ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 36,No. 3,May 2021,pp. 391－417
DOI:10. 16337/j. 1004-9037. 2021. 03. 001

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

小目标检测研究进展

高新波，莫梦竟成，汪海涛，冷佳旭

（重庆邮电大学计算机科学与技术学院，重庆 400065）

摘　要：小目标检测长期以来是计算机视觉中的一个难点和研究热点。在深度学习的驱动下，小目标检测已取得了重大突破，并成功应用于国防安全、智能交通和工业自动化等领域。为了进一步促进小目标检测的发展，本文对小目标检测算法进行了全面的总结，并对已有算法进行了归类、分析和比较。首先，对小目标进行了定义，并概述小目标检测所面临的挑战。然后，重点阐述从数据增强、多尺度学习、上下文学习、生成对抗学习以及无锚机制等方面来提升小目标检测性能的方法，并分析了这些方法的优缺点和关联性。之后，全面介绍小目标数据集，并在一些常用的公共数据集上对已有算法进行了性能评估。最后本文对小目标检测技术的未来发展方向进行了展望。

关键词：小目标检测；数据增强；多尺度学习；上下文学习；生成对抗学习；无锚机制

中图分类号：TP391.4　　**文献标志码：**A

Recent Advances in Small Object Detection

GAO Xinbo, MO Mengjingcheng, WANG Haitao, LENG Jiaxu

(College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Small object detection has long been a difficult and hot topic in computer vision. Driven by deep learning, small object detection has achieved a major breakthrough and has been successfully applied in national defense security, intelligent transportation, industrial automation, and other fields. In order to further promote the development of small target detection, this paper makes a comprehensive summary of small target detection algorithms, and makes a reasonable classification, analysis and comparison of existing algorithms. Firstly, this paper defines the small object and summarizes the challenges of small object detection. Then, this paper focuses on the algorithms to improve the performance of small object detection from the aspects of data augmentation, multi-scale learning, context learning, generative adversarial learning, anchor-free mechanism, and analyzes the advantages and disadvantages, and relevance of these algorithms. Finally, this paper looks forward to the future development directions of small object detection.

Key words: small object detection; data augmentation; multi-scale learning; context learning; generative adversarial learning; anchor-free mechanism

---

基金项目：国家自然科学基金(62036007，62050175)资助项目。
收稿日期：2021-04-20；**修订日期：**2021-05-10

引 言

目标检测是计算机视觉领域中的一个重要研究方向,也是其他复杂视觉任务的基础。作为图像理解和计算机视觉的基石,目标检测是解决分割、场景理解、目标跟踪、图像描述和事件检测等更高层次视觉任务的基础。小目标检测长期以来是目标检测中的一个难点,其旨在精准检测出图像中可视化特征极少的小目标(32像素×32像素以下的目标)。在现实场景中,由于小目标是的大量存在,因此小目标检测具有广泛的应用前景,在自动驾驶、智慧医疗、缺陷检测和航拍图像分析等诸多领域发挥着重要作用。近年来,深度学习技术的快速发展为小目标检测注入了新鲜血液,使其成为研究热点。然而,相对于常规尺寸的目标,小目标通常缺乏充足的外观信息,因此难以将它们与背景或相似的目标区分开来。在深度学习的驱动下,尽管目标检测算法已取得了重大突破,但是对于小目标的检测仍然是不尽人意的。在目标检测公共数据集MS COCO[1]上,小目标和大目标在检测性能上存在显著差距,小目标的检测性能通常只有大目标的一半。由此可见,小目标检测仍然是充满挑战的。此外,真实场景是错综复杂的,通常会存在光照剧烈变化、目标遮挡、目标稠密相连和目标尺度变化等问题,而这些因素对小目标特征的影响是更加剧烈的,进一步加大了小目标检测的难度。事实上,小目标检测具有重要的研究意义和应用价值。对于机场跑道,路面上会存在微小物体,如螺帽、螺钉、垫圈、钉子和保险丝等,精准地检测出跑道的这些小异物将避免重大的航空事故和经济损失。对于自动驾驶,从汽车的高分辨率场景照片中准确地检测出可能引起交通事故的小物体是非常有必要的。对于工业自动化,同样需要小目标检测来定位材料表面可见的小缺陷。对于卫星遥感图像,图像中的目标,例如车、船,可能只有几十甚至几个像素。精确地检测出卫星遥感图像中的微小目标将有助于政府机构遏制毒品和人口贩运,寻找非法渔船并执行禁止非法转运货物的规定。综上所述,小目标检测具有广泛的应用价值和重要的研究意义。对小目标检测展开研究将有助于推动目标检测领域的发展,扩宽目标检测在现实世界的应用场景,提高中国的科技创新水平和加快中国全面步入智能化时代的步伐。

目标检测作为计算机视觉的基础研究,已有许多优秀的综述发表。Zou等[2]梳理了400多篇关于目标检测技术发展的论文,包括历史上的里程碑检测器、检测框架、评价指标、数据集、加速技术和检测应用等诸多内容,系统而全面地展现了目标检测这个领域的现状。Oksuz等[3]则从目标检测中存在的类别不平衡、尺度不平衡、空间不平衡以及多任务损失优化之间的不平衡等四大不平衡问题出发,对现有的目标检测算法进行了深入的总结。Zhao等[4]在对比总结目标检测中提及了小目标检测所面临的挑战。Agawal等[5]则在目标检测任务的主要挑战中简要介绍了几种常用的小目标检测方法。Chen等[6]立意于小目标检测的4大支柱性方法,详细描述了多尺度表示、上下文信息、超分辨率、区域建议以及其他方法等5类具代表性的网络,并介绍了部分小目标数据集。Tong等[7]从多尺度学习、数据增强、训练策略、基于上下文的检测和基于生成对抗网络的检测等5个维度全面回顾了基于深度学习的小目标检测方法,并在一些流行的小目标检测数据集上,对当前经典的小目标检测算法进行了比较分析。Liu等[8]在总结对比最近用于小目标检测的深度学习方法的基础上,还简单阐述了常规目标检测、人脸检测、航空图像目标检测以及图像分割等4个研究领域的相关技术。此外,还有文献[9-10]等中文综述中对小目标检测这一领域做了一定的总结工作。然而,文献[2]主要对一般目标检测算法进行了回顾,而对小目标检测方法的介绍甚少。文献[3]则主要关注于目标检测领域中存在的不平衡问题。文献[4-5]对目标检测领域进行了全面的综述总结,虽然有所涉及小目标检测问题,但是并没有进行全面的总结和深入的分析。文献[6-8]是针对小目标这一问题的综述,对小目标检测方法与性能评估进行了较为全面的总结,但是在对小目标的定义、难点分析和性能评估等方面仍有所欠缺。文献[9-10]作为中文的小目标检测综述,分别对小目标检测这一领域进行了总结综述,但是对于小目标检测方法的归类与分析

仍不够深入。

与以往将小目标与常规目标等同对待或只关注特定应用场景下的目标检测综述不同,本文对小目标检测这一不可或缺且极具挑战性的研究领域进行了系统且深入的分析与总结。本文不仅对小目标的定义进行了解释,也对小目标检测领域存在的挑战进行了详细地分析和总结,同时重点阐述了小目标检测优化思路,包括数据增强、多尺度学习、上下文学习、生成对抗学习以及无锚机制以及其他优化策略等。此外,本文还在常用的小目标数据集上分析对比了现有算法的检测性能。最后,对本文内容进行了简要的总结,并讨论了小目标检测未来可能的研究方向和发展趋势。

1 小目标定义及难点分析

1.1 小目标定义

不同场景对于小目标的定义各不相同,目前尚未形成统一的标准。现有的小目标定义方式主要分为以下两类,即基于相对尺度的定义与基于绝对尺度的定义。

(1)基于相对尺度定义。即从目标与图像的相对比例这一角度考虑来对小目标进行定义。Chen等[11]提出一个针对小目标的数据集,并对小目标做了如下定义:同一类别中所有目标实例的相对面积,即边界框面积与图像面积之比的中位数在0.08%~0.58%之间。文中对小目标的定义也给出了更具体的说法,如在640像素×480像素分辨率图像中,16像素×16像素到42像素×42像素的目标应考虑为小目标。除了Chen等对小目标的定义方式以外,较为常见的还有以下几种:(1)目标边界框的宽高与图像的宽高比例小于一定值,较为通用的比例值为0.1;(2)目标边界框面积与图像面积的比值开方小于一定值,较为通用的值为0.03;(3)根据目标实际覆盖像素与图像总像素之间比例来对小目标进行定义。

但是,这些基于相对尺度的定义存在诸多问题,如这种定义方式难以有效评估模型对不同尺度目标的检测性能。此外,这种定义方式易受到数据预处理与模型结构的影响。

(2)基于绝对尺度定义。则从目标绝对像素大小这一角度考虑来对小目标进行定义。目前最为通用的定义来自于目标检测领域的通用数据集——MS COCO数据集[1],将小目标定义为分辨率小于32像素×32像素的目标。对于为什么是32像素×32像素,本文从两个方向进行了思考。一种思路来自于Torralba等[12]的研究,人类在图像上对于场景能有效识别需要的彩色图像像素大小为32像素×32像素,即小于32像素×32像素的目标人类都难以识别。另一种思路来源于深度学习中卷积神经网络本身的结构,以与MS COCO数据集第一部分同年发布的经典网络结构VGG-Net[13]为例,从输入图像到全连接层的特征向量经过了5个最大池化层,这导致最终特征向量上的“一点”对应到输入图像上的像素大小为32像素×32像素。于是,从特征提取的难度不同这一角度考虑,可以将32像素×32像素作为区分小目标与常规目标的一个界定标准。除了MS COCO之外,还有其他基于绝对尺度的定义,如在航空图像数据集DOTA[14]与人脸检测数据集WIDER FACE[15]中都将像素值范围在[10, 50]之间的目标定义为小目标。在行人识别数据集CityPersons[16]中,针对行人这一具有特殊比例的目标,将小目标定义为了高度小于75像素的目标。基于航空图像的小行人数据集TinyPerson[17]则将小目标定义为像素值范围在[20, 32]之间的目标,而且近一步将像素值范围在[2, 20]之间的目标定义为微小目标。

1.2 小目标检测面临的挑战

前文中已简要阐述小目标的主流定义,通过这些定义可以发现小目标像素占比少,存在覆盖面积小、包含信息少等基本特点。这些特点在以往综述或论文中也多有提及,但是少有对小目标检测难点进行分析与总结。接下来本文将试图对造成小目标检测难度高的原因以及其面临的挑战进行分析与总结。

