

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2025)11-3479-27

论文引用格式: Gao G S, Shang Y Q and Dong Y. 2025. Review of deep learning algorithms for small object detection in optical remote sensing images. Journal of Image and Graphics, 30(11):3479-3505(高广帅, 尚云琦, 董燕. 2025. 光学遥感图像小目标检测深度学习算法综述. 中国图象图形学报, 30(11):3479-3505)[DOI:10.11834/jig.240740]

光学遥感图像小目标检测深度学习算法综述

高广帅^{1*}, 尚云琦¹, 董燕^{1,2}

1. 中原工学院信息与通信工程学院, 郑州 450007; 2. 电子科技大学自动化工程学院, 成都 610000

摘要: 遥感目标检测(remote sensing object detection, RSOD)是遥感领域中备受关注的核心任务之一,其复杂性和基础性使其一直受到广泛关注和研究。在遥感领域,小目标检测的挑战在于其尺寸较小、信息量有限,并且容易受到背景干扰的影响,这使得传统的目标检测方法难以有效应对。因此,遥感小目标检测逐渐成为该领域研究的重要难点和焦点。本文旨在对基于深度学习的遥感小目标检测领域的研究现状和发展趋势进行全面回顾。首先,针对小目标检测问题,本文确定了7大挑战,包括但不限于可用特征少、度量不匹配、图像幅面宽、背景干扰大、分布不均匀、方向不确定和数据集稀缺;接着,针对这些挑战,依次提出相应的解决方案,并对其进行了深入分析和探讨;然后,概述了遥感小目标检测领域广泛使用的基准数据集和评估指标以及潜在的应用场景;最后,对遥感小目标检测的未来方向进行了深入分析与展望,以探讨潜在的研究方向和技术创新,以解决当前挑战和限制,提高遥感小目标检测的性能和适用性。

关键词: 光学遥感图像; 目标检测; 小目标检测(SOD); 深度学习; 多尺度特征融合

Review of deep learning algorithms for small object detection in optical remote sensing images

Gao Guangshuai^{1*}, Shang Yunqi¹, Dong Yan^{1,2}

1. School of Information and Communication Engineering, Zhongyuan University of Technology, Zhengzhou 450007, China;

2. School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 610000, China

Abstract: Remote sensing object detection (RSOD) has emerged as a critical research direction, and it attracts considerable attention due to its complexity and fundamental importance. The primary objective of RSOD is to accurately classify and regress the position of the objects of interest in remote sensing imagery (RSI). According to the image acquisition platform, RSI can be classified into three classes: satellite-, aerial-, and drone-based. The sensors equipped on these platforms capture ground information from different altitudes and views, which fulfills diverse observational needs and applications

收稿日期: 2024-12-12; 修回日期: 2025-03-31; 预印本日期: 2025-04-06

* 通信作者: 高广帅 6911@zut.edu.cn

基金项目: 国家自然科学基金项目(62301623, 62472463, 61873293); 河南省科技攻关项目(232102211002); 河南省重点研发项目(241111220700); 中原工学院青年硕士生导师培育计划(SD202416); 河南省国际科技合作一般项目(252102521049); 河南省高等学校重点科研项目(25A620001); 中原工学院自然科学基金重点项目(K2026ZD021)

Supported by: National Natural Science Foundation of China(62301623, 62472463, 61873293); Henan Province Key Science and Technology Research(232102211002); Henan Key Research and Development Project(241111220700); Incubation Program for Young Master Supervisors of Zhongyuan University of Technology(SD202416); General International Scientific and Technological Cooperation Project of Henan Province(252102521049); Key Scientific Research Project of Colleges and Universities in Henan Province(25A620001); Key Project of the Natural Science Foundation of Zhongyuan University of Technology(K2026ZD021)

ranging from urban planning to environmental monitoring. With the advent of deep learning, especially the convolutional neural networks (CNNs) and Transformers, the detection accuracies obtain substantive improvement relying on the powerful feature representative capability and better adaptivity. However, small object detection (SOD), as a subfield of object detection, limits the performance improvement in recent years. This study focuses on the remote sensing small object detection, of which we summarize seven major challenges.

- 1) Less available features. Small objects typically occupy only a few pixels with absolutely small sizes in RSI, which leads to limited appearance information and detailed spatial feature information. In addition, with multiple convolution and pooling operations in deep CNNs, the information loss becomes severe as the deepening feature layers. In general, the detailed feature information of small objects resides in shallow layers. To address this issue, super-resolution and multi-scale feature fusion are usually employed. Super-resolution methods enhance the details of small objects by reconstructing high-frequency information and refining spatial resolution through advanced upsampling techniques and feature restoration algorithms. This approach rectifies structural distortions and improves the clarity of features. In addition, the incorporation of shallow CNN layers can retain rich spatial details, which facilitates precise localization. The fusion of shallow and deep features enables networks to capture fine-grained details of small objects and global semantic information, which ultimately improves detection accuracy and robustness.
- 2) Misaligned evaluation metrics. The evaluation metric of object detection often relies on intersection over union (IoU), but it can be problematic for small objects, which are highly sensitive to even minor localization errors. Such errors can lead to a significant drop in IoU and an increase in false negatives. To address this issue, several alternative evaluation metrics, such as mean average precision (mAP) at lower IoU thresholds, have been proposed to capture the performance of models on small objects effectively. Meanwhile, the assignment of anchor boxes further complicates small object detection, as a smaller number of anchors match small objects, which makes precise localization more difficult. Therefore, improving loss functions and evaluation metrics specifically tailored for small object detection is essential. Moreover, the sensitivity of IoU during non-maximum suppression can result in the erroneous suppression of small objects, which negatively impacts overall detection performance. Developing robust evaluation strategies that account for the unique characteristics of small objects is critical for enhancing detection accuracy.
- 3) Large image coverage. Remote sensing images typically have high resolution and cover wide areas. As a result, small objects appear smaller, which makes them harder to detect. Directly processing entire images leads to high computational costs and can result in missed detections. Many existing detection frameworks are optimized for medium-sized images and objects. Thus, novel methods that effectively enhance object detection within large-area RSI are necessary. Strategies such as image segmentation or region proposal networks can help mitigate these challenges by allowing models to focus on specific areas of interest rather than processing entire images indiscriminately.
- 4) Complex background interference. Remote sensing images encompass a diverse array of terrains, man-made structures, and natural elements. This condition complicates the separation of small objects from their complex backgrounds. Environmental factors, including atmospheric conditions, terrain variations, lighting, and shadows, can further increase false positive and false negative rates. To tackle these challenges, strategies such as context-aware learning and label noise suppression have been developed. Contextual learning utilizes spatial and semantic relationships to enhance object detection accuracy. Meanwhile, label noise suppression techniques help mitigate the impact of labeling uncertainty on model performance, which improves the robustness.
- 5) Non-uniform data distribution. Objects in remote sensing images are often sparsely distributed, with certain regions densely populated while others are not. This non-uniform distribution necessitates a focus on clustered areas to enhance detection efficiency. However, traditional uniform or random crop processing methods may prove inefficient for detecting dense objects, which leads to wasted computational resources. Developing mechanisms that localize and prioritize these dense areas represents a critical challenge in improving detection performance.
- 6) Uncertain orientation information. Owing to the bird's-eye perspective characteristic of RSI, objects present arbitrary rotated orientations. Standard horizontal bounding boxes often fail to accurately describe these rotated objects. By contrast, oriented bounding boxes facilitate highly precise localization by incorporating rotation angles. However, several issues remain unexplored, including label mismatching, positive-negative sample imbalance, and limited feature supervision for geometrically complex small objects. These challenges can lead to suboptimal learning of key object features, which requires further research and innovation in bounding box representation.
- 7) Scarcity of specialized datasets. Existing

remote sensing datasets predominantly focus on medium to large objects, with relatively few designed specifically for small object detection. Although some datasets such as AI-TOD, TinyPerson, and SODA exist, they are often limited in terms of the number of object categories and instances available. Furthermore, class frequency imbalances exacerbate the issues related to model performance in specific applications. The lack of comprehensive small object datasets hinders the development of robust models tailored to this task, which necessitates the construction of new datasets that address these limitations. This study presents a comprehensive review of deep learning-based techniques for small object detection in RSI. We analyze the primary challenges associated with this task and highlight recent advancements in key areas such as data augmentation, super-resolution techniques, multi-scale feature fusion, anchor box mechanisms, contextual information integration, and label assignment strategies. Furthermore, we provide an overview of commonly used benchmark datasets, evaluation metrics, and various applications related to small object detection. Despite of significant progress that has been made in the field of RSOD, numerous challenges remain, particularly concerning small object detection. Future research directions should focus on developing innovative solutions to address these challenges, which provides insights that will advance the field and enhance the performance of small object detection in remote sensing applications. By addressing these critical areas, researchers can contribute to more effective and accurate remote sensing methodologies, which ultimately enhance our understanding and monitoring of the world.

Key words: optical remote sensing image; object detection; small object detection(SOD); deep learning; multi-scale feature fusion

0 引言

随着传感器技术和地球观测能力的不断提升,遥感卫星航拍技术不断革新,获取大量遥感图像变得更加便捷。作为遥感图像智能解译的核心技术之一,遥感目标检测(remote sensing object detection, RSOD)旨在判断遥感图像(remote sensing images, RSIs)中感兴趣的类别并进行定位(Zhu等,2017)。根据成像平台类型和拍摄高度的不同,RSIs主要包括遥感卫星成像、航拍成像和无人机成像。这些成像方式可以从不同的角度获取地面信息,以满足不同的观测需求。得益于大量遥感数据和深度学习技术强大的特征学习能力,遥感目标检测获得巨大的性能提升。遥感小目标检测作为遥感目标检测的重要分支之一,在无人机场景分析、交通管理和城市安全等多个领域有着不可或缺的理论研究和实际应用意义。

尽管遥感目标检测取得一定的性能进展,但是遥感小目标检测依然存在巨大的性能差距。本文对基于深度学习的遥感小目标检测方法进行全面综述。结合遥感和小目标内在特性,本文梳理了遥感小目标检测的7个主要挑战,具体如图1所示。

1)可用特征少。与其他自然图像相比,遥感图像中的许多目标在高分辨率图像中仅占据极少的像

素,使得特征提取变得异常困难。对于遥感图像而言,小目标的外观信息有限,尤其是尺寸极小、分辨率较低时,通用检测器对小目标的敏感度较差。此外,小目标通常存在于浅层特征层,缺乏高级语义信息,经过多层卷积和池化层后某些小目标特征可能出现信息丢失,这进一步限制了检测的性能。

2)度量不匹配。在回归任务中,模型通常使用交并比(intersection over union, IoU)(Yu等,2016)作为评估指标。然而,由于小目标的尺寸较小,预测框的微小偏移会显著影响IoU值。即使是一个像素的偏差,也可能导致IoU急剧下降,并将预测结果判定为负样本。现有的IoU度量方法在评估小目标检测效果时存在局限性,难以准确反映模型在小目标定位和回归任务中的表现。

3)图像幅面宽。相比自然图像,遥感图像分辨率高、成像范围广,超大的图像尺寸使得小目标的相对尺寸更小。为精准检测幅面宽大的遥感图像中的小目标,直接对全图进行处理,会造成巨大的计算负担。此外,当前流行的目标检测器通常针对尺寸适中的图像和目标,因此,如何使得遥感图像中小尺寸的目标和大尺寸的图像之间能够实现尺度匹配是需要研究的关键问题。

4)背景干扰大。遥感图像覆盖范围广和场景丰富,包括各种地形、建筑、自然景观和人造结构。除了以上因素外,遥感图像还受到诸如大气条件、地形

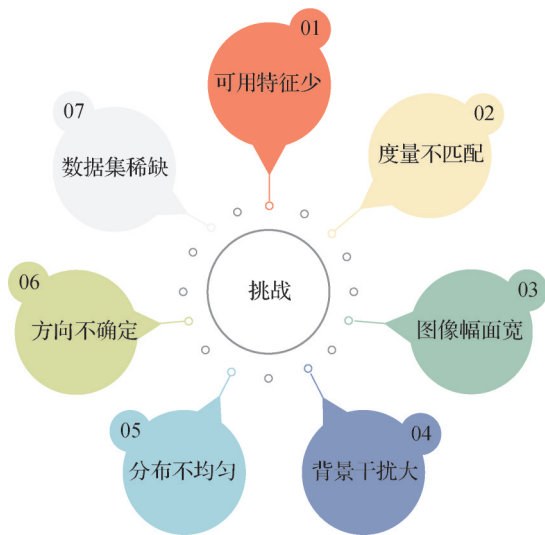


图1 遥感小目标检测的挑战

Fig. 1 Challenges of small object detection in remote sensing

起伏、光照变化和阴影等多重因素的影响,使得背景信息更加复杂且难以预估。地物和环境信息丰富的复杂背景使得小目标往往被掩盖或混淆。例如,建筑、道路和树木等背景干扰因素,可能与目标具有相似的颜色、纹理或形状特征,从而增加了目标检测的难度。这种复杂背景干扰不仅提高误检率,还容易导致小目标的漏检,从而影响小目标的检测精度。

5)分布不均匀。在航拍图像中,目标通常呈现稀疏、不均匀的分布,并且在某些特定区域内往往会高度聚集。为了提升检测效率,一个有效的策略是将检测器的计算资源集中于这些目标密集的区域。找到密集区域的准确方法对于优化目标检测过程至关重要。传统的全图处理方法在处理稀疏目标时效率低下,尤其当大量计算资源被浪费在目标稀少的区域时,检测性能和效率都会受到影响。然而,如何精确定位这些密集区域,确保检测器能够有效地捕捉到潜在的目标簇,是当前面临的主要挑战之一。

6)方向不确定。遥感图像从鸟瞰视角进行拍摄,使得目标可能以任意方向出现。因此,传统的水平边界框(horizontal bounding box, HBB)无法准确地定位这些旋转的目标。相比之下,定向边界框(oriented bounding box, OBB)通过引入旋转角度,可以更精确地描述目标的实际形状,显著减少背景区域的干扰。然而,对于极端几何形状和有限特征的小目标,仍然会存在严重的标签匹配和正负样本不平衡问题,具体包括位置先验和正样本特征的偏差以

及实例数量的不匹配。此外,由于缺乏充足的特征监督,对这些极端形状目标的学习会产生特征学习偏差和样本数量不均衡,从而影响检测的准确性。

7)数据集稀缺。现有遥感目标数据集大多针对大、中型目标,专门用于遥感小目标检测的数据集比较稀缺。尽管已经有一些遥感小目标检测数据集,例如 AI-TOD (tiny object detection in aerial images) (Wang 等, 2020),但是它们类别少、实例少。与此同时,航拍图像中的目标实例频率往往不均衡,不同类别的目标在出现频率上存在显著差异。虽然这些数据集在推动目标检测技术的发展方面做出了重要贡献,但普遍缺乏对遥感图像中小目标及不平衡问题的专门设计。现有的通用数据集未能为小目标检测和实例样本不平衡提供足够的挑战性和多样性。因此,模型在面对这些特定应用场景时,表现通常受到限制。

近年来,一些关于小目标检测的综述陆续发表,这些工作从不同角度探讨了小目标检测领域的最新进展和挑战,如表1所示。Liu 等人(2021)回顾了深度学习在通用目标检测、人脸检测和航空图像检测中的应用,通过性能评估分析了各种方法在小目标检测中的准确性和效率。Tong 和 Wu(2022)综合介绍了30个现有的微小目标检测数据集,并系统综述了基于深度学习的小型和微小目标检测方法,强调了处理微小目标复杂特征时的独特挑战。袁翔等人(2023)系统总结了光学遥感、合成孔径雷达(synthetic aperture radar, SAR)和红外图像的小目标检测任务和方法,并详细介绍了小目标检测的评价指标和数据集。Rekavandi 等人(2023)探索了Transformer 技术在小目标检测中的优势,对比了最新的基于Transformer 的方法,并提供了基准评测。潘晓英等人(2023)从小目标的可视化特征、目标分布和检测环境等方面,系统总结了小目标检测的难点和方法,整理了单阶段小目标检测的技术。Cheng 等人(2023)对小目标检测进行了较为系统的回顾,并提出两个大规模小目标检测数据集(small object detection datasets, SODA),专注于驾驶(SODA-D)和航拍场景(SODA-A)。

