



基于深度学习的小目标检测算法研究进展

廖龙杰¹,吕文涛¹,叶冬²,郭庆³,鲁竞³,刘志伟¹

(1. 浙江理工大学, a. 信息科学与工程学院; b. 浙江省智能织物与柔性互联重点实验室, 杭州 310018;
2. 浙江移动信息系统集成有限公司, 杭州 311217; 3. 浙江省技术创新服务中心, 杭州 310007)

摘要: 基于深度学习的小目标检测算法可以有效提高小目标检测性能和检测速率,在图像处理领域得到了广泛应用。首先概述了小目标检测的难点,分别对基于锚框优化、基于网络结构优化、基于特征增强的小目标检测算法进行了分析,总结了各算法的优缺点;然后介绍了用于小目标检测的公共数据集和小目标检测算法的评价指标,对检测算法的性能指标进行了分析;最后对小目标检测算法已经解决的难点进行了总结,并对有待后续研究方向进行了展望。深度学习在小目标检测领域仍有较大的发展空间,在模型通用性、耗时与精度和特定场景的小目标检测等方面有待深入研究。

关键词: 深度学习;神经网络;图像处理;目标检测;小目标检测

中图分类号: TP391.4

文献标志码: A

文章编号: 1673-3851(2023)05-0331-13

引文格式: 廖龙杰,吕文涛,叶冬,等. 基于深度学习的小目标检测算法研究进展[J]. 浙江理工大学学报(自然科学),2023,49(3):331-343.

Reference Format: LIAO Longjie, LÜ Wentao, YE Dong, et al. Research progress of small target detection based on deep learning[J]. Journal of Zhejiang Sci-Tech University, 2023, 49(3): 331-343.

Research progress of small target detection based on deep learning

LIAO Longjie¹, LÜ Wentao¹, YE Dong², GUO Qing³, LU Jing³, LIU Zhiwei¹

(1a. School of Information Science and Engineering; 1b. Key Laboratory of Intelligent Textile and Flexible Interconnection of Zhejiang Province, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2. Zhejiang Mobile Information System Integration Co., Ltd., Hangzhou 311217, China; 3. Zhejiang Technology Innovation Service Center, Hangzhou 310007, China)

Abstract: The small target detection algorithm based on deep learning can effectively improve the detection performance and detection rate of small targets, and has been widely used in the field of image processing. Firstly, the difficulties of small target detection are summarized, and the small target detection algorithms based on anchor frame optimization, network structure optimization and feature enhancement are analyzed respectively, and the advantages and disadvantages of each algorithm are summarized. Then, the common data set for small target detection and the evaluation index of small target detection algorithm are introduced, and the performance index of the detection algorithm is analyzed. Finally, the difficulties that have been solved by the small target detection algorithm are summarized, and the future research directions are prospected. Deep learning still has a large development space in the field of small target detection, and needs to be further studied in the aspects of model universality, time

收稿日期: 2022-11-18 网络出版日期: 2023-03-01

基金项目: 国家自然科学基金项目(61601410); 浙江省科技厅重点研发计划项目(2021C01047, 2022C01079); 东北大学流程工业综合自动化国家重点实验室联合基金项目(2021-KF-21-03, 2021-KF-21-06)

作者简介: 廖龙杰(1998—), 男, 湖南衡阳人, 硕士研究生, 主要从事计算机视觉与模式识别方面的研究。

通信作者: 吕文涛, E-mail: alvinlwt@zstu.edu.cn

consuming and precision, and small target detection in specific scenes.

Key words: deep learning; neural network; image processing; object detection; small target detection

0 引言

在图像处理中,小目标检测是一个非常重要的研究方向。小目标检测算法的发展过程可以大致分为两个阶段,即传统的小目标检测算法和基于深度学习的小目标检测算法^[1]。传统的目标检测方法主要有3个步骤:选择候选目标区域、提取特征和利用分类器分类^[2]。传统的目标检测算法有很多不足,如目标区域选择算法一般采用滑动窗口实现,针对性较差,时间复杂度较高,冗余窗口较多,并且鲁棒性也不强,检测的效率和准确性都很低。因此,传统的目标检测算法并不适用于小目标检测。然而,随着深度学习技术在图像处理中的应用,小目标检测技术已经取得了长足的进步。

基于深度学习的小目标检测算法,通常在两阶段目标检测算法和单阶段目标检测算法的基础上改进。这两种算法可以根据是否需要生成候选区域进行划分^[3]。两阶段目标检测算法首先将目标可能出现的区域(即候选区域)筛选出来,然后对候选区域中的目标进行分类和回归。目前两阶段目标检测算法的典型代表为 R-CNN 系列算法^[4],如 Sparse R-CNN^[5]、DeFRCN^[6]等。相较于两阶段目标检测算法,单阶段目标检测算法无需生成候选区域,直接计算物体的类别概率和位置坐标,经过一次检测即可得到最终的检测结果。现阶段较为常用的单阶段目标检测算法有 YOLO 系列目标检测算法^[7-9]和 SSD 系列目标检测算法^[10-13]。因为两阶段目标检测算法分为两步进行,通常效率较低,但检测效果较好;而单阶段目标检测算法只有一步操作,因此它较两阶段目标检测算法速度更快,但检测的精度略有下降。

目前基于深度学习的目标检测算法在大、中型目标上已取得了较好的结果,但由于小目标在图像中面积占比小,难以获取有效的特征信息,检测性能并不理想^[14]。目前,在 Microsoft common objects in context(以下简称“MS COCO”)等公共数据集^[15]中,小目标检测算法的精度远不及大目标和中目标检测算法,并且经常出现漏检和误检。虽然小目标的检测十分困难,但是小目标大量存在于各个场景中,有着重要应用,这也让小目标检测成为图像处理领域的研究热点。

本文检索了近几年来国内外典型的相关文献,

对各类基于深度学习的小目标检测算法进行了综述。首先,概述了小目标检测算法的难点;其次,将基于深度学习的小目标检测算法分为基于锚框优化的目标检测算法、基于网络结构优化的目标检测算法、基于特征增强的目标检测算法3类,分析了小目标检测算法的研究现状,并总结了各算法的优缺点;再次,介绍了用于小目标检测的公共数据集和小目标检测算法的评价指标,并对检测算法的性能指标进行了分析;最后,对小目标检测算法的研究进展进行了总结,并对未来研究进行了展望。