(1) 可利用特征少

无论是从基于绝对尺度还是基于相对尺度的定义,小目标相对于大/中尺度尺寸目标都存在分辨率低的问题。低分辨率的小目标可视化信息少,难以提取到具有鉴别力的特征,并且极易受到环境因素的干扰,进而导致了检测模型难以精准定位和识别小目标。

(2) 定位精度要求高

小目标由于在图像中覆盖面积小,因此其边界框的定位相对于大/中尺度尺寸目标具有更大的挑战性。在预测过程中,预测边界框框偏移一个像素点,对小目标的误差影响远高于大/中尺度目标。此外,现在基于锚框的检测器依旧占据绝大多数,在训练过程中,匹配小目标的锚框数量远低于大/中尺度目标,如图1所示,这进一步地导致了检测模型更侧重于大/中尺度目标的检测,难以检测小目标。图中IoU(Intersection over union)为交并比。

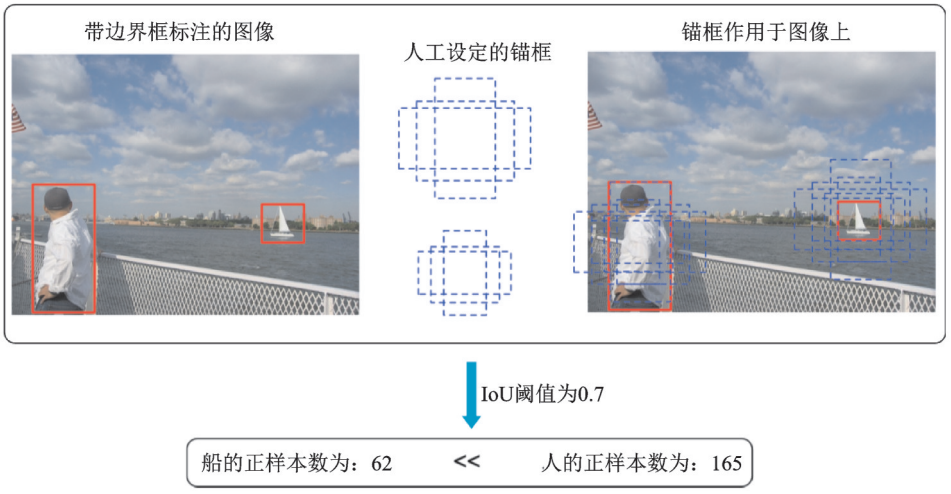


图1 小目标匹配的锚框数量相对大/中尺度的目标更少

Fig.1 Small-size objects match with fewer anchors than large/medium objects

(3) 现有数据集中小目标占比少

在目标检测领域中,现有数据集大多针对大/中尺度尺寸目标,较少关注小目标这一特别的类型。MS COCO中虽然小目标占比较高,达31.62%,但是每幅图像包含的实例过多,小目标分布并不均匀。同时,小目标不易标注,一方面来源于小目标在图像中不易被人类关注,很难标全;另一方面是小目标对于标注误差更为敏感。另外,现有的小目标数据集往往针对特定场景,例如文献[14]针对空中视野下的图像、文献[15]针对人脸、文献[16-17]针对行人、文献[18]针对交通灯、文献[19]针对乐谱音符,使用这些数据集训练的网络不适用于通用的小目标检测。总的来说,大规模的通用小目标数据集尚处于缺乏状态,现有的算法没有足够的先验信息进行学习,导致了小目标检测性能不足。

(4) 样本不均衡问题

为了定位目标在图像中的位置,现有的方法大多是预先在图像的每个位置生成一系列的锚框。在训练的过程中,通过设定固定的阈值来判断锚框属于正样本还是负样本。这种方式导致了模型训练过程中不同尺寸目标的正样本不均衡问题。当人工设定的锚框与小目标的真实边界框差异较大时,小目标的训练正样本将远远小于大/中尺度目标的正样本,这将导致训练的模型更加关注大/中尺度目标的检测,而忽略小目标的检测。如何解决锚框机制导致的小目标和大/中尺度目标样本不均衡问题也是

当前面临的一大挑战。

(5) 小目标聚集问题

相对于大/中尺度目标,小目标具有更大概率产生聚集现象。当小目标聚集出现时,聚集区域相邻的小目标通过多次降采样后,反应到深层特征图上将聚合成一个点,导致检测模型无法区分。当同类小目标密集出现时,预测的边界框还可能会因后处理的非极大值抑制操作将大量正确预测的边界框过滤,从而导致漏检情况。另外,聚集区域的小目标之间边界框距离过近,还将导致边界框难以回归,模型难以收敛。

(6) 网络结构原因

在目标检测领域,现有算法的设计往往更为关注大/中尺度目标的检测性能。针对小目标特性的优化设计并不多,加之小目标自身特性所带来的难度,导致现有算法在小目标检测上普遍表现不佳。虽然无锚框的检测器设计是一个新的发展趋势,但是现有网络依旧是基于锚框的检测器占据主流,而锚框这一设计恰恰对小目标极不友好。此外,在现有网络的训练过程中,小目标由于训练样本占比少,对于损失函数的贡献少,从而进一步减弱了网络对于小目标的学习能力。

2 小目标检测研究思路

2.1 数据增强

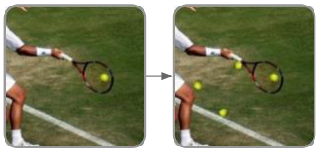
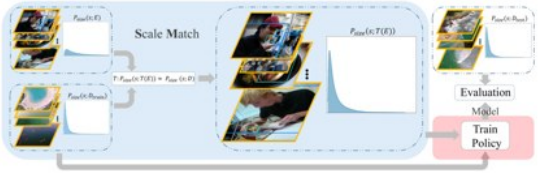
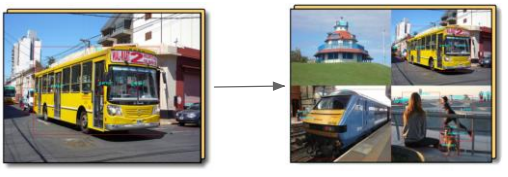
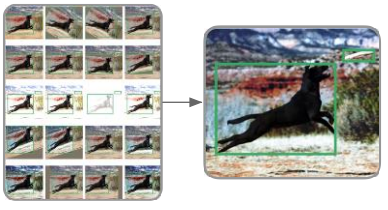
数据增强是一种提升小目标检测性能的最简单和有效的方法,通过不同的数据增强策略可以扩充训练数据集的规模,丰富数据集的多样性,从而增强检测模型的鲁棒性和泛化能力。在相对早期的研究中,Yaeger等[20]通过使用扭曲变形、旋转和缩放等数据增强方法显著提升了手写体识别的精度。之后,数据增强中又衍生出了弹性变形[21]、随机裁剪[22]和平移[23]等策略。目前,这些数据增强策略已被广泛应用于目标检测中。

近些年来,基于深度学习的卷积神经网络在处理计算机视觉任务中获得了巨大的成功。深度学习的成功很大程度上归功于数据集的规模和质量,大规模和高质量的数据能够大幅度提升模型的泛化能力。数据增强策略在目标检测领域有着广泛应用,例如Fast R-CNN[24]、Cascade R-CNN[25]中使用的水平翻转,YOLO[26]、YOLO9000[27]中使用的调整图像曝光和饱和度,还有常被使用的CutOut[28]、MixUp[29]、CutMix[30]等方法。最近,更是有诸如马赛克增强(YOLOv4[31])、保持增强[32]等创新策略提出,但是这些数据增强策略主要是针对常规目标检测。

聚焦到小目标检测领域,小目标面临着分辨率低、可提取特征少、样本数量匮乏及分布不均匀等诸多挑战,数据增强的重要性愈发显著。近些年来,出现了一些适用于小目标的数据增强方法(表1)。Yu等[17]在对数据的处理中,提出了尺度匹配策略,根据不同目标尺寸进行裁剪,缩小不同大小目标之间的差距,从而避免常规缩放操作中小目标信息易丢失的情形。Kisantal等[33]针对小目标覆盖的面积小、出现位置缺乏多样性、检测框与真值框之间的交并比远小于期望的阈值等问题,提出了一种复制增强的方法,通过在图像中多次复制粘贴小目标的方式来增加小目标的训练样本数,从而提升了小目标的检测性能。在Kisantal等的基础上,Chen等[34]在RRNet中提出了一种自适应重采样策略进行数据增强,这种策略基于预训练的语义分割网络对目标图像进行考虑上下文信息的复制,以解决简单复制过程中可能出现的背景不匹配和尺度不匹配问题,从而达到较好的数据增强效果。Chen等[35]则从小目标数量占比小、自身包含信息少等问题出发,在训练过程中对图像进行缩放与拼接,将数据集中的大尺寸目标转换为中等尺寸目标,中等尺寸目标转换为小尺寸目标,并在提高中/小尺寸目标的数量与质量的同时也兼顾考虑了计算成本。在针对小目标的特性设计对应的数据增强策略之外,Zoph等[36]超越了目标特性限制,提出了一种通过自适应学习方法例如强化学习选择最佳的数据增强策略,在小目标检

表1 适用于小目标的5种数据增强方法

Table 1 Five data augmentation methods for small objects

编号	增强策略	主要内容	年份	发表	引用量
1	复制增强[33] Artificial augmentation by copy pasting the small objects	 通过对图像中的小目标的复制与粘贴操作进行数据增强	2019	arXiv	68
2	自适应采样[34] AdaResampling	 在文献[33]的基础上，考虑上下文信息进行复制，避免出现尺度不匹配和背景不匹配的问题	2019	ICCV	10
3	尺度匹配[17] Scale match	 通过尺度匹配策略对图像进行尺度变换，用作额外的数据补充	2020	WACV	14
4	缩放与拼接[35] Component stitching	 通过缩放拼接操作增加中/小尺寸目标的数量与质量	2020	arXiv	6
5	自学习数据增强[36] Learning data augmentation strategies	 通过强化学习选择最佳数据增强策略	2020	ECCV	105

测上获得了一定的性能提升。

数据增强这一策略虽然在一定程度上解决了小目标信息量少、缺乏外貌特征和纹理等问题，有效提高了网络的泛化能力，在最终检测性能上获得了较好的效果，但同时带来了计算成本的增加。而且在实际应用中，往往需要针对目标特性做出优化，设计不当的数据增强策略可能会引入新的噪声，损害特征提取的性能，这也给算法的设计带来了挑战。