与现有的工作相比,本文在以下几个方面进行了创新和扩展:首先,本文针对光学遥感图像中的小目标检测,提出更加全面和深入的挑战与难点分析,覆盖了比以往研究更广泛的挑战类别;其次,本文对

表 1 小目标检测相关的综述总结
Table 1 Summary of reviews on small object detection

序号	论文标题	发表方式	主要内容
1	A survey on deep domain adaptation and tiny object detection challenges, techniques and datasets(Muzam-mul 和 Li, 2021)	arXiv	概述深度领域自适应在小目标检测中的应用,讨论面临的挑战、采用的技术和相关数据集,强调在不同域间传递学习能力以提高检测精度的关键方法。
2	A survey and performance evaluation of deep learning methods for small object detection (Liu 等, 2021)	ESWA	讨论通用目标检测、人脸检测以及航空图像中的目标检测和分割等相关技术,综述深度学习在小目标检测中的各种方法,并通过性能评估分析这些方法在准确性和效率上的表现及其局限性。
3	Deep learning-based detection from the perspective of small or tiny objects: A survey(Tong 和 Wu, 2022)	IVC	全面介绍现有的 30 个微小目标数据集,综述基于深度学习的小和微小目标检测技术,分析现有方法的优势和局限性,并探讨在处理微小目标复杂特征时的独特挑战和未来发展方向。
4	A guide to image and video based small object detec-tion using deep learning: case study of maritime sur-veillance(Rekavandi 等, 2025)	arXiv	总结利用深度学习技术在图像和视频中进行小目标检测技术,重点探讨在海上监视领域的应用案例、方法论和特有的挑战,探索如何提高海洋环境中小目标检测的性能。
5	Transformers in small object detection: a benchmark and survey of state-of-the-art(Rekavandi 等, 2023)	arXiv	探索 Transformer 所提供的性能优势,对比当前最先进的基于 Transformer 的小目标检测方法,提供了基准评测并全面调查了这些方法在提升检测效果方面的最新进展和挑战。
6	小目标检测研究综述(潘晓英 等, 2023)	中国图象图形学报	分别从小目标可视化特征、目标分布情况和检测环境等角度对小目标检测的难点进行分析,系统总结小目标检测方法,并整理了较为成熟的单阶段小目标检测方法。
7	遥感影像小目标检测研究进展(袁翔 等, 2023)	中国图象图形学报	对光学遥感图像小目标检测、SAR 图像小目标检测和红外图像小目标检测任务和方法进行系统性总结,并对小目标检测的评价指标以及可用于小目标检测的各类数据集进行详细介绍。
8	Towards large-scale small object detection: survey and benchmarks(Cheng 等, 2023)	TPAMI	首先对小目标检测进行了彻底的回顾,并构建两个大规模小目标检测数据集 SODA,即 SODA-D 和 SODA-A,分别关注驾驶和航拍场景。

基于深度学习的光学图像下的微小目标检测方法进行了更系统的分析和总结,不仅回顾了遥感小目标检测的现有方法,还综合了国内外最新的研究成果,并对这些方法进行了更为细致的优缺点分析。此外,本文探讨了当前综述中未充分涵盖的前沿技术,并提出具有前瞻性的发展方向,为未来研究提供了切实可行的建议。

1 小目标定义

各种数据集包含的场景多种多样,不同场景下的小目标尺寸各有不同,目前对于小目标的定义还未达成统一的标准。现有小目标的定义主要分为两

类,即基于绝对尺度的定义和基于相对尺度的定义。

1.1 基于绝对尺度的定义

从目标的绝对像素大小角度来考量小目标的定义。目前,最通用的小目标定义源自目标检测领域的主流数据集——MS COCO (Microsoft common objects in context)数据集(Lin 等, 2014),将小目标界定为尺寸低于 32×32 像素的目标。基于绝对尺寸的定义方式,在不同分辨率图像中,小目标的相对视觉效果存在差异,即在低分辨率图像中,由于目标相对占比增大,特征更加突出,检测算法更容易识别和提取相关特征;而在高分辨率图像中,尽管目标尺寸未变,但其相对占比变小,细节特征较为隐蔽,且容易与复杂背景混淆,从而导致检测难度增加。

在 Tiny Person (Yu 等, 2020) 中, 目标的大小定义为目标边界框面积的平方根。目标的绝对尺寸 AS 和相对尺寸 RS 计算为

$$AS(G_{ij}) = \sqrt{w_{ij} \times h_{ij}} \tag{1}$$

$$RS(G_{ij}) = \sqrt{\frac{w_{ij} \times h_{ij}}{W_i \times H_i}} \tag{2}$$

式中, $G_{ij} = (x_{ij}, y_{ij}, w_{ij}, h_{ij})$ 描述数据集中第 i 个图像 I_i 的第 j 个目标的边界框, 其中 (x_{ij}, y_{ij}) 表示左上角点的坐标, w_{ij}, h_{ij} 是边界框的宽度和高度。 W_i, H_i 分别表示 I_i 的宽度和高度。

Cheng 等人 (2023) 采用 $S = W \times H$ 表示目标的面积, 并以此定义“微小”和“小”。采用绝对面积准则, 将面积小于 1 024 像素的实例 (即 $S \leq 1\,024$) 定义为小目标。同时, 面积在 1 024 到 2 000 像素之间的目标被注释为普通目标。并将小目标分为 3 个子集: 极小 (extremely small, eS)、相对小 (relatively small, rS) 和一般小 (generally small, gS), 如表 2 所示。

表 2 SODA 基准中目标的区域子集和相应的区域范围
Table 2 Area subsets and corresponding area ranges of objects in SODA benchmark

区域子集	面积范围/像素
极小	(0, 144]
相对小	(144, 400]
一般小	(400, 1 024]
常规	(1 024, 2 000]

1.2 基于相对尺度的定义

从目标与图像的相对比例角度来考虑定义小目标。据了解 SOD (small object detection) (Chen 等, 2017) 可能为第一个专用于小目标检测的数据集, 阐明了小目标的定义: 同一类别中所有实例的相对面积的中位数 (边界框面积与图像面积之比) 在 0.08%~0.58% 之间即为小目标。采用相对尺寸的定义方式难以全面评估模型对不同尺度目标的检测性能。这种定义方式可能会忽视目标在实际尺度下的重要特征, 从而影响模型的性能评估结果。

2 遥感小目标检测技术

遥感小目标检测是遥感图像目标检测领域的前沿问题, 因其在处理尺寸较小、分辨率较低以及目标

密集的场景中具有显著挑战性而备受研究者关注。

当前基于深度学习的目标检测任务通常包括两个核心任务: 分类和边界框回归。根据是否生成候选区域, 目标检测算法可以分为双阶段算法和单阶段算法两类。双阶段检测算法以 Faster R-CNN (faster region-based convolutional neural network) (Ren 等, 2015) 为代表, 通过由粗到细的检测策略, 首先利用区域建议网络 (region proposal network, RPN) 生成候选框, 然后基于候选框进行下一步的分类和边界框回归。双阶段算法检测精度较高, 但是由于需要预先生成候选区域, 使得检测效率低下。单阶段检测算法如 YOLO (you only look once) 系列、SSD (single shot multibox detection) (Liu 等, 2016) 等, 直接预测分类分数和坐标, 绕过了生成候选框的步骤, 因而在检测效率上具有优势。单阶段检测器具有较高的计算效率, 但在准确性方面往往落后。目前存在的检测方法对于自然场景的目标检测表现良好, 但是对小目标的检测欠佳, 尤其在遥感场景中, 由于背景更复杂、方向不确定等因素, 使得遥感小目标更具挑战性。

本文针对遥感小目标检测, 针对不同挑战总结遥感小目标检测算法, 并从多个方面分析小/微小目标的检测技术, 即数据增强、超分辨率技术、尺度判别、基于上下文的信息、多尺度表示学习、锚框机制、基于损失函数的改进、标签分配以及密集、旋转小目标检测等。在遥感小目标检测领域中, 各种技术方法并非完全独立运作, 而是相互交叉、互为补充, 以实现更优的检测性能, 图 2 列举了近年来遥感小目标检测的关键方法。针对前述 7 大挑战与难点, 本文将分别描述现有代表性方法, 并分析它们的优缺点, 以便更好地理解其可借鉴的思路以及可能的改进方向。

2.1 针对可用特征少的小目标检测算法

无论是绝对尺度还是相对尺度, 小目标相较于大/中尺度目标, 外观信息有限, 难以提取到具有判别力的特征, 导致可用特征少。此外, 卷积和池化操作会进一步压缩小目标的特征, 使其信息在网络中逐渐丢失。这种特征稀疏性导致模型难以提取出具有鉴别力的有效特征, 影响对小目标的精准定位和识别。

2.1.1 基于超分辨率的方法

小目标在图像中占据的像素数量有限, 导致细

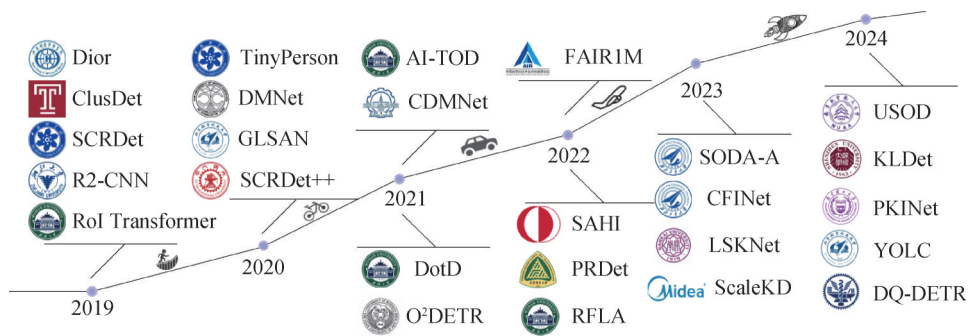


图2 遥感小目标检测算法的年表

Fig. 2 Chronology of small object detection algorithms in remote sensing

节特征难以充分展现,视觉上表现为轮廓模糊、特征不明显,这使得小目标的识别和检测极具挑战性。提高图像分辨率的策略旨在通过特定算法增强小目标的细节信息,修复可能存在的结构畸变,从而提升小目标的可辨识度和视觉清晰度。这个过程不仅仅是简单地提高图像的分辨率,而是通过有针对性的细节恢复和增强,帮助小目标在高分辨率图像中展现更多的细节,提高检测效果。

一些方法采用超分辨率策略作为检测通道的预处理步骤,以放大输入图像的分辨率。Ferdous 等人(2019)结合了 SSD 检测器上游的超分辨率网络,用于卫星图像上的车辆检测,仅对第一个 SSD 层进行了轻微修改,处理超分辨率图像的 SSD 性能明显优化。Rabbi 等人(2020)强调了边缘信息对于微小目标检测的重要性,并提出一种边缘增强的超分辨率生成对抗网络(generative adversarial network, GAN)来生成具有详细边缘信息的高分辨率 RSIs。Wang 等人(2022)提出一种具有残差特征聚合和辅助 YOLOv5 检测器的新型多类循环超分辨率生成对抗网络,用于对基于超分辨率的图像目标检测进行基准测试,并且提出一个大规模的、公开的基准遥感超分辨率目标检测数据集 RSSOD(remote sensing super-resolution object detection)。

超分辨率技术通过提升图像的分辨率改善目标的边缘和纹理细节,显著增强了遥感图像中小目标的可见性和特征。但是,超分辨率方法在重建图像时可能引入伪影或人工增强的细节,这些并不存在的细节可能混淆检测模型,使其误识别为真实特征,导致检测性能的不稳定和潜在的过拟合风险。

2.1.2 基于多尺度特征融合的方法

卷积神经网络(convolutional neural network,

CNN)通常采用多层次结构,不同的层次具有不同的特征。浅层特征通常包含细粒度的特征(例如目标的边缘、纹理和其他细节),并能够提供更加精确的空间位置信息,更适合小目标的检测和定位。由于小目标尺寸较小、对比度较低,它们往往依赖浅层特征进行准确的定位;而高层特征则更多包含语义信息和上下文特征。因此,高层和浅层特征的整合是一个关键步骤,可以在保留细节的同时结合语义信息,通过这种特征融合,网络能够同时捕捉小目标的精细细节和大目标的全局信息,实现更全面的目标检测效果。

为了结合来自不同层的信息并生成多尺度表示,一些研究人员引入了多层特征融合方法,将多个层的特征集成到单个特征图上,并在此重建的特征图上执行检测(Ding 等,2018;Liu 等,2018;Zheng 等,2020a)。Liang 等人(2020)在 SSD 中引入了一个额外的缩放分支,该分支包括反卷积模块和平均池化层,旨在增强对小型车辆的检测性能。Gong 等人(2022)修改了通用的一级检测器 YOLOv5 以适应卫星图像,从浅层添加新的特征融合层和预测头用于小目标检测。Ma 等人(2022)设计了一种适用于小目标的稀疏特征提取网络,以及针对小目标特征映射和背景抑制的多层联合上采样特征映射网络。Wei 等人(2020)提出高分辨率特征金字塔网络(high-resolution feature pyramid network, HRFPN),旨在充分利用高分辨率特征表示,以实现合成孔径雷达(SAR)船舶的精确且鲁棒的检测。Lu 等人(2020)提出一种全局特征门控过程,实现局部特征学习的信道注意力机制,以增强微小目标的特征判别力。

特征金字塔网络(feature pyramid network, FPN)

(Lin 等, 2017a)引入了一种自上而下的连接路径,以便将深层语义信息传播到浅层。然而,即使在骨干网络中低层的空间信息经过了长距离传播,在顶层仍然可能会丢失。为解决这个问题, Huang 等人(2022a)和 Chalavadi 等人(2022)提出一种双向 FPN, 该网络通过可学习的参数来确定不同级别特征的重要性,并通过迭代的自上而下和自下而上的路径来融合多级特征。与上述的顺序增强途径不同, Cheng 等人(2021)采用了跨级特征融合的方法,通过特征串联操作实现了跨尺度特征融合,这种方法考虑到了不同层次特征在融合过程中的不同贡献。

不同尺度的特征包含着不同层次的信息,通过多尺度特征融合,可以充分整合低层细节信息和高层语义信息,从而提升目标检测的精度,并增强模型在复杂场景下的适应性。然而,不当的融合可能导致特征冗余或信息失真,影响模型对目标的精准定位。此外,不同尺度特征之间存在分布差异,直接融合可能会引入噪声,降低检测的鲁棒性。

2.2 针对度量不匹配的小目标检测算法

对于小目标检测,锚框的数量、尺寸和比例等超参数对性能影响较大,需要精细调优。Rekavandi 等人(2025)指出,当前流行的检测方法在微小目标检测上存在局限,优化损失函数是提升小目标检测性能的有效途径。基于锚框的检测器仍占主导地位,但匹配小目标的锚框数量远少于大/中尺度目标。由于小目标占据图像的面积较小,其边界框定位更具挑战性。在预测中,边界框的微小偏移对小目标的影响远大于对大目标的影响,因此优化损失函数有助于小目标的精准定位。