1 小目标检测算法的难点

各类基于深度学习的目标检测算法在对于大目标检测时效果相对较好,而对于小目标的检测的效果相对较差,主要原因有如下几点。

a)小目标的特征信息难以被充分提取。在图像中,小目标不像大、中目标占较大的面积,一般并不清晰,且不如大、中目标那样携带丰富的特征信息,因此小目标易受噪音干扰,检测模型无法精确定位小目标,且不易对目标类型进行分类^[16]。

b)小目标样本难以在数据集中平均分布。目前,公共数据集普遍存在一个问题,即大、中目标的数量远远多于小目标的数量。如 MS COCO 数据集^[15],小目标所占比例只有 31.62%,且每幅图像的目标数量庞大,小目标分散于图像的各个位置。又比如 TinyPerson 数据集^[17],每一幅图像中小目标数量差异较大,有的图像中小目标数量非常的多,超过 100 个,也有的图像中小目标数量非常的少,不足 20 个。这样目标不均匀分布的样本,在使用特定方法训练时难以提升模型的检测性能。同时,模型训练时更加关注大、中目标,小目标样本数量不均也加大了模型训练的困难程度。

c)小目标检测模型的先验框难以设置。预先设定的先验框,通常都有固定的尺寸和比例,对于大、中目标的检测一般比较有效。但对于小目标,其尺寸较小且比例可能不像大、中目标那样易于预估,这样导致先验框几乎不能正确框选小目标,甚至有可能框选的大目标中包含了小目标,造成了正负样本不平衡,使模型训练变得困难^[18]。

d)小目标检测模型的损失函数难以设置。对于采用深度学习的目标检测算法来说,每轮训练完成

后需要计算损失并进行回归来优化参数的设置,而最初的损失函数是为大、中目标而设计的。损失函数由定位损失和类别损失组成,有些算法在这两种损失的基础上额外增加了交并比损失。在进行回归优化时,有些损失忽略了小目标样本的特殊性,使得部分检测模型的效果较差。

e)小目标检测模型的正负样本难以匹配。现阶段,正负样本的划分是按照检测器生成的边界框与真实框之间的交并比决定的。通常交并比大于50%所对应的锚框中的目标判定为正样本,其余为负样本^[7]。但是,一般小目标在图像中所占的面积较小,有时交并比很难达到50%,小目标匹配的情况不如大、中目标那么理想,存在包含很多小目标的正样本遗漏的情况。

f)小目标检测模型难以通用。现阶段,小目标的数据集通常是某一类物体,例如织物瑕疵数据集,小目标样本全都是织物瑕疵,TinyPerson数据集的小目标样本都是一些小型人物目标。这样造成训练出来的模型只能专门检测某一类物体,模型的通用性会非常差。

2 基于锚框优化的小目标检测算法

在目标检测中,锚框(Anchor box)是以图像的每个像素点为中心生成的多个大小和宽高比不同的边界框。目前绝大多数基于深度学习的目标检测算法都是基于锚框机制(Anchor based)设计的。但基于锚框机制的目标检测算法主要用来检测大目标或者中目标,对于小目标的检测效果不好,因此有不少研究人员对基于锚框机制的小目标检测算法进行研究,且近几年基于无锚框机制(Anchor free)的小目标检测也是研究的热点。

2.1 锚框机制

锚框机制最早出现于两阶段目标检测算法Faster R-CNN^[4]中的Region proposal network(RPN),RPN使用特征提取的形式生成预选框的位置,从而降低了Selective search算法带来的计算时间上的开销。而在单阶段目标检测算法如YOLO V3中,锚框是通过使用K-means聚类算法从真实框中得到一些不同宽高比的框。现阶段大部分目标检测算法都是使用锚框机制来选择候选区域,显著提升小目标检测效果的方法之一就是改善锚框的设置机制。

Yang等^[19]针对预先设定锚框的尺寸比例过于死板的问题,提出了MetaAnchor,这是一种动态锚框生成机制,MetaAnchor可以自定义任意的锚框,

并从中选择合适的动态生成。Zhang等^[20]提出了单镜头尺度不变人脸检测器,对不同尺度的人脸采用不同尺度的锚框,并通过尺度补偿锚框匹配机制提高小人脸的召回率,显著提高了锚框对于小目标匹配的成功率。Wang等^[21]提出一种称为引导锚框的锚框生成机制,其中锚框的选定借助上下文信息引导实现,同时该机制还推算出目标中心点可能出现的坐标和不同坐标处的尺度和长宽比,并通过自适应的方法来解决锚框形状特征不相符的问题。

以上几种方法都采用动态的方法来产生锚框,这些方法能有效解决锚框预先设定后无法改动的问题,能有效提高基于锚框机制的小目标检测算法的性能。

此外,由于小目标一般数量规模较大且密集,锚框在该问题上表现的效果非常差,以下几种方法是较为有效的改进方法。Zhang等^[22]提出了一种关于锚框密度的策略,让在同一幅图像上不同类型的锚框具有相同的密度,从而大幅提高了小人脸的召回率。Zhu等^[23]提出了一种锚框设计方案,引入了新的预期最大重叠分数,该分数可以从理论上解释锚框与小目标人脸低重叠问题,使基于锚框的小目标人脸检测获得更好的性能。Wang等^[24]提出了一种称为SFace的算法,有效整合了基于锚框和不基于锚框的方法,以解决高分辨率图像和视频中广泛存在的尺度变化较大的问题,该算法能有效提高具有超大尺度变化的人脸检测算法的性能。

2.2 无锚框机制

锚框在目标检测中起到了重要的作用,目前大部分的目标检测算法都是基于锚框设计的,但是该类算法也有不足之处。第一,锚框通常需要预先设定大小和比例,这对于检测小目标不利。相较于大目标,小目标不易被锚框框选,这会造成正负样本不均衡,使模型难以针对小目标进行训练。第二,锚框引入超参数过多,如锚框数量、大小、长宽比等,提高了算法设计难度。因此,最近几年来目标检测领域逐渐将研究的主要方向转移到无锚框机制上,在小目标检测应用上显著提升了检测效果。

无锚框机制的一种研究思路是将原先用锚框来框选目标的操作改成根据关键点来定位目标。根据关键点的目标检测主要包含两种:一种是根据角点来定位,另一种是根据中心点来定位。DeNet通过估计4个角点来替代锚框定位,这4个角点分别为左上、右上、左下、右下角点,具体流程为:a)让模型训练带有标签的数据集;b)用训练好的模型预测角点的位置分布;c)将4个角点所包围的区域定为候

选区域^[25]。Law等^[26]提出了一种根据角点来定位目标的网络模型 CornerNet, CornerNet 通过两个点来定位:左上和右下角点,具体流程为:首先将所有目标的左上和右下角点的位置预测出来,并在每两个角点中嵌入距离向量,通过判断距离向量,让属于同一个目标的两个角点两两配对,然后利用这两个角点生成目标的边界框。