2.2 多尺度学习

小目标与常规目标相比可利用的像素较少，难以提取到较好的特征，而且随着网络层数的增加，小

目标的特征信息与位置信息也逐渐丢失,难以被网络检测。这些特性导致小目标同时需要深层语义信息与浅层表征信息,而多尺度学习将这两种相结合,是一种提升小目标检测性能的有效策略。

早期的多尺度检测有两个思路。一种是使用不同大小的卷积核通过不同的感受野大小来获取不同尺度的信息,但这种方法计算成本很高,而且感受野的尺度范围有限,Simonyan和Zisserman[13]提出使用多个小卷积核代替大卷积核具备巨大优势后,使用不同大小卷积核的方法逐渐被弃用。之后,Yu等[37]提出的空洞卷积和Dai等[38]提出的可变卷积又为这种通过不同感受野大小获取不同尺度信息的方法开拓了新的思路。另一种来自于图像处理领域的思路——图像金字塔[39],通过输入不同尺度的图像,对不同尺度大小的目标进行检测,这种方法在早期的目标检测中有所应用[40-41](见图2(a))。但是,基于图像金字塔训练卷积神经网络模型对计算机算力和内存都有极高的要求。近些年来,图像金字塔在实际研究应用中较少被使用,仅有文献[42-43]等方法针对数据集目标尺度差异过大等问题而使用。

目标检测中的经典网络如Fast R-CNN[24]、Faster R-CNN[44]、SPPNet[45]和R-FCN[46]等大多只是利用了深度神经网络的最后层来进行预测。然而,由于空间和细节特征信息的丢失,难以在深层特征图中检测小目标。在深度神经网络中,浅层的感受野更小,语义信息弱,上下文信息缺乏,但是可以获得更多空间和细节特征信息。从这一思路出发,Liu等[47]提出一种多尺度目标检测算法SSD(Single shot multibox detector),利用较浅层的特征图来检测较小的目标,而利用较深层的特征图来检测较大的目标,如图2(b)所示。Cai等[48]针对小目标信息少,难以匹配常规网络的问题,提出统一多尺度深度卷积神经网络,通过使用反卷积层来提高特征图的分辨率,在减少内存和计算成本的同时显著提升了小目标的检测性能。

针对小目标易受环境干扰问题,Bell等[49]为提出了ION(Inside-outside network)目标检测方法,通过从不同尺度特征图中裁剪出同一感兴趣区域的特征,然后综合这些多尺特征来预测,以达到提升检测性能的目的。与ION的思想相似,Kong等[50]提出了一种有效的多尺度融合网络,即HyperNet,通过综合浅层的高分辨率特征和深层的语义特征以及中间层特征的信息显著提高了召回率,进而提高了小目标检测的性能(见图2(c))。这些方法能有效利用不同尺度的信息,是提升小目标特征表达的一种有效手段。但是,不同尺度之间存在大量重复计算,对于内存和计算成本的开销较大。

为节省计算资源并获得更好的特征融合效果,Lin等[51]结合单一特征映射、金字塔特征层次和综合特征的优点,提出了特征金字塔FPN(Feature Pyramid network)。FPN是目前最流行的多尺度网络,它引入了一种自底向上、自顶向下的网络结构,通过将相邻层的特征融合以达到特征增强的目的(见图2(d))。在FPN的基础上,Liang等[52]提出了一种深度特征金字塔网络,使用具有横向连接的特征金字塔结构加强小目标的语义特征,并辅以特别设计的锚框和损失函数训练网络。为了提高小目标的检测速度,Cao等[53]提出一种多层次特征融合算法,即特征融合SSD,在SSD的基础上引入上下文信息,较好地平衡了小目标检测的速度与精度。但是基于SSD的特征金字塔方法需要从网络的不同层中抽取不同尺度的特征图进行预测,难以充分融合不同尺度的特征。针对这一问题,Li和Zhou[54]提出一种特征融合单次多箱探测器,使用一个轻量级的特征融合模块,联系并融合各层特征到一个较大的尺度,然后在得到的特征图上构造特征金字塔用于检测,在牺牲较少速度的情形下提高了对小目标的检测性能。针对机场视频监控中的小目标识别准确率较低的问题,韩松臣等[55]提出了一种结合多尺度特征融合与在线难例挖掘的机场路面小目标检测方法,该方法采用ResNet-101作为特征提取网络,并在该网络基础上建立了一个带有上采样的“自顶向下”的特征融合模块,以生成语义信息更加丰富的高分辨率特征图。

最近,多尺度特征融合这一方法又有了新的拓展,如Nayan等[56]针对小目标经过多层网络特征信息易丢失这一问题,提出了一种新的实时检测算法,该算法使用上采样和跳跃连接在训练过程中提取

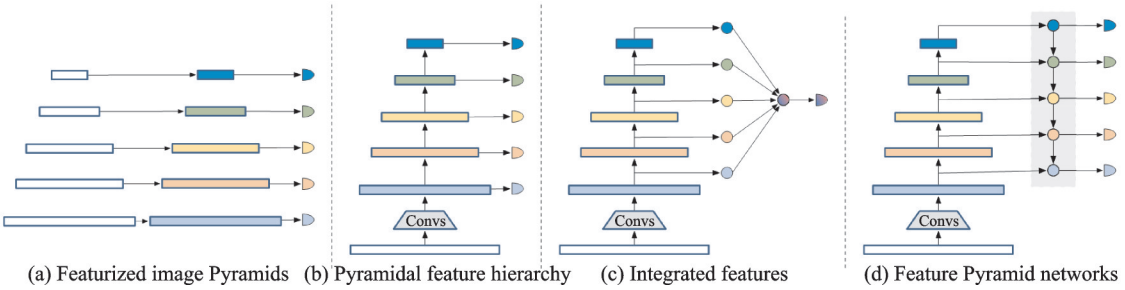


图2 多尺度学习的4种方式

Fig.2 Four ways of multi-scale learning

不同网络深度的多尺度特征，显著提高了小目标检测的检测精度与速度。Liu等[57]为了降低高分辨率图像的计算成本，提出了一种高分辨率检测网络，通过使用浅层网络处理高分辨率图像和深层网络处理低分辨率图像，在保留小目标尽可能多的位置信息同时提取了更多的语义信息，在降低计算成本的情形下提升了小目标的检测性能。Deng等[58]发现虽然多尺度融合可以有效提升小目标检测性能，但是不同尺度的特征耦合仍然会影响性能，于是提出了一种扩展特征金字塔网络，使用额外的高分辨率金字塔级专门用于小目标检测。

总体来说，多尺度特征融合同时考虑了浅层的表征信息和深层的语义信息，有利于小目标的特征提取，能够有效地提升小目标检测性能。然而，现有多尺度学习方法在提高检测性能的同时也增加了额外的计算量，并且在特征融合过程中难以避免干扰噪声的影响，这些问题导致了基于多尺度学习的小目标检测性能难以得到进一步提升。

2.3 上下文学习

在真实世界中，“目标与场景”和“目标与目标”之间通常存在一种共存关系，通过利用这种关系将有助于提升小目标的检测性能。在深度学习之前，已有研究[59]证明通过对上下文进行适当的建模可以提升目标检测性能，尤其是对于小目标这种外观特征不明显的目标。随着深度神经网络的广泛应用，一些研究也试图将目标周围的上下文集成到深度神经网络中，并取得了一定的成效。以下将从基于隐式上下文特征学习和基于显式上下文推理的目标检测两个方面对国内外研究现状及发展动态进行简要综述。

(1)基于隐式上下文特征学习的目标检测。隐式上下文特征是指目标区域周围的背景特征或者全局的场景特征。事实上，卷积神经网络中的卷积操作在一定程度上已经考虑了目标区域周围的隐式上下文特征。为了利用目标周围的上下文特征，Li等[60]提出一种基于多尺度上下文特征增强的目标检测方法，该方法首先在图像中生成一系列的目标候选区域，然后在目标周围生成不同尺度的上下文窗口，最后利用这些窗口中的特征来增强目标的特征表示(见图3(a))。随后，Zeng等[61]提出一种门控双向卷积神经网络，该网络同样在目标候选区域的基础上生成包含不同尺度上下文的支撑区域，不同之处在于该网络让不同尺度和分辨率的信息在生成的支撑区域之间相互传递，从而综合学习到最优的特征。为了更好地检测复杂环境下的微小人脸，Tang等[62]提出一种基于上下文的单阶段人脸检测方法，该方法设计了一种新的上下文锚框，在提取人脸特征的同时考虑了其周围的上下文信息，例如头部信息和身体信息。郑晨斌等[63]提出一种强化上下文模型网络，该网络利用双空洞卷积结构来节省参数量的同时，通过扩大有效感受野来强化浅层上下文信息，并在较少破坏原始目标检测网络的基础上灵活作用于网络中浅预测层。然而，这些方法大多依赖于上下文窗口的设计或受限于感受野的大小，可能会导

致重要上下文信息的丢失。

为了更加充分地利用上下文信息,一些方法尝试将全局的上下文信息融入到目标检测模型中(见图3(b))。对于早期的目标检测算法,一种常用的集成全局上下文方法是通过构成场景元素的统计汇总,例如Gist[64]。Torralba等[65]提出通过计算全局场景的低级特征和目标的特征描述符的统计相关性来对视觉上下文建模。随后,Felzenszwalb等[66]提出一种基于混合多尺度可变形部件模型的目标检测方法。该方法通过引入上下文来对检测结果进行二次评分,从而进一步提升检测结果的可靠性。对于目前的基于深度学习的目标检测算法,主要通过较大的感受野、卷积特征的全局池化或把全局上下文看作一种序列信息3种方式来感知全局上下文。Bell等[49]提出基于循环神经网络的上下文传递方法,该方法利用循环神经网络从4个方向对整个图像中的上下文信息进行编码,并将得到的4个特征图进行串联,从而实现对全局上下文的感知。然而,该方法使模型变得复杂,并且模型的训练严重依赖于初始化参数的设置。Ouyang等[67]通过学习图像的分类得分,并将该得分作为补充的上下文特征来提升目标检测性能。为了提升候选区域的特征表示,Chen等[68]提出一种上下文微调网络,该网络首先通过计算相似度找到与目标区域相关的上下文区域,然后利用这些上下文区域的特征来增强目标区域特征。随后,Barnea等[69]将上下文的利用视为一个优化问题,讨论了上下文或其他类型的附加信息可以将检测分数提高到什么程度,并表明简单的共现性关系是最有效的上下文信息。此外,Chen等[70]提出一种层次上下文嵌入框架,该框架可以作为一个即插即用的组件,通过挖掘上下文线索来增强候选区域的特征表达,从而提升最终的检测性能。最近,张瑞琰等[71]提出了面向光学遥感目标的全局上下文检测模型,该模型通过全局上下文特征与目标中心点局部特征相结合的方式生成高分辨率热点图,并利用全局特征实现目标的预分类。此外,一些方法通过语义分割来利用全局上下文信息。He等[72]提出一种统一的实例分割框架,利用像素级的监督来优化检测器,并通过多任务的方式联合优化目标检测和实例分割模型。尽管通过语义分割可以显著提高检测性能,但是像素级的标注是非常昂贵的。鉴于此,Zhao等[73]提出一种生成伪分割标签的方法,通过利用伪分割标签来于优化检测器,并取得了不错的效果。进一步地,Zhang等[74]提出一种无监督的分割方法,在无像素级的标注下通过联合优化目标检测和分割来增强用于目标检测的特征图。目前,基于全局上下文的方法在目标检测上已经取得了较大的进展,但如何从全局场景中找到有利于提升小目标检测性能的上下文信息仍然是当前的研究难点。