2.2.1 基于锚框机制的方法

由于小目标的真实框和锚框尺寸较小,任何微小的偏移都会显著影响两者的重叠度,使得原本可能匹配的锚框被判定为负样本,因此如何合理设置锚框是小目标检测的一大难点。

Faster R-CNN(Ren 等, 2015)引入了区域建议网络 RPN,以生成候选目标边界框。Wang 等人(2018)设计了一个级联掩膜生成框架,该框架接受多尺度图像作为输入,并按照尺度大小的顺序逐步处理这些图像。为追求高召回率,需要在高分辨率特征图上放置大量密集的锚点。然而,这会导致类别分布的极度不平衡,从而严重影响了检测框架中的分类任务。Duan 等人(2020)提出一种新颖的基于通道

感知反卷积网络(channel-aware deconvolutional network, CADNet)的基于锚点中心点平移策略(matching strategy based on the center point translation, CPT-Matching),以在训练阶段更鲁棒地选择更多延伸的锚框作为正样本,从而实现更好的锚框匹配。该策略基于对锚框中心点进行平移的方法,使得在训练过程中更有效地捕获目标,从而提高模型对不同尺度和形状目标的检测能力。

无锚方法因其简单的体系结构和优越的性能,近年来成为研究热点。与基于锚框的方法不同,无锚机制直接预测目标的位置,而不需要估计锚框的偏移。FBR-Net(feature balancing and refinement network)(Fu 等, 2021)提出注意力引导的平衡金字塔,扩展了无锚检测器,平衡了不同金字塔层的特征,并增强了复杂情况下小目标的学习。Law 和 Deng(2018)提出 CornerNet,通过使用单个 CNN 将目标边界框检测为配对关键点,即左上角和右下角。Zhou 等人(2019)所提出的检测器 CenterNet 与上述方法不同,他们将目标建模为单一点,通常为其边界框的几何中心。利用关键点估计目标的中心位置,并进行全面的属性回归,包括尺寸、位置和方向等。Wang 等人(2020)提出一种名为 M-CenterNet 的检测器,采用多个中心点来精确定位目标的中心,从而提高遥感小目标的定位性能。

锚框机制简化了目标匹配和定位,通过在图像中生成密集的锚框,将检测和定位转化为分类和回归问题,有效捕捉不同大小和形状的目标,提升检测精度。然而,锚框依赖预定义框集,可能无法精确覆盖目标,导致样本不匹配、负样本过多以及候选框冗余等问题。相比之下,无锚方法通过直接在特征图上预测目标位置和类别,避免了无效候选框的生成和处理,减少了对目标形状和尺寸的先验约束。然而,无锚方法对特征图的质量和对齐精度依赖较高,特征对齐不准确可能导致定位偏差,尤其是在高分辨率和复杂背景的遥感图像中。

2.2.2 基于训练优化的方法

微小的边界框位置变化可能导致小目标的 IoU 值发生显著波动,这对小目标检测的性能产生不利影响。此外,在非极大值抑制(non-maximum suppression, NMS)过程中, IoU 用于判断预测边界框是否被视为误检。然而, IoU 对小目标的高度敏感性可能导致 NMS 错误地将一些真实的边界框预测视

为冗余,进而影响整体检测性能。

Rezatofghi 等人(2019)提出的广义交并比(generalized IoU, GIoU)是在 IoU 的基础上引入了两个边界框的最小封闭区域,以更准确地描述两个边界框之间的位置关系,但每次都要额外计算最小外接矩阵,收敛速度受限,并且预测框重叠的情况下,无法反映出实际情况。Zheng 等人(2020b)提出距离交并比(distance IoU, DIoU)和完整性交并比(complete IoU, CIoU)。DIoU 通过考虑两个边界框之间的中心点距离和最小封闭区域对角线的长度,进一步改进了 IoU。虽然有效改善了 GIoU 的缺点,但只考虑了中心点和重叠面积,并没有考虑长宽比。CIoU 则在 DIoU 的基础上,考虑了边界框的宽度和高度,添加了高和宽的损失,使回归更稳定。但 CIoU 的高和宽不能同时增大或减小,这显然是不合理的。为解决 CIoU 的问题,Zhang 等人(2022)将 CIoU 中预测框和真实框的影响因子拆开,分别计算高和宽,高宽损失极大降低了两者的差异,并提出高效交并比(efficient IoU, EIoU)。之后,对于 IoU 的改进还有 α IoU (α -IoU) (He 等, 2021)、SIoU (Seylla IoU) (Gevorgyan, 2022)、Wise IoU (Tong 等, 2023)等。

然而,尽管这些指标对 IoU 进行了微调,并在一定程度上改进了位置关系的描述,但它们仍然受限于非归一化取值范围的数学特性。此外,这些改进主要被设计为损失函数,在解决微小目标对 IoU 敏感的问题方面并没有本质上的突破。由于前景和背景存在类不平衡的问题,Chen 等人(2019c)通过将排序任务代替单级检测器中的分类任务,选择平均精度损失(average precision loss, AP loss)以解决排序问题,对基于 AP 损失的单级检测器的性能有明显改善。Lin 等人(2017b)为降低分配给分类良好的示例的损失权重,设计一个新颖的密集检测器 RetinaNet,并提出新焦点损失(focal loss),将训练集中在一组稀疏的困难示例上。Wang 等人(2019)提出一种焦点区域损失函数(focal-area loss),通过关注道路上小目标的面积变化来进行学习。

He 等人(2019)引入 KL (Kullback-Leibler) 散度作为边界框回归损失,用于边界框变换和定位。Xu 等人(2021a)基于微小边界框的绝对尺寸和相对尺寸远小于中型或大型边界框的特征提出点距离(dot distance, DotD),即两个边界框中心点之间的归一化欧几里得距离。Wang 等人(2021)使用 Wasser-

stein 距离代替 IoU,提出一种归一化 Wasserstein 距离的新基准(normalized Wasserstein distance, NWD)。Wasserstein 距离最大的优点是即使两个边界框无重叠或相互包含,也可以测量分布的相似性。并且 NWD 对不同尺度的目标不敏感,因此更适合测量极小目标间的相似性。Yang 等人(2021b)计算高斯分布之间的 KLD 作为回归损失,通过目标的特性动态地调整参数梯度。Fu 等人(2023)针对小边界框和预定义锚框之间相似度测量不准确的问题,提出基于高斯概率分布的模糊相似性度量(Gaussian probabilistic distribution-based fuzzy similarity metric, GPM),为标签分配提供更准确的基础。KLDet (Zhou 和 Zhu, 2024)将参数化边界框转换为二维高斯分布。通过使用分布之间的相似度 KLD 作为定位度量,能准确反映边界框的微小偏移量,提供了更多潜在的正样本。

新的评估指标能够专门针对小目标的检测性能进行细致评估,表3列出了损失函数方案的优缺点。它们为模型优化提供了更加明确的方向,有助于提升模型对小目标的检测精度和鲁棒性。然而,与传统的评估方法相比,新指标往往需要更复杂的计算过程,如对目标进行尺寸分类、过滤和加权处理,这增加了评估过程的复杂性和操作的烦琐性。此外,某些新指标的设计可能针对特定的应用场景或任务,导致其泛化能力不足,在不同的应用场景中可能不具备广泛的适用性。

2.3 针对图像幅面宽的小目标检测算法

遥感图像因其高分辨率和广阔的成像范围,导致小目标的相对尺寸显著缩小。因此,直接处理整幅遥感图像可能会导致巨大的计算负担,从而影响推断速率。当前主流的目标检测算法多为大/中等尺寸图像和目标设计,如何在这些大型图像中实现小目标与大幅面之间的有效尺度匹配,不仅是提升检测精度的关键,也是加快推断速度的必要前提。

2.3.1 基于尺度匹配的方法

遥感图像由于其高分辨率和广阔的成像范围,使得图像中的小目标相对于整幅图像的尺寸变得更小,造成了尺度不匹配的问题。现有的目标检测算法大多是为处理中等尺寸的目标和图像设计的,在面对遥感图像中小目标时,难以精准定位和识别。

Yu 等人(2020)为促进极小目标检测的特征表示,提出一种尺度匹配(scale match, SM)来调整两

表3 基于损失函数的方案优缺点总结
Table 3 Summary of advantages and disadvantages of loss function-based approaches

方法	优势	不足
AP loss	通过设计或优化损失函数来缓解小目标类别不平衡的问题,从而提高小目标的检测性能。	AP loss 中采用的排序方案并不稳定, Focal loss 并没有考虑样本间的相关性。
Focal loss		
Focal-area loss		
IoU loss	随着优化改进,降低了宽高之间的差异,提高目标边界框精度,有利于目标定位并加速检测。在一定程度上可以改善小目标检测。	基于 IoU 的损失优化方案难以平衡精度和召回率,并且不易收敛,其有效性过度依赖正负样本的平衡。
GIoU loss		
DIoU loss		
CIoU loss		
EIoU loss		
α IoU loss		
SIoU loss		
WIoU loss		
KL loss	有助于测量分布的相似性,改善边界框和预定义锚框之间相似度测量不准确的问题。	这些方法使得操作更加复杂,在训练过程中,超参数选择困难,并导致训练不稳定。
DotD loss		
NWD loss		
GWD loss		
KLD loss		
GPM loss		

个数据集间的目标尺度,为长距离和复杂场景下的极小目标检测开辟了新方向。但SM是针对整个图像进行操作分析,对于具体的目标和细节关注较少。作为SM的改进工作, Jiang 等人(2021)提出的SM+从实例入手,将尺度匹配从图像级扩展到实例级,将图像分为标注的实例和背景,提升了预训练数据集和目标数据集之间的相似性,并通过概率结构修复方法动态地修复图像。

Li 等人(2019)提出一种新型框架 YOLT(you only look twice),根据目标的尺度将目标分为3类,通过不同的训练策略来训练不同尺度目标的分类器和检测器。最后,将多尺度检测图像块进行拼接和融合,以获得更精确的定位和分类结果。Hong 等人(2021)提出尺度选择模块(scale selection module, SSM)来指导深层将合适的特征传递给浅层,解决不同层级之间的梯度计算不一致问题。此外, Zhu 等人(2023)设计了尺度感知知识蒸馏模型(scale-aware knowledge distillation, ScaleKD),将复杂的教

师模型的知识转移到紧凑的学生模型,设计了两个模块来提高知识转移质量,即尺度解耦的特征蒸馏模块,实现学生模型在小目标上显式模仿,并进一步细化杂乱且无信息的边界框预测学生模型。此外建立多尺度交叉注意层来捕捉多尺度语义信息以改进学生模型。

精准的尺度匹配能让模型对不同大小的目标特征提取更加高效。通过精准匹配不同大小目标的特征,避免漏检或误检,并提高资源利用效率,减少冗余计算。然而,该策略的优化需要大量的先验知识和实验调优,不同场景下需要定制化的方案,增加了算法设计和部署的复杂性。

2.3.2 基于推断加速的方法

遥感图像幅面宽大,这意味着在数据处理和传输过程中可能会面临速度较慢的挑战。在平均分块的过程中,处理的图像块数量较多。每个块都需要单独进行特征提取和分类,这导致了计算量的显著增加。此外,由于块数的增加,模型需要更长的时间来逐一处理每个块的内容,从而增加了整体的推断时间。这个问题在处理高分辨率遥感图像或复杂场景时尤为明显,因为需要分割出更多的块来覆盖所有的细节,从而进一步拖慢了推断速度。

SSD-MSN(multi-scale object detection network based on SSD)(Chen 等, 2019d)通过从剪裁区域中选择包含一个或多个目标的区域建议,从原始图像中剪裁的放大区域中学习小目标的更丰富特征。并提出一种有效的分割图像策略,即从原始图像中生成 3×3 个裁剪区域。该策略不仅可以生成更多的区域建议,还可以确保每个裁剪区域中包含更多目标。Van Etten(2019)提出SIMRDWN(satellite imagery multiscale rapid detection with windowed networks)框架,为多个目标检测框架提供一种统一的方法,这些框架可以对任意大小的图像进行推理,该流程以大于 $0.2 \text{ km}^2/\text{s}$ 的速率评估原始分辨率下任意大尺寸的卫星图像,有效提高检测速度。

Akyon 等人(2022)设计了一个名为切片辅助超推理(slicing aided hyper inference, SAHI)的开源框架,为小目标检测提供了通用的切片辅助推理和微调框架。通过将图像分成重叠的图像块可以使小目标相对于图像的像素区域变大,然后送入网络,并在微调数据集的图像中进行提取图像块扩充。Zhang 等人(2023)提出一种自适应切片辅助超推理(adap-

tive slicing aided hyper inference, ASHAI)的新型自适应切片方法。ASAHI 关注的是切片数量而不是切片大小,它自适应地调整切片大小以根据图像分辨率控制切片数量。

通过提升推断速度,可以减少每个图像或数据块的处理时间,从而降低整体计算资源的消耗,使得对大量遥感影像的分析和处理变得更加可行。但在加速推断速度的过程中,通常需要简化模型结构或降低模型复杂性、采用较大的步幅或较低的分辨率来进行特征提取,这可能会导致检测精度的下降。特别是在处理复杂的遥感图像时,分辨率降低会严重影响对小目标的识别和定位,这可能导致对小目标的细节捕捉不足,进而丢失特征。

2.4 针对背景干扰大的小目标检测算法

光学遥感图像通常包含多样且复杂的地表特征,这些特征与目标在颜色和纹理上可能相似,增加了识别的难度。此外,近期的研究主要致力于改进定向边界框表达,特别是在遥感目标检测领域。然而,这些研究忽略了遥感场景中的独特先验知识。在这种情况下,可能会误检测微小的遥感目标,并且不同类型的目标可能需要不同的远距离上下文信息来进行小目标的准确检测。

2.4.1 基于上下文学习的方法

遥感图像中目标的准确检测往往需要大量的上下文信息,并且不同目标类型所需的上下文信息范围有很大差异。上下文信息通常是指目标与周围环境或场景之间的空间和语义关系,可以为无法清晰区分的目标提供辅助特征表示。因此,上下文感知的方法在遥感小目标检测中能够通过利用周围环境的信息,提高目标定位精度、辨别能力和检测稳定性。

Liu 等人(2019a)提出多分量融合网络,旨在优化遥感图像中小目标检测的性能。首先,通过构建双金字塔融合网络,密集地融合空间和语义线索,利用编码—解码操作提取小目标的特征。随后,引入相对区域建议网络以充分捕获小目标样本和部分的特征。最后,在最终检测阶段,采用上下文信息融合策略,增强对背景干扰的鲁棒性。Zhang 等人(2020)通过将候选的初始区域缩放为3个不同大小的区域,生成多个局部上下文区域,并引入了上下文双向增强模块来有效融合局部上下文特征和目标特征。Yang 等人(2020)开发了一种初始并行注意网

络,进一步加强检测遥感图像中小目标的有效权重。Liang 等人(2020)提出一种基于特征融合和缩放的SSD(feature fusion and feature enhancement SSD, FS-SSD)用于无人机图像中的小目标检测。Cheng 等人(2021)设计上下文特征增强模块来利用全局上下文线索,并通过使用指示目标类存在或不存在的图像级上下文信息来选择性地增强类别感知特征。Chen 等人(2019b)采用RoI-Align 操作从全局图像特征中提取场景上下文信息,并将其与目标级感兴趣区域(region of interest, RoI)特征进行融合,以增强目标与场景之间的关联性。

