Tiny3):12~20像素的微小目标3的AP_{50} AP_{50}(Small):20~32像素的小目标的AP_{50} AP_{25}(Tiny):2~20像素的微小目标的AP_{25} AP_{75}(Tiny):2~20像素的微小目标的AP_{75}	SM[22], RetinaNet[69], SSPNet[66], SM+[63], S-α[65], SFRF[64]
	MR+目标定义:MR_{50}(Tiny, Tiny1, Tiny2, Tiny3, Small), MR_{25}(Tiny), MR_{75}(Tiny) MR_{50}(Tiny):2~20像素的微小目标的MR_{50} MR_{50}(Tiny1):2~8像素的微小目标1的MR_{50} MR_{50}(Tiny2):8~12像素的微小目标2的MR_{50} MR_{50}(Tiny3):12~20像素的微小目标3的MR_{50} MR_{50}(Small):20~32像素的小目标的MR_{50} MR_{25}(Tiny):2~20像素的微小目标的MR_{25} MR_{75}(Tiny):2~20像素的微小目标的MR_{75}	SM[22], RetinaNet[69], SSPNet[66], S-α[65]

续表

数据集	评估指标具体描述	代表性小目标检测算法
MS-COCO	AP+目标定义：AP，AP_S，AP_M，AP_L AP：在0.5和0.95之间的10个IoU阈值（间隔为0.05）上对AP取平均值 AP_S：面积小于或等于32^2像素的小目标的AP AP_M面积在32^2和96^2像素之间的中等目标的AP AP_L面积大于96^2像素的大目标的AP	LocalNet[68]，IMFRE[61]，FDL[71]，QueryDet[72]，IENet[67]，CPT-Matching[52]，GDL[73]，RHFNet[62]，IPGNet[59]，MT-GAN[74]，Focal Loss[69]
	DoR-AP-SM，DoR-AP-SL	
SODA-D	AP+目标定义：AP，AP_{50}，AP_{75}，AP_T，AP_{eT}，AP_{rT}，AP_{gT}，AP_S AP描述同上述MS-COCO；AP_{50}和AP_{75}分别以单个IoU阈值0.5和0.75计算 AP_T：面积小于1 024像素的微小目标的AP AP_{eT}：面积小于256像素的极其微小目标的AP AP_{rT}：面积介于256和576像素之间的相对微小目标的AP AP_{gT}：面积介于576和1 024像素之间的一般微小目标的AP AP_S：面积介于1 024和2000像素之间的小目标的AP	ATSS[56]，RetinaNet[69]，RepPoints[54]，FCOS[55]，Sparse RCNN[75]，Deformable-DETR[76]，Cascade RCNN[51]
	AP+单目标类别：人、骑手、自行车、汽车、车辆、交通标志、交通灯、摄像头、警示锥	
SODA-A	AP+目标定义：AP，AP_{50}，AP_{75}，AP_T，AP_{eT}，AP_{rT}，AP_{gT}，AP_S（具体说明同上述SODA-D数据集）	Rotated RetinaNet[69]，S^2A-Net[77]，Gliding Vertex[78]，Oriented RCNN[79]，DODet[80]，RoI Transformer[81]
	AP+单目标类别：飞机、直升机、小型车辆、大型车辆、船、集装箱、储罐、游泳池、风车	

4.1 代表性小目标检测算法

本节从5个方面对代表性小目标检测算法进行阐述，分别为锚框机制、尺度感知与融合、上下文信息、超分辨率技术、以及其他改进思路.

(1)锚框机制

文献[51]提出了一种多阶段检测结构Cascade RCNN，它由一系列不断增加的IoU阈值训练的检测器构成，以解决负样本过拟合以及检测器和测试假设之间的推理时间质量不匹配问题. 为了更好地锚框匹配，文献[52]提出了一种基于锚框中心点平移的匹配策略(Matching strategy based on the Center Point Translation，CPT-Matching)，以在训练阶段选择更多扩展的锚框作为正样本. 该策略有助于准确的小目标检测. 文献[53]引入了一种称为Grid R-CNN的检测框架，它使用网格引导定位机制来精确检测目标. 它通过设计多点监督公式编码更多的线索，以减少对特定点不准确预测的影响；为了充分利用网格中点的相关性，构建两阶段融合策略来融合相邻网格点的特征图，以实现高质量的目标定位. 此外，文献[12]提出了一种基于多个中心点的学习网络(Multiple Center points based learning Network，M-CenterNet)，它使用多个中心点来定位精确的目标中心，以提高航空图像中微小目标检测的定位性能. 文献[54]提出了一种更精细的目标表示方法，称为代表点(Representative Points，RepPoints). 给定用于训练的真值定位和识别的目标，RepPoints通过限制目标空间范围并揭示语义上重要的局部区域的方式自动排列自己. 不同于上述检测器，文献[55]提出了一种全卷积的单阶段(Fully Convolutional One Stage，FCOS)检测器，用于以每个像素预测的方式进行目标检测. 通过消除预定义的一组锚框，FCOS完全避免了所有关于锚框的超参数以及与锚框相关的复杂计算，一定程度上有利于对小目标的检测.

将高质量的锚框分配给小目标边界框并不容易. 常见做法是在选择正样本时降低IoU阈值，以使小目标与更多的锚框匹配，但这会导致训练样本的整体质量恶化. 文献[56]提出了一种自适应训练样本选择(Adaptive Training Sample Selection，ATSS)，通过一组锚框的IoU统计值自动计算每个真值的正负阈值. 然而，IoU度量对小目标的位置偏差非常敏感. 为此，文献[48]提出了一种称为点距离的评估指标来克服传统IoU度量在小目标上的弱点. 在基于锚框的检测器中使用IoU会大大降低标签分配的质量. 为缓解该问题，文献[13]提出了一种基于排序的分配(RanKing-based Assigning，RKA)策略，并将其与新指标NWD相结合，以充分利用新指标在微小目标上的优势，该方法可以显著改善标签分配，并为网络训练提供足够的监督信息. 文献[57]引入了一种基于高斯感受野的标签分配(Receptive Field based Label Assignment，RFLA)策略：利用特

征感受野遵循高斯分布的先验信息;提出一种新的感受野距离来直接测量真值与高斯感受野的相似性,并基于感受野距离,设计了分层标签分配模块,以实现对小目标性能的改善. 不同于上述方法,文献[58]设计了两个新的组件:高斯概率分布的模糊相似性度量(Gaussian Probabilistic distribution-based fuzzy similarity Metric,GPM)以及自适应动态锚框挖掘策略(Adaptive Dynamic Anchor mining Strategy,ADAS). GPM 旨在解决小边界框和预定义锚框之间不准确的相似性测量问题. 而 ADAS 采用动态调整的标签分配策略来解决正负样本间的分布偏差,确保标签分配与图像中目标的分布一致. ADAS-GPM 可以实现对小目标的良好检测.

(2)尺度感知与融合

文献[59]提出了一个图像金字塔引导网络(Image Pyramid Guidance Network,IPGNet),以确保每一层的空间和语义信息都丰富. 它的主要思想是将 IPG 引入主干网络以处理信息不平衡问题,从而缓解小目标特征的消失. 然而,由于内存消耗和推理时间的快速增加,这种方法的计算成本很高. 受人类视觉系统感受野结构的启发,文献[60]引入了一种三叉戟网络(Trident Network,TridentNet),以生成具有统一表示能力的尺度特定的特征图. 该方法构建了一个并行的多分支结构(每个分支共享相同的参数但却具有不同的感受野),可以更好地检测小目标. 不同于上述两种方法,文献[61]提出了一种交互式多尺度特征表示增强(Interactive Multi-scale Feature Representation Enhancement,IMFRE)策略. 其中,多尺度辅助增强网络将输入缩放到对应于预测层的多个尺度,并且只通过轻量级模块提取更详细的特征以增强原始特征;自适应交互模块旨在聚合相邻层的特征,在未改变原有网络结构情形下,实现了对小目标检测的改进. 考虑到在实时嵌入式设备上检测小目标,文献[62]提出了一种递归混合融合金字塔网络(Recursive Hybrid Fusion pyramid Network,RHFNet). RHFNet 中的双向融合模块融合自上而下和自下而上方向的特征图,以产生灵活的特征金字塔,用于小目标检测. 递归连接和重塑模块不仅可以递归地连接来自深层的高级语义特征,还可以从较浅的层重塑空间上更丰富的特征,以防止小目标的消失. 由于采用低成本计算以及特征保留操作,RHFNet 在嵌入式设备上实现高效且准确的检测.