(2)基于显式上下文推理的目标检测。显示上下文推理是指利用场景中明确的上下文信息来辅助推断目标的位置或类别,例如利用场景中天空区域与目标的上下文关系来推断目标的类别。上下文关系通常指场景中目标与场景或者目标与目标之间的约束和依赖关系(见图3(c))。为了利用上下文关系,Chen等[75]提出一种自适应上下文建模和迭代提升的方法,通过将一个任务的输出作为另一个任务的上下文来提升目标分类和检测性能。此后,Gupta等[76]提出一种基于空间上下文的目标检测方法。该方法能够准确地捕捉到上下文和感兴趣目标之间的空间关系,并且有效地利用了上下文区域的外观特征。进一步地,Liu等[77]提出一种结构推理网络,通过充分考虑场景上下文和目标之间的关系来提升目标的检测性能。为了利用先验知识,Xu等[78]在Faster R-CNN[44]的基础上提出了一种Reasoning-RCNN,通过构建知识图谱来编码上下文关系,并利用先验的上下文关系来影响目标检测。Chen等[79]提出了一种空间记忆网络,空间记忆实质上是将目标实例重新组合成一个伪图像表示,并将伪图像表示输入到卷积神经网络中进行目标关系推理,从而形成一种顺序推理体系结构。在注意力机制的基础上,Hu等[80]提出一种轻量级目标关系网络,通过引入不同物体之间的外观和几何结构关系来做约束,实现物体之间的关系建模。该网络无需额外的监督,并且易于嵌入到现有的网络中,可以有效地过滤冗余框,从而提升目标的检测性能。

近年来,基于上下文学习的方法得到了进一步发展。Lim等[81]提出一种利用上下文连接多尺度特

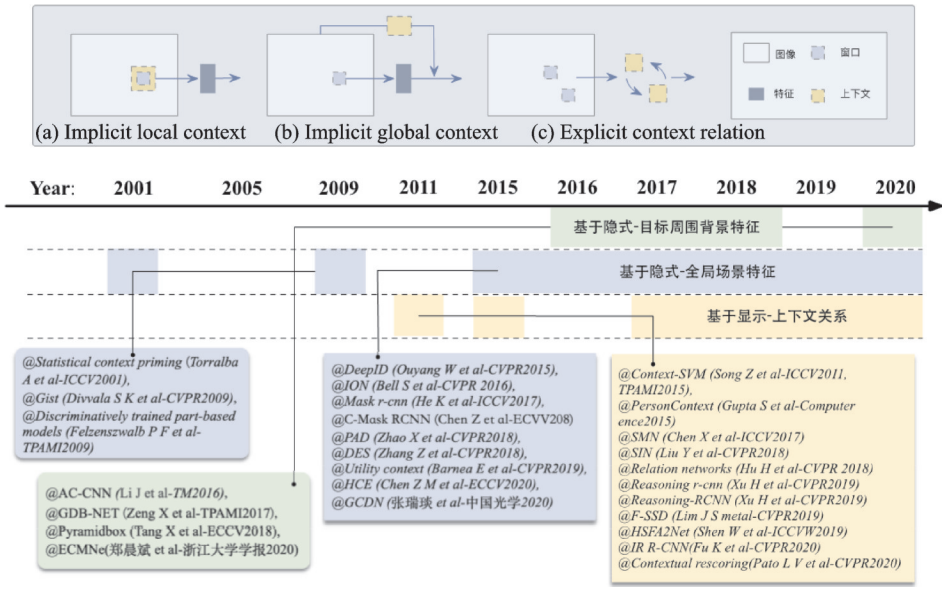


图3 上下文在目标检测中的探索历程

Fig.3 Exploration of context in object detection

征的方法，该方法中使用网络不同深度层级中的附加特征作为上下文，辅以注意力机制聚焦于图像中的目标，充分利用了目标的上下文信息，进而提升了实际场景中的小目标检测精度。针对室内小尺度人群检测面临的目标特征与背景特征重叠且边界难以区分的问题，Shen等[82]提出了一种室内人群检测网络框架，使用一种特征聚合模块（Feature aggregation module，FAM）通过融合和分解的操作来聚合上下文特征信息，为小尺度人群检测提供更多细节信息，进而显著提升了对于室内小尺度人群的检测性能。Fu等[83]提出了一种新颖的上下文推理方法，该方法对目标之间的固有语义和空间布局关系进行建模和推断，在提取小目标语义特征的同时尽可能保留其空间信息，有效解决了小目标的误检与漏检问题。为了提升目标的分类结果，Pato等[84]提出一种基于上下文的检测结果重打分方法，该方法通过循环神经网络和自注意力机制来传递候选区域之间的信息并生成上下文表示，然后利用得到的上下文来对检测结果进行二次评估。

基于上下文学习的方法充分利用了图像中与目标相关的信息，能够有效提升小目标检测的性能。但是，已有方法没有考虑到场景中的上下文信息可能匮乏的问题，同时没有针对性地利用场景中易于检测的结果来辅助小目标的检测。鉴于此，未来的研究方向可以从以下两个角度出发考虑：(1)构建基于类别语义池的上下文记忆模型，通过利用历史记忆的上下文来缓解当前图像中上下文信息匮乏的问题；(2)基于图推理的小目标检测，通过图模型和目标检测模型的结合来针对性地提升小目标的检测性能。

2.4 生成对抗学习

生成对抗学习的方法旨在通过将低分辨率小目标的特征映射成与高分辨率目标等价的特征，从而达到与尺寸较大目标同等的检测性能。前文所提到的数据增强、特征融合和上下文学习等方法虽然可以有效地提升小目标检测性能，但是这些方法带来的性能增益往往受限于计算成本。针对小目标分辨率低问题，Haris等[85]提出一种端到端的联合训练超分辨率和检测模型的方法，该方法一定程度上提升了低分辨率目标的检测性能。但是，这种方法对于训练数据集要求较高，并且对小目标检测性能的提

升不足。

目前,一种有效的方法是通过结合生成对抗网络(Generative adversarial network, GAN)[86]来提高小目标的分辨率,缩小小目标与大/中尺度目标之间的特征差异,增强小目标的特征表达,进而提高小目标检测的性能。在Radford等[87]提出了DCGAN(Deep convolutional GAN)后,计算视觉的诸多任务开始利用生成对抗模型来解决具体任务中面临的问题。针对训练样本不足的问题,Sixt等[88]提出了RenderGAN,该网络通过对抗学习来生成更多的图像,从而达到数据增强的目的。为了增强检测模型的鲁棒性,Wang等[89]通过自动生成包含遮挡和变形特征的样本,以此提高对困难目标的检测性能。随后,Li等[90]提出了一种专门针对小目标检测的感知GAN方法,该方法通过生成器和鉴别器相互对抗的方式来学习小目标的高分辨率特征表示。在感知GAN中,生成器将小目标表征转换为与真实大目标足够相似的超分辨表征。同时,判别器与生成器对抗以识别生成的表征,并对生成器施加条件要求。该方法通过生成器和鉴别器相互对抗的方式来学习小目标的高分辨率特征表示。这项工作将小目标的表征提升为"超分辨"表征,实现了与大目标相似的特性,获得了更好的小目标检测性能。

近年来,基于GAN对小目标进行超分辨率重建的研究有所发展,Bai等[91]提出了一种针对小目标的多任务生成对抗网络(Multi-task generative adversarial network, MTGAN)。在MTGAN中,生成器是一个超分辨率网络,可以将小模糊图像上采样到精细图像中,并恢复详细信息以便更准确地检测。判别器是多任务网络,区分真实图像与超分辨率图像并输出类别得分和边界框回归偏移量。此外,为了使生成器恢复更多细节以便于检测,判别器中的分类和回归损失在训练期间反向传播到生成器中。MTGAN由于能够从模糊的小目标中恢复清晰的超分辨目标,因此大幅度提升了小目标的检测性能。进一步地,针对现有的用于小目标检测的超分辨率模型存在缺乏直接的监督问题,Noh等[92]提出一种新的特征级别的超分辨率方法,该方法通过空洞卷积的方式使生成的高分辨率目标特征与特征提取器生成的低分辨率特征保持相同的感受野大小,从而避免了因感受野不匹配而生成错误超分特征的问题。此外,Deng等[58]设计了一种扩展特征金字塔网络,该网络通过设计的特征纹理模块生成超高分辨率的金字塔层,从而丰富了小目标的特征信息。

基于生成对抗模型的目标检测算法通过增强小目标的特征信息,可以显著提升检测性能。同时,利用生成对抗模型来超分小目标这一步骤无需任何特别的结构设计,能够轻易地将已有的生成对抗模型和检测模型相结合。但是,目前依旧面临两个无法避免的问题:(1)生成对抗网络难以训练,不易在生成器和鉴别器之间取得好的平衡;(2)生成器在训练过程中产生样本的多样性有限,训练到一定程度后对于性能的提升有限。

2.5 无锚机制

锚框机制在目标检测中扮演着重要的角色。许多先进的目标检测方法都是基于锚框机制而设计的,但是锚框这一设计对于小目标的检测极不友好。现有的锚框设计难以获得平衡小目标召回率与计算成本之间的矛盾,而且这种方式导致了小目标的正样本与大目标的正样本极度不均衡,使得模型更加关注于大目标的检测性能,从而忽视了小目标的检测。极端情况下,设计的锚框如果远远大于小目标,那么小目标将会出现无正样本的情况。小目标正样本的缺失,将使得算法只能学习到适用于较大目标的检测模型。此外,锚框的使用引入了大量的超参,比如锚框的数量、宽高比和大小等,使得网络难以训练,不易提升小目标的检测性能。近些年无锚机制的方法成为了研究热点,并在小目标检测上取得了较好效果。