CornerNet 有效地解决了锚框预先固定尺寸和高宽比对小目标检测带来的困难,但 CornerNet 也有明显的缺点。CornerNet 在生成目标边界框时,有时不能准确框选目标的边界,有的边界框虽然框选了目标,但存在边界框远远大于目标的情况;此外还会框选错误的目标,出现了很多冗余和错误的目标框。针对这个问题,Duan等^[27]提出了一种借助中心点提高判别能力的检测框架 CenterNet。CenterNet 使用3个关键点,即左上角点、中心点、右下角点,这让网络能通过中心点来加强网络对物体的辨别能力,能有效减少冗余和错误的目标框。

全卷积神经网络(Fully convolutional networks, FCN)通过对图像每个像素点都进行分类,解决了很多语义分割的问题,特别是对于密集型的预测任务,FCN 的完成情况相对较好。受到 FCN 的启发,Tian等^[28]在目标检测框架上运用语义分割的方法,提出了一种基于全卷积的单阶段目标检测框架(Fully convolutional one-stage object detection, FCOS),解决了由锚框过多带来的超参数计算困难的问题;同时,该研究表明将 FCOS 应用于两阶段检测模型的第一阶段,也能显著提高检测效率。

3 基于网络结构优化的目标检测算法

用于小目标检测算法的一般网络模型的结构由骨干网络(Backbone)、特征融合与特征增强(Neck)和检测头(Head)3部分组成。优化 Backbone 和 Neck 可显著提升小目标检测性能,此外在模型上加入注意力机制也能起到较好的效果。下面将从这几个角度介绍相关研究进展。

3.1 Backbone 优化

目标检测模型通常使用骨干网络进行特征提取操作,典型的骨干网络有 AlexNet、VGGNet、ResNet、GoogLeNet 等^[29]。由于特征提取网络通过多层卷积层进行下采样来提取特征,在这过程中目标特别是小目标的特征容易丢失,众多研究人员提出了引入 CSP 模块和残差模块等改进方案。

YOLO V3 模型采用 DarkNet-53 作为骨干网络。

Bochkovskiy 等^[30]受到 Cross stage partial network (CSPNet)思想的启发,在 YOLO V3 的基础上提出 YOLO V4 模型,该模型构造 CSPDarkNet-53 作为骨干网络进行特征提取操作,在减少了计算量的同时可以保证准确率。崔文靓等^[31]针对 YOLO V3 算法检测公路车道线准确率低和漏检率高的问题,提出了一种基于改进 YOLO V3 网络结构的公路车道线检测方法,该方法去掉了3个 YOLO 层前的两组卷积层,降低了车道线小目标在复杂背景下漏检的概率。Pan等^[32]对 YOLO-Tiny 的骨干网络进行改进,将 DenseNet 作为骨干网络。DenseNet 可以将本层特征层的输出作为下一层的输入,使每一层特征信息的利用率上升,消除了部分冗余参数,提升了网络训练的效率。Fu等^[11]提出反卷积单步检测模型(Deconvolutional single shot detector, DSSD),该模型在 SSD 的基础上将骨干网络由 VGG 更换为 ResNet-101,并在此基础上增加了反卷积和预测模块,让模型增强了识别和分辨小目标的能力。

由此可见,通过引入 CSP 模块和残差模块,可以有效提高模型的性能,其主要思想都是降低卷积操作对小目标的影响。此外,关于 DenseNet 和 DSSD,第一个模型是建立特征通道使前面的特征能传递到后面,第二个模型是采用反卷积模块,两个模型都是通过提高特征的利用率来提升小目标的检测性能。

3.2 Neck 优化

在骨干网络提取特征时,分布在浅层的特征因下采样次数少,特征信息损失较少,具有较高的分辨率和大量位置特征信息,但语义特征信息不足;深层特征经过多层下采样,特征损失较多,分辨率也下降较多,但是语义特征信息得到了增强。引入特征融合和特征增强(Neck),可将浅层特征和深层特征结合,也是一种提高小目标检测性能的方式。

遥感图像目标通常会遇到目标间尺度差距较大和种类间近似度较高的问题,多尺度特征融合是一种解决该问题的方法,但目前基本上都采用不变的权重系数来融合不同尺度的特征,忽略了目标尺度对特征融合的影响。针对该问题,谢星星等^[33]提出了一种动态特征融合网络,通过引入特征门控和动态融合模块实现了多尺度特征的动态融合,其中特征门控模块在特征融合前针对性地对部分特征进行了加强或者削弱,减少了背景信息对之后融合的影响,动态融合模块可以根据目标的尺度动态调整融合的权重,该网络在大规模遥感目标检测数据集上验证了其有效性。

针对之前用来特征融合的 Feature pyramid networks(FPN)层都是由手工设计的,Ghiasi 等^[34]提出了特征金字塔网络 NAS-FPN,该网络主要由自下而上和自上而下的连接组成,可以跨范围地融合特征,优化了小目标的特征融合。

3.3 注意力机制优化

除了改进 Backbone 和 Neck 之外,在网络模型中加入注意力机制也是一种改进方式。添加注意力机制后,网络模型能重点关注候选区域的局部特征,提高了模型的检测效率。对于小目标检测来说,引入注意力机制可以使网络更加关注小目标所在的位置,增强了小目标的特征表达能力。

Lim 等^[35]提出一种上下文注意力机制,该机制可以聚焦于图像中的目标,并且还可以聚焦来自目标层的上下文信息,能有效降低浅层环境噪声的影响,能使检测器聚焦于小目标。Hu 等^[36]提出了 Squeeze-and-excitation networks (SENet),该网络对通道维度上的特征进行信息权重分配,对重要程度不同的信息分配不同的权重,以此在特征通道上加入注意力机制。Wang 等^[37]认为 SENet 对通道注意力机制的预测带来了副作用,捕获所有通道的依赖关系是低效并且是不必要的,认为卷积具有良好的跨通道信息获取能力,并提出了 ECA-Net;该网络将 SE 模块中的全连接层去除,直接在全局平均池化之后的特征上通过一个 1D 卷积进行学习。Shen 等^[38]提出了一种群体多尺度注意力金字塔网络(Group multi-scale attention pyramid network, GMSAPN),通过引入一个多尺度注意力模块,聚集了不同尺度的特征,并且抑制了背景中的杂乱信息。Li 等^[39]提出了一种跨层注意力网络,该网络设计了一个跨层注意力模块,获得了每个层中小目标的非局部关联,并通过跨层集成和平衡进一步增强小目标的特征表达能力。