注意力机制强调了有价值的区域,同时抑制了那些无价值的区域。Lu 等人(2021)结合了空间注意力和通道注意力,引入了双路径注意力模块,用于消除背景噪声并突出小目标。Ran 等人(2021)采用多尺度上下文和增强通道注意力模型(multiscale context and enhanced channel attention, MSCCA)增强小目标特征。Li 等人(2021)设计了一个跨层注意力模块以获得小目标更强的响应。DETR(detection TRansformer)作为一种端到端目标检测框架,采用Transformer 进行全局关系建模,在小目标检测任务中表现良好。O²DETR(oriented object DETR)(Ma 等,2021)采用局部卷积替换注意力模块,并结合多尺度特征图的集成,证明在定向目标检测背景下可以提高检测性能。Dynamic DETR(Dai 等,2021)在解码器中设计了一种基于ROI的动态注意力机制,以多尺度方式关注感兴趣的区域。Deformable DETR(Zhu 等,2020)提出一个注意力模块,仅关注参考点周围的几个采样点。Gong 等人(2022)采用YOLOv5,并首次从浅层添加了新的特征融合层和预测头来检测小目标,将原始卷积预测头替换为Swin Transformer 预测头,其移位窗口设计可以将计算复杂度降低到线性。AO2-DETR(arbitrary-oriented object DETR)(Dai 等,2023)采用一种定向建议生成机制,可显式生成定向建议以提供更好的位置先验。DQ-DETR(dynamic query DETR)(Huang 等,2024)使用分类计数模块的预测和密度图动态调整目标查询的数量并改善查询的位置信息,以精确检测微小目标。

Chen 等人(2023)提出一种名为MDCT(multi-kernel dilated convolution and Transformer)的新型单阶段目标检测模型,将Transformer 块集成到一级目

标检测模型的颈部网络中,以防止复杂背景和密集场景中目标信息的丢失。Li 等人(2023)考虑到先验并提出选择性大核网络 LSKNet (large selective kernel network),这是首次在遥感目标检测领域探索选择性大核机制。通过在特征提取模块动态调整感受野,能够更有效地处理不同目标所需的背景信息差异。这一动态感受野的实现依赖于空间选择机制,该机制对通过一系列深度卷积核处理的特征进行了有效加权和空间融合。这些卷积核的权重根据输入动态确定,使模型能够自适应地选择不同大小的核,并相应地调整感受野,以更好地模拟遥感场景中各种目标的上下文信息。

上下文信息学习策略在遥感小目标检测中,通过整合目标周围的背景特征来辅助目标检测,有助于模型更好地理解复杂的背景环境,从而有效减少误检和漏检的情况。此策略降低了模型对单一目标特征的依赖,使得在面对目标的形状变化和部分遮挡时,仍能保持较高的检测性能。然而,这种方法也可能引入不必要的冗余信息。特别是当小目标与其背景高度相似或部分重叠时,上下文信息可能造成目标与背景之间的混淆,从而增加检测过程中的误检或漏检风险。

2.4.2 基于标签噪声鲁棒学习的方法

微小目标缺乏外观信息,使得标注过程费力,而且容易引入标签噪声。噪声标签的产生主要是由于人为因素或数据本身的问题造成。噪声标签会引入错误信息,加剧计算成本,造成定位不准确,严重影响机器学习模型的性能和效果。

为降低微小目标检测的现实世界标注成本,交互式分割方法(Xu 等,2016)专注于减少单个前景目标分割中的交互数量。目标计数方法(Ranjan 等,2021)通过用户几次点击来计算多个实例。Lee 等人(2022)提出一种交互式多类微小目标检测 C3Det (class-wise collated correlation detection),分别通过后期融合和特征相关,以局部和全局方式将完整图像上下文与注释器输入相关联。一旦用户单击一些目标并提供其类信息,C3Det 就会将这些信息作为输入并检测许多目标的边界框,甚至包括用户未指定的目标类,有效减少了标注成本。

目标检测中标签噪声的类型多种多样,并且不同类型的标签噪声会以不同的方式降低检测性能。Zhu 等人(2024)将标签噪声分为4类:标签丢失、额

外标签、类别转移和不确定的边界框,并提出一种去噪微小目标检测器 DN-TOD (denoising tiny object detector),它结合了类感知标签校正方案来解决类转移问题,采用趋势引导学习策略来处理边界框噪声。广义焦点损失(Li 等,2020c)将边界框建模为任意分布而不是狄拉克分布,通过边界框的不确定性有效指导定位质量估计。Xu 等人(2021c)采用了一种基于元学习的模型(meta-refine-net, MRNet),通过交替执行噪声修正和模型训练来解决标签噪声问题。Bernhard 和 Schubert (2021)假设所有类型的噪声都会发生,将空间误差视做目标的位置噪声,设计一种辅助校正技术尝试同时处理所有类型的噪声。DN-DETR (denoising DETR) (Li 等,2022a)引入去噪训练来降低二分图匹配的难度。Liu 等人(2022)利用分类作为改进定位结果专注于处理边界框噪声。Wu 等人(2023)使用基于多实例学习的方法在框噪声下进行鲁棒学习,提出一种基于空间自蒸馏的目标检测器 SSD-Det 来挖掘空间信息,以自蒸馏的方式细化不准确的边界框。

有效的标签噪声处理策略可以显著提高模型的鲁棒性和泛化能力,确保模型在面对带有噪声的标签时,可以自适应地应对不同程度的标签噪声。通过引入噪声修正或噪声过滤方法,模型可以减少由于错误标签带来的负面影响,从而得到更准确的预测结果。这不仅有助于模型更好地学习目标与背景之间的差异,还能增强模型对异常数据的容忍度。但模型如果过度关注噪声数据,可能会造成过拟合,导致模型无法正确泛化到准确的数据上。此外,在面对大规模数据集,标签噪声处理策略会增加模型复杂度,尤其是那些涉及动态调整或多阶段训练的策略,会引入不稳定性,导致模型难以收敛,并显著增加训练的计算复杂度和时间成本。

2.5 针对分布不均匀的小目标检测算法

在航拍图像中,目标的分布呈现出显著的特征,不仅稀疏且分布不均匀,还常常在特定区域高度聚集。稀疏的小目标在广袤的背景中散落分布,仅占据极少的像素区域,这使得它们极易被背景噪声所掩盖,极大地增加了检测的难度。针对大尺度场景中的稀疏分布和尺度变化问题,通过自适应算法或人工手动设定的方式,先将原始图像裁剪成较小的区域/碎片,再将这些小裁剪图调整为与原始图像相同的大小,以更好地聚焦和处理小目标的检测任务。

2.5.1 基于密集区域搜索的方法

按照图像裁剪的思想,如何找到合理的裁剪区域对于遥感图像目标检测至关重要。当前研究者通常采用先识别密集区域的方法进行裁剪,以避免产生分割过多的密集区域或目标,并减少生成不包含目标的纯背景子区域,从而节约计算资源。

早期的检测方法采用了随机或均匀裁剪。Özge Ünel 等人(2019)证明了均匀裁剪对于小目标检测的有效性。R²-CNN(remote sensing region-based CNN)(Pang 等,2019)在对高分辨率图像进行检测之前将其划分为小的重叠块。但这些方法无法利用语义信息进行裁剪,因此导致大多数裁剪图只有背景。此外,这些策略可能会将大目标切成两个或更多不同的裁剪图,进而影响性能。

ClusDet(clustered detection)(Yang 等,2019a)通过聚类提议子网络 CPNet(cluster proposal sub-network)生成目标簇区域,尺度估计子网络 ScaleNet(scale estimation sub-network)估计这些区域的目标比例,最后每个尺度归一化的聚类区域都被输入到专用检测网络 DetecNet(dedicated detection network),进行目标检测。为了解决在高分辨率图像上进行分块再检测导致推断速度低下的问题,UFPMP-Det(multi-proxy detection network with unified foreground packing)(Huang 等,2022b)引入了统一前景组装模块,该模块首先将粗糙检测器输出的子区域进行聚类合并,以抑制背景干扰。接着,利用 Mosaic 技术(Zhang 等,2016)将子区域进行组装,通过单个推理过程完成检测任务,从而显著降低总体时间成本。

DMNet(density-map guided object detection network)(Li 等,2020)受图像目标密度图可呈现目标依像素强度分布这一现象的启发,首先预测密度图,然后引入滑动窗口,根据窗口中的密度值判断该区域是否包含目标,根据密度强度学习尺度信息以形成裁剪区域。CDMNet(coarse-grained density map network)(Duan 等,2021)将密度图定义为粗粒度的形式,并设计了一个轻量的对偶密度估计网络。粗粒度的密度图不仅能描述目标的分布,而且可以聚集目标、量化尺度和减少计算量。Liu 等人(2024)提出 YOLC(you only look clusters),一个基于无锚目标检测器 CenterNet 构建的高效且有效的框架,通过在生成的热图上,添加一个卷积层和两个转置卷积层,将热图放大到与输入图像相同的大小,得到更高分辨

率的热图。引入了局部尺度模块(local scale module, LSM),通过密度排序定位前 K 个密集区域,自适应地搜索聚类区域并放大以实现准确检测。

通过先定位密集区域进行目标识别和定位,可以有效排除大量不含目标或仅含少量目标的背景区域,使模型专注于处理密集区域内的目标特征,从而优化计算资源分配,提升小目标检测的精度。在背景复杂或干扰较多的场景中,这一策略能够帮助模型更精确地区分目标与背景。然而,该方法对区域划分的准确性依赖较大,若密集区域识别不精确,可能导致漏检重要目标或误将无关区域识别为密集区域,从而影响检测效果。此外,即便聚焦于密集区域,一些小而密集的目标仍然难以准确识别。

2.5.2 基于低置信度区域挖掘的方法

在粗检测结果中,高置信度区域往往受到过度关注,尽管这些区域可能存在误判或背景干扰;而低置信度区域则潜藏着被忽视的真实目标线索。通过筛选低置信度区域,可以有效补充漏检信息,提升检测结果的完整性。同时,低置信度区域的筛选有助于规避高置信度区域的误标注与不相关背景干扰,聚焦关键内容,从而提高检测效率和准确性。

DREN(difficult region estimation network)(Zhang 等,2019)加强了困难区域的检测。Wang 等人(2020)利用聚类算法根据整个图像的检测结果生成初始困难区域。GLSAN(global-local self-adaptive network)(Deng 等,2021)使用全局—局部融合策略和渐进的尺度变化网络来进行准确检测。对目标边界框的预测分两个阶段,首先在原始图像上进行全局粗略检测,在降采样的整幅图像上预测轮廓线索;将裁剪子图像通过超分辨率处理后,进行局部精细检测,以预测更精确的优化结果。两个阶段的检测结果通过 NMS 合并得出最终最优结果。Koyun 等人(2022)提出一个两阶段的目标检测框架,称为“聚焦和检测”。通过高斯混合模型监督的目标检测器生成构成聚焦区域的目标簇,采用向量表示图像中均匀采样点,位置信息则由边界框距离向量组成,并通过可变形卷积增加簇的可迁移性。

基于低置信度区域的筛选能够有效补充漏检信息,提升检测结果的完整性,并避免高置信度区域的误标注与背景干扰,从而提高检测效率和准确性。但该方法未充分利用目标间的上下文信息,可能导致在复杂场景中的误判或漏判。此外,依赖初始粗

糙检测结果的准确性存在风险,尤其在初始检测性能较差时,无法有效处理假阴性和假阳性问题,进而影响整体检测效果。

2.6 针对方向不确定的小目标检测算法

遥感图像中独特的俯视视角使得目标通常具有多样的朝向和角度变化,如船只、车辆、建筑物等,这些目标的姿态和方向不固定,传统的水平边界框可能无法准确地包围和识别这些旋转目标,地理空间目标在图像中通常呈现任意方向的分布,这种旋转不定的特性显著增加了遥感小目标检测的难度。传统基于水平框的方法在处理方向估计时存在准确性有限的问题,而精确的方向估计对于遥感图像中的信息提取至关重要。

2.6.1 基于旋转目标表征的方法

传统目标检测方法在处理旋转目标时存在显著局限,旋转角度的变化使得基于固定坐标体系的检测方法容易出现漏检和误检,影响检测精度。此外,紧密排列、遮挡或尺寸小的旋转目标更难以准确识别。旋转目标表征的方法通过建模目标的旋转角度并融入旋转坐标体系,能够全面捕捉目标的几何特征,从而提升旋转目标和遮挡目标的检测精度,减少漏检和误检。

Yang 等人(2019b)提出 SCRDet (small, cluttered, and rotated object detection)来解决微小、杂乱和旋转目标,通过采样特征融合网络 SF-Net(sampling and feature fusion network)结构解决锚框采样不足的问题,并通过抑制噪声和突出目标特征,联合探索监督像素注意网络和通道注意网络构建了 MDA-Net (multi-dimensional attention network),最后将 IoU 常数因子添加到平滑 L1 损失中,以解决旋转边界框的边界问题。而改进工作 SCRDet++(Yang 等, 2023)将去噪的思想加入检测中,在特征图上进行实例级去噪,通过有监督的分割减小噪声和干扰。边界不连续性及其与最终检测指标的不一致一直是旋转检测回归损失设计的瓶颈, Yang 等人(2021a)将旋转的边界框转换为二维高斯分布,这使得能够通过高斯 Wasserstein 距离 (Gaussian Wasserstein distance, GWD)近似不可微的旋转 IoU 引起的损失,而高斯 Wasserstein 距离可以通过梯度反向传播有效地学习,基于高斯 Wasserstein 距离的新型回归损失有效地解决了这个问题。

Zand 等人(2022)通过一种纯分类、无需对 OBB

进行额外计算的方法 DarkNet-RI (darknet-rotation invariant)对旋转小目标进行检测,对每个尺度应用旋转不变特征表示,并施加正则化约束以强制覆盖训练样本的 360° 平面旋转范围进行特征共享,通过优化新的目标函数约束训练样本在每个尺度上对旋转的泛化。

旋转小目标检测可以为目标生成具有任意角度的旋转边界框,这使得模型能够更紧密和准确地包围旋转的目标,并通过引入旋转不变性,模型可以更好地适应不同角度和姿态的目标变化,能减少背景和目标的混淆,减少由于目标旋转而引入的检测误差。但细微角度误差可能影响检测精度,尤其对微小或形状复杂的目标。此外,额外的旋转参数增加了计算开销,可能导致推理速度下降。

2.6.2 基于标签匹配的方法

现有的标签分配范式,无论是基于锚框还是无锚框的,标签匹配指标高度依赖于域的重叠,在实际应用中可能会产生大量异常小尺寸的真实样本。对于旋转小目标,其本身尺寸小且具有旋转角度,目标的有效特征更加难以被准确捕捉,检测器在训练过程中对这些样本的关注度会显著降低,从而导致检测器出现漏检或误检的情况。

Xu 等人(2022b)认为现有的先验框与先验点及其相应的度量策略在处理微小目标时是次优的,并会阻碍标签匹配的过程,他们提出将单个的先验框和点作为实例进行处理,提出标签分配策略 RFLA (Gaussian receptive field based label assignment),通过设计一个新的高斯感受野距离 (receptive field distance, RFD)直接测量高斯感受野和真实边界框之间的相似度。高斯感受野不是阶跃变化的,每个单独先验的范围是整个图像,其中每个位置的权重从中心往外以一个大于 0 的值逐步衰减,采用高斯感受野作为先验信息,可以有效获得不同目标大小的平衡正样本,有效缓解了感受野不匹配问题。Hou 等人(2022)提出用于定向目标检测的新型灵活形状自适应选择 (shape-adaptive selection, SA-S)和形状自适应测量策略 (shape-adaptive measurement, SA-M),根据目标的形状和方向准确地分配标签。Xu 等人(2023)认为特征先验不匹配和正样本不平衡是阻碍定向微小目标标签分配的两个障碍,提出一种动态先验以及由粗到细的分配器 (dynamic coarse-to-fine label assignment, DCFL),以动态方式对先验、标签