一种摆脱锚框机制的思路是将目标检测任务转换为关键点的估计,即基于关键点的目标检测方法。基于关键点的目标检测方法主要包含两个大类:基于角点的检测和基于中心的检测。基于角点的检测器通过对从卷积特征图中学习到的角点分组来预测目标边界框。DeNet[93]将目标检测定义为估计

目标4个角点的概率分布,包括左上角、右上角、左下角和右下角(见图4(a))。首先利用标注数据来训练卷积神经网络,然后利用该网络来预测角点分布。之后,利用角点分布和朴素贝叶斯分类器来确定每个角点对应的候选区域是否包含目标。在DeNet之后,Wang等[94]提出了一种新的使用角点和中心点之间的连接来表示目标的方法,命名为PLN(Point linking network)。PLN首先回归与DeNet相似的4个角点和目标的中心点,同时通过全卷积网络预测关键点两两之间是否相连,然后将角点及其相连的中心点组合起来生成目标边界框。PLN对于稠密目标和具有极端宽高比率目标表现良好。但是,当角点周围没有目标像素时,PLN由于感受野的限制将很难检测到角点。继PLN之后,Law等[95]提出了一种新的基于角点的检测算法,命名为CornerNet。CornerNet将目标检测问题转换为角点检测问题,首先预测所有目标的左上和右下的角点,然后将这些角点进行两两匹配,最后利用配对的角点生成目标的边界框。CornetNet的改进版本——CornerNet-Lite[96],从减少处理的像素数量和减少在每个像素上进行的计算数量两个角度出发进行改进,有效解决了目标检测中的两个关键用例:在不牺牲精度的情况下提高效率以及实时效率的准确性。与基于锚框的检测器相比,CornerNet系列具有更简洁的检测框架,在提高检测效率的同时获得了更高的检测精度。但是,该系列仍然会因为错误的角点匹配预测出大量不正确的目标边界框。

为了进一步提高目标检测性能,Duan等[97]提出了一种基于中心预测的目标检测框架,称为CenterNet(见图4(b))。CenterNet首先预左上角和右下角的角点以及中心关键点,然后通过角点匹配确定边界框,最后利用预测的中心点消除角点不匹配引起的不正确的边界框。与CenterNet类似,Zhou等[98]通过对极值点和中心点进行匹配,提出了一种自下而上的目标检测网络,称为ExtremeNet。ExtremeNet首先使用一个标准的关键点估计网络来预测最上面、最下面、最左边、最右边的4个极值点和中心点,然后在5个点几何对齐的情况下对它们进行分组以生成边界框。但是ExtremeNet和CornerNet等基于关键点的检测网络都需要经过一个关键点分组阶段,这降低了算法整体的速度。针对这一问题,Zhou等[99]将目标建模为其一个单点,即边界框中心点,无需对构建点进行分组或其他后处理操作。然后在探测器使用关键点估计来查找中心点,并回归到所有其他对象属性,如大小、位置等。这一方法很好地平衡了检测的精度与速度。

近年来,基于关键点的目标检测方法又有了新的扩展。Yang等[100]提出了一种名为代表点(RepPoints)的检测方法,提供了更细粒度的表示方式,使得目标可以被更精细地界定。同时,这种方法能够自动学习目标的空间信息和局部语义特征,一定程度上提升了小目标检测的精度(见图4(c))。更进一步地,Kong等[101]受到人眼的中央凹(视网膜中央区域,集中了绝大多数的视锥细胞,负责视力的高清成像)启发,提出了一种直接预测目标存在的可能性和边界框坐标的方法,该方法首先预测目标存在的可能性,并生成类别敏感语义图,然后为每一个可能包含目标的位置生成未知类别的边界框。由于摆脱了锚框的限制,FoveaBox对于小目标等具有任意横纵比的目标具备良好的鲁棒性和泛化能力,并在检测精度上也得到了较大提升。与FoveaBox相似,Tian等[102]使用语义分割的思想来解决目标检测问题,提出了一种基于全卷积的单级目标检测器FCOS(Fully convolutional one-stage),避免了基于锚框机制的方法中超参过多、难以训练的问题(见图4(d))。此外,实验表明将两阶段检测器的第一阶段任务换成FCOS来实现,也能有效提升检测性能。而后,Zhu等[103]将无锚机制用于改进特征金字塔中的特征分配问题,根据目标语义信息而不是锚框来为目标选择相应特征,同时提高了小目标检测的精度与速度。Zhang等[104]则从基于锚框机制与无锚机制的本质区别出发,即训练过程中对于正负样本的定义不同,提出了一种自适应训练样本选择策略,根据对象的统计特征自动选择正反样本。针对复杂的场景下小型船舶难以检测的问题,Fu等[105]提出了一种新的检测方法——特征平衡与细化网络,采用直接学习编码边界框的一般无锚策略,消除锚框对于检测性能的负面影响,并使用基于语义信息的注意力

机制平衡不同层次的多个特征,达到了最先进的性能。为了更有效地处理无锚框架下的多尺度检测,Yang等[106]提出了一种基于特殊注意力机制的特征金字塔网络,该网络能够根据不同大小目标的特征生成特征金字塔,进而更好地处理多尺度目标检测问题,显著提升了小目标的检测性能。

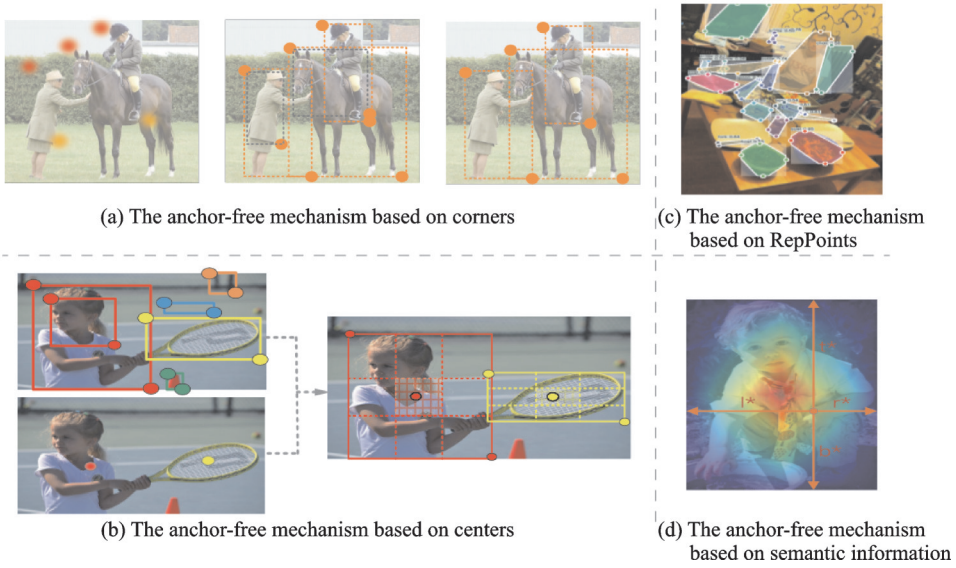


图4 无锚机制的4种形式

Fig.4 Four ways of anchor-free methods

2.6 其他优化策略

在小目标检测这一领域,除了前文所总结的几个大类外,还有诸多优秀的方法。针对小目标训练样本少的问题,Kisantal等[33]提出了一种过采样策略,通过增加小目标对于损失函数的贡献,以此提升小目标检测的性能。除了增加小目标样本权重这一思路之外,另一种思路则是通过增加专用于小目标的锚框数量来提高检测性能。Zhang等[107]提出了一种密集锚框策略,通过在一个感受野中心设计多个锚框来提升小目标的召回率。与密集锚框策略相近,Zhang等[108]设计了一种基于有效感受野和等比例区间界定锚框尺度的方法,并提出一种尺度补偿锚框匹配策略来提高小人脸目标的召回率。增加锚框数量对于提升小目标检测精度十分有效,同时也额外增加了巨大的计算成本。Eggert等[109]从锚框尺度的优化这一角度入手,通过推导小目标尺寸之间的联系,为小目标选择合适的锚框尺度,在商标检测上获得了较好的检测效果。之后,Wang等[110]提出了一种基于语义特征的引导锚定策略,通过同时预测目标中心可能存在的位置及目标的的尺度和纵横比,提高了小目标检测的性能。此外,这种策略可以集成到任何基于锚框的方法中。但是,这些改进没有实质性地平衡检测精度与计算成本之间的矛盾。

近些年来,随着计算资源的增加,越来越多的网络使用级联思想来平衡目标漏检率与误检率。级联这一思想来源已久[111],并在目标检测领域得到了广泛的应用。它采用了从粗到细的检测理念:用简单的计算过滤掉大多数简单的背景窗口,然后用复杂的窗口来处理那些更困难的窗口。随着深度学习时代的到来,Cai等[25]提出了经典网络Cascade R-CNN,通过级联几个基于不同IoU阈值的检测网络达到不断优化预测结果的目的。之后,Li等[112]在Cascade R-CNN的基础上进行了扩展,进一步提升了小目标检测性能。受到级联这一思想的启发,Liu等[113]提出了一种渐近定位策略,通过不断增加IoU阈值来提升行人检测的检测精度。另外,文献[114-116]展现了级联网络在困难目标检测上的应用,也一定程度上提升了小目标的检测性能。

另外一种思路则是分阶段检测，通过不同层级之间的配合平衡漏检与误检之间的矛盾。Chen等[117]提出一种双重探测器，其中第一尺度探测器最大限度地检测小目标，第二尺度探测器则检测第一尺度探测器无法识别的物体。进一步地，Drenkow等[118]设计了一种更加高效的目标检测方法，该方法首先在低分辨率下检查整个场景，然后使用前一阶段生成的显著性地图指导后续高分辨率下的目标检测。这种方式很好地权衡了检测精度和检测速度。此外，文献[119-121]针对空中视野图像中的困难目标识别进行了前后景的分割，区分出重要区域与非重要区域，在提高检测性能的同时也减少了计算成本。