为更好地检测 TinyPerson 中微小的人,文献[22]提出了尺度匹配(Scale Match,SM)策略. 该策略根据不同的目标大小,对图像进行裁剪以缩小不同大小目标之间的差距,避免常规缩放操作中容易丢失小目标信息的情况. 随后,文献[63]全面分析 TinyPerson 数据集中目标的尺度信息,进一步提出了细化尺度匹配方案(SM+). 与 SM 只考虑整个图像不同,SM+关注每个实例,有效地促进了预训练数据集与目标数据集的相似性,大大提高了检测性能. 不同于 SM 和 SM+策略,文献[64]引入了一种基于特征重新缩放和融合的特定特征(Specific characteristics based Feature Rescaling and Fusion,SFRF)方法. 通过设计一种非参数自适应稠密感知算法,它可以自动选择并生成具有微小目标高密度分布的新调整大小的特征图. 为了增强特征表示,SFRF 还使用多对一策略进行特征金字塔网络层的特征融合. 文献[65]认为特征金字塔网络中相邻层之间的自上而下的连接对微小目标的检测产生了负面影响,为此引入了一种称为融合因子的新概念来控制从深层到浅层的信息传输,并探索了如何通过统计策略估计特定数据集的融合因子的有效值. 一系列的实验和分析后发现估计值依赖于分布在每一层的目标的数量. 当使用适当的融合因子配置特征金字塔网络时,模型可以在微小目标上实现显著的性能改进. 大多数现有方法采用特征金字塔网络,通过组合深层的上下文特征来丰富浅层特征. 但由于不同层之间梯度计算的不一致性,特征金字塔网络中的浅层并未充分被利用来检测微小目标. 鉴于此,文献[66]提出了一种由上下文注意力模块、尺度增强模块以及尺度选择模块这三个模块构建的尺度选择金字塔网络,用于微小行人的检测. 通过同时考虑网络中的多尺度特征信息以及上下文信息,该网络能够更准确地检测小目标.

(3)上下文信息

为了同时考虑小尺度人脸和日常生活中的小目标,文献[67]构建了一种新的内外部网络(Internal-External Network,IENet),以利用目标的外观和上下文信息进行鲁棒检测. 双向特征融合模块(Bidirectional Feature Fusion Module,BiFFM)、上下文推理模块(Context Reasoning Module,CRM)和上下文特征增强模块(Context Feature Augmentation Module,CFAM)这三个模块分别用于小目标的特征提取、提议定位和分类. BiFFM 通过将卷积神经网络中更深级别层的语义特征转移到更低级别层,将更低级别层的细节特征转移到更深级别层来捕获目标的内部特征,不仅利用卷积特征的层次结构,还通过上下文关系来促进其预测. CRM 通过上下文推理改善区域提议的质量,上下文推理使用容易检测到的目标帮助理解困难的目标. CFAM 学习通过 CRM 生成的区域提议之间的成对关系,并利用这种关系生成与区域提议相关联的全局特征信息,以进行准确的分类. 文献[68]提出了一种更加关注详细信息建模的单阶段检测器 LocalNet. 该检测器目的是在早期阶段保留更详细的信息,以增强小目标的表示. 此外,文献[68]通过设计局部细节上下文模块,以增强检

测层的语义信息,从而重新引入网络中丢失的细节,并在有限的感受野范围内利用局部上下文. 专注于微小行人检测,文献[66]提出了一个尺度选择金字塔网络(Scale Selection Pyramid Network,SSPNet),它包括上下文注意力模块(Context Attention Module,CAM)、尺度增强模块(Scale Enhancement Module,SEM)和尺度选择模块(Scale Selection Module,SSM). CAM考虑上下文信息生成层次化的注意力热图. SEM在不同层上突出特定尺度的特征,使检测器聚焦于特定尺度的目标,而不是广阔的背景. SSM利用相邻层之间的关系来实现浅层和深层之间合适的特征共享,从而避免不同层之间梯度计算的不一致. 由这3个模块构建的SSPNet能够获得良好的小目标检测性能.

(4)超分辨率技术

文献[74]提出了一种新的多任务生成对抗网络(Multi-Task Generative Adversarial Network,MTGAN)以检测小尺度人脸和常见的小目标. 其生成器是一个超分辨率网络,它将小的模糊图像上采样为精细尺度图像,并恢复详细信息以实现更准确的检测. 其鉴别器是一个多任务网络,在训练过程中,鉴别器中的分类和回归损失被反向传播到生成器中,以使生成器获得更多细节以帮助检测. 针对小目标边界框预测的局限性,文献[73]通过生成判别学习(Generative and Discriminative Learning,GDL)设计了一个检测框架. 该框架通过生成器网络来重建低频到高频的映射,以便于锚框预测. 检测器模块从生成的结果中提取感兴趣区域,并通过感兴趣头来预测目标类别和细化边界框. 为了加快基于特征金字塔的目标检测器的推理速度,文献[72]提出了一种称为QueryDet的查询机制,在低分辨率特征中预测小目标可能存在的位置(查询键),并在这些位置使用高分辨率特征构建稀疏特征图(查询值). 该方案采用的级联形式能够快速准确检测小目标.

(5)其他改进思路

文献[70]提出的DyHead通过将尺度感知、空间感知以及任务感知的注意力统一在一个框架中以更灵活高效的检测目标. 焦距损失(Focal Loss,FL)[69]通过两个权重因子将训练集中在一组稀疏的难样本上,以防止大量的容易负样本在训练过程中损害检测器. 文献[71]提出的反馈驱动损失函数(Feedback-Driven Loss,FDL),利用损失分布信息作为反馈信号,并以更平衡的方式训练模型,从而更有效地监督小目标信息,获得良好的检测性能. 基于查询的检测器Sparse RCNN[75]提供学习到的目标提议的固定稀疏集合,通过动态头执行定位和分类,无需非极大值抑制后处理,直接预测输出. 尽管利用多尺度可变形注意力来减少编码器中的高计算量并允许访问高分辨率特征,但Deformable-DETR[76]性能仍落后其他检测器,这一定程度上表明稀疏查询范式无法充分覆盖小目标. 除了针对水平边界框的处理,Gliding Vertex[78]和DODet[80]为有方向的目标设计了新颖的表示方法,前者学习到对应边的4个滑动偏移量,而后者利用纵横比和面积来表示目标. S^2A-Net[77]设计特征对齐模块和方向检测模块来缓解锚框与卷积特征错位问题,以实现对目标的更好检测. RoI Transformer[81]的成功归因于其强大的候选生成器. 在该生成器中,由旋转的感兴趣区域学习器生成的旋转提议可以保证小目标的高召回率. 由于高效的定向区域提议网络,Oriented RCNN[79]可以生成高质量的提议区,且参数增长很小. 从实验中RoI Transformer和Oriented RCNN的结果也可以看出,高质量的提议对小目标检测至关重要.