分配和目标表示进行建模,以缓解不匹配问题。同时,利用粗先验匹配和更精细的后验约束来动态分配标签,为不同实例提供适当且相对平衡的监督。

有效的标签分配策略可以准确地将先验框与真实目标匹配,并能够为模型提供更准确的训练信号,使模型能够更好地学习到区分目标与背景的特征。通过对小目标进行更加细致的标签匹配和分配,模型可以得到更多关于小目标的有效训练数据,提升对小目标的识别能力。而动态标签分配方法可以根据实际数据进行自适应调整,减少对预定义锚框的先验假设的依赖。但动态或复杂的标签分配策略可能会大幅增加模型训练的计算复杂度和时间成本。此外,在标签分配过程中,如果匹配策略选择不合适,可能会导致先验框与错误的目标匹配,从而引入标签噪声。

2.7 针对数据集稀缺的小目标检测算法

人们普遍认为,随着数据集规模的扩展,深度学习模型的性能通常会显著提升(Halevy等,2009)。这是因为更大规模的数据集能够提供更丰富和多样化的样本,使模型能够更有效地学习数据的潜在分布及复杂特征,从而提升其泛化能力和准确性。然而,构建这样的大规模数据集并非易事。数据的收集和标注往往需要投入大量的时间、人力和资源,尤其当数据要求精确标注或来自复杂场景时,手动标注的难度和成本会显著增加。此外,数据的多样性和质量对模型性能也至关重要,简单增加数据量并不一定能确保模型的优化效果。

2.7.1 基于数据扩增的方法

数据是深度学习模型训练和推理的核心,决定其学习与预测能力。数据增强作为提升小目标检测性能的有效手段,可扩充数据规模、增强多样性,从而提升模型的泛化能力和鲁棒性(Kaur等,2021)。通过缩小训练、验证及测试集之间的差距,数据增强有助于降低泛化误差。模型性能受数据质量与数量影响,充分利用高质量大规模数据是提升检测效果的关键。

早期的数据增强策略主要依赖于几何变换,以扩展训练数据的分布空间并提升模型泛化能力。其中,常见方法包括目标实例的几何变形(Simard等,2003)、旋转与缩放(Yaeger等,1996)、裁剪(Krizhevsky等,2012)以及平移(Wan等,2013)。这些变换在保持图像语义不变的前提下,有效增加了数据

的多样性,从而增强目标检测的稳健性。Stitcher(Chen等,2020)引入一种方便的数据增强拼贴方式,然后由运行时的反馈来指导。该方法侧重于利用动态反馈引导的拼接图像处理比例变化。Auto-Augment(Cubuk等,2018)通过自动增强在一定程度上缓解了一个问题,通过创建一个数据增强策略的搜索空间,直接在感兴趣的数据集上评估特定策略的质量,但训练速度过慢。而Fast-AutoAugment(Lim等,2019)算法使用基于密度匹配的更有效的搜索策略学习增强策略,有效加快了训练速度。Yun等人(2019)提出的CutMix数据增强方法通过从数据集中选择两幅图像,裁剪其中一幅图像的部分区域,并将其叠加在另一幅图像之上,生成新的输入图像用于训练。受此方法启发,YOLOv4(Bochkovskiy等,2020)引入一种名为Mosaic的数据增强方法。Mosaic的主要思想是随机裁剪4幅图像块并将它们拼接在1幅图像上,使得目标以更小的比例出现,从而作为训练数据。

数据增强通过对原始训练数据进行多种变换,能够有效增加样本多样性,从而提升模型的泛化能力并减少过拟合的风险,尤其在数据稀缺的情况下表现突出。对于小目标检测任务,数据增强有助于模拟不同场景中的目标变化,提升模型的鲁棒性。然而,数据增强虽然能提升模型性能,但也会增加计算开销和训练时间,尤其在大规模数据集上。若增强策略不当,可能引入噪声,降低模型准确性。同时,过度增强可能导致效果饱和,难以进一步提升性能。

2.7.2 基于生成合成的方法

小目标通常在实际场景中难以捕捉,且受背景干扰较大,导致传统方法在检测时容易出现漏检或误检。通过生成合成目标,可以有效地增加小目标实例的数量,扩充数据集,从而使模型在训练过程中能够接触到更多种类的目标和场景。这种方法不仅增加了数据集的多样性,还能在目标的位置、尺度和背景方面进行更精准的调整,使合成目标更加符合实际场景的特征。基于生成合成的数据增强方法有助于提高模型的泛化能力,使其在面对不同方向、尺度和复杂背景的目标时,仍能保持较高的检测精度。

生成合成对象的一个简单方法是通过随机复制粘贴增加小目标的实例数量。Kisantant等人(2019)通过复制小目标来对感兴趣的小目标进行过采样提

升检测性能。这种方法的问题有两个方面:1)目标的特征保持不变;2)目标的位置和尺度可能与上下文不符。Chen等人(2019a)通过一种自适应增强策略 AdaResampling 解决第2个问题,该策略通过使用分割 CNN 生成先验上下文图,并根据目标的尺度和位置放置目标。另一种解决方案是通过对抗训练学习可能的增强空间。Wei等人(2018)引入了 PTGAN (person transfer generative adversarial network),通过将目标从一个数据集转移到另一个数据集来解决经典的领域间差距问题。DetectorGAN(Liu等,2019b)基于 CycleGAN,执行图像到图像的转换,将无目标的图像转化为包含目标的图像。Bosquet等人(2023)除了对象分割、图像介入和图像混合的技术之外,设计了一种新型的数据增强管道,以生成小目标的高质量合成数据。通过一种生成的对抗网络架构(downsmpling GAN, DS-GAN),将大尺寸目标转换为高质量的小目标,即从大目标中生成小目标。

生成对抗网络(GAN)的出现为训练体系结构开辟了新的数据增强可能性,而无需为小目标注释大量数据集的昂贵任务。它能够通过生成合成目标有效扩充训练数据,增强模型对各种目标特征的学习,提升检测精度。但是生成的目标可能无法完全反映真实场景的复杂性,尤其在处理遮挡或复杂背景时,合成图像的准确性可能不足。此外,合成目标的特征与真实目标之间存在差异,可能导致模型过于依赖合成数据,进而影响对真实场景的适应能力。

3 数据集和评估指标

3.1 数据集选择与介绍

在遥感目标检测领域,数据集扮演着不可或缺的角色。作为评估和比较检测器性能的共同基础,数据集的质量直接影响着该领域的发展。本节将划分为两个主要部分,即遥感领域的常规目标检测数据集和专门的遥感小目标检测数据集,以下是几个具有代表性的数据集,详细介绍参见表4。

3.1.1 遥感领域常规目标检测数据集

1) NWPU VHR-10 (Northwestern Polytechnical University very-high-resolution 10-class dataset)(Cheng等,2014)是西北工业大学模式识别与智能系统课题组于2014年提出的一个专门为高分辨率(very high resolution, VHR)遥感图像中的目标检测任务设计的

数据集,包含800幅高分辨率图像,共10个类别,包括飞机、船舶、储油罐、棒球场、网球场、篮球场、田径场、港口、桥梁和直升机等。该数据集的提出对高分辨率遥感小目标检测有很高的指导意义。

2)x View(Lam等,2018)数据集是2018年美国国防部创新实验部门举行“x View探测挑战赛”时提出,是RSOD中最大的公开数据集之一,包含60个类别、大约100万个标记目标。该数据集覆盖了超过1 400 km²的区域,具有丰富多样的场景。

3) UAVDT (unmanned aerial vehicle detection and tracking dataset)(Du等,2018)数据集是中国科学院大学于2018年提出,包括无人机拍摄的100多个视频序列,涵盖城市街道、高速公路以及停车场等多种场景。该数据集的目标类别主要包括行人、车辆等,数据集中包含8 416个目标实例,重点在于无人机拍摄视角下的目标检测和跟踪,旨在推动无人机监控和识别技术的发展。

4) DIOR(detection in optical remote sensing images)(Li等,2019)数据集中的图像来源于不同的遥感平台和传感器,拍摄时间跨度大,覆盖的地理区域广泛,共23 463幅光学图像,空间分辨率为0.5 m至30 m不等。数据集中的目标总数为192 472个。2022年,Cheng等人(2022)构建DIOR-R数据集,用于旋转目标检测任务。

5) DOTA (dataset for object detection in aerial images)(Xia等,2018)数据集由武汉大学遥感国家重点实验室和华科电信学院等联合发布。它的版本演变包括DOTA 1.0、DOTA 1.5和DOTA 2.0。DOTA 1.0于2018年提出,包含15个目标类别:飞机、轮船、储罐、棒球场、网球场、游泳池、田径场、港口、桥梁、大型车辆、小型车辆、直升机、环岛、足球场和篮球场,共计2 806幅航空图像,188 282个标注框。DOTA 1.5于2019年提出,在DOTA 1.0的基础上,DOTA 1.5增加了更加细化的目标类别,并引入了更多极小目标的标注,使得检测任务更加复杂。DOTA 2.0进一步扩展了数据集的规模,涵盖18个目标类别,共计11 268幅图像,标注框数量超过1 793 658个。该版本的数据集更加多样化,目标尺度和复杂度显著增加,适用于更广泛的场景下的目标检测任务。

6) VisDrone(Zhu等,2018)由天津大学机器学习和数据挖掘实验室2018年提出,数据集包含了成千

表 4 RSOD 领域广泛使用的数据集的比较

Table 4 Comparison of widely used datasets in the RSOD field

名称	标注	空间分辨率	类别	图像数量/幅	实例数量/幅
UCAS-AOD(Zhu 等,2015)	OBB	0.3~2 m	2	910	6 029
VEDAI(Razakarivony 和 Jurie,2016)	OBB	0.5 m	9	1 268	2 950
NWPU VHR-10(Cheng 等,2014)	HBB	0.8~2 m	10	800	3 775
SSDD(Li 等,2017)	HBB	1~15 m	1	1 160	2 456
VisDrone(Zhu 等,2018)	HBB	-	10	10 209	542 000
x View(Lam 等,2018)	HBB	0.3 m	60	1 413	1 000 000
DOTA-v1.0(Xia 等,2018)	HBB/OBB	0.1~1 m	15	2 806	188 282
DOTA-v1.5(Xia 等,2018)	HBB/OBB	0.1~1 m	16	2 806	472 089
DIOR(Li 等,2019)	HBB	0.5~30 m	20	23 463	190 288
Tiny Person(Yu 等,2020)	HBB	-	1	1 610	72 651
VISO(Yin 等,2021)	HBB	-	4	17 730	1 646 038
DOTA-v2.0(Xia 等,2018)	HBB/OBB	0.1~4.5 m	18	11 268	1 793 658
AI-TOD(Wang 等,2020)	HBB	-	8	28 036	700 621
FAIR1M(Sun 等,2022)	OBB	0.3~0.8 m	37	42 796	1 020 579
AI-TOD v2(Xu 等,2022a)	HBB	-	8	28 036	752 745
SODA-A(Cheng 等,2023)	OBB	-	9	2 510	800 203
USOD(Zhang 等,2024)	HBB	0.4 m	1	3 000	43 378
DTOD(Zhao 等,2024)	HBB	-	2	11 600	1 019 800

注:HBB 和 OBB 分别指水平边界框和定向边界框。“-”表示无法考据结果。

上万幅高分辨率图像和视频片段,主要拍摄于真实世界中的复杂场景,如城市街道、乡村道路、停车场、建筑物等。VisDrone 数据集包括多个版本,其中 VisDrone2018、VisDrone2019 和 VisDrone2021 是比较常见的版本。不同版本的数据集在规模和标注上有所扩展。目前 VisDrone2021 提供 10 209 幅图像、超过 90 000 个标注框以及 263 个视频片段。该数据集结合了无人机拍摄的高分辨率图像与复杂场景的挑战,推动了无人机图像中目标检测和跟踪技术的发展。

7) FAIR1M (fine-grained object recognition in high-resolution remote sensing imagery)(Sun 等,2022) 是中国科学院空天信息创新研究院 2022 年发布的 RSI 细粒度目标识别数据集,包括 5 个类别和 37 个细粒度类别,超过 40 000 幅图像和超过 100 万个由定向边界框注释的对象,场景覆盖全球上百个典型城市、乡镇等。考虑到不同时间拍摄的遥感图像中蕴含了更为丰富的目标关联信息,提供了针对同一

地区在不同时间段获取的影像数据。

3. 1. 2 遥感小目标数据集

在目标检测领域,小目标数据集的稀缺性以及数据集中多样性不足是一个广泛存在的挑战。这种问题不仅限制了模型在多样化场景下的泛化能力,也对提升小目标检测算法的性能带来了诸多困难。由于小目标数据集的数量有限,且通常涵盖的场景和目标类别较为单一,在开发和评估检测算法时,常常面临数据不足的问题。这种数据的不平衡性进一步导致了模型在处理复杂背景或不同尺度变化时的表现不佳,现有的数据资源显得尤为不足。由于训练中的样本数量较少,使得模型难以充分学习目标特征,从而影响了小目标检测算法的性能评估和泛化能力。

1) Tiny Person(Yu 等,2020)数据集是由中国科学院自动化研究所的计算机视觉与模式识别课题组于 2020 年提出。该数据集专注于极小目标的行人检测任务,只针对微小人物。该数据集从高分辨率

视频中进行采样,包含大量高分辨率图像,图像中标注了极小尺度的行人目标,手工注释 72 651 个带有框的目标。

2) AI-TOD(Wang 等, 2020)航空图像数据集由武汉大学夏桂松课题组于 2021 年提出,包含 28 036 幅航拍图像,涵盖 8 个类别,总共有 700 621 个目标实例。与现有的航拍图像目标检测数据集相比, AI-TOD 中目标的平均尺寸约为 12.8 像素,远小于其他数据集。该数据于 2023 年发布第 2 个版本 AI-TOD v2,新增了多个目标类别:小型无人机、小型船只(如渔船、快艇等)、动物(涵盖野生动物和家畜)、小型车辆(如摩托车、三轮车)及基础设施(如路标、交通信号灯、小型建筑物等)。此外, AI-TOD v2 还特别引入了在极端环境下的目标(如夜间和雾天等极端环境下的目标),以提升数据集在不良天气和暗弱光照条件下的适用性。

3) SODA-A(Cheng 等, 2023)是西北工业大学程堪课题组于 2023 年发布的数据集,专为 RSI 中的微小目标检测而设计。该数据集包含 2 510 幅图像,平均图像大小为 $4\,761 \times 2\,777$ 像素和 800 203 个带有 OBB 注释的目标。该数据集中有 9 个类别,包括飞机、直升机、小型车辆、大型车辆、船舶、集装箱、储罐、游泳池和风车。此外,他们还提出一个应用于自动驾驶场景的数据集 SODA-D,专注于在复杂驾驶场景下检测尺寸极小的目标。

4) USOD(unmanned aerial imagery small object detection dataset)(Zhang 等, 2024)由南京航空航天大学闫钧华课题组于 2024 年提出,共包含 3 000 幅图像,包含 43 378 个车辆实例。尺寸小于 16×16 像素的目标占比为 96.3%,尺寸小于 32×32 像素的目标占比为 99.9%。

5) DTOD(dense tiny object detection)(Zhao 等, 2024)数据集由安徽大学多模态认知计算重点实验室于 2024 年发布,包含 11 600 幅图像和 1 019 800 个实例,对象的平均绝对尺寸为 12.9 像素,每幅图像平均包含 88 个对象。该数据集包含大量注释良好且密集目标实例,目前为密集小目标比例最高、规模最大的数据集,单幅图像甚至包含 11 112 个实例。

3.2 评估指标

给定真实边界框(GT box)和预测框,通过计算交并比(IoU)衡量二者重叠程度。若 IoU 预测标签与 GT Box 标签一致且大于预设阈值,当前检测框为

正样本(true positive, TP),否则为负样本(false positive, FP)。未被检测到的 GT(ground truth)判定为假阴性(false negative, FN),即漏检。统计 TP、FP 和 FN 数量,计算平均精度均值(mean average precision, mAP)以评估检测器性能。