4 基于特征增强的小目标检测算法

由于小目标的特征信息较为缺乏,在检测过程经常会受到抑制,而增强特征能有效提高小目标检测性能。本文总结并分析了以下几种特征增强的算法,分别为基于生成对抗网络的小目标检测算法、基于数据增强的小目标检测算法、基于多尺度学习的小目标检测算法、基于上下文学习的小目标检测算法和基于其他优化模型的小目标检测算法。

4.1 生成对抗网络

Goodfellow 等^[40]于 2014 年提出了生成对抗网

络(Generative adversarial network, GAN)。该网络是由生成器网络和鉴别器网络这两个相互对抗的子网络构成,在两者竞争对抗的过程中,将各自的参数进行交替学习和更新。在小目标检测中,一般把分辨率较低的小目标通过 GAN 转换成对应的分辨率高的特征,让小目标与其他较大目标的特征差异缩小,使小目标的特征信息表达得更清晰,从而提高小目标检测效果。

Li 等^[41]提出了一种新的生成对抗网络模型,该模型通过生成对抗学习,将感知到的低分辨率小目标转换为超分辨率表示,缩小小目标与大目标的表示差异,以改进小目标检测算法。Bai 等^[42]提出了多任务生成对抗网络(Multi-task generative adversarial network, MTGAN),它将生成器网络提升到超分辨率,该超分辨率网络可以通过上采样将不清晰的小目标采样到清晰的图像中,让不清晰的小目标还原为清晰的状态,从而提供更多特征信息来获得更精确的检测结果。Noh 等^[43]提出一种特征超分辨率方法,该方法在检测时通过对抗生成网络将普通特征转化为超分辨率的特征,在对抗生成网络转化小目标时,会充分利用上下文信息,扩大了感受野。

基于生成对抗网络的超分辨率图像增强模型具有较好的图像增强性能,但在重构图像时常常丢失边缘信息,尤其是遥感小目标。由此,Rabbi 等^[44]提出了端到端的增强型超分辨率生成对抗网络(End-to-end enhanced super-resolution generative adversarial network, EESRGAN),该网络借鉴了 EEGAN 和 ESRGAN,以端到端的方式运用不同的检测网络,并将得到的损失反馈回去,提高了小目标尤其是遥感目标的检测性能。Zhao 等^[45]针对红外小目标的基本特征,提出了一种基于生成对抗网络的目标检测模型,该模型根据红外小目标其独特的分布特征,添加了对抗性损失以提高定位能力,构建了一个生成对抗网络模型来自动学习目标特征并直接预测目标的类别概率。

4.2 数据增强

数据增强一般用于训练集,指的是让原本少量有限的的数据通过某种方法变得更多,扩充可训练的样本量,并且通过增加类别不同的样本使其多样性得以提升,从而减少模型对参数的依赖,可以防止模型过拟合。数据增强的方法一般有:将数据上下左右各种方向平移,将数据旋转不同角度,变换颜色、调整亮度、调整饱和度和色调等。对于小目标像素低、特征信息不足、样本数量少等困难,数据增强方

法也可提高小目标检测算法的性能。

Kisantal等^[46]提出了一种复制增强的方法,这个方法将小目标复制出很多个副本,增加小目标的数量,让训练集有更多的小目标样本,解决了小目标在图像中面积占比低、在图像中分布不合理等问题,有效提升了小目标的检测效果。针对复制时有可能发生的背景和尺度无法配对的情况,Chen等^[47]提出了一种称为RRNet的混合检测器,引入了重采样策略,在逻辑上进行数据增强;该策略是自适应的,能够根据当前语义环境在复制数据时将数据的周围环境信息也考虑进去,表现出较好的数据增强效果。

在一般情况下,网络的预训练数据集和检测器学习的数据集在尺度上要相适配,如果出现不匹配情况,可能会对特征表示和检测器造成负面影响。针对这个问题,Yu等^[17]提出了一种尺度匹配方法,使两个数据集之间的目标尺度相适配。该方法按照目标大小进行相应的裁剪,使目标之间的尺度差异尽可能地减少,让小目标在缩放操作时的特征丢失大幅减少。Chen等^[48]引入了一种由反馈驱动的数据提供器Stitcher,用平衡的方式训练目标检测器。在训练过程中,Stitcher对数据进行压缩和拼接,将数据集中的目标尺度缩小一个量级,解决了小目标在数据集中占比不平衡导致的问题,提高了小目标检测性能。此外,强化学习也可以应用于数据增强,如Zoph等^[49]提出了一种数据增强方法,该方法通过强化学习将最合适的数据选出来,让模型选择最优的数据进行训练,提高了小目标检测性能。

4.3 多尺度学习

目标检测网络的最后一层通常都是对前面获得的特征信息进行回归和预测,但随着网络的深入,一些较小的目标信息在经过大量的下采样后逐渐损失,不容易被网络识别出。小目标在浅层网络有较多的坐标信息,在深层网络中有较多上下文信息,兼顾这两种信息才能更好地检测小目标,而多尺度学习在一定程度上可以完成这一任务。

Liu等^[10]提出了单步多框目标检测算法(Single shot multibox detector, SSD),该算法对不同尺度的特征图进行分层检测,浅层的特征图感受野小,适合检测小目标,深层特征图感受野大,适合检测大目标。Bell等^[50]提出了Inside-outside net (ION),该网络首先收集了待检测目标可能出现区域的浅层到深层的不同尺度的特征图,并在其中提取出所需的特征,将这些不同尺度的特征信息相结合,消除了背景环境对小目标检测造成的影响。

以上两种多尺度学习方法通过对不同层次采取不同的操作进行特征提取,此外还可以采用不同尺度特征融合的方法。Zeng等^[51]提出了一种增强多尺度特征的融合方法,即空洞空间金字塔池平衡特征金字塔网络(Atrous spatial pyramid pooling-balanced-feature pyramid network, ABFPN),该网络采用具有不同膨胀率的空洞卷积算子来充分利用上下文信息,应用跳跃连接来实现特征的充分融合。Lin等^[52]提出了特征金字塔网络(Feature pyramid networks, FPN),该网络在特征提取网络之后,将特征提取获得的最后一层特征层经过上采样后与相邻的同尺寸的特征层进行融合,可以在不浪费较多计算资源的情况下获得更好的特征增强效果。Han等^[53]提出了多尺度残差块(Multiscale residual block, MRB),通过在级联残差块中使用扩张卷积来捕获多尺度上下文信息,从而提高卷积神经网络的特征表示能力。Cao等^[54]提出了Feature-fused SSD,该算法在SSD上改进,特征融合时增加了上下文信息,减少了小目标检测的计算时间。Li等^[13]提出了特征融合单阶段多框检测器(Feature fusion single shot multi-box detector, FSSD),该算法在特征融合阶段选择小型模块,先将不同尺度的特征提取出来,然后将它们融合到一个较大尺度的特征图中,之后在这个大尺度的特征图上建立FPN,解决了在不同特征层上提取的特征尺度不同导致融合困难的问题。