优化损失函数也是一种提升小目标检测性能的有效方法。Redmon等[26]发现，在网络的训练过程中，小目标更容易受到随机误差的影响。随后，他们针对这一问题进行了改进[27]，提出一种依据目标尺寸设定不同权重的损失函数，实现了小目标检测性能的提升。Lin等[122]则针对类别不均衡问题，在RetinaNet中提出了焦距损失，有效解决了训练过程中存在的前景-背景类不平衡问题。进一步地，Zhang等[123]将级联思想与焦距损失相结合，提出了Cascade RetinaNet，进一步提高了小目标检测的精度。针对小目标容易出现的前景与背景不均衡问题，Deng等[58]则提出了一种考虑前景-背景之间平衡的损失函数，通过全局重建损失和正样本块损失提高前景与背景的特征质量，进而提升了小目标检测的性能。

为了权衡考虑小目标的检测精度和速度，Sun等[124]提出了一种多接受域和小目标聚焦弱监督分割网络，通过使用多个接收域块来关注目标及其相邻背景，并依据不同空间位置设置权重，以达到增强特征可辨识性的目的。此外，Yoo等[125]将多目标检测任务重新表述为边界框的密度估计问题，提出了一种混合密度目标检测器，通过问题的转换避免了真值框与预测框匹配以及启发式锚框设计等繁琐过程，也一定程度上解决了前景与背景不平衡的问题。

3 数据集介绍及性能评估

在常规目标检测数据集上，现有研究对大/中尺寸的目标已取得了不错的成效。但是，小目标的检测仍然是不尽人意的，一方面是由小目标自身特性所导致的的，另一方面是因为常规目标检测数据集中小目标存在占比少、分布不均匀等问题。接下来本文将按照时间顺序简要介绍现有的小目标数据集(见表2)，并在一些公用数据集上对现有算法进行性能评估(见表3～6)。这些数据可供研究人员参考，希望可以为小目标检测的研究发展贡献微薄之力。

3.1 数据集介绍

(1)BIRDSAI数据集[126]。BIRDSAI寓意鸟的眼睛(bird's-eye)，由Bondi等在WACV 2020(Winter Conference on Applications of Computer Vision 2020)上提出。该数据集使用带有红外摄像机的固定翼无人机收集，是第1个覆盖多个非洲保护区的大型数据集。主要由人类和动物的红外图像视频组成，总共包含10个类别：－1：未知，0：人类，1：大象，2：狮子，3：长颈鹿，4：狗，5：鳄鱼，6：河马，7：斑马，8：犀牛。其中涉及几个具有挑战性的场景，如尺度变化、热反射导致的背景杂波、大尺度旋转和运动模糊等。此外，该数据集还包含使用微软开源的AirSim模拟平台，即使用非洲热带草原的3D模型和TIR相机模型合成的虚拟视频。随着航空图像用于监测/监视场景的普及，该数据集将有助于推动基于航空红外视频图像的目标检测、目标跟踪以及领自适应等领域的研究。除了促进相关领域研究外，这个数据集也将有助于野生动物保护，成功的算法可以用来有效计数或跟踪保护区内的野生动物，进而避免野生动物偷猎。

(2)TinyPerson数据集[17]。随着深度卷积神经网络的兴起，视觉目标检测取得了前所未有的进展。然而，在大尺度图像中检测小于20像素的极小目标仍然没有得到很好的研究。对于极小目标的检测，

一方面的挑战来自于其特征表示微弱,另一方面是复杂背景中存在大量相似特征增加了误报的风险。为了促进对于极小目标检测的研究,Yu等提出该数据集——TinyPerson,这是第1个远距离和大背景下进行人员检测的基准,为极小目标检测开辟了一个新的前景方向。该数据集由1 610幅图像构成,每幅图像包含超过200个人员,其中目标分为5个类别,共有72 651个手工标注的极小目标。

(3)EuroCity Persons数据集[127]。EuroCity Persons数据集由Braun等提出,该数据集主要为城市交通场景,包含大量种类繁多、准确且详细的目标,如行人、骑自行车者和其他乘客等。其中图像由一辆移动车辆在12个欧洲国家的31个城市收集。EuroCity Persons这一数据集包含47 300多张图像,含有手工标记的超过238 200个人员实例,比以前用于基准测试的人员数据集几乎大了一个数量级。特别地,该数据集还包含超过211 200条标明人员朝向的注释。总的来说,该数据集数量大、种类多、细节详尽,将城市交通场景中的人员注释提升到了一个新的水平。

(4)WiderPerson数据集[128]。WiderPerson是一个户外密集行人检测基准数据集,其中的图像不局限于交通场景而包含了更广泛的较拥挤场景。该数据集由13 382张图像组成,涉及5种类型的注释,共包含约400K条带有多种遮挡信息的标注,平均每幅图像标注29.87个目标,这意味着该数据集包含了各种遮挡下的密集行人。在该数据集中,训练集、验证集和测试集由随机选择的8 000/1 000/4 382张图像分别构成。与后文将提到的CityPersons和WIDER FACE数据集相似,WiderPerson数据集不发布测试图像的标注文件。

(5)DOTA数据集[14]。为了促进“Earth Vision”中的目标检测研究,Xia等提出了用于航空图像中目标检测的大型数据集DOTA。该数据集包含从不同传感器和平台上收集的2 806幅航拍图像。每幅图像的大小约为4 000像素×4 000像素,包含了各种尺度、方向和形状的对象。这些DOTA图像由航拍图像解译方面的专家使用15种常见的目标类别进行注释。完整注释的DOTA图像包含188 282个实例,每个实例都由一个任意四边形标记。

(6)Nighttowls数据集[129]。Nighttowls是一个用于夜间行人检测的公共数据集。不同于常规的白天场景,夜间的行人检测,由于存在更复杂的低光照、反射、模糊和变化的图像对比度等问题,更具挑战性。该数据集由行业标准相机跨越3个国家,在不同的季节和天气条件下拍摄,包含40个序列,共279 000帧的夜间影像。所有的图像都有详尽的注释,其中目标类别分为行人、骑自行车者、骑摩托车者和忽略区域4类。此外,注释汇总还包含了目标的额外属性,如遮挡、姿势和难度等,以及用于在多个帧中识别相同对象的跟踪信息。

(7)DeepScores数据集[19]。DeepScores是由Lukas等提出一个十分特别的小目标数据集,包含高质量的乐谱图像,由30万张包含不同形状和大小音乐符号的图像组成,共接近一亿个小目标,是最大的公共数据集。该数据集中提供了用于目标分类、目标检测和语义分割的真值标注,而且前10%的类含有整个数据集中85%的标志,可以用来模拟异常检测中的真实世界数据流。DeepScores通过将对象识别问题置于场景理解的背景下,意图促进小目标识别领域的研究,同时也对计算机视觉,尤其是光学音乐识别研究提出了相关挑战。

(8)Bosch小交通灯数据集[18]。Bosch小交通灯数据集是一个基于视觉图像的交通灯检测的精准数据集。该数据集由13 427幅分辨率为1 280像素×720像素的摄像机图像组成,其中包含约24 000个带标注的交通信号灯。标注信息包括交通灯的边框以及每个交通灯的当前状态。该数据集图像包含摄像机拍摄的原始12位HDR图像和重构的8位RGB彩色图像。RGB图像可用于训练和测试,但由于原始图像的压缩转换问题,RGB图像可能颜色异常或包含伪像。

(9)CityPersons数据集[16]。为了更好地训练数据,CityPersons这一数据集由Zhang等基于

Cityscape数据集[130]提出。Cityscape数据集是一个大型数据集,包含来自50个不同城市街道场景中记录的多种立体视频序列,除了20 000个弱注释帧以外,还包含5 000帧高质量像素级注释。Citypersons数据集基于Cityscapes数据集为27个城市的5 000幅图像提供了30个视觉类的精细像素级注释,精细的标注包括人员和车辆的实例标签。另外来自其他23个城市的20 000张图片用粗糙的语义标签标注,没有实例标签。

(10)Tsinghua-Tencent 100K数据集[131]。Tsinghua-Tencent 100K是由Zhu等从中国5个城市的腾讯街景全景图中创建的一个大型交通标志基准。该数据集由100 000幅分辨率为2 048像素×2 048像素的图像组成,涵盖了不同光线和天气状况。在该数据集中,包含3万个交通标志实例,45个类别,其中每个交通标志都带有一个类别标签、边界框以及像素蒙版。此外,Tsinghua-Tencent 100K这一基准使用与MS COCO基准相同的检测指标进行性能评估。

(11)WIDER FACE数据集[15]。WIDER FACE是由香港中文大学发布的大型人脸数据集,包含32 203图像,393 703标注人脸,涉及问题全面,难度较大。该数据集中以60个事件类别为基础进行划分,每个事件类别中随机选择40%/10%/50%的数据分别作为训练集/验证集/测试集。WIDER FACE考虑到通用目标的检测率和人眼的辨别能力,以图像的高将人脸分成3个尺度:小(10~50像素)、中(50~300像素)、大(大于300像素)。除尺度之外,该数据集中还标注了遮挡和姿态等信息用于对事件进行描述,并将事件分为了简单、中等、困难3类。

(12)MS COCO数据集[1]。MS COCO的全称是Microsoft Common Objects in Context,起源于微软于2014年出资标注的Microsoft COCO数据集,与ImageNet竞赛一样,被视为是计算机视觉领域最受关注和最权威的比赛之一。其中包括91类目标,328 000幅图像和2 500 000个标签。该数据集通过大量使用Amazon Mechanical Turk来收集数据,以场景理解为目标,主要从复杂的日常场景中截取。图像中的目标通过精确的分割标注进行位置的标定。现在有3种标注类型:目标实例、目标上的关键点和看图说话。

(13)Caltech行人检测数据集[132]。Dollar等提出的Caltech行人检测基准提供25万帧分辨率为640像素×480像素的图像序列,这些序列主要在城市环境中拍摄。Caltech数据集中注释了350 000个边界框和2 300个独立行人,包括边界框和详细的遮挡标签之间的时间对应关系,比同年的其他任何数据集都大两个数量级。此外,该数据集包含彩色视频序列,并包含了比典型行人数据集尺度范围更大、姿态变化更多的行人,也是第一个将边界框与详细遮挡时间对应的数据集。

(14)Penn-Fudan行人检测与分割数据库[133]。Penn-Fudan Database是由Wang等提出的1个图像数据库,由用于行人检测的图像组成。该图像数据库中包含170张取自校园周围和城市街道场景的图片,其中96张来自宾夕法尼亚大学周围,74张来自复旦大学周围。这些图片中共有345个带有标记的行人,而且每张图片中至少有一个行人。在Penn-Fudan Database中,所有带标记的行人都是直立行走姿态,行人的高度范围为180~390像素。