4.2 性能评估与分析

本文在表7和表8中报告了12种不同算法在AI-TOD-v2验证集上的检测结果,每个评估指标的最优和次优结果分别以粗体和下划线突出显示. 观察可知,ADAS-GPM[58]和NWD[13]获得了更好的检测性能. 具体来说,DetectoRS w/ADAS-GPM[58]在AP和AP_{50}指标上的性能分别为25.0%和57.5%,稳步优于其他检测算法. 对于特定的AP_t,AP_s和AP_m指标,DetectoRS w/ADAS-GPM[58]同样取得最好结果,性能分别为24.8%,30.8%和41.7%. 此外,DetectoRS /NWD[13]在AP_{75}和AP_{vt}这两个指标上实现了最优性能,检测结果分别达到17.4%和7.6%. 由此可见,专注于小目标评估指标改进的NWD算法能够实现对小目标更准确的定位(高的AP_{75}值)以及获得更高的分类精度(高的AP_{vt}值). 除了报告AI-TOD-v2验证集上的检测结果,本文还在表9和表10中进一步报告了AI-TOD-v2测试集上的检测性能,每个评估指标的最优和次优结果分别以粗体和下划线突出显示.

表9为检测器在AI-TOD-v2测试集上所有类AP的检测结果. 表10则为这些检测器在单个目标类别上的AP结果. 通过观察表9和表10可知,将检测器直接用于AI-TOD-v2时,它们的性能比在MS-COCO等自然场景数据集上的性能差很多,特别是AP_{vt}接近于0(如TridentNet[60]和ATSS[56]). 这也表明航空图像中微小目标极其有限的外观信息和复杂背景给现有检测器带来巨大挑战. 总的来看,检测器NWD-RKA[13]和DotD[48]的性能要明显优于其他检测器. DetectoRS w/NWD-RKA[13]获得了24.7%的AP. 当深入分析具体指标时,我们发现AP_m的性能从最高的39.3%下降到AP_{vt}的9.7%,这也表明当目标尺寸变得非常小时,检测难度陡增.

上述AI-TOD-v2与AI-TOD共享相同的图像但却包含更精确的标注,可以看成是无噪声数据集. 然而现实世界中存在标签噪声问题. 因此AI-TOD可以视为有噪声数据集,这将激励噪声鲁棒的小目标检测方法的设

表7　不同方法在AI-TOD-v2验证集上的性能　　单位:%

年份	方法	主干网络	AP	AP_{50}	AP_{75}	AP_{vt}	AP_{t}	AP_{s}	AP_{m}
2023	Faster R-CNN w/ADAS-GPM[58]	ResNet50	22.3	53.7	13.5	7.1	21.9	27.5	35.1
2023	Cascade R-CNN w/ADAS-GPM[58]	ResNet50	24.2	54.2	17.0	6.0	24.0	29.3	40.0
2023	DetectoRS w/ADAS-GPM[58]	ResNet50	**25.0**	**57.5**	17.3	7.1	**24.8**	**30.8**	**41.7**
2022	SSPNet[66]	ResNet50	13.1	30.3	8.8	0.0	9.7	27.1	37.6
2022	Faster R-CNN /NWD[13]	ResNet50	20.5	51.5	12.4	5.8	20.3	25.4	35.7
2022	DetectoRS /NWD[13]	ResNet50	24.5	56.4	**17.4**	**7.6**	24.3	29.9	41.4
2022	RFLA[57]	ResNet50	20.9	51.8	11.8	6.0	20.9	26.0	35.9
2021	Cascade RCNN[51]	ResNet50	14.4	32.7	10.6	0.0	9.9	28.3	39.9
2021	DyHead[70]	ResNet50	14.0	32.0	9.5	1.7	10.9	22.9	37.9
2020	SM[22]	ResNet50	19.8	46.1	12.7	5.0	19.9	26.3	37.1
2020	ATSS[56]	ResNet50	15.5	36.5	9.6	1.9	12.7	24.6	36.2
2020	RetinaNet[69]	ResNet50	7.4	21.1	3.5	2.5	6.5	13.1	22.9
2019	RepPoints[54]	ResNet50	10.6	27.8	5.6	2.0	10.1	16.0	21.8
2019	Grid R-CNN[53]	ResNet50	14.7	31.7	11.4	0.0	11.5	27.4	38.0
2019	TridentNet[60]	ResNet50	9.7	23.3	6.5	0.0	5.2	20.5	32.7

表8　不同方法在AI-TOD-v2验证集上的类别AP结果　　单位:%

年份	方法	飞机	桥	储罐	船舶	游泳池	车辆	人	风力发电机
2023	Faster R-CNN w/ADAS-GPM[58]	19.0	15.6	38.6	37.0	27.8	27.9	7.0	5.6
2023	Cascade R-CNN w/ADAS-GPM[58]	20.2	**19.0**	**41.1**	40.0	**30.6**	**29.4**	7.9	4.8
2023	DetectoRS w/ADAS-GPM[58]	**25.9**	18.4	40.9	**42.5**	29.6	28.8	8.2	**5.8**
2022	SSPNet[66]	15.5	3.8	22.8	20.1	23.4	15.8	3.5	0.2
2022	Faster R-CNN /NWD[13]	17.8	11.0	38.1	34.7	24.2	27.7	7.1	3.7
2022	DetectoRS /NWD[13]	24.2	17.4	40.5	42.0	29.5	28.6	**8.3**	**5.8**
2022	RFLA[57]	16.1	14.0	38.2	33.8	26.6	27.0	7.3	4.0
2021	Cascade RCNN[51]	13.8	5.5	22.6	24.5	28.2	17.4	3.4	0.0
2021	DyHead[70]	11.4	5.8	25.1	22.9	22.7	19.6	3.8	0.4
2020	SM[22]	13.4	10.2	38.2	33.8	25.3	27.4	7.9	2.3
2020	ATSS[56]	13.7	3.4	30.2	25.6	24.2	22.5	3.9	0.1
2020	RetinaNet[69]	2.4	7.5	13.0	18.8	2.9	12.3	2.3	0.1
2019	RepPoints[54]	0.0	0.6	26.0	25.4	9.9	19.6	3.1	0.0
2019	Grid R-CNN[53]	13.3	10.8	22.7	25.0	26.1	16.2	3.2	0.1
2019	TridentNet[60]	12.2	0.0	17.9	13.5	20.0	11.9	1.9	0.0

表9　检测器在AI-TOD-v2测试集上的性能　　单位:%

年份	方法	主干网络	AP	AP_{50}	AP_{75}	AP_{vt}	AP_{t}	AP_{s}	AP_{m}
2022	RetinaNet w/ NWD-RKA[13]	ResNet50	10.5	28.5	5.2	3.5	11.0	15.2	24.2
2022	Faster R-CNN w/ NWD-RKA[13]	ResNet50	21.4	53.2	12.5	7.7	20.7	26.8	35.2
2022	Faster R-CNN w/ NWD-RKA[13]	ResNet101	20.8	52.4	12.3	8.5	20.4	24.9	35.1
2022	Cascade R-CNN w/ NWD-RKA[13]	ResNet50	22.2	52.5	15.1	7.8	21.8	28.0	37.2
2022	DetectoRS w/ NWD-RKA[13]	ResNet50	**24.7**	**57.4**	**17.1**	**9.7**	**24.2**	**29.8**	**39.3**
2021	DotD[48]	ResNet50	20.4	51.4	12.3	8.5	21.1	24.6	30.4
2020	FR w/ ATSS[56]	ResNet50	12.8	29.6	9.2	0.0	9.2	24.9	36.6
2020	RetinaNet[69]	ResNet50	8.9	24.2	4.6	2.7	8.4	13.1	20.4
2019	TridentNet[60]	ResNet50	10.1	24.5	6.7	0.1	6.3	19.8	31.9