除了将不同尺度特征融合的方法外,提升特征图尺度和增强图像分辨率也是有效的方法。Nayan等^[55]提出了一种目标检测算法,在特征融合阶段,将特征图向上采样,使特征图尺度增大,并使用跳跃连接,可以获得不同尺度下的多种特征,在对小目标检测上取得了不错的提升。Liu等^[56]提出了一种高分辨率检测网络(High-resolution detection network, HRDNet),通过将高分辨率图像输入浅层网络,以保留更多位置信息并降低计算成本,将低分辨率图像输入深层网络以提取更多语义信息,不仅能节约计算成本还能增强检测效果。Deng等^[57]提出了扩展特征金字塔网络(Extended feature pyramid network, EFPN),EFPN的特征金字塔具有超高分辨率,适合检测小目标,能够有效解决特征金字塔中不同尺度的特征耦合影响小目标检测性能的不足。

4.4 上下文学习

由于小目标分辨率较低,特征信息较少,小目标检测较易受周围背景环境信息影响,因此将小目标

周围的背景环境信息即上下文信息作为额外信息能有效地提升小目标检测性能。

Feng 等^[58]提出了三重上下文感知网络(Triple context-aware network, TCANet),该网络引入了全局上下文感知增强(Global context-aware enhancement, GCAE)模块,通过捕获全局视觉场景上下文来激活整个目标的特征。Lim 等^[35]提出了一种基于上下文连接多尺度特征的目标检测方法,该方法通过连接多尺度特征,将各层的附加特征作为语义特征,并引入了注意力机制,使模型关注图像中的目标,更好地利用语义信息,提高了模型在实际场景下对小目标的检测性能。Shen 等^[59]提出了一种基于特征聚合模块和混合注意力选择模块 HSFA2Net 的室内多人检测网络模型,其中特征聚合模块使用聚合和离散的思想来融合环境语义信息,给小目标多人检测提供了更多的详细信息,而提出的改进混合注意力选择模块将选择机制与混合注意模块相结合,解决了室内群体特征和背景特征重叠且分类边界不明显的问题,有效地提高了室内小目标多人的检测效果。Cui 等^[60]提出了一种上下文感知块网络(Context-aware block net, CAB Net),通过构建高分辨率和强语义特征图来提高小目标检测性能。Leng 等^[61]提出了一种基于内部-外部网络(Internal-external network, IENet)的检测器,它使用目标的外观和上下文信息进行鲁棒检测,从特征提取、定位和分类等方面改进了小目标检测算法。Guan 等^[62]提出了语义上下文感知网络(Semantic context aware network, SCAN),通过金字塔池化融合多级上下文信息,以构建上下文感知特征,有效提高了对小目标检测的性能。

由此可见,小目标与背景场景存在一定的联系,通过研究这种关系可以有效提升小目标的检测

性能。

4.5 其他优化模型

损失函数对小目标检测的效果有很大影响,将其优化也是一种不错的思路。Chen 等^[48]发现由于小目标在图像中面积占比较少,在模型训练时小目标对损失函数基本没有有效的反馈,因此提出了一种目标检测算法 Sticher,将训练损失作为反馈,指导下一次训练的迭代更新,大幅提高了小目标检测精度。Kim 等^[63]提出了类不确定性感知(Class uncertainty-aware, CUA)损失,CUA 损失在调制损失函数时考虑了预测模糊性以及分类分数的预测。Shuang 等^[64]提出了尺度平衡损失(Scale-balanced loss, SBL),SBL 通过重新加权策略显著提高了多尺度目标尤其是小目标的检测精度。

分阶段检测也是一种提升小目标检测性能的方法。Chen 等^[65]提出一种双重检测器 Dubox,该检测器分两次检测,第一个检测器尽可能地检测小目标,第二个检测器帮第一个检测器检测遗漏或者无法检测的小目标。

尺度自适应也能提高小目标的检测性能。Gao 等^[66]采用从粗略到精细的检测方式,对低分辨的图像进行粗略检测,对目标有可能出现的位置进行放大,并在该高分辨的区域进行精细搜索,通过不断循环以上操作,提高网络对于小目标的检测性能。

平衡小目标检测的准确率和计算时间也是一种改进思路。Sun 等^[67]提出了模块化特征融合检测器(Modular feature fusion detector, MFFD),该检测器包含了两个基本模块用于高效计算,前模块减少了原始输入图像的信息损失,后模块减少了模型大小和计算成本,同时确保了检测精度。

以上基于深度学习模型的小目标检测算法的优缺点见表1。

表1 基于深度学习的小目标检测算法优缺点

算法种类	主要算法类型	优点	缺点	相关文献
基于锚框优化的 小目标检测算法	锚框机制	Bounding Box 回归简单	超参数多,难以调整	文献[19-24]
	无锚框机制	几乎没有超参数	Bounding Box 回归困难	文献[25-28]
基于网络结构优化的 小目标检测算法	Backbone 优化	降低小目标的特征损失	模型复杂、泛化能力弱	文献[11,29-32]
	Neck 优化	浅层特征与深层特征结合	网络结构变得复杂	文献[33-34]
	注意力机制优化	小目标特征表达能力增强	具有较大的计算复杂度	文献[35-39]
基于特征增强的小 目标检测算法	生成对抗网络	缩小与大目标的特征差异	模型训练较为困难	文献[40-45]
	数据增强	增加小目标的样本数量	增加计算成本	文献[17,46-49]
基于特征增强的小 目标检测算法	多尺度学习	结合浅层位置信息与深层语义信息	难以避免干扰噪声的影响	文献[10,13,50-57]
	上下文学习	增加小目标的数量与质量	需要有足够的相同特征区域进行匹配,速度较慢	文献[35,58-62]

5 小目标数据集、检测算法评价指标和各类检测算法的性能

目前,大多数目标检测算法在常规数据集上已经得到了较好的效果,但对于小目标数据集的效果还较差。对于一个算法,需要合适的数据集进行训练和检验,因此适合小目标检测的数据集十分重要。以下介绍一些适合用来检测小目标的公共数据集和小目标检测算法的评价指标,并对一些算法在小目标数据集上的检测性能进行分析。