3.2 性能评估

为了便于研究人员更好地了解小目标的发展现状,本文在几个常用的小目标数据集上对现有算法的性能进行了评估。

(1)MS COCO数据集。表3给出了较为先进的检测算法在COCO数据数据集上的检测结果。其中,AP表示平均精准率(Average precision),AP^{50}、AP^{75}分别表示IoU设为0.5、0.75时的平均精准率,AP^{S}、AP^{M}、AP^{L}分别表示小目标、中等尺寸目标、大尺寸目标的平均精准率。可以发现,大目标的检测

性能是远远高于小目标的,小目标的检测性能只有大目标的一半。在所有比较算法中,Scaled-YOLOv4[134]取了最好的检测性能,将小目标的检测性能提升到了38.1%。Scaled-YOLOv4的成功主要归功于大量先进思想的集合,包括数据增强、特征融合、上下文学习和多尺度学习等。

(2)WIDER FACE数据集。表4给出了较为先进的检测算法在WIDER FACE数据集上的检测结果。在这些比较的算法中,IENet[135]取得了最好的检测性能,在Easy、Medium和Hard测试集上的AP分别为96.1%、94.7%和89.6%。在IENet中,特征融合和上下文被得到了充分利用。SRFACE(Super resolving face)[136]通过利用超分的思想也取得了不错的检测效果,在Hard测试集上的AP能达到87.3%。

(3)TinyPerson数据集。表5给出了较为先进的检测算法在TinyPerson数据数据集上的检测结果。

表2 小目标检测数据集

Table 2 Small object detection datasets

名称	简介	年份	发表期刊	尺寸/(像素×像素)	类别	图像数量	标注数/10^3	各部分占比/%		
								训练	验证	测试
BIRDSAI[126]	行人、动物检测	2020	WACV	640×480	8		164			
TinyPerson[17]	小行人检测	2020	WACV		5	1 610	72.6	50		50
EuroCity Persons[127]	城市行人检测	2019	TPAMI			47 300	238	60	10	30
WiderPerson[128]	高密度行人检测	2019	TMM			13 382	400	60	7	33
DOTA[14]	空中图像检测	2018	CVPR	4 000×4 000		1 806	188	50	33	17
NightOwls[129]	夜间行人检测	2018	ACCV	1 024×640	4	40	279			
DeepScores[19]	乐谱音符检测	2018	ICPR	220×120	123	30 000	80 000			
Bosch Small Traffic Lights[18]	小交通灯检测	2017	ICRA	1 280×720		13 427	24	38		62
CityPersons[16]	行人检测	2017	CVPR			5 000	25	59.50	10	31.50
Tsinghua-Tencent 100K[131]	交通信号灯检测	2016	CVPR	2 000×2 000	45	100 000		66.67		33.33
WIDER FACE[15]	人脸检测	2016	CVPR			32 203	393.7	40	10	50
MS COCO[1]	复杂场景下的大型数据集	2014	ECCV		91	328 000		50	25	25
Caltech Pedestrian[132]	行人检测	2012	TPAMI	640×480			350			
Penn-Fudan Database[133]	行人检测	2007	ACCV			170	0.3			

表3 MS COCO数据集上的简要性能评估

Table 3 Performance evaluation on MS COCO dataset

方法名称	主干网络	AP	AP^{50}	AP^{75}	AP^{S}	AP^{M}	AP^{L}	年份
SSD[47]	Res101	31.2	50.4	33.3	10.2	34.5	49.8	2016
RetinaNet[122]	Res101-FPN	39.1	59.1	42.3	21.8	42.7	50.2	2017
FPN[51]	Res101-FPN	36.2	59.1	39.0	18.2	39.0	48.2	2017
Mask R-CNN[72]	Res101-FPN	38.2	60.3	41.7	20.1	41.1	50.2	2017
Deformable R-FCN[38]	Aligned-Inception-ResNet	37.5	58.0	40.8	19.4	40.1	52.5	2017
Cascade R-CNN[25]	Res101-FPN	42.8	62.1	46.3	23.7	45.5	55.2	2018
YOLOv3[137]	Darknet-53	33.0	57.9	34.4	18.3	35.4	41.9	2018
FCOS[102]	ResNeXt-101	44.7	64.1	48.4	27.6	47.5	55.6	2019
DCNv2[138]	Res101-DeformableV2	46.0	67.9	50.8	27.8	49.1	59.5	2019
TridentNet[139]	Res101	42.7	63.6	46.5	23.9	46.6	56.6	2019
Cascade+Rank-NMS[140]	Res101-FPN	43.2	61.8	47.0	24.6	46.2	55.4	2019
ATSS[104]	ResNeXt-101 + DCN	50.7	68.9	56.3	33.2	52.9	62.4	2020
TSD[141]	SENet154 + DCN	51.2	71.9	56.0	33.8	54.8	64.2	2020
Deformable DETR[142]	ResNeXt-101 + DCN	52.3	71.9	58.1	34.4	54.4	65.6	2020
HCE Cascade R-CNN[70]	Res101-FPN	46.5	65.6	50.6	27.4	49.9	59.4	2020
EfficientDet[143]	EfficientNet	55.1	**74.3**	59.9				2020
Scaled-YOLOv4[134]	CSP-P7	**55.4**	73.3	**60.7**	**38.1**	**59.5**	**67.4**	2020

表4 WiderFace数据集上的简要性能评估

Table 4 Performance evaluation on WIDER FACE dataset

方法名称	AP			年份	方法名称	AP			年份
	Easy	Medium	Hard			Easy	Medium	Hard	
Faceness-WIDER[144]	71.3	63.4	34.5	2015	Face R-FCN[149]	94.7	93.5	84.7	2017
Faster R-CNN[44]	84.0	72.4	34.7	2015	FAN[150]	95.3	94.2	88.8	2017
MSCNN[15]	69.1	64.0	42.4	2016	PSDNN[151]	60.5	60.5	39.6	2019
MTTCNN[145]	84.8	82.5	59.8	2016	FDNet[152]	95.3	94.2	88.8	2018
CMS-RCNN[146]	89.9	87.4	62.4	2017	SRFACE[136]	94.4	93.3	87.3	2018
HR[147]	92.5	91.0	80.6	2017	LSC-CNN[153]	57.3	70.1	68.9	2020
SSH[148]	93.1	92.1	84.5	2017	IENet[135]	**96.1**	**94.7**	**89.6**	2021
S3FD[108]	93.7	92.4	85.2	2017	Crowd-SDNet[154]	75.8	71.0	64.4	2021

其中，MR_{50}^{samll}表示小目标在IoU设置为0.5时的漏检率(Miss rate)，MR_{50}^{tiny}、MR_{25}^{tiny}、MR_{75}^{tiny}分别表示极小目标在IoU设置为0.5、0.25、0.75时的漏检率；AP_{50}^{samll}表示小目标在IoU设置为0.5时的平均精确率，AP_{50}^{tiny}、AP_{25}^{tiny}、AP_{75}^{tiny}分别表示极小目标在IoU设置为0.5、0.25、0.75时的平均精确率。在这些比较的算法中，FCOS[102]在MR_{50}^{tiny}上以96.28%取得了最好的检测结果。尽管如此，在表4中可以发现它在AP_{50}^{tiny}上的性能不尽人意，仅有17.90%，完全不能达到实际应用的需求。对于极小目标，RetinaNet with S-α[155]设计一种专门针对极小目标的特征融合的方法，对FPN进行了改进，在AP_{50}^{tiny}上以48.48%取得了最高的检测精度。

(4)Tsinghua-Tencent 100K数据集。表6给出了较为先进的检测算法在Tsinghua-Tencent 100K数据数据集上的检测结果。在这些比较的算法中,YOLOv3-Final[156]取得了最好的检测性能,在小目标的召回率和精确率上均取得了91%。Perceptual GAN[90]通过生成对抗网络将小目标的特征映射成与大目标等价的特征,显著提升了小目标的检测性能,取得了89%和84%的召回率和精确率。

表5 TinyPerson数据集上的简要性能评估

Table 5 Performance evaluation on TinyPerson dataset

方法名称	MR_{50}^{tiny}	MR_{50}^{small}	MR_{25}^{tiny}	MR_{75}^{tiny}	AP_{50}^{tiny}	AP_{50}^{small}	AP_{25}^{tiny}	AP_{75}^{tiny}	年份
RetinaNet-SM[15]	88.87	71.82	77.88	98.57	**48.48**	63.01	69.41	5.83	2016
RetinaNet[122]	92.66	82.84	81.95	99.13	33.53	48.26	61.51	2.28	2017
FPN[51]	87.57	72.56	76.59	98.39	47.35	**63.18**	68.43	5.83	2017
FCOS[102]	**96.28**	**84.16**	**90.34**	**99.56**	17.90	40.54	41.95	1.50	2019
Libra R-CNN[157]	89.22	74.86	82.44	98.78	44.68	62.65	64.77	6.26	2019
Grid R-CNN[158]	87.96	73.16	78.27	98.21	47.14	62.48	**68.89**	**6.38**	2019
FreeAnchor[159]	89.66	73.88	79.61	98.78	44.26	60.28	67.06	4.35	2021
RetinaNet with S-α[155]	87.73	72.82	74.85	98.57	48.34	61.73	**71.18**	5.34	2021

表6 Tsinghua-Tencent 100K数据集上的性能评估

Table 6 Performance evaluation on Tsinghua-Tencent 100K dataset

方法名称	Recall				Precision				年份
	All	Small	Medium	Large	All	Small	Medium	Large	
Fast R-CNN[24]	56	24	73	86	50	45	50	55	2015
Faster R-CNN[44]		50	84	**91**		24	66	81	2015
Zhu et al.[130]	91	87	94	88	88	82	91	91	2016
Perceptual GAN[90]		89	**96**	89		84	91	91	2017
Song et al.[160]		88	93	89		85	91	**92**	2019
YOLOv3-Final[156]	92	**91**	94	89	94	**91**	**96**	**92**	2020

4 结束语

本文对小目标检测算法进行了详尽的回顾,并对已有的算法进行了归类分析和比较。首先,本文对小目标检测定义进行了解释,并对小目标检测面临的挑战进行了分析和总结。然后,本文重点阐述了小目标检测优化思路,包括数据增强、多尺度学习、上下文学习、生成对抗学习、无锚机制以及其他优化策略等,同时对采用统一思路来提升小目标检测性能的算法进行了性能比较和分析。最后,本文全面介绍了已有的小目标检测数据集,并在这些数据集上对现有的算法进行了性能比较和分析。尽管在大数据和深度学习的驱动下,小目标检测算法得到了快速的发展。但是,小目标的检测性能仍不能满足实际应用的需求,还有很多方面值得进一步研究:

(1)特征融合方面。现有的方法通常通过融合深度神经网络中同层的多尺度特征来提升小目标的特征表达能力。尽管这种方式一定程度提升了小目标的检测性能,但是在特征融合的过程中没有考虑到语义间隔和噪声干扰的问题。因此,如何消除特征融合中的语义间隔和噪声干扰问题是未来的一个研究方向。

(2)上下文学习方面。尽管上下文在目标检测中已经得到了充分的重视,并在众多目标检测算法

中得到了充分利用。但是,场景中并不是所有上下文信息都是有价值的,无效的上下文信息将可能破坏目标区域的原始特征,如何从图像中挖掘有利于提升小目标区域特征表示的上下文信息是未来的一个研究方向。此外,现有的上下文建模方法对于不同尺度目标是同等对待,并没有针对小目标而做相应的设计。因此,如何在检测模型中利用易于检测目标来辅助小目标的检测是未来的一个重要研究方向。

(3)超分辨率重构方面。尽管已有一些方法通过生成对抗的方式来提升小目标的特征,以此获得与大目标等价的特征表示,并取得了一定的成效。但是,这一类方法研究还尚少,仍有较大的研究空间。超分辨率重构是一种最直接的、可解释的提升小目标检测性能的方法。如何将超分辨率重构中先进技术与目标检测技术深度结合是未来的一个可行研究思路。

参考文献:

[1] 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, 2014: 740-755.