表 10 检测器在 AI-TOD-v2 测试集上的类别 AP 结果 单位:%

年份	方法	飞机	桥	储罐	船舶	游泳池	车辆	人	风力发电机
2022	RetinaNet w/ NWD-RKA[13]	0.1	13.8	14.3	28.5	5.4	16.5	4.7	0.7
2022	Faster R-CNN w/ NWD-RKA[13]	26.7	16.9	35.1	33.6	12.8	26.0	10.3	6.0
2022	Cascade R-CNN w/ NWD-RKA[13]	28.5	17.5	36.9	38.3	13.7	26.6	10.4	5.7
2022	DetectoRS w/ NWD-RKA[13]	**32.0**	**20.2**	**37.8**	**43.4**	**16.6**	**27.3**	**11.6**	**9.1**
2021	DotD[48]	18.7	17.5	34.7	37.0	12.4	25.4	10.3	7.4
2020	FR w/ ATSS[56]	24.7	5.5	20.5	19.8	12.0	15.1	4.8	0.0
2020	RetinaNet[69]	1.3	11.8	14.3	23.6	5.8	11.4	2.3	0.5
2019	TridentNet[60]	19.3	0.1	17.2	16.2	12.4	12.5	3.4	0.0

计. 表 11 报告了 AI-TOD 数据集上不同检测器的 AP 性能,每个评估指标的最优和次优结果分别以粗体和下划线突出显示.

从表 11 可知,TridentNet[60],RetinaNet[69]和 ATSS[56]检测器性能较差,不能很好地适应 AI-TOD. 对于指标 AP_{75},TridentNet[60]和 RetinaNet[69]的性能低于 5%,意味着检测器定位性能较差. 对于 AP_{vt}指标,检测器的最好性能都没达到 10%,这也表明当目标尺寸非常小时检测难度急剧增加,同时无法在现实世界中使用. 总体来看,RFLA[57]的性能远超其他检测器. 具体而言,DetectoRS w/RFLA[57]在 6 个指标中获得最佳性能,包括 AP,AP_{50},AP_{75},AP_t,AP_s和 AP_m. Faster R-CNN w/RFLA[57]在 AP_{vt}指标上获得了最优,性能达到 9.5%. 此外,表 12 和表 13 分别报告了 AI-TOD 数据集上每个目标类别的 AP 和 OLRP 结果,每个目标类别的最优和次优结果分别以粗体和下划线突出显示. 相较于其他检测器,M-CenterNet[12]在桥、储罐、车辆、人和风力发电机 5 个类别上取得了最优结果. 表 14 和表 15 分别显示了 TinyPerson 基准上不同方法在 AP 和 MR 指标上的性能,每个指标的最优和次优结果分别用粗体和下划线突出显示.

表 11 AI-TOD 测试集上不同检测器的性能 单位:%

年份	方法	主干网络	AP	AP_{50}	AP_{75}	AP_{vt}	AP_t	AP_s	AP_m
2022	RetinaNet w/ RFLA[57]	ResNet50	9.1	23.1	5.2	4.1	10.5	10.5	12.3
2022	AutoAssign w/ RFLA[57]	ResNet50	14.2	37.8	6.9	6.4	14.9	17.4	21.8
2022	Faster R-CNN w/ RFLA[57]	ResNet50	21.1	51.6	13.1	**9.5**	21.2	26.1	31.5
2022	Cascade R-CNN w/ RFLA[57]	ResNet50	22.1	51.6	15.6	8.2	22.0	27.3	35.2
2022	DetectoRS w/ RFLA[57]	ResNet50	**24.8**	**55.2**	**18.5**	9.3	**24.8**	**30.3**	**38.2**
2022	Faster R-CNN with NWD-RKA[13]	ResNet50	19.5	49.2	11.7	8.3	19.6	24.5	31.9
2022	Cascade R-CNN with NWD-RKA[13]	ResNet50	20.5	48.7	13.8	8.1	20.6	25.6	34.0
2022	DetectoRS with NWD-RKA[13]	ResNet50	23.4	53.5	16.8	8.7	23.8	28.5	36.0
2021	M-CenterNet[12]	DLA34	14.5	40.7	6.4	6.1	15.0	19.4	20.4
2021	Cascade RPN w/ DotD[48]	ResNet50	13.7	34.0	8.7	6.9	14.8	15.8	24.6
2021	Faster R-CNN w/ DotD[48]	ResNet50	14.9	38.5	9.3	7.2	16.1	17.9	23.7
2021	Cascade R-CNN w/ DotD[48]	ResNet50	16.1	39.2	10.6	8.3	17.6	18.1	22.1
2020	ATSS[56]	ResNet50	12.8	30.6	8.5	1.9	11.6	19.5	29.2
2020	RetinaNet[69]	ResNet50	8.7	22.3	4.8	2.4	8.9	12.2	16.0
2019	TridentNet[60]	ResNet50	7.5	20.9	3.6	1.0	5.8	12.6	14.0

表 12 不同算法在 AI-TOD 测试集上的类别 AP 结果 单位:%

年份	方法	飞机	桥	储罐	船舶	游泳池	车辆	人	风力发电机
2022	FCOS[55]	14.30	4.75	19.77	22.24	0.65	12.51	3.98	0.17
2021	Cascade RCNN[51]	**25.57**	7.47	23.33	23.55	**10.81**	14.09	5.34	0.00
2021	M-CenterNet[12]	18.59	**10.58**	**27.55**	22.27	7.53	**18.60**	**9.17**	**2.03**
2019	TridentNet[60]	9.67	0.77	12.28	17.11	3.20	11.87	3.98	0.94
2019	RepPoints[54]	2.92	2.34	21.37	**26.40**	0.00	15.16	5.39	0.00
2019	Grid R-CNN[53]	22.55	8.59	18.93	21.99	7.28	12.94	4.81	0.35

表13　不同算法在AI-TOD测试集上的类别OLRP结果　　单位:%

年份	方法	飞机	桥	储罐	船舶	游泳池	车辆	人	风力发电机
2022	FCOS[55]	86.46	94.83	82.89	80.97	98.29	88.10	95.62	99.57
2021	Cascade RCNN[51]	**77.62**	92.87	79.07	79.69	**89.75**	86.80	94.55	100.00
2021	M-CenterNet[12]	83.00	**89.23**	**74.50**	79.47	92.06	**81.19**	**90.49**	**96.73**
2019	TridentNet[60]	89.84	98.56	88.00	85.00	97.00	88.66	95.80	98.38
2019	RepPoints[54]	96.18	97.32	80.92	**77.23**	100.00	85.90	94.53	100.00
2019	Grid R-CNN[53]	78.59	91.46	82.74	81.21	92.72	87.68	94.99	99.28

表14　TinyPerson基准上不同方法的AP值　　单位:%

年份	方法	AP_{50}					AP_{25}	AP_{75}
		Tiny	Tiny1	Tiny2	Tiny3	Small	Tiny	Tiny
2022	RetinaNet-SSPNet[66]	54.66	42.72	60.16	61.52	65.24	77.03	6.31
2022	Cascade R-CNN-SSPNet[66]	58.59	45.75	62.03	65.83	**71.80**	78.72	8.24
2022	Faster R-CNN-SSPNet[66]	**59.13**	47.56	62.36	66.15	71.17	**79.47**	**8.62**
2021	Faster RCNN-FPN-MSM+[63]	52.61	34.20	57.60	63.61	67.37	72.54	6.72
2021	Faster RCNN-FPN-RSM+[63]	51.46	33.74	55.32	62.95	66.68	72.38	6.62
2021	RetinaNet with S-α[65]	48.34	28.61	54.59	59.38	61.73	71.18	5.34
2021	Faster RCNN-FPN with S-α[65]	48.39	31.68	52.20	60.01	65.15	69.32	5.78
2021	RetinaNet+SM with S-α[65]	52.56	33.90	58.00	63.72	65.69	73.09	6.64
2021	RetinaNet+MSM with S-α[65]	51.60	33.21	56.88	62.86	64.39	72.60	6.43
2021	Faster RCNN-FPN+SM with S-α[65]	51.76	34.58	55.93	62.31	66.81	72.19	6.81
2021	Faster RCNN-FPN+MSM with S-α[65]	51.41	34.64	55.73	61.95	65.97	72.25	6.69
2021	Faster R-CNN with SFRF[64]	57.24	**51.49**	**64.51**	**67.78**	65.33	78.65	6.42
2020	RetinaNet-SM[22]	48.48	29.01	54.28	59.95	63.01	69.41	5.83
2020	RetinaNet-MSM[22]	49.59	31.63	56.01	60.78	63.38	71.24	6.16
2020	Faster R-CNN-FPN-SM[22]	51.33	33.91	55.16	62.58	66.96	71.55	6.46
2020	Faster R-CNN-FPN-MSM[22]	50.89	33.79	55.55	61.29	65.76	71.28	6.66
2020	RetinaNet[69]	33.53	12.24	38.79	47.38	48.26	61.51	2.28