5.1 主流小目标数据集介绍

a) MS COCO 数据集^[15]。该数据集由微软构建,是一个庞大的、样本充足的目标检测、语义分割和关键点检测数据集。该数据集共有 330 万幅图像,包括 80 个目标类别和 91 个物体类别,有 20 多万幅已标注的图像。相比于 PASCAL VOC 数据集和 ImageNet 数据集,MS COCO 数据集的小目标数量更多。

b) DOTA 数据集^[68]。该数据集是由航拍图像构成的小目标检测数据集,由武汉大学于 2017 年 11 月建立。该数据集共有 2806 幅图像,共有 15 种标签的类型,尺寸由 800×800 到 4000×4000 不等,共有 35 Gi。完全标注的 DOTA 图像包含 188282 个样本,每个样本都由一个四边形标注。

c) WIDER FACE 数据集^[69]。该数据集是一个大规模人脸图像的 benchmark 数据集,包括 32203 幅图像,共有 393703 个带有标签的人脸。该数据集被分为训练集、验证集和测试集,其中:训练集包含 158989 个带有标签的人脸,验证集包含 39496 个带有标签的人脸,测试集包含了 195218 个带有标签的人脸。WIDER FACE 数据集考虑了通用目标的检测效率以及对人眼的识别程度,按照像素高低把人脸分为 3 个尺度:大尺度(>300 像素)、中尺度($50 \sim 300$ 像素)、小尺度($10 \sim 50$ 像素)。每种人脸的大小、位置以及光线、动作、遮挡角度等各类情况较为全面。

d) WiderPerson 数据集^[70]。该数据集是比较拥挤场景的行人检测数据集,其图像由多种场景组成,不只有交通场景。数据集共有 13382 幅图像,有 5 个不同种类的标签,共有 40 万多条遮挡注释。在该数据集中,训练集、验证集、测试集图像数量之比大约为 8:1:4。

e) TinyPerson 数据集^[17]。该数据集为海上快速救援背景下的小目标人群数据集,主要关注海边

的人。TinyPerson 数据集在海上和海滩场景中人物图像的分辨率很低,一般少于 20 个像素。在 TinyPerson 数据集中,人像的纵横比有较大差异,人的姿势和拍摄视角更加复杂,在多样性方面对现有数据集进行了有效补充。在 TinyPerson 数据集中,有很多带有密集目标的图像(每个图像超过 200 个人)。

f) BIRDSAI 数据集^[71]。该数据集由 Bondi 等^[71]在 WACV 2020 上提出。该数据集通过在无人机上安装红外摄像头以模仿鸟的视觉拍摄获得的,是一个大自然场景的数据集。该数据集一共包含 10 个种类,有各种动物、人类还有未知事物,其中还有一些特殊图像,如图像尺寸剧烈变化、热反射造成的背景混乱、摄像机旋转抖动造成的图像模糊等。

5.2 小目标检测算法的评价指标

小目标检测算法的评价指标主要有精确率、召回率、平均精度、平均精确度均值和每秒处理帧数。待检测目标分为正样本(Positive)和负样本(Negative)两类,而检测目标样本分为如下 4 类: a) True positives (TP) 是正样本被正确预测为正样本; b) False positives (FP) 是负样本被错误地预测为正样本; c) True negatives (TN) 是负样本被正确预测为负样本; d) False negatives (FN) 是正样本被错误预测为负样本。小目标检测的评价指标分为如下 5 种:

a) 精确率(Precision): 是指正确被预测为正样本的样本个数与所有被预测为正样本的样本个数的比值,计算公式为: $p = n_{TP} / (n_{TP} + n_{FP})$, 其中: p 表示精确率, n_{TP} 表示 TP 样本数, n_{FP} 表示 FP 样本数。

b) 召回率(Recall): 是指正确被预测的正样本个数与所有正样本的比值,计算公式为: $r = n_{TP} / (n_{TP} + n_{FN})$, 其中 r 表示召回率。

c) 平均精度(Average precision, AP): 在多类物体检测中,单个类别以召回率为横坐标,以精确率为纵坐标,与在根据一定阈值计算基础形成的曲线所围成的面积称为 AP。对于单类目标检测效果越好,AP 值越高。

d) 平均精确度均值(Mean average precision, mAP): 在检测多类目标时,分别计算一次每一类目标的 AP,再做平均运算即为 mAP。mAP 一般用来衡量模型在检测多类目标时表现的优劣程度。

e) 每秒处理帧数(Frames per second, FPS): 对于评价一个目标检测算法的检测速度,通常采用

FPS 每秒处理帧数这一指标。FPS 值越大,表示算法检测速度越快。

5.3 典型基于深度学习的小目标检测算法的性能

以 MS COCO 数据集作为检测数据集,近几年较为典型的目标检测算法性能对比情况见表 2。表 2 中 AP_S 、 AP_M 、 AP_L 分别表示小目标、中目标、大目标的平均精度。通过对比分析可以得出,各种算法对于小目标的平均精度都远不及大目标,基本上小目标的平均精度都只有大目标的 50%,这说明小目标检测算法目前还不够完善,仍有待提升。在表 2 中, Scaled-YOLOv4^[72] 是效果最好的检测算法,这主要因为该算法集成了众多优秀的改进方法,例如特征融合、数据增强、多尺度学习、上下文学习等。

表 2 MS COCO 数据集上的检测算法性能

算法名称	主干网络	平均精度/%			
		AP	AP_S	AP_M	AP_L
SSD ^[10]	Res101	31.2	10.2	34.5	49.8
FPN ^[52]	Res101-FPN	36.2	18.2	39.0	48.2
YOLOv3 ^[7]	Darknet-53	33.0	18.3	35.4	41.9
FCOS ^[28]	ResNeXt-101	44.7	27.6	47.5	55.6
Scaled-YOLOv4 ^[72]	CSP-P7	55.4	38.1	59.5	67.4