平均精度(average precision, AP)首次在 VOC2007(visual object classes 2007)挑战赛引入,给定置信度和 IoU 阈值可计算出召回率和精度,对不同召回率下精度分数求平均得 mAP。MS COCO 数据集推出后,引入新指标,即 0.5 到 0.95(间隔 0.05)多个 IoU 阈值上的平均精度(AP)。按目标尺度划分采用不同评价指标, AP_{50} 和 AP_{75} 分别是 IoU 为 0.5 和 0.75 时的平均精度,小、中、大尺度目标平均精度表示为 AP_s 、 AP_m 、 AP_l 。针对极小目标,用 AP_{eS} 、 AP_{rS} 和 AP_{gS} 评估,不同数据集对“极小”目标定义有差异。这些指标在检测领域已得到广泛应用。

3.3 遥感小目标检测结果对比

为进一步展示现阶段具有代表性小目标检测方法在各数据集上的检测效果,对部分方法在 VisDrone2021、UAVDT、AI-TOD 和 SODA-D 数据集上进行性能对比。

表 5 为不同方法在不同数据集上的性能比较。综合来看,不同方法在不同数据集上性能差异明显,没有一种方法能在所有数据集上都表现最佳。CDMNet(Duan 等, 2021)在 UAVDT 数据集表现最佳。Liu 等人(2024)提出的 YOLC 在 VisDrone 和 UAVDT 数据集平均性能突出,检测能力均衡,持续提高小型、中型和大型目标的准确性。通过 LSM 模块, YOLC 在 VisDrone 数据集以 39.6% 的 mAP 取得最佳成绩,还能检测出许多一般检测器易遗漏的小目标。

可以看到, YOLC 在这两个数据集都有不错的表现,但不同方法在两个数据集的性能表现仍存在差异。究其原因, VisDrone 数据集目标种类繁多,尺度变化大,且有大量小目标; UAVDT 数据集目标类型则相对集中在特定任务相关的目标,目标尺度变化相对较小。数据特性的明显区别,使得不同方法在这两个数据集上难以同时达到最佳性能。

Xu 等人(2022b)提出的 RFLA 在 AI-TOD 数据集上表现出色,达到了最佳性能。该方法利用特征感受野遵循高斯分布的先验信息,提出一个新的高斯感受野,并基于此设计了一种标签分配方法,有效实

表 5 不同方法在 VisDrone, UAVDT, AI-TOD 和 SODA-D 数据集上的性能比较
Table 5 Performance comparison of different methods on the VisDrone, UAVDT, AI-TOD and SODA-D datasets

方法	VisDrone			UAVDT			AI-TOD			SODA-D		
	AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅
Faster R-CNN(Ren, 2015)	21.8	41.8	20.1	5.8	17.4	2.5	12.8	29.9	9.4	28.9	59.7	24.2
RetinaNet(Lin 等, 2017b)	14.4	24.1	15.5	-	-	-	8.9	24.2	4.6	28.2	57.6	23.7
Cascade RCNN(Cai 和 Vasconcelos, 2018)	16.0	16.0	15.0	-	-	-	15.1	34.2	11.2	31.2	59.9	27.8
CenterNet(Duan 等, 2020)	27.8	14.2	1.7	16.4	29.7	16.6	13.4	39.2	5.0	21.5	48.8	15.6
FCOS(Deng 等, 2021)	16.6	28.8	16.7	-	-	-	12.6	30.4	8.1	23.9	49.5	19.9
ClusDet(Yang 等, 2019a)	28.4	53.2	26.4	13.7	26.5	12.5	-	-	-	-	-	-
CDMNet(Duan 等, 2021)	30.7	51.3	32.0	20.7	35.5	22.4	-	-	-	-	-	-
RFLA(Xu 等, 2022b)	-	-	-	-	-	-	25.7	58.9	18.8	29.7	60.2	25.2
YOLC(Liu 等, 2024)	39.6	63.7	41.6	19.3	30.9	20.1	-	-	-	-	-	-

注:加粗字体表示各列最优结果。“-”表示无法考据结果。

现了微小目标的平衡学习。相比之下, Cascade RCNN(Cai 和 Vasconcelos, 2018)在 SODA-D 数据集上表现最佳。SODA-D 数据集与 AI-TOD 数据集的显著区别在于, SODA-D 包含了大量的目标遮挡情况,这使得检测任务更加复杂和具有挑战性。Cascade RCNN 通过多阶段的回归方法有效应对了这一问题,尤其在高遮挡的情况下,能够较好地恢复目标的位置。总的来说, RFLA 在 AI-TOD 数据集上优势明显,得益于其对微小目标的精确学习,而 Cascade RCNN 在 SODA-D 数据集集中的表现则得益于其良好的遮挡处理能力。

Faster R-CNN(Ren 等, 2015)作为有锚目标检测方法,依赖预定义锚框进行区域提议和定位,在 SODA-D 数据集表现优于 CenterNet(Duan 等, 2020),但在其他 3 个数据集性能差,暴露了其在密集小目标和复杂背景下的局限,更适用于简单、目标分布均匀的场景。CenterNet 则在稀疏目标和动态场景有优势。锚框机制在小目标检测中表现受限,小目标尺寸、形状多变,与固定尺寸和比例的锚框匹配度低,难以准确覆盖,影响特征提取和检测精度、召回率,还会产生大量冗余候选框,增加计算量并干扰判断。

为更直观呈现检测结果,本文列举了不同算法在各类数据集上的可视化结果,具体如图 3 所示。通过对图中结果的分析,能够清晰地观察到,这些算

法有效地应对了遥感小目标检测中可用特征少、分布不均匀和大量小目标等挑战,显著提升了检测性能与准确性。

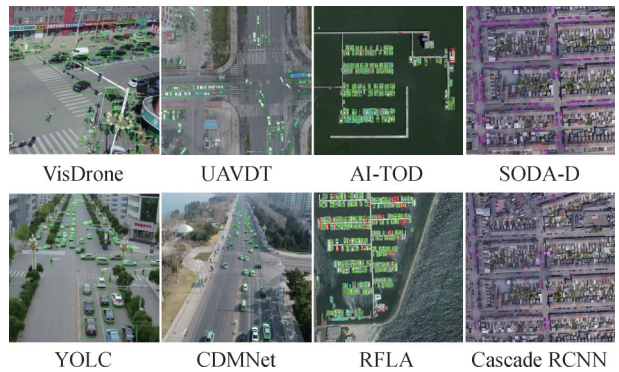


图 3 不同方法在不同数据集上的可视化结果
Fig. 3 Visualization results of different methods on different datasets

在复杂环境与多样化目标面前,单一模型在某些特定场景下依然表现出色,优势十分有限。目前科学研究重点聚焦于大模型的设计与应用领域。这主要是因为大模型具备更深的网络结构,能够执行更丰富的特征提取,完成更复杂的推理过程,从而极大提升对复杂目标和多变场景的适应能力。大模型不仅能敏锐捕捉图像中细微的特征,还能通过多层次的表达,增强模型在面对不同尺寸、角度以及遮挡条件下目标时的鲁棒性。为了进一步探究不同方法

的性能差异,对部分方法在DOTA 1.0数据集上展开了性能对比,如表6所示。

在遥感图像目标检测领域,高分辨率图像中多个小规模目标的识别是一大关键特征,这对轻质骨干网络构成了严峻挑战。Lu等人(2024)提出的轻量级大模型骨干网络DecoupleNet,创新性地融合了特征集成降采样模块与多支路特征解耦模块,展现出强大的通用性,能够高效完成分类、分割以及目标

检测等多种任务。

在DOTA 1.0目标检测任务中,DecoupleNet检测精度高达78.0%,远超同类模型。同时,该模型在参数量和计算复杂度控制上表现出色,参数量仅为23.3 M,每秒浮点运算次数(floating point operations per second, FLOPs)为142.4 G,远低于其他模型,在保证检测精度的同时,极大地提升了运算效率。

表6 不同方法在DOTA 1.0数据集上的性能比较
Table 6 Performance comparison of different methods on the DOTA 1.0 dataset

方法	AP/%	Params/M	FLOPs/G
SCRDet(Yang等,2020)	72.6	—	—
ROI Trans(Ding等,2019)	74.0	55.1	225.3
G.V.(Xu等,2021b)	75.0	41.1	211.3
AOPG(Cheng等,2022)	75.3	—	—
ReDet(Han等,2021)	76.2	—	—
Oriented RepPoints(Li等,2022b)	77.6	37.3	200.5
Oriented R-CNN-DecoupleNet D0(Lu等,2024)	77.3	18.6	124.2
Oriented R-CNN-DecoupleNet D2(Lu等,2024)	78.0	23.3	142.4

注:加粗字体表示各列最优结果。“—”表示无法考据结果。

4 结 语

遥感小目标检测一直是遥感领域的一个基本但具有挑战性的研究课题。得益于深度学习技术的快速发展,遥感小目标检测受到广泛关注并取得令人瞩目的成就。本文回顾了130余篇研究论文,全面系统地阐述基于深度学习的遥感小目标检测任务的研究现状。首先,根据地理空间目标的特点总结遥感小目标检测的7大挑战,即可用特征少、度量不匹配、图像幅面宽、背景干扰大、分布不均匀、方向不确定和数据集稀缺。其次,小目标的定义在不同数据集中可能会有所不同,这取决于数据集的特性、任务需求以及目标检测的具体应用场景。因此,通过不同数据集的角度论述小目标的定义。接下来,依次对挑战提出相应的解决方案,对每一类的方法进行回顾和总结,包括传统的目标检测算法和基于深度学习的目标检测算法,并根据不同挑战对现有小目标算法成果进行全面阐述。最后,介绍小目标检测领域的典型基准数据集、评估指标以及在不同数据

集上的性能分析。

考虑到现有小目标检测方法的局限性,本节讨论一些有前景的进一步研究方向。小目标检测在未来有着广阔的展望,其发展方向主要包括但不限于以下6个方面:

- 1)模型轻量化。现有遥感小目标检测模型获得不错的检测精度,大多依赖于复杂的网络设计,计算复杂度高,参数量大。模型轻量化对于遥感小目标检测至关重要,它能够在计算资源受限的环境中实现实时数据处理,提升处理速度和能效,适应多样化硬件平台,同时降低部署和运行成本,从而满足广域观测和高分辨率成像等实际应用的需求。轻量化的核心在于,在不显著牺牲深度神经网络精度的前提下,优化其存储需求和计算效率。这种优化不仅可以加速模型的推理过程,提高响应速度,还能有效降低设备的功耗,拓展其在多样化应用场景中的适用性。在遥感小目标检测领域,模型轻量化是应对目标尺寸微小、分布密集和背景复杂等挑战的关键策略。
- 2)多/跨模态融合。现有遥感小目标检测多是

由单个传感器进行拍摄,获得的信息单一,不能完全涵盖目标的真实特性。不同模态(如光学、雷达、红外等)捕捉到的目标信息各有侧重,通过融合这些信息,可以弥补单一模态的局限性,提供更加全面和准确的目标特征。多模态信息融合到小目标检测中,包括图像、文本和声音等多种数据类型,可以提供更全面、更准确的目标检测结果。这样的融合将带来更丰富的语义理解和更准确的目标定位,进一步提升小目标检测的性能。基于单模态目标检测的探测器在实际应用中往往存在性能偏差,并难以满足需求。相比之下,多模态遥感图像集成了不同传感器的特点。多模态RSI的集成处理可以提高对场景的解释能力,实现对地理空间目标更客观、更全面的理解,从而为进一步提高小目标检测的性能提供了可能。

3)小样本学习。遥感领域中,获取高质量的大规模标注数据既费时又昂贵,并且遥感任务中经常会出现新的目标类型或环境变化,需要模型提高对未见过的数据或少量数据的泛化能力,增强模型的实用性。小样本目标检测(few-shot object detection, FSOD)是指仅使用有限数量(不超过30个)样本来检测新类别。遥感领域的FSOD和小目标检测进一步融合,能有效推动在复杂背景中高效识别微小目标的能力。随着元学习和迁移学习技术的不断进步,FSOD模型将能够从有限的标注数据中快速学习并适应新类别的小目标检测任务。这些技术进步将使模型在基础训练阶段积累的丰富特征和先验知识,能够更好地支持小样本的微调阶段,提升对新类小目标的检测精度和鲁棒性。此外,结合自适应数据增强、上下文信息增强等方法,未来的FSOD模型将更加高效地在资源受限的环境中运行,实现更广泛的应用,如无人机监控、自然灾害监测和环境保护等领域。

4)低监督学习。由于小目标尺寸较小,常被背景掩盖,且细节容易受到复杂背景和噪声的干扰,增加了检测的难度,并且海量未标注数据无法直接用于训练,导致数据资源浪费。此外,标注数据的稀缺性在特定任务中限制了模型的学习能力。而且传统监督学习方法依赖大量标注数据,难以泛化到新场景或目标,而低监督学习通过减少对标注数据的依赖,能更好地利用现有数据资源。通过少量标注数据或无标注数据进行训练,可以大幅减少对标注数

据的依赖,显著降低数据标注的成本和时间。通过充分利用未标注数据,低监督学习有助于提升模型的泛化能力,使模型在不同类型的遥感图像中表现优异。然而,低监督学习也面临标注不均衡、噪声和误差传递等挑战,通过多模态融合、自适应学习和人机协同标注等技术的进步,低监督学习在遥感小目标检测中的应用前景将更加广阔。

5)合适的指标。在遥感图像中的小目标检测任务中,选择合适的评价指标至关重要。传统目标检测任务通常采用平均精度均值(mAP)作为评价标准。mAP综合考虑了精度(precision)和召回率(recall),通过计算每个类别的AP,再对所有类别的AP取平均得出。而在计算精度和召回率时,通常需要借助IoU判断预测框是否与真实框匹配。然而,对于小目标检测,特别是极小目标,mAP和IoU的适用性存在显著局限性。IoU对边界框的精确匹配要求较高,而在小目标检测中,由于目标尺寸非常小,即使是像素级的偏移,都会导致IoU显著下降。这种情况会放大mAP对模型性能的要求,从而无法准确反映模型的实际检测能力。这种对边界框精度的苛刻要求,使得mAP在小目标检测中的适用性受到限制,尤其是在人工标注误差较大的情况下。为了更准确地评估小目标检测模型的性能,未来需要探索 and 引入更适合的评价指标。

6)数据集构建。目前,遥感小目标检测领域面临的一个主要挑战是现有数据集的稀缺和单一性。为了突破这一限制,推动技术的进一步发展,构建更多且多样化的微小目标数据集显得尤为重要。数据集的构建应涵盖广泛的目标类型,如各种微小的车辆、船只、建筑物、基础设施,以及农业、环保、海洋监测等特定领域的小目标。同时,这些数据集还应涵盖多样化的场景和复杂的背景。通过集成高分辨率和多光谱数据,数据集将能够提供丰富的光谱和空间信息,支持细致的目标特征提取和识别。此外,借助自动化标注技术和生成对抗网络等工具,可以大幅度提高数据集的标注效率和多样性。

参考文献(References)

- Akyon F C, Altinuc S O and Temizel A. 2022. Slicing aided hyper inference and fine-tuning for small object detection//Proceedings of 2022 IEEE International Conference on Image Processing (ICIP).