表15　TinyPerson基准上不同方法的MR值　　单位:%

年份	方法	MR_{50}					MR_{25}	MR_{75}
		Tiny	Tiny1	Tiny2	Tiny3	Small	Tiny	Tiny
2022	RetinaNet-SSPNet[66]	85.30	82.87	76.73	77.20	72.37	69.25	98.63
2022	Cascade R-CNN-SSPNet[66]	83.47	82.80	75.02	73.52	62.06	68.93	98.27
2022	Faster R-CNN-SSPNet[66]	**82.79**	**81.88**	**73.93**	**72.43**	**61.26**	**66.80**	**98.06**
2021	RetinaNet with S-α[65]	87.73	89.51	81.11	79.49	72.82	74.85	98.57
2021	Faster RCNN-FPN with S-α[65]	87.29	87.69	81.76	78.57	70.75	76.58	98.42
2021	RetinaNet+SM with S-α[65]	87.00	87.62	79.47	77.39	69.25	74.72	98.41
2021	RetinaNet+MSM with S-α[65]	87.07	88.34	79.76	77.76	70.35	75.38	98.41
2021	Faster R-CNN-FPN+SM with S-α[65]	85.96	86.57	79.14	77.22	69.35	73.92	98.30
2021	Faster R-CNN-FPN+MSM with S-α[65]	86.18	86.51	79.05	77.08	69.28	73.90	98.24
2020	RetinaNet-SM[22]	88.87	89.83	81.19	80.89	71.82	77.88	98.57
2020	RetinaNet-MSM[22]	88.39	87.80	79.23	79.77	72.18	76.25	98.57
2020	Faster R-CNN-FPN-SM[22]	86.22	87.14	79.60	76.14	68.59	74.16	98.28
2020	Faster R-CNN-FPN-MSM[22]	85.86	86.54	79.20	76.86	68.76	74.33	98.23
2020	RetinaNet[69]	88.31	89.65	81.03	81.08	74.05	76.33	98.76

观察发现，使用SFRF[64]的Faster R-CNN在Tiny1，Tiny2和Tiny3目标上都取得了最佳的AP_{50}结果．此外，检测器（如RetinaNet，Faster R-CNN和Cascade R-CNN）通过使用SSPNet[66]获得性能进一步提升．Cascade R-CNN-SSPNet[66]和Faster R-CNN-SSPNet[66]，这两个检测器分别获得了对小目标和微小目标的最佳检测性能，AP_{50}的结果分别为71.80%和59.13%．此外，在表15中可以发现Faster R-CNN-SSPNet[66]在MR指标上以较大优势超过了其他方法．不难发现，采用多尺度表示学习和基于上下文信息的SSPNet可以很好地检测各种大小的目标，例如微小目标和小目标．

表16报告了MS-COCO test-dev数据集上不同方法的检测结果（最佳和次优结果分别用粗体和下划线标记），涉及的主干网络有VGG16，DarkNet53，ResNet50，ResNet101，ResNeXt101，LN-ResNet，ResNet101-FPN和IPGNet101. IENet[67]在AP，AP_S，AP_M和AP_L这4个指标上获得最佳性能．对于AP_S指标，IENet[67]以5.4%的优势超过了次优检测器．此外使用ResNet101-FPN的GDL[73]检测器在DoR-AP-SM和DoR-AP-SL指标上表现突出，这表明该检测器能够相对更好地平衡对不同尺度目标的检测．通过整体对比指标AP_S，AP_M和AP_L，不难发现，小目标的检测效果最差．DoR指标的结果也再次表明了检测小目标的困难．

表16 MS-COCO test-dev数据集上不同方法的检测结果 单位：%

年份	方法	主干网络	AP	DoR-AP-SM	DoR-AP-SL	AP_S	AP_M	AP_L
2021	LocalNet512[68]	LN-ResNet	34.4	22.0	26.8	20.3	42.3	47.1
2021	Feedback-driven loss[71]	ResNet101-FPN	43.9	21.7	32.6	25.1	46.8	57.7
2021	IMFRE512[61]	ResNet101	37.3	21.1	26.3	22.9	44.0	49.2
2021	QueryDet[72]	ResNet101	43.8	18.9	25.5	27.5	46.4	53.0
2021	QueryDet[72]	ResNeXt101	44.7	<u>18.4</u>	24.0	<u>29.1</u>	47.5	53.1
2021	IENet[67]	ResNet101	**51.2**	19.3	29.1	**34.5**	**53.8**	**63.6**
2020	CPT-Matching[52]	VGG16	30.5	23.6	33.4	11.4	35.0	44.8
2020	GDL[73]	ResNet50	34.8	20.7	<u>12.2</u>	23.5	44.2	35.7
2020	GDL[73]	ResNet101-FPN	39.2	**18.0**	**10.0**	28.8	46.8	38.8
2020	RHFNet416[62]	DarkNet53	35.2	21.6	33.0	15.9	37.5	48.9
2020	RHFNet512[62]	ResNet101	37.7	23.0	31.6	19.9	42.9	51.5
2020	IPG RCNN[59]	IPG-Net101	<u>45.7</u>	22.0	31.7	26.6	<u>48.6</u>	<u>58.3</u>
2020	MTGAN[74]	ResNet101	41.4	19.5	27.9	24.7	44.2	52.6
2020	Focal loss[69]	ResNet101-FPN	39.1	20.9	28.4	21.8	42.7	50.2

表17为SODA-D测试集上不同算法的结果，最优和次优结果分别用粗体和下划线标记．可以看出，Cascade RCNN[51]在所有指标上都获得了最好的性能．得益于级联结构，Cascade RCNN[51]获得了46.9%的AP_S和31.2%的AP_T，表明其在检测小目标和微小目标方面的优势．通过对比AP_S和AP_{eT}，我们发现后者性能比前者低26.5%，这一结果表明当目标尺寸变得非常小时，检测难度急剧增加．

表17 SODA-D测试集上的基线结果（模型主干网络均采用ResNet50） 单位：%

年份	方法	AP	AP_{50}	AP_{75}	AP_T	AP_{eT}	AP_{rT}	AP_{gT}	AP_S
2022	FCOS[55]	28.7	55.1	26.0	23.9	11.9	25.6	32.8	40.9
2021	Cascade RCNN[51]	**35.7**	**64.6**	**33.8**	**31.2**	**20.4**	**32.5**	**39.0**	**46.9**
2021	Sparse RCNN[75]	28.3	55.8	25.5	24.2	14.1	25.5	31.7	39.4
2021	Deformable-DETR[76]	23.4	50.6	18.8	19.2	10.1	20.0	26.5	34.2
2020	RetinaNet[69]	29.2	58.0	25.3	25.0	15.7	26.3	31.8	39.6
2020	ATSS[56]	30.1	59.5	26.3	26.1	<u>17.0</u>	27.4	32.8	40.5
2019	RepPoints[54]	<u>32.9</u>	<u>60.8</u>	<u>30.9</u>	<u>28.0</u>	16.2	<u>29.6</u>	<u>36.8</u>	<u>45.3</u>

表18列出了目标类别的检测结果（最优和次优结果分别用粗体和下划线标记），观察可知，骑手、自行车和交通摄像头的AP明显低于其他类别，我们认为这种现象的根本原因在于类别的不平衡，即与其他类别相比，这些类别包含的实例样本较少．此外大多交通摄像头目标尺寸小于256个像素，这对检测也带来巨大挑战．

表19报告了SODA-A测试集上的检测结果（粗体和下划线分别表示最优和次优结果）．不难发现，RoI Transformer[81]以37.7%的AP，36.0%的AP_T和39.5%的AP_S实现了最高性能．这归因于其强大的候选生成器，保证了小

目标的高召回率. Oriented RCNN[79]通过生成高质量的候选区,在8个指标上获得了第二好的性能. 从上述结果可以看出,高质量的候选区对小目标检测具有重要意义.