6 结 语

本文对基于深度学习的小目标检测算法进行了总结,介绍了小目标检测算法的难点与评价指标,综述了基于锚框优化、网络结构优化、特征增强的小目标检测算法,介绍了小目标数据集并分析了各类检测算法的性能。对于小目标检测的难点,目前已经解决的有如下几点:

a) 对于小目标的特征信息难以被充分提取的问题,可以采用网络结构优化中的基于 Neck 优化的方法。将特征融合与特征增强模块引入小目标检测模型,可以有效解决小目标在特征信息不足的情况下,经过特征提取网络后损失大量特征而导致检测效果低下的问题。

b) 对于小目标样本在数据集中难以平均分布的问题,可以采用数据增强的方法。如果是样本数量不足,可以采用复制增强的方法。如果是样本多样性不足,数据集过于单一导致模型过拟合,可以采用颜色变换、随机翻转、随机裁剪等方法增加样本的多样性。

c) 对于小目标检测模型的先验框难以设置的问题,可以采用锚框优化的方法。针对小目标尺寸较

小且比例不统一不易被框选的问题,目前主要的解决思路是采用自适应的动态锚框。

d) 对于小目标检测模型的损失函数难以设置的问题,可以采用优化损失函数的方法。目前大多目标检测模型的损失函数都是为大、中目标检测算法设置的,并不适合小目标检测算法,因此可以优化损失函数,使其对小目标也能起作用。

e) 对于小目标检测模型的正负样本难以匹配的问题,可以采用 Anchor Free 方法。由于 Anchor Based 的检测模型一般将锚框与物体的交并比大于 50% 的当作正样本,但小目标一般难以满足这一要求,导致正负样本失衡。采用 Anchor Free 方法后,不需要满足交并比大于 50% 这一要求去寻找正样本,通过关键点或者语义分割的方法可以更加准确地找到正样本,并且比起 Anchor Based 方法正样本的遗漏率更低。

目前,随着深度学习的快速发展,小目标检测算法也随之完善,但仍有待提升。本文提及的数据增强、多尺度学习等方法可以在一定程度上解决小目标样本分布不均、难以提取充足特征信息等问题,但小目标检测仍面临许多困难,以下方面还需进一步研究:

a) 小目标检测模型难以通用。目前,小目标检测模型基本上都是专用于检测某一类物体,如专门用于检测织物瑕疵或者针对医学影像,这样很难将一个模型迁移到另一个模型上。对于通用性差的问题,在目前还没有较好的解决方案。而且,不仅仅是小目标检测模型的通用性差,小目标数据集的通用性也不足。研究通用的小目标检测模型是未来重要的发展方向,建立大体量、多种类的小目标数据集也是一项非常重要的工作。

b) 小目标检测模型难以较好地平衡耗时与精度。现阶段大多数小目标检测模型的规模都比较大,冗余程度比较高,如何在维持高精度的检测结果下简化检测模型,还需再进一步研究。目前该方向较为主流的思路是采用知识蒸馏的方法,将大模型在尽可能保持精度的情况下转化为轻量级的模型。

c) 特定场景的小目标检测。在遥感航拍等特定场景中,小目标非常密集,这非常不利于检测。由于大、中目标在一幅图中所占面积比较大,而图像总面积有限,不会出现密集的大规模检测情况,因此该问题只针对小目标。特定场景的小目标检测也将是研究的热点之一。

参考文献:

- [1] Zou Z X, Shi Z W, Guo Y H, et al. Object detection in 20 years: A survey[EB/OL]. (2019-05-13)[2022-11-18]. <https://arxiv.org/abs/1905.05055>.
- [2] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
- [3] 陈科圻, 朱志亮, 邓小明, 等. 多尺度目标检测的深度学习研究综述[J]. 软件学报, 2021, 32(4): 1201-1227.
- [4] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [5] 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). Nashville, USA. IEEE, 2021: 14449-14458.
- [6] Qiao L M, Zhao Y X, Li Z Y, et al. DeFRCN: Decoupled faster R-CNN for few-shot object detection [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada. IEEE, 2021: 8681-8690.
- [7] Redmon J, Farhadi A. YOLOv3: An incremental improvement [EB/OL]. (2018-04-08)[2022-11-18]. <https://arxiv.org/abs/1804.02767>.
- [8] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications [EB/OL]. (2022-09-07)[2022-11-18]. <https://arxiv.org/abs/2209.02976>.
- [9] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [EB/OL]. (2022-07-06)[2022-11-18]. <https://arxiv.org/abs/2207.02696>.
- [10] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [C]// Proceedings of the 14th European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21-37.
- [11] Fu C Y, Liu W, Ranga A, et al. DSSD: Deconvolutional single shot detector[EB/OL]. (2017-01-23)[2022-11-18]. <https://arxiv.org/abs/1701.06659>.
- [12] Zhou S R, Qiu J. Enhanced SSD with interactive multi-scale attention features for object detection [J]. Multimedia Tools and Applications, 2021, 80(8): 11539-11556.
- [13] Li Z X, Zhou F Q. FSSD: feature fusion single shot multi-box detector[EB/OL]. (2017-12-04)[2022-11-18]. <https://arxiv.org/abs/1712.00960>.
- [14] 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.
- [15] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context[C]//Proceedings of European Conference on Computer Vision. Cham: Springer International Publishing, 2014: 740-755.
- [16] 宁欣, 田伟娟, 于丽娜, 等. 面向小目标和遮挡目标检测的脑启发 CIRA-DETR 全推理方法[J]. 计算机学报, 2022, 45(10): 2080-2092.
- [17] Yu X H, Gong Y Q, Jiang N, et al. Scale match for tiny person detection [C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Snowmass, USA. IEEE, 2020: 1246-1254.
- [18] 聂光涛, 黄华. 光学遥感图像目标检测算法综述[J]. 自动化学报, 2021, 47(8): 1749-1768.
- [19] Yang T, Zhang X Y, Li Z M, et al. MetaAnchor: Learning to detect objects with customized anchors[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 318-328.
- [20] Zhang S F, Zhu X Y, Lei Z, et al. S³FD: Single shot scale-invariant face detector [C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 192-201.
- [21] Wang J Q, Chen K, Yang S, et al. Region proposal by guided anchoring[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA. IEEE, 2019: 2960-2969.
- [22] Zhang S F, Zhu X Y, Lei Z, et al. FaceBoxes: A CPU real-time face detector with high accuracy[C]//2017 IEEE International Joint Conference on Biometrics (IJCB). Denver, USA. IEEE, 2018: 1-9.
- [23] Zhu C C, Tao R, Lu K, et al. Seeing small faces from robust anchor's perspective [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA. IEEE, 2018: 5127-5136.
- [24] Wang J F, Yuan Y, Li B X, et al. SFace: An efficient network for face detection in large scale variations[EB/OL]. (2018-04-18)[2022-11-18]. <https://arxiv.org/abs/1804.06559>.
- [25] Tychsen-Smith L, Petersson L. DeNet: Scalable real-time object detection with directed sparse sampling[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 428-436.
- [26] Law H, Deng J. CornerNet: Detecting objects as