- Bordeaux, France: IEEE: 966-970 [DOI: 10.1109/ICIP46576.2022.9897990]
- Bernhard M and Schubert M. 2021. Correcting imprecise object locations for training object detectors in remote sensing applications. *Remote Sensing*, 13(24): #4962 [DOI: 10.3390/rs13244962]
- Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2004.10934.pdf>
- Bosquet B, Cores D, Seidenari L, Brea V M, Mucientes M and Del Bimbo A. 2023. A full data augmentation pipeline for small object detection based on generative adversarial networks. *Pattern Recognition*, 133: #108998 [DOI: 10.1016/j.patcog.2022.108998]
- Cai Z W and Vasconcelos N. 2018. Cascade R-CNN: delving into high quality object detection//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 6154-6162 [DOI: 10.1109/CVPR.2018.00644]
- Chalavadi V, Jeripothula P, Datla R, Ch S B and Mohan C K. 2022. mSODANet: a network for multi-scale object detection in aerial images using hierarchical dilated convolutions. *Pattern Recognition*, 126: #108548 [DOI: 10.1016/j.patcog.2022.108548]
- Chen C R, Zhang Y, Lv Q X, Wei S, Wang X R, Sun X and Dong J Y. 2019a. RRNet: a hybrid detector for object detection in drone-captured images//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshops*. Seoul, Korea (South): IEEE: #18 [DOI: 10.1109/iccvw.2019.00018]
- Chen C, Liu M Y, Tuzel O and Xiao J. 2017. R-CNN for small object detection//*Proceedings of the 13th Asian Conference on Computer Vision*. Taipei, China: Springer: 14-230 [DOI: 10.1007/978-3-319-54193-8_14]
- Chen C Y, Gong W G, Chen Y L and Li W H. 2019b. Object detection in remote sensing images based on a scene-contextual feature pyramid network. *Remote Sensing*, 11(3): #339 [DOI: 10.3390/rs11030339]
- Chen J J, Hong H S, Song B, Guo J, Chen C and Xu J J. 2023. MDCT: multi-kernel dilated convolution and transformer for one-stage object detection of remote sensing images. *Remote Sensing*, 15(2): #371 [DOI: 10.3390/rs15020371]
- Chen K A, Li J G, Lin W Y, See J, Wang J, Duan L Y, Chen Z B, He C W and Zou J N. 2019c. Towards accurate one-stage object detection with AP-loss//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 5114-5122 [DOI: 10.1109/cvpr.2019.00526]
- Chen Y K, Zhang P Z, Li Z M, Li Y W, Zhang X Y, Meng G F, Xiang S M, Sun J and Jia J Y. 2020. Stitcher: feedback-driven data provider for object detection [EB/OL]. [2024-12-12]. <http://arxiv.org/pdf/2004.12432.pdf>
- Chen Z G, Wu K H, Li Y B, Wang M J and Li W. 2019d. SSD-MSN: an improved multi-scale object detection network based on SSD. *IEEE Access*, 7: 80622-80632 [DOI: 10.1109/ACCESS.2019.2923016]
- Cheng G, Han J W, Zhou P C and Guo L. 2014. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 98: 119-132 [DOI: 10.1016/j.isprsjprs.2014.10.002]
- Cheng G, Lang C B, Wu M X, Xie X X, Yao X W and Han J W. 2021. Feature enhancement network for object detection in optical remote sensing images. *Journal of Remote Sensing*, 2021: #9805389 [DOI: 10.34133/2021/9805389]
- Cheng G, Wang J B, Li K, Xie X X, Lang C B, Yao Y Q and Han J W. 2022. Anchor-free oriented proposal generator for object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5625411 [DOI: 10.1109/TGRS.2022.3183022]
- Cheng G, Yuan X, Yao X W, Yan K B, Zeng Q H, Xie X X and Han J W. 2023. Towards large-scale small object detection: survey and benchmarks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11): 13467-13488 [DOI: 10.1109/TPAMI.2023.3290594]
- Cubuk E D, Zoph B, Mane D, Vasudevan V and Le Q V. 2018. Auto-Augment: learning augmentation policies from data [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/1805.09501.pdf>
- Dai L H, Liu H, Tang H, Wu Z W and Song P H. 2023. AO2-DETR: arbitrary-oriented object detection transformer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(5): 2342-2356 [DOI: 10.1109/TCSVT.2022.3222906]
- Dai X Y, Chen Y P, Yang J W, Zhang P C, Yuan L and Zhang L. 2021. Dynamic DETR: end-to-end object detection with dynamic attention//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 2968-2977 [DOI: 10.1109/icc48922.2021.00298]
- Deng S T, Li S, Xie K, Song W F, Liao X, Hao A M and Qin H. 2021. A global-local self-adaptive network for drone-view object detection. *IEEE Transactions on Image Processing*, 30: 1556-1569 [DOI: 10.1109/TIP.2020.3045636]
- Ding J, Xue N, Long Y, Xia G S and Lu Q K. 2019. Learning RoI transformer for oriented object detection in aerial images//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 2844-2853 [DOI: 10.1109/cvpr.2019.00296]
- Ding P, Zhang Y, Deng W J, Jia P and Kuijper A. 2018. A light and faster regional convolutional neural network for object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 141: 208-218 [DOI: 10.1016/j.isprsjprs.2018.05.005]
- Du D W, Qi Y K, Yu H Y, Yang Y F, Duan K W, Li G R, Zhang W G, Huang Q M and Tian Q. 2018. The unmanned aerial vehicle benchmark: object detection and tracking//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 375-391 [DOI: 10.1007/978-3-030-01249-6_23]
- Duan C Z, Wei Z W, Zhang C, Qu S Y and Wang H P. 2021. Coarse-grained density map guided object detection in aerial images//*Proceedings of 2021 IEEE/CVF International Conference on Computer*

- Vision. Montreal, Canada: IEEE: 2789-2798 [DOI: 10.1109/iccwv54120.2021.00313]
- Duan K W, Du D W, Qi H G and Huang Q M. 2020. Detecting small objects using a channel-aware deconvolutional network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(6): 1639-1652 [DOI: 10.1109/TCSVT.2019.2906246]
- Ferdous S N, Mostofa M and Nasrabadi N M. 2019. Super resolution-assisted deep aerial vehicle detection//*Proceedings of 2019 SPIE Conference on Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*. Baltimore, USA: SPIE: 1100617 [DOI: 10.1117/12.2519045]
- Fu J M, Sun X, Wang Z R and Fu K. 2021. 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*, 59(2): 1331-1344 [DOI: 10.1109/TGRS.2020.3005151]
- Fu R H, Chen C C, Yan S, Heidari A A, Wang X C, Escorcía-Gutiérrez J, Mansour R F and Chen H L. 2023. Gaussian similarity-based adaptive dynamic label assignment for tiny object detection. *Neurocomputing*, 543: #126285 [DOI: 10.1016/j.neucom.2023.126285]
- Gevorgyan Z. 2022. Siou loss: more powerful learning for bounding box regression [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2205.12740.pdf>
- Gong H, Mu T K, Li Q X, Dai H S, Li C L, He Z P, Wang W J, Han F, Tuniyazi A, Li H Y, Lang X C, Li Z Y and Wang B. 2022. Swin-transformer-enabled YOLOv5 with attention mechanism for small object detection on satellite images. *Remote Sensing*, 14(12): #2861 [DOI: 10.3390/rs14122861]
- Halevy A, Norvig P and Pereira F. 2009. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2): 8-12 [DOI: 10.1109/mis.2009.36]
- Han J M, Ding J, Xue N and Xia G S. 2021. ReDet: a rotation-equivariant detector for aerial object detection//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, Tennessee, USA: IEEE: 2785-2794 [DOI: 10.1109/CVPR46437.2021.00281]
- He J B, Erfani S, Ma X J, Bailey J, Chi Y and Hua X S. 2021. Alpha-IoU: a family of power intersection over union losses for bounding box regression//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. [s.l.]: Curran Associates Inc.: #1547 [DOI: 10.5555/3540261.3541808]
- He Y H, Zhu C C, Wang J R, Savvides M and Zhang X Y. 2019. Bounding box regression with uncertainty for accurate object detection//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 2883-2892 [DOI: 10.1109/cvpr.2019.00300]
- Hong M B, Li S W, Yang Y C, Zhu F Y, Zhao Q J and Lu L. 2021. SSPNet: scale selection pyramid network for tiny person detection from UAV images. *IEEE Geoscience and Remote Sensing Letters*, 19: #8018505 [DOI: 10.1109/lgrs.2021.3103069]
- Hou L P, Lu K, Xue J and Li Y Q. 2022. Shape-adaptive selection and measurement for oriented object detection//*Proceedings of the 36th AAAI Conference on Artificial Intelligence*. [s.l.]: AAAI Press: 923-932 [DOI: 10.1609/aaai.v36i1.19975]
- Huang W, Li G Y, Jin B H, Chen Q Q, Yin J R and Huang L. 2022a. Scenario context-aware-based bidirectional feature pyramid network for remote sensing target detection. *IEEE Geoscience and Remote Sensing Letters*, 19: #6505005 [DOI: 10.1109/lgrs.2021.3135935]
- Huang Y C, Chen J X and Huang D. 2022b. UFPMP-Det: toward accurate and efficient object detection on drone imagery//*Proceedings of the 36th AAAI Conference on Artificial Intelligence*. [s.l.]: AAAI Press: 1026-1033 [DOI: 10.1609/aaai.v36i1.19986]
- Huang Y X, Liu H I, Shuai H H and Cheng W H. 2024. DQ-DETR: DETR with dynamic query for tiny object detection [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2404.03507.pdf>
- Jiang N, Yu X H, Peng X K, Gong Y Q and Han Z J. 2021. SM+: refined scale match for tiny person detection//*Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing*. Toronto, Canada: IEEE: 1815-1819 [DOI: 10.1109/icassp39728.2021.9414162]
- Kaur P, Khehra B S and Mavi B S. 2021. Data augmentation for object detection: a review//*Proceedings of 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*. Lansing, MI, USA: IEEE: 537-543 [DOI: 10.1109/MWSCAS47672.2021.9531849]
- Kisantal M, Wojna Z, Murawski J, Naruniec J and Cho K. 2019. Augmentation for small object detection [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2004.12178.pdf>
- Koyun O C, Keser R K, Akkaya İ B and Töreyin B U. 2022. Focus-and-Detect: a small object detection framework for aerial images. *Signal Processing: Image Communication*, 104: #116675 [DOI: 10.1016/j.image.2022.116675]
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90 [DOI: 10.1145/3065386]
- Lam D, Kuzma R, McGee K, Dooley S, Laielli M, Klaric M, Bulatov Y and McCord B. 2018. xView: objects in context in overhead imagery [EB/OL]. [2024-12-12]. <http://arxiv.org/pdf/1802.07856.pdf>
- Law H and Deng J. 2018. CornerNet: detecting objects as paired keypoints//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 765-781 [DOI: 10.1007/978-3-030-01264-9_45]
- Lee C, Park S, Song H, Ryu J, Kim S, Kim H, Pereira S and Yoo D. 2022. Interactive multi-class tiny-object detection//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 14116-14125 [DOI: 10.1109/CVPR52688.2022.01374]
- Li C L, Yang T J N, Zhu S J, Chen C and Guan S Y. 2020. Density map guided object detection in aerial images//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*

- Workshops. Seattle, USA: IEEE: 737-746 [DOI: 10.1109/cvprw50498.2020.00103]
- Li F, Zhang H, Liu S L, Guo J, Ni L M and Zhang L. 2022a. DN-DETR: accelerate DETR training by introducing query denoising//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 13609-13617 [DOI: 10.1109/cvpr52688.2022.01325]
- Li J W, Qu C W and Shao J Q. 2017. Ship detection in SAR images based on an improved faster R-CNN//Proceedings of 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA). Beijing, China: IEEE: 1-6 [DOI: 10.1109/BIGSAR DATA.2017.8124934]
- Li W T, Chen Y J, Hu K X and Zhu J K. 2022b. Oriented RepPoints for aerial object detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 1819-1828 [DOI: 10.1109/cvpr52688.2022.00187]
- Li W T, Li W Y, Yang F and Wang P. 2019. Multi-scale object detection in satellite imagery based on YOLT//Proceedings of 2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama, Japan: IEEE: 162-165 [DOI: 10.1109/IGARSS.2019.8898170]
- Li X, Wang W H, Wu L J, Chen S, Hu X L, Li J, Tang J H and Yang J. 2020c. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #1763 [DOI: 10.5555/3495724.3497487]
- Li Y X, Hou Q B, Zheng Z H, Cheng M M, Yang J and Li X. 2023. Large selective kernel network for remote sensing object detection//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 16748-16759 [DOI: 10.1109/icc51070.2023.01540]
- Li Y Y, Huang Q, Pei X, Chen Y Q, Jiao L C and Shang R H. 2021. Cross-layer attention network for small object detection in remote sensing imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 2148-2161 [DOI: 10.1109/JSTARS.2020.3046482]
- Liang X, Zhang J, Zhuo L, Li Y Z and Tian Q. 2020. Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis. IEEE Transactions on Circuits and Systems for Video Technology, 30(6): 1758-1770 [DOI: 10.1109/tcsvt.2019.2905881]
- Lim S, Kim I, Kim T, Kim C and Kim S. 2019. Fast AutoAugment//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: #598 [DOI: 10.5555/3454287.3454885]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017a. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017b. Focal loss for dense object detection//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2999-3007 [DOI: 10.1109/iccv.2017.324]
- Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]
- Liu C G, Gao G S, Huang Z Y, Hu Z H, Liu Q J and Wang Y H. 2024. YOLC: you only look clusters for tiny object detection in aerial images. IEEE Transactions on Intelligent Transportation Systems, 25(10): 13863-13875 [DOI: 10.1109/TITS.2024.3386928]
- Liu C X, Wang K W, Lu H, Cao Z G and Zhang Z M. 2022. Robust object detection with inaccurate bounding boxes//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 53-69 [DOI: 10.1007/978-3-031-20080-9_4]
- Liu J, Yang S J, Tian L, Guo W, Zhou B Y, Jia J Q and Ling H B. 2019a. Multi-component fusion network for small object detection in remote sensing images. IEEE Access, 7: 128339-128352 [DOI: 10.1109/ACCESS.2019.2939488]
- Liu L L, Muelly M, Deng J, Pfister T and Li L J. 2019b. Generative modeling for small-data object detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6072-6080 [DOI: 10.1109/ICCV.2019.00617]
- Liu S, Qi L, Qin H F, Shi J P and Jia J Y. 2018. Path aggregation network for instance segmentation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8759-8768 [DOI: 10.1109/cvpr.2018.00913]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot MultiBox detector//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]
- Liu Y, Sun P, Wergeles N and Shang Y. 2021. A survey and performance evaluation of deep learning methods for small object detection. Expert Systems with Applications, 172: #114602 [DOI: 10.1016/j.eswa.2021.114602]
- Lu W, Chen S B, Shu Q L, Tang J and Luo B. 2024. DecoupleNet: a lightweight backbone network with efficient feature decoupling for remote sensing visual tasks. IEEE Transactions on Geoscience and Remote Sensing, 62: #4414613 [DOI: 10.1109/TGRS.2024.3465496]
- Lu X C, Ji J, Xing Z Q and Miao Q G. 2021. Attention and feature fusion SSD for remote sensing object detection. IEEE Transactions on Instrumentation and Measurement, 70: #5501309 [DOI: 10.1109/TIM.2021.3052575]
- Lu X Q, Zhang Y L, Yuan Y and Feng Y C. 2020. Gated and axis-concentrated localization network for remote sensing object detection. IEEE Transactions on Geoscience and Remote Sensing, 58(1): 179-192 [DOI: 10.1109/TGRS.2019.2935177]