此外,该数据集上每个目标类别的AP结果如表20所示(最优和次优结果分别用粗体和下划线标记). 由于实例数量有限,直升机的AP明显低于其他类别. 值得注意的是,最优检测器RoI Transformer[81]在小型车辆类别中得分最低,AP为26.1%.

基于上述分析,我们进一步在表21中呈现了不同数据集上最优方法的分析.

表18　算法在SODA-D测试集上的类别AP(模型主干网络均采用ResNet50)　　单位:%

年份	方法	人	骑手	自行车	汽车	车辆	交通标志	交通灯	交通摄像头	警示锥
2022	FCOS[55]	36.5	16.3	14.9	22.7	48.0	44.1	37.3	11.4	27.3
2021	Cascade RCNN[51]	**45.5**	**21.8**	**21.1**	**27.1**	**54.1**	**52.8**	**44.7**	**16.7**	**37.3**
2021	Sparse RCNN[75]	38.1	15.5	11.6	21.3	46.4	46.9	37.8	10.1	27.2
2021	Deformable-DETR[76]	30.0	12.5	10.6	17.7	37.3	39.7	30.3	9.0	24.0
2020	RetinaNet[69]	36.7	15.1	12.1	20.1	47.9	48.3	39.0	13.8	29.3
2020	ATSS[56]	37.8	18.0	15.3	22.9	48.2	47.1	38.5	12.9	30.2
2019	RepPoints[54]	42.9	19.7	16.1	24.4	52.5	51.2	42.6	14.5	32.6

表19　SODA-A测试集上的基线结果(模型主干网络均采用ResNet50)　　单位:%

年份	方法	AP	AP_{50}	AP_{75}	AP_T	AP_{eT}	AP_{rT}	AP_{gT}	AP_S
2022	S^2A-Net[77]	29.6	72.4	14.0	28.3	15.6	29.1	33.8	29.5
2022	DODet[80]	32.4	69.5	24.4	30.9	17.7	32.0	36.6	32.9
2021	Gliding Vertex[78]	33.2	73.2	24.1	31.7	18.6	32.6	38.6	33.8
2021	Oriented RCNN[79]	36.0	73.2	30.4	34.4	19.5	35.6	41.2	36.7
2020	Rotated RetinaNet[69]	28.1	66.1	17.4	26.8	14.9	28.3	32.8	28.2
2019	RoI Transformer[81]	**37.7**	**75.5**	**32.1**	**36.0**	**20.7**	**37.3**	**43.3**	**39.5**

表20　算法在SODA-A测试集上的类别AP(模型主干网络均采用ResNet50)　　单位:%

年份	方法	飞机	直升机	小型车辆	大型车辆	船	集装箱	储罐	游泳池	风车
2022	S^2A-Net[77]	42.1	19.8	31.2	18.8	36.8	26.1	30.4	37.5	24.2
2022	DODet[80]	50.3	19.4	31.3	19.9	40.3	24.1	43.3	38.1	24.8
2021	Gliding Vertex[78]	48.1	12.4	33.3	26.9	43.4	29.8	44.3	34.8	25.7
2021	Oriented RCNN[79]	52.8	19.8	**34.3**	30.9	45.1	32.0	44.0	40.0	25.3
2020	Rotated RetinaNet[69]	42.3	16.6	30.1	14.1	35.6	23.1	35.8	34.3	20.6
2019	RoI Transformer[81]	**54.2**	**21.7**	26.1	**31.7**	**46.5**	**35.7**	**45.7**	**40.8**	**26.7**

表21　不同数据集上最优方法的分析(T和V分别为测试集和验证集的缩写)

数据集	方法	具体策略
AI-TOD-v2(V)	ADAS-GPM[58]	高斯概率分布的模糊相似性度量与自适应动态锚框挖掘相结合,实现了良好的小目标检测
AI-TOD-v2(T)	NWD-RKA[13]	基于排序的分配策略结合新指标NWD,有效改善标签分配,以提高小目标检测性能
AI-TOD(T)	RFLA[57]	基于高斯感受野的标签分配策略实现对小目标性能的改善
TinyPerson(T)	SSPNet[66]	上下文注意力、尺度增强、尺度选择三个模块的有效结合,助力小目标检测
MS-COCO(T)	IENet[67]	双向特征融合、上下文推理、上下文特征增强联合使用,以提升小目标检测性能
SODA-D(T)	Cascade RCNN[51]	多阶段级联结构有效改善小目标检测性能
SODA-A(T)	RoI Transformer[81]	高质量的区域提议保证小目标的高召回率

具体来说,ADAS-GPM[58]在AI-TOD-v2验证集上,通过结合高斯概率分布的模糊相似性度量和自适应动态锚框挖掘策略,在特定指标(如AP_t,AP_s,AP_m)上取得优异性能. 对AI-TOD-v2测试集而言,NWD-RKA[13]提出的集成指标NWD的排序分配策略有效改善了小目标检测性能,在评估指标AP_{vt},AP_t,AP_s和AP_m上都获得了最好的性能. 此外,在AI-TOD测试集上,RFLA[57]提出的基于高斯感受野的标签分配策略也实现了对目标性能的改善. 对于TinyPerson数据集,SSPNet[66]构建的上下文注意力、尺度增强和尺度选择模块有效提升了目标检测器在特定评估指标上的性能,如AP_{50}(Tiny),AP_{25}(Tiny),AP_{75}(Tiny)和MR_{50}(Small)等. AP_S,AP_M和

AP_L是MS-COCO数据集上的特定评估指标，IENet[67]设计的双向特征融合、上下文推理以及上下文特征增强模块在这三个评估指标上均表现出优势. 就数据集SODA-D和SODA-A而言，Cascade RCNN[51]和RoI Transformer[81]在特定评估指标AP_T, AP_{eT}, AP_{rT}, AP_{gT}和AP_S上均取得优异成绩，这也再次表明多阶段级联结构以及高质量的区域提议对小目标检测性能的改善具有促进作用.

5 总结及未来发展趋势

小目标检测是计算机视觉中极具挑战性的问题. 本文综述了小目标数据集和定义，小目标检测评估指标，以及典型评估指标下代表性小目标检测算法的分析比较. 虽然小目标检测已经取得了可喜的进展，但当前先进的小目标检测方法与一般尺寸目标的检测方法之间仍存在巨大性能差距. 我们认为以下6个方面仍需进一步探究.

(1)小目标检测新基准

基准测试对基于深度学习的小目标检测方法来说至关重要. 最近，文献[9]构建了2个大规模小目标检测数据集SODA-D和SODA-A，分别关注驾驶场景和空中场景. 然而，它们仍然遵循传统基于IoU的评估标准. 在基于IoU的评估指标中，即使目标边界框的微小移动也会导致IoU值的巨大差异，这将直接损害对小目标或微小目标的检测. 因此，建立一个全新评估标准的大规模小目标检测基准是值得期待的研究方向.