- paired keypoints [C] // Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2018: 765-781.
- [27] Duan K W, Bai S, Xie L X, et al. CenterNet: Keypoint triplets for object detection [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2020: 6568-6577.
- [28] Tian Z, Shen C H, Chen H, et al. FCOS: Fully convolutional one-stage object detection [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2020: 9626-9635.
- [29] Khan A, Sohail A, Zahoor U, et al. A survey of the recent architectures of deep convolutional neural networks[J]. Artificial Intelligence Review, 2020, 53(8): 5455-5516.
- [30] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23) [2022-11-18]. <https://arxiv.org/abs/2004.10934>.
- [31] 崔文靓, 王玉静, 康守强, 等. 基于改进 YOLOv3 算法的公路车道线检测方法[J]. 自动化学报, 2022, 45: 1-9.
- [32] Pan Z H, Chen Y. Object detection algorithm based on dense connection [C] // 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference. Chengdu. IEEE, 2020: 1558-1562.
- [33] 谢星星, 程堃, 姚艳清, 等. 动态特征融合的遥感图像目标检测[J]. 计算机学报, 2022, 45(4): 735-747.
- [34] Ghiasi G, Lin T Y, Le Q V. NAS-FPN: Learning scalable feature pyramid architecture for object detection [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2020: 7029-7038.
- [35] Lim J S, Astrid M, Yoon H J, et al. Small object detection using context and attention [C] // 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC). Jeju Island, Korea (South). IEEE, 2021: 181-186.
- [36] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [37] Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 11531-11539.
- [38] Shen L L, You L, Peng B, et al. Group multi-scale attention pyramid network for traffic sign detection[J]. Neurocomputing, 2021, 452: 1-14.
- [39] 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.
- [40] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [41] Li J N, Liang X D, Wei Y C, et al. Perceptual generative adversarial networks for small object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA. IEEE, 2017: 1951-1959.
- [42] Bai Y C, Zhang Y Q, Ding M L, et al. SOD-MTGAN: Small object detection via multi-task generative adversarial network [C] // Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2018: 210-226.
- [43] Noh J, Bae W, Lee W, et al. Better to follow, follow to be better: Towards precise supervision of feature super-resolution for small object detection [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2020: 9724-9733.
- [44] Rabbi J, Ray N, Schubert M, et al. Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network [J]. Remote Sensing, 2020, 12(9): 1432.
- [45] Zhao B, Wang C P, Fu Q, et al. A novel pattern for infrared small target detection with generative adversarial network [J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(5): 4481-4492.
- [46] Kisantal M, Wojna Z, Murawski J, et al. Augmentation for small object detection [EB/OL]. (2019-02-19) [2022-11-18]. <https://arxiv.org/abs/1902.07296>.
- [47] Chen C R, Zhang Y, Lv Q X, et al. RRNet: A hybrid detector for object detection in drone-captured images [C] // 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, Korea (South). IEEE, 2020: 100-108.
- [48] Chen Y K, Zhang P Z, Li Z M, et al. Stitcher:

- Feedback-driven data provider for object detection[EB/OL]. (2020-04-26)[2022-11-18]. <https://arxiv.org/abs/2004.12432>.
- [49] Zoph B, Cubuk E D, Ghiasi G, et al. Learning data augmentation strategies for object detection [C] // Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer International Publishing, 2020: 566-583.
- [50] Bell S, Lawrence Zitnick C, Bala K, et al. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA. IEEE, 2016: 2874-2883.
- [51] Zeng N Y, Wu P S, Wang Z D, et al. A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-14.
- [52] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA. IEEE, 2017: 936-944.
- [53] Han W X, Kuerban A, Yang Y C, et al. Multi-vision network for accurate and real-time small object detection in optical remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [54] Cao G M, Xie X M, Yang W Z, et al. Feature-fused SSD: Fast detection for small objects [C] // Ninth International Conference on Graphic and Image Processing (ICGIP 2017). Qingdao: SPIE, 2018: 381-388.
- [55] Nayan A A, Saha J, Mozumder A N. Real time detection of small objects[J]. International Journal of Innovative Technology and Exploring Engineering, 2020, 9(5): 837-839.
- [56] Liu Z M, Gao G Y, Sun L, et al. HRDNet: High-resolution detection network for small objects [C] // 2021 IEEE International Conference on Multimedia and Expo (ICME). Shenzhen, China. IEEE, 2021: 1-6.
- [57] Deng C F, Wang M M, Liu L, et al. Extended feature pyramid network for small object detection[J]. IEEE Transactions on Multimedia, 2022, 24: 1968-1979.
- [58] Feng X X, Han J W, Yao X W, et al. TCANet: Triple context-aware network for weakly supervised object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(8): 6946-6955.
- [59] Shen W X, Qin P L, Zeng J C. An indoor crowd detection network framework based on feature aggregation module and hybrid attention selection module[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, Korea(South). IEEE, 2020: 82-90.
- [60] Cui L S, Lv P, Jiang X H, et al. Context-aware block net for small object detection[J]. IEEE Transactions on Cybernetics, 2022, 52(4): 2300-2313.
- [61] 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.
- [62] Guan L T, Wu Y, Zhao J Q. SCAN: Semantic context aware network for accurate small object detection[J]. International Journal of Computational Intelligence Systems, 2018, 11(1): 951.
- [63] Kim J U, Kim S T, Lee H J, et al. CUA loss: Class uncertainty-aware gradient modulation for robust object detection [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(9): 3529-3543.
- [64] Shuang K, Lyu Z H, Loo J, et al. Scale-balanced loss for object detection[J]. Pattern Recognition, 2021, 117: 107997.
- [65] Chen S, Li J P, Yao C Q, et al. DuBox: No-prior box objection detection via residual dual scale detectors[EB/OL]. (2019-04-16)[2022-11-18]. <https://arxiv.org/abs/1904.06883>.
- [66] Gao M F, Yu R C, Li A, et al. Dynamic zoom-in network for fast object detection in large images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA. IEEE, 2018: 6926-6935.
- [67] Liu Y Z, Cao S, Lasang P, et al. Modular lightweight network for road object detection using a feature fusion approach[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(8): 4716-4728.
- [68] 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. Salt Lake City, USA. IEEE, 2018: 3974-3983.
- [69] 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). Las Vegas, NV, USA. IEEE, 2016: 5525-5533.

- [70] Zhang S F, Xie Y L, Wan J, et al. Widerperson: A diverse dataset for dense pedestrian detection in the wild[J]. IEEE Transactions on Multimedia, 2020, 22 (2): 380-393.
- [71] Bondi E, Jain R, Aggrawal P, et al. BIRDSAI: A dataset for detection and tracking in aerial thermal infrared videos[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass, USA. IEEE, 2020: 1736-1745.
- [72] Wang C Y, Bochkovskiy A, Liao H Y M. Scaled-YOLOv4: Scaling cross stage partial network[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA. IEEE, 2021: 13024-13033.

(责任编辑:康 锋)