- Ma T L, Mao M Y, Zheng H H, Gao P, Wang X D, Han S M, Ding E R, Zhang B C and Doermann D. 2021. Oriented object detection with transformer [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2106.03146.pdf>
- Ma T L, Yang Z, Wang J Q, Sun S Y, Ren X Y and Ahmad U. 2022. Infrared small target detection network with generate label and feature mapping. *IEEE Geoscience and Remote Sensing Letters*, 19: #6505405 [DOI: 10.1109/LGRS.2022.3140432]
- Muzammul M and Li X. 2021. A survey on deep domain adaptation and tiny object detection challenges, techniques and datasets [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2107.07927.pdf>
- Özge Ünel F, Özkalayci B O and Çiğla C. 2019. The power of tiling for small object detection//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Long Beach, USA: IEEE: 582-591 [DOI: 10.1109/cvprw.2019.00084]
- Pan X Y, Jia N X, Mu Y Z and Gao X R. 2023. Survey of small object detection. *Journal of Image and Graphics*, 28(9): 2587-2615 (潘晓英, 贾凝心, 穆元震, 高炫蓉. 2023. 小目标检测研究综述. *中国图象图形学报*, 28(9): 2587-2615) [DOI: 10.11834/jig.220455]
- Pang J M, Li C, Shi J P, Xu Z H and Feng H J. 2019. R²-CNN: fast Tiny object detection in large-scale remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8): 5512-5524 [DOI: 10.1109/TGRS.2019.2899955]
- Rabbi J, Ray N, Schubert M, Chowdhury S and Chao D. 2020. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network. *Remote Sensing*, 12(9): #1432 [DOI: 10.20944/preprints202003.0313.v2]
- Ran Q, Wang Q, Zhao B Y, Wu Y F, Pu S L and Li Z J. 2021. Light-weight oriented object detection using multiscale context and enhanced channel attention in remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 5786-5795 [DOI: 10.1109/JSTARS.2021.3079968]
- Ranjan V, Sharma U, Nguyen T and Hoai M. 2021. Learning to count everything//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 3393-3402 [DOI: 10.1109/cvpr46437.2021.00340]
- Razakarivony S and Jurie F. 2016. Vehicle detection in aerial imagery: a small target detection benchmark. *Journal of Visual Communication and Image Representation*, 34: 187-203 [DOI: 10.1016/j.jvcir.2015.11.002]
- Rekavandi A M, Rashidi S, Boussaid F, Hoefs S, Akbas E and Benna-moun M. 2023. Transformers in small object detection: a benchmark and survey of state-of-the-art [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2309.04902.pdf>
- Rekavandi A M, Xu L, Boussaid F, Seghouane A K, Hoefs S and Ben-namoun M. 2025. A guide to image- and video-based small object detection using deep learning: case study of maritime surveillance. *IEEE Transactions on Intelligent Transportation Systems*, 26(3): 2851-2879 [DOI: 10.1109/ITITS.2025.3530678]
- Ren S Q, He K M, Girshick R and Sun J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks//*Proceedings of the 29th International Conference on Neural Information Processing Systems*. Montreal, Canada: MIT Press: #2969250 [DOI: 10.5555/2969239.2969250]
- Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I and Savarese S. 2019. Generalized intersection over union: a metric and a loss for bounding box regression//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 658-666 [DOI: 10.1109/cvpr.2019.00075]
- Simard P Y, Steinkraus D and Platt J C. 2003. Best practices for convolutional neural networks applied to visual document analysis//*Proceedings of the 7th International Conference on Document Analysis and Recognition*. Edinburgh, UK: IEEE: 958-963 [DOI: 10.1109/ICDAR.2003.1227801]
- Sun X, Wang P J, Yan Z Y, Xu F, Wang R P, Diao W H, Chen J, Li J B, Feng Y C, Xu T, Weinmann M, Hinz S, Wang C and Fu K. 2022. FAIR1M: a benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184: 116-130 [DOI: 10.1016/j.isprsjprs.2021.12.004]
- Tong K and Wu Y Q. 2022. Deep learning-based detection from the perspective of small or tiny objects: a survey. *Image and Vision Computing*, 123: #104471 [DOI: 10.1016/j.imavis.2022.104471]
- Tong Z J, Chen Y H, Xu Z W and Yu R. 2023. Wise-IoU: bounding box regression loss with dynamic focusing mechanism [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2301.10051.pdf>
- Van Etten A. 2019. Satellite imagery multiscale rapid detection with windowed networks//*Proceedings of 2019 IEEE Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 735-743 [DOI: 10.1109/WACV.2019.00083]
- Wan L, Zeiler M, Zhang S X, LeCun Y and Fergus R. 2013. Regularization of neural networks using dropconnect//*Proceedings of the 30th International Conference on Machine Learning*. Atlanta, Georgia, USA: JMLR.org: 1058-1066 [DOI: 10.5555/3042817.3043055]
- Wang G T, Xiong Z W, Liu D and Luo C. 2018. Cascade mask generation framework for fast small object detection//*Proceedings of 2018 IEEE International Conference on Multimedia and Expo*. San Diego, USA: IEEE: 1-6 [DOI: 10.1109/ICME.2018.8486561]
- Wang J W, Xu C, Yang W and Yu L. 2021. A normalized Gaussian Wasserstein distance for tiny object detection [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2110.13389.pdf>
- Wang Y, Bashir S M A, Khan M, Ullah Q, Wang R, Song Y L, Guo Z and Niu Y L. 2022. Remote sensing image super-resolution and object detection: benchmark and state of the art. *Expert Systems with Applications*, 197: #116793 [DOI: 10.1016/j.eswa.2022.116793]
- Wang Y, Yang Y L and Zhao X. 2020. Object detection using clustering algorithm adaptive searching regions in aerial images//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 651-664 [DOI: 10.1007/978-3-030-66823-5_39]
- Wang Z J, Fang J W, Dou J and Xue J R. 2019. Small object detection

- on road by embedding focal-area loss//Proceedings of the 10th International Conference on Image and Graphics. Beijing, China: Springer: 657-665 [DOI: 10.1007/978-3-030-34120-6_53]
- Wei L H, Zhang S L, Gao W and Tian Q. 2018. Person transfer GAN to bridge domain gap for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 79-88 [DOI: 10.1109/CVPR.2018.00016]
- Wei S J, Su H, Ming J, Wang C, Yan M, Kumar D, Shi J and Zhang X L. 2020. Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet. *Remote Sensing*, 12 (1) : #167 [DOI: 10.3390/rs12010167]
- Wu D, Chen P F, Yu X H, Li G R, Han Z J and Jiao J B. 2023. Spatial self-distillation for object detection with inaccurate bounding boxes//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 6832-6842 [DOI: 10.1109/ICCV51070.2023.00631]
- Xia G S, Bai X, Ding J, Zhu Z, Belongie S J, Luo J B, Datcu M H, Pelillo M and Zhang L P. 2018. DOTA: a large-scale dataset for object detection in aerial images//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3974-3983 [DOI: 10.1109/CVPR.2018.00418]
- Xu C, Ding J, Wang J W, Yang W, Yu H, Yu L and Xia G S. 2023. Dynamic coarse-to-fine learning for oriented tiny object detection//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 7318-7328 [DOI: 10.1109/cvpr52729.2023.00707]
- Xu C, Wang J W, Yang W, Yu H, Yu L and Xia G S. 2022a. Detecting tiny objects in aerial images: a normalized Wasserstein distance and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190: 79-93 [DOI: 10.1016/j.isprsjprs.2022.06.002]
- Xu C, Wang J W, Yang W, Yu H, Yu L and Xia G S. 2022b. RFLA: Gaussian receptive field based label assignment for tiny object detection//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer Nature Switzerland: 526-543 [DOI: 10.1007/978-3-031-20077-9_31]
- Xu C, Wang J W, Yang W and Yu L. 2021a. Dot distance for tiny object detection in aerial images//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 1192-1201 [DOI: 10.1109/cvprw53098.2021.00130]
- Xu N, Price B, Cohen S, Yang J M and Huang T S. 2016. Deep interactive object selection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 373-381 [DOI: 10.1109/cvpr.2016.47]
- Xu Y C, Fu M T, Wang Q M, Wang Y K, Chen K, Xia G S and Bai X. 2021b. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43 (4) : 1452-1459 [DOI: 10.1109/TPAMI.2020.2974745]
- Xu Y J, Zhu L C, Yang Y and Wu F. 2021c. Training robust object detectors from noisy category labels and imprecise bounding boxes. *IEEE Transactions on Image Processing*, 30: 5782-5792 [DOI: 10.1109/tip.2021.3085208]
- Yaeger L, Lyon R and Webb B. 1996. Effective training of a neural network character classifier for word recognition//Proceedings of the 10th International Conference on Neural Information Processing Systems. Denver Colorado, USA: MIT Press: 807-813 [DOI: 10.5555/2998981.2999095]
- Yang F, Fan H, Chu P, Blasch E and Ling H B. 2019a. Clustered object detection in aerial images//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 8310-8319 [DOI: 10.1109/iccv.2019.00840]
- Yang S J, Tian L, Zhou B Y, Chen D, Zhang D, Xu Z N, Guo W and Liu J. 2020. Inception parallel attention network for small object detection in remote sensing images//Proceedings of 2020 Chinese Conference on Pattern Recognition and Computer Vision. Nanjing, China: Springer: 469-480 [DOI: 10.1007/978-3-030-60633-6_39]
- Yang X, Yan J C, Liao W L, Yang X K, Tang J and He T. 2023. SCRDet++: detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2384-2399 [DOI: 10.1109/TPAMI.2022.3166956]
- Yang X, Yan J C, Ming Q, Wang W T, Zhang X P and Tian Q. 2021a. Rethinking rotated object detection with Gaussian Wasserstein distance loss//Proceedings of the 38th International Conference on Machine Learning. [s.l.]: PMLR: 11830-11841
- Yang X, Yang J R, Yan J C, Zhang Y, Zhang T F, Guo Z, Sun X and Fu K. 2019b. SCRDet: towards more robust detection for small, cluttered and rotated objects//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 8231-8240 [DOI: 10.1109/iccv.2019.00832]
- Yang X, Yang X J, Yang J R, Ming Q, Wang W T, Tian Q and Yan J C. 2021b. Learning high-precision bounding box for rotated object detection via Kullback-Leibler divergence//Proceedings of the 35th International Conference on Neural Information Processing Systems. [s. l.]: Curran Associates Inc.: #1405 [DOI: 10.5555/3540261.3541666]
- Yin Q, Hu Q Y, Liu H, Zhang F, Wang Y Q, Lin Z P, An W and Guo Y L. 2021. Detecting and tracking small and dense moving objects in satellite videos: a benchmark. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5612518 [DOI: 10.1109/TGRS.2021.3130436]
- Yu J H, Jiang Y N, Wang Z Y, Cao Z M and Huang T. 2016. UnitBox: an advanced object detection network//Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam, the Netherlands: ACM: 516-520 [DOI: 10.1145/2964284.2967274]
- Yu X H, Gong Y Q, Jiang N, Ye Q X and Han Z J. 2020. Scale match for tiny person detection//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass, USA: IEEE: 1246-1254 [DOI: 10.1109/WACV45572.2020.9093394]

- Yuan X, Cheng G, Li G, Dai W, Yin W X, Feng Y C, Yao X W, Huang Z L, Sun X and Han J W. 2023. Progress in small object detection for remote sensing images. *Journal of Image and Graphics*, 28(6): 1662-1684 (袁翔, 程焱, 李戈, 戴威, 尹文昕, 冯瑛超, 姚西文, 黄钟铃, 孙显, 韩军伟. 2023. 遥感影像小目标检测研究进展. *中国图象图形学报*, 28(6): 1662-1684) [DOI: 10.11834/jig.221202]
- Yun S, Han D, Chun S, Oh S J, Yoo Y and Choe J. 2019. CutMix: regularization strategy to train strong classifiers with localizable features//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 6022-6031 [DOI: 10.1109/iccv.2019.00612]
- Zand M, Etemad A and Greenspan M. 2022. Oriented bounding boxes for small and freely rotated objects. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #4701715 [DOI: 10.1109/tgrs.2021.3076050]
- Zhang D F, Shi L Y, Leung S C H and Wu T. 2016. A priority heuristic for the guillotine rectangular packing problem. *Information Processing Letters*, 116(1): 15-21 [DOI: 10.1016/j.ipl.2015.08.008]
- Zhang H, Hao C Y, Song W R, Jiang B and Li B Z. 2023. Adaptive slicing-aided hyper inference for small object detection in high-resolution remote sensing images. *Remote Sensing*, 15(5): #1249 [DOI: 10.3390/rs15051249]
- Zhang J, Xie C M, Xu X, Shi Z W and Pan B. 2020. A contextual bidirectional enhancement method for remote sensing image object detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 4518-4531 [DOI: 10.1109/JSTARS.2020.3015049]
- Zhang J Y, Huang J Y, Chen X K and Zhang D Y. 2019. How to fully exploit the abilities of aerial image detectors//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshops*. Seoul, Korea (South): IEEE: #7 [DOI: 10.1109/iccvw.2019.00007]
- Zhang Y, Ye M, Zhu G Y, Liu Y, Guo P Y and Yan J H. 2024. FFCA-YOLO for small object detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5611215 [DOI: 10.1109/TGRS.2024.3363057]
- Zhang Y F, Ren W Q, Zhang Z, Jia Z, Wang L and Tan T N. 2022. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing*, 506: 146-157 [DOI: 10.1016/j.neucom.2022.07.042]
- Zhao Z C, Du J X, Li C L, Fang X, Xiao Y and Tang J. 2024. Dense tiny object detection: a scene context guided approach and a unified benchmark. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #5606913 [DOI: 10.1109/TGRS.2024.3357706]
- Zheng Z, Zhong Y F, Ma A L, Han X B, Zhao J, Liu Y F and Zhang L P. 2020a. HyNet: hyper-scale object detection network framework for multiple spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166: 1-14 [DOI: 10.1016/j.isprsjprs.2020.04.019]
- Zheng Z H, Wang P, Liu W, Li J Z, Ye R G and Ren D W. 2020b. Distance-IOU loss: faster and better learning for bounding box regression//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York, USA: AAAI: 12993-13000 [DOI: 10.1609/aaai.v34i07.6999]
- Zhou X Y, Wang D Q and Krähenbühl P. 2019. Objects as points [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/1904.07850.pdf>
- Zhou Z Z and Zhu Y Y. 2024. KLDet: detecting tiny objects in remote sensing images via Kullback-Leibler divergence. *IEEE Transactions on Geoscience and Remote Sensing*, 62: #4703316 [DOI: 10.1109/TGRS.2024.3382099]
- Zhu H G, Chen X G, Dai W Q, Fu K, Ye Q X and Jiao J B. 2015. Orientation robust object detection in aerial images using deep convolutional neural network//*Proceedings of 2015 IEEE International Conference on Image Processing*. Quebec City, Canada: IEEE: 3735-3739 [DOI: 10.1109/ICIP.2015.7351502]
- Zhu H R, Xu C, Yang W, Zhang R X, Zhang Y and Xia G S. 2024. Robust tiny object detection in aerial images amidst label noise [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2401.08056.pdf>
- Zhu P F, Wen L Y, Bian X, Ling H B and Hu Q H. 2018. Vision meets drones: a challenge [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/1804.07437.pdf>
- Zhu X X, Tuia D, Mou L C, Xia G S, Zhang L P, Xu F and Fraundorfer F. 2017. Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4): 8-36 [DOI: 10.1109/MGRS.2017.2762307]
- Zhu X Z, Su W J, Lu L W, Li B, Wang X G and Dai J F. 2020. Deformable DETR: deformable transformers for end-to-end object detection [EB/OL]. [2024-12-12]. <https://arxiv.org/pdf/2010.04159.pdf>
- Zhu Y C, Zhou Q Q, Liu N, Xu Z Y, Ou Z C, Mou X F and Tang J. 2023. ScaleKD: distilling scale-aware knowledge in small object detector//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 19723-19733 [DOI: 10.1109/CVPR52729.2023.01889]

作者简介

高广帅,男,讲师,主要研究方向为图像处理、模式识别、遥感图像分析及机器学习。E-mail:6911@zut.edu.cn

尚云琦,男,硕士研究生,主要研究方向为计算机视觉和遥感目标检测。E-mail:shangyunqi114@163.com

董燕,女,教授,主要研究方向为人工智能、图像处理和机器学习。E-mail:dy@zut.edu.cn