(2)小目标定义的统一

当前定义小目标主要有绝对尺度和相对尺度两种方式. 然而，这两种定义方式都存在一定的局限性. 对于相对尺度(如目标与图像的面积比)来说，在两幅不同分辨率的图像中出现的尺寸相似的目标将被分成两组，这不利于性能的评估. 对于绝对尺度(如目标的像素面积)来说，虽然一定程度上有利于性能的定量评估，但当两幅图像的分辨率差距很大时，模型可能无法检测出极大分辨率图像中的固定像素大小的目标. 因此，如何平衡相对尺度和绝对尺度，给出公认且统一的小目标定义，将有助于小目标检测的发展.

(3)小目标检测新框架

现有的深度学习小目标检测方法在当前目标检测范式中设计了巧妙的结构，以帮助检测小目标. 除了基于生成对抗网络以及卷积神经网络这2个主流范式之外，基于Transformer[82,83]的范式也在计算机视觉领域大放光彩. 然而基于Transformer的范式拥有数十亿量级的参数、数周的训练时间以及需要在超大规模数据集上进行预训练. 这极大阻碍了下游检测任务的训练和推理. 因此高效学习的新范式和架构的轻量化也许是未来推动小目标检测的一个突破口.

(4)多模态小目标检测算法

小目标检测算法性能较差的根源在于可用信息很少. 通过不同模态数据之间的互补及其协同应用，有效丰富信息来源，能一定程度缓解模型缺乏判别信息的问题. 因此，开发基于多模态的小目标检测算法是个值得探索的方向.

(5)旋转小目标检测

当前小目标检测研究大多集中于水平边界框标注的目标. 然而，现实场景中的目标不仅随机分散在图像各处，而且形状方向各异. 因此，水平分布排列的目标只是其中的一部分，还有很多带有角度方向的目标. 所以，设计针对有方向的旋转小目标检测方法对实际场景应用具有重要作用.

(6)高精度且实时的小目标检测

精度和速度一定程度上反应了目标检测算法的成功与否. 对于一些特定的应用场景，小目标的高精度且实时检测就十分必要. 例如在智能交通场景中，小交通标志、小交通灯、小障碍物和行人的快速检测，对无人驾驶系统准确并快速做出决策起着举足轻重的作用. 因此，高精度且实时的小目标检测对于智能驾驶的安全应用至关重要，值得进一步研究.

参考文献

[1] 刘颖, 刘红燕, 范九伦, 等. 基于深度学习的小目标检测研究与应用综述[J]. 电子学报, 2020, 48(3): 590-601.
LIU Y, LIU H Y, FAN J L, et al. A survey of research and application of small object detection based on deep learning[J]. Acta Electronica Sinica, 2020, 48(3): 590-601. (in Chinese)

[2] TONG K, WU Y Q, ZHOU F. Recent advances in small object detection based on deep learning: A review[J]. Image and Vision Computing, 2020, 97: 103910.

[3] 高新波, 莫梦竟成, 汪海涛, 等. 小目标检测研究进展[J]. 数据采集与处理, 2021, 36(3): 391-417.
GAO X B, MO M J C, WANG H T, et al. Recent advances in small object detection[J]. Journal of Data Acquisition and Processing, 2021, 36(3): 391-417. (in Chinese)

[4] 李红光, 于若男, 丁文锐. 基于深度学习的小目标检测研究进展[J]. 航空学报, 2021, 42(7): 024691.
LI H G, YU R N, DING W R. Research development of small object traching based on deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(7): 024691. (in Chinese)

[5] TONG K, WU Y Q. Deep learning-based detection from the perspective of small or tiny objects: A survey[J]. Image

and Vision Computing, 2022, 123: 104471.

[6] 袁翔, 程塨, 李戈, 等. 遥感影像小目标检测研究进展[J]. 中国图象图形学报, 2023, 28(6): 1662-1684.
YUAN X, CHENG G, LI G, et al. Progress in small object detection for remote sensing images[J]. Journal of Image and Graphics, 2023, 28(6): 1662-1684. (in Chinese)

[7] LIU Y, SUN P, WERGELES N, et al. A survey and performance evaluation of deep learning methods for small object detection[J]. Expert Systems with Applications, 2021, 172: 114602.

[8] CHEN G, WANG H T, CHEN K, et al. A survey of the four Pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022, 52(2): 936-953.

[9] CHENG G, YUAN X, YAO X W, et al. Towards large-scale small object detection: Survey and benchmarks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13467-13488.

[10] LI Y Y, HUANG Q, PEI X, et al. Cross-layer attention network for small object detection in remote sensing imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 14: 2148-2161.

[11] LI K, WAN G, CHENG G, et al. Object detection in optical remote sensing images: A survey and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.

[12] WANG J W, YANG W, GUO H W, et al. Tiny object detection in aerial images[C]//2020 25th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE, 2021: 3791-3798.

[13] XU C, WANG J W, YANG W, et al. Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 190: 79-93.

[14] DU D W, QI Y K, YU H Y, et al. The unmanned aerial vehicle benchmark: Object detection and tracking[C]//European Conference on Computer Vision. Cham: Springer, 2018: 375-391.

[15] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3974-3983.

[16] ROBICQUET A, SADEGHIAN A, ALAHI A, et al. Learning social etiquette: Human trajectory understanding in crowded scenes[C]//European Conference on Computer Vision. Cham: Springer, 2016: 549-565.

[17] LIU K, MATTYUS G. Fast multiclass vehicle detection on aerial images[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(9): 1938-1942.

[18] BEHRENDT K, NOVAK L, BOTROS R. A deep learning approach to traffic lights: Detection, tracking, and classification[C]//2017 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE, 2017: 1370-1377.

[19] ZHU Z, LIANG D, ZHANG S H, et al. Traffic-sign detection and classification in the wild[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 2110-2118.

[20] HOUBEN S, STALLKAMP J, SALMEN J, et al. Detection of traffic signs in real-world images: The German traffic sign detection benchmark[C]//The 2013 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE, 2013: 1-8.

[21] MOGELMOSE A, TRIVEDI M M, MOESLUND T B. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13(4): 1484-1497.

[22] YU X H, GONG Y Q, JIANG N, et al. Scale match for tiny person detection[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2020: 1246-1254.

[23] BRAUN M, KREBS S, FLOHR F, et al. EuroCity persons: A novel benchmark for person detection in traffic scenes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1844-1861.

[24] ZHANG S S, BENENSON R, SCHIELE B. CityPersons: A diverse dataset for pedestrian detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 4457-4465.

[25] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 3213-3223.

[26] DOLLÁR P, WOJEK C, SCHIELE B, et al. Pedestrian detection: An evaluation of the state of the art[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4): 743-761.

[27] YANG S, LUO P, LOY C C, et al. WIDER FACE: A

face detection benchmark[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 5525-5533.

[28] YAN J J, ZHANG X Z, LEI Z, et al. Face detection by structural models[J]. Image and Vision Computing, 2014, 32(10): 790-799.

[29] SUN X, LV Y X, WANG Z R, et al. SCAN: Scattering characteristics analysis network for few-shot aircraft classification in high-resolution SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5226517.

[30] WEI S J, ZENG X F, QU Q Z, et al. HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation[J]. IEEE Access, 2020, 8: 120234-120254.

[31] DAI Y M, WU Y Q, ZHOU F, et al. Asymmetric contextual modulation for infrared small target detection[C]// 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2021: 949-958.

[32] TONG K, WU Y Q. Rethinking PASCAL-VOC and MS-COCO dataset for small object detection[J]. Journal of Visual Communication and Image Representation, 2023, 93: 103830.

[33] JI Z, KONG Q K, WANG H R, et al. Small and dense commodity object detection with multi-scale receptive field attention[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 1349-1357.

[34] FANG P C, SHI Y J. Small object detection using context information fusion in faster R-CNN[C]//2018 IEEE 4th International Conference on Computer and Communications (ICCC). Piscataway: IEEE, 2018: 1537-1540.

[35] CHEN C Y, LIU M Y, TUZEL O, et al. R-CNN for small object detection[C]//Asian Conference on Computer Vision. Cham: Springer, 2017: 214-230.

[36] XIAO J X, EHINGER K A, HAYS J, et al. SUN database: Exploring a large collection of scene categories[J]. International Journal of Computer Vision, 2016, 119(1): 3-22.

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

[38] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[39] WANG Z Z, XIE K, ZHANG X Y, et al. Small-object detection based on YOLO and dense block via image super-resolution[J]. IEEE Access, 2021, 9: 56416-56429.

[40] BOSQUET B, MUCIENTES M, BREA V M. STDnet: Exploiting high resolution feature maps for small object detection[J]. Engineering Applications of Artificial Intelligence, 2020, 91: 103615.

[41] BOSQUET B, MUCIENTES M, BREA V M. STDnet: A convnet for small target detection[C]//British Machine Vision Conference. Newcastle: BMVA Press, 2018: 253.

[42] TUGGENER L, ELEZI I, SCHMIDHUBER J, et al. DeepScores-A dataset for segmentation, detection and classification of tiny objects[C]//2018 24th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE, 2018: 3704-3709.

[43] PINGGERA P, RAMOS S, GEHRIG S, et al. Lost and Found: Detecting small road hazards for self-driving vehicles[C]//2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE, 2016: 1099-1106.

[44] KRISHNA H, JAWAHAR C V. Improving small object detection[C]//2017 4th IAPR Asian Conference on Pattern Recognition (ACPR). Piscataway: IEEE, 2017: 340-345.

[45] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 658-666.

[46] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[47] HAN B, WANG Y H, YANG Z, et al. Small-scale pedestrian detection based on deep neural network[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(7): 3046-3055.

[48] XU C, WANG J W, YANG W, et al. Dot distance for tiny object detection in aerial images[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2021: 1192-1201.

[49] OKSUZ K, CAM B C, AKBAS E, et al. Localization recall precision (LRP): A new performance metric for object detection[C]//European Conference on Computer Vision. Cham: Springer, 2018: 521-537.

[50] OKSUZ K, CAM B C, KALKAN S, et al. One metric to measure them all: Localisation recall precision (LRP) for evaluating visual detection tasks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44 (12): 9446-9463.

[51] CAI Z W, VASCONCELOS N. Cascade R-CNN: High quality object detection and instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(5): 1483-1498.

[52] DUAN K W, DU D W, QI H G, et al. Detecting small objects using a channel-aware deconvolutional network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(6): 1639-1652.

[53] LU X, LI B Y, YUE Y X, et al. Grid R-CNN[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 7355-7364.

[54] YANG Z, LIU S H, HU H, et al. RepPoints: point set representation for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 9656-9665.

[55] TIAN Z, SHEN C H, CHEN H, et al. FCOS: A simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 1922-1933.

[56] ZHANG S F, CHI C, YAO Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9756-9765.

[57] XU C, WANG J W, YANG W, et al. RFLA: Gaussian receptive field based label assignment for tiny object detection[C]//European Conference on Computer Vision. Cham: Springer, 2022: 526-543.

[58] FU R H, CHEN C C, YAN S, et al. Gaussian similarity-based adaptive dynamic label assignment for tiny object detection[J]. Neurocomputing, 2023, 543: 126285.

[59] LIU Z M, GAO G Y, SUN L, et al. IPG-net: Image pyramid guidance network for small object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2020: 4422-4430.

[60] LI Y H, CHEN Y T, WANG N Y, et al. Scale-aware trident networks for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 6053-6062.

[61] ZHENG Q Y, CHEN Y. Interactive multi-scale feature representation enhancement for small object detection[J]. Image and Vision Computing, 2021, 108: 104128.

[62] CHEN P Y, HSIEH J W, WANG C Y, et al. Recursive hybrid fusion pyramid network for real-time small object detection on embedded devices[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2020: 1612-1621.

[63] JIANG N, YU X H, PENG X K, et al. SM: refined scale match for tiny person detection[C]//2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE, 2021: 1815-1819.

[64] LIU J W, GU Y, HAN S M, et al. Feature rescaling and fusion for tiny object detection[J]. IEEE Access, 2021, 9: 62946-62955.

[65] GONG Y Q, YU X H, DING Y, et al. Effective fusion factor in FPN for tiny object detection[C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2021: 1159-1167.

[66] HONG M B, LI S W, YANG Y C, et al. SSPNet: Scale selection pyramid network for tiny person detection from UAV images[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 8018505.

[67] LENG J X, REN Y H, JIANG W, et al. Realize your surroundings: Exploiting context information for small object detection[J]. Neurocomputing, 2021, 433: 287-299.

[68] YAN Z W, ZHENG H C, LI Y, et al. Detection-oriented backbone trained from near scratch and local feature refinement for small object detection[J]. Neural Processing Letters, 2021, 53(3): 1921-1943.

[69] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.

[70] DAI X Y, CHEN Y P, XIAO B, et al. Dynamic head: Unifying object detection heads with attentions[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 7369-7378.

[71] LIU G, HAN J, RONG W Z. Feedback-driven loss function for small object detection[J]. Image and Vision Computing, 2021, 111(2): 104197.

[72] YANG C, HUANG Z H, WANG N Y. QueryDet: cascaded sparse query for accelerating high-resolution small object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscat-

away: IEEE, 2022: 13658-13667.
[73] GU Y, LI J, WU C T, et al. Small object detection by generative and discriminative learning[C]//2020 25th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE, 2021: 1926-1933.
[74] ZHANG Y Q, BAI Y C, DING M L, et al. Multi-task generative adversarial network for detecting small objects in the wild[J]. International Journal of Computer Vision, 2020, 128(6): 1810-1828.
[75] SUN P Z, ZHANG R F, JIANG Y, et al. Sparse R-CNN: End-to-end object detection with learnable proposals[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 14449-14458.
[76] Zhu X, Su W, Lu L, et al. Deformable DETR: Deformable transformers for end-to-end object detection[C]//9th International Conference on Learning Representations. Virtual: OpenReview, 2021: 1-16.
[77] HAN J M, DING J, LI J, et al. Align deep features for oriented object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2048, 60: 5602511.
[78] XU Y C, FU M T, WANG Q M, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(4): 1452-1459.
[79] XIE X X, CHENG G, WANG J B, et al. Oriented R-CNN for object detection[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 3500-3509.
[80] CHENG G, YAO Y Q, LI S Y, et al. Dual-aligned oriented detector[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5618111.
[81] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 2844-2853.
[82] HAN K, WANG Y H, CHEN H T, et al. A survey on vision transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 87-110.
[83] 张智, 易华挥, 郑锦. 聚焦小目标的航拍图像目标检测算法[J]. 电子学报, 2023, 51(4): 944-955.
ZHANG Z, YI H H, ZHENG J. Focusing on small objects detector in aerial images[J]. Acta Electronica Sinica, 2023, 51(4): 944-955. (in Chinese)

作者简介

童 康 男,1992年2月出生于江苏省南京市. 现为南京航空航天大学电子信息工程学院博士研究生. 主要研究方向为计算机视觉、模式识别、小目标检测.
E-mail: tkangcv@nuaa.edu.cn

吴一全 男,1963年1月出生于江苏省启东市. 现为南京航空航天大学电子信息工程学院教授、博士生导师. 主要研究方向为遥感图像处理与理解、红外小目标检测与识别、视觉检测与图像测量、视频处理与智能分析等.
E-mail: nuaaimage@163.com