[2] ZOU Z, SHI Z, GUO Y, et al.Object detection in 20 years: A survey[EB/OL].(2019-05-13)[2019-05-16].https://arxiv.org/abs/1905.05055.

[3] OKSUZ K, CAM B C, KALKAN S, et al. Imbalance problems in object detection: A review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.DOI:10.1109/TPAMI.2020.2981890.

[4] ZHAO Z Q, ZHENG P, XU S, et al.Object detection with deep learning: A review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232.

[5] AGARWAL S, TERRAIL J O D, JURIE F. Recent advances in object detection in the age of deep convolutional neural networks[EB/OL].(2018-09-10)[2019-08-20].https://arxiv.org/abs/1809.03193.

[6] CHEN G, WANG H, 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, 2020, 99: 1-18.

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

[8] LIU Y, SUN P, WERGELES N, et al.A survey and performance evaluation of deep learning methods for small object detection [J].Expert Systems with Applications, 2021, 172(4): 114602.

[9] 梁鸿,王庆玮,张千,等.小目标检测技术研究综述[J].计算机工程与应用,2021,57(1): 17-28.
LIANG Hong, WANG Qingwei, ZHANG Qian, et al. Small object detection technology: A review[J]. Computer Engineering and Applications, 2021, 57(1): 17-28.

[10] 刘颖,刘红燕,范九伦,等.基于深度学习的小目标检测研究与应用综述[J].电子学报,2019,48(3): 590-601.
LIU Ying, LIU Hongyan, FAN Jiulun, et al. A survey of research and application of small object detection based on deep learning [J]. Acta Electronica Sinica, 2019, 48(3): 590-601.

[11] CHEN C, LIU M Y, TUZEL O, et al. R-CNN for small object detection[C]//Proceeding of Asian Conference on Computer Vision. Cham: Springer, 2016: 214-230.

[12] TORRALBA A, FERGUS R, FREEMAN W T. 80 million tiny images: A large data set for nonparametric object and scene recognition[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11): 1958-1970.

[13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL].(2014-09-04) [2015-04-10]. https://arxiv.org/abs/1409.1556.

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

[15] YANG S, LUO P, LOY C C, et al. Wider face: A face detection benchmark[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 5525-5533.

[16] ZHANG S, BENENSON R, SCHIELE B. Citypersons: A diverse dataset for pedestrian detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 3213-3221.

[17] YU X, GONG Y, JIANG N, et al. Scale match for tiny person detection[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Los Alamitos: IEEE, 2020: 1257-1265.

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

[19] LUKAS T, ELEZI I, SCHMIDHUBER J, et al. Deepscores-a dataset for segmentation, detection and classification of tiny objects[C]//Proceedings of 2018 24th International Conference on Pattern Recognition (ICPR). New York: IEEE, 2018: 3704-3709.

[20] YAEGER L, LYON R, WEBB B.Effective training of a neural network character classifier for word recognition[J].Advances in Neural Information Processing Systems, 1996, 9: 807-816.

[21] SIMARD P Y, STEINKRAUS D, PLATT J C. Best practices for convolutional neural networks applied to visual document analysis[C]//Proceedings of ICDAR. [S.l.]: IEEE, 2003, 3(2003).

[22] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.

[23] WAN L, ZEILER M, ZHANG S, et al. Regularization of neural networks using dropconnect[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2013: 1058-1066.

[24] GIRSHICK R. Fast R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2015: 1440-1448.

[25] CAI Z, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 6154-6162.

[26] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 779-788.

[27] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 7263-7271.

[28] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[EB/OL]. (2017-08-15) [2017-11-29].https://arxiv.org/abs/1708.04552.

[29] ZHANG H, CISSE M, DAUPHIN Y N, et al.Mixup: Beyond empirical risk minimization[EB/OL].(2017-10-25)[2018-04-27]. https://arxiv.org/abs/1710.09412.

[30] YUN S, HAN D, OH S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 6023-6032.

[31] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2020-04-23].https://arxiv.org/abs/2004.10934.

[32] GONG C, WANG D, LI M, et al. KeepAugment: A simple information-preserving data augmentation approach[EB/OL]. (2020-11-23)[2020-11-23].https://arxiv.org/abs/2011.11778.

[33] KISANTAL M, WOJNA Z, MURAWSKI J, et al. Augmentation for small object detection[EB/OL]. (2019-02-19) [2019-02-19]. https://arxiv.org/abs/1902.07296.

[34] CHEN C, ZHANG Y, LV Q, et al. RRNet: A hybrid detector for object detection in drone-captured images[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Los Alamitos: IEEE, 2019: 100-108.

[35] CHEN Y, ZHANG P, LI Z, et al. Stitcher: Feedback-driven data provider for object detection[EB/OL]. (2020-04-26) [2021-03-14]. https://arxiv.org/abs/2004.12432.

[36] ZOPH B, CUBUK E D, GHIASI G, et al. Learning data augmentation strategies for object detection[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2020: 566-583.

[37] YU F, KOLTUN V.Multi-scale context aggregation by dilated convolutions[EB/OL].(2015-11-23)[2016-04-30].https://arxiv.org/abs/1511.07122.

[38] DAI J, QI H, XIONG Y, et al.Deformable convolutional networks[C]// Proceedings of the IEEE International Conference on

Computer Vision. New York: IEEE, 2017: 764-773.

[39] ADELSON E H, ANDERSON C H, BERGEN J R, et al.Pyramid methods in image processing[J].RCA Engineer, 1984, 29(6): 33-41.

[40] LOWE D G.Distinctive image features from scale-invariant keypoints[J].International Journal of Computer Vision, 2004, 60(2): 91-110.

[41] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of IEEE Computer Society Conference on Computer Vision & Pattern Recognition. [S.l.]: IEEE, 2005.

[42] SINGH B, DAVIS L S. An analysis of scale invariance in object detection snip[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 3578-3587.

[43] SINGH B, NAJIBI M, DAVIS L S.Sniper: Efficient multi-scale training[EB/OL].(2018-05-23)[2018-12-13].https://arxiv.org/abs/1805.09300.

[44] REN S, HE K, GIRSHICK R, et al.Faster R-CNN: Towards real-time object detection with region proposal networks[EB/OL].(2015-06-04)[2016-01-06].https://arxiv.org/abs/1506.01497.

[45] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[46] DAI J, LI Y, HE K, et al.R-FCN: Object detection via region-based fully convolutional networks[EB/OL].(2016-05-20)[2016-06-21].https://arxiv.org/abs/1605.06409.

[47] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[48] CAI Z, FAN Q, FERIS R S, et al. A unified multi-scale deep convolutional neural network for fast object detection[C]// Proceedings of European Conference on Computer Vision. Cham: Springer, 2016: 354-370.

[49] BELL S, ZITNICK C L, BALA K, et al. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 2874-2883.

[50] KONG T, YAO A, CHEN Y, et al. Hypernet: Towards accurate region proposal generation and joint object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 845-853.

[51] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 2117-2125.

[52] LIANG Z, SHAO J, ZHANG D, et al. Small object detection using deep feature pyramid networks[C]//Proceedings of Pacific Rim Conference on Multimedia. Cham: Springer, 2018: 554-564.

[53] CAO G, XIE X, YANG W, et al. Feature-fused SSD: Fast detection for small objects[C]//Proceedings of Ninth International Conference on Graphic and Image Processing (ICGIP 2017). Bellingham: SPIE-int SOC Optical Engineering, 2018: 106151E.

[54] LI Z, ZHOU F.FSSD: Feature fusion single shot multibox detector[EB/OL].(2017-12-04)[2018-05-17].https://arxiv.org/abs/1712.00960.

[55] 韩松臣, 张比浩, 李炜, 等. 基于改进Faster-RCNN的机场场面小目标物体检测算法[J]. 南京航空航天大学学报, 2019, 51(6): 735-741.
HAN Songchen, ZHANG Bihao, LI Wei, et al.Small target detection in airport scene via modified faster-RCNN[J].Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(6): 735-741.

[56] NAYAN A A, SAHA J, MOZUMDER A N, et al.Real time detection of small objects[EB/OL].(2020-03-17)[2020-04-14].https://arxiv.org/abs/2003.07442.

[57] LIU Z, GAO G, SUN L, et al.HRDNet: High-resolution detection network for small objects[EB/OL].(2020-06-13)[2020-06-13].https://arxiv.org/abs/2006.07607.

[58] DENG C, WANG M, LIU L, et al.Extended feature pyramid network for small object detection[EB/OL].(2020-05-16)[2020-04-09].https://arxiv.org/abs/2003.07021.

[59] OLIVA A, TORRALBA A.The role of context in object recognition[J].Trends in Cognitive Sciences, 2007, 11(12): 520-527.

[60] LI J, WEI Y, LIANG X, et al. Attentive contexts for object detection[J]. IEEE Transactions on Multimedia, 2016, 19(5):

944-954.

[61] ZENG X, OUYANG W, YAN J, et al. Crafting gbd-net for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(9): 2109-2123.

[62] TANG X, DU D K, HE Z, et al. Pyramidbox: A context-assisted single shot face detector[C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 797-813.

[63] 郑晨斌,张勇,胡杭,等.目标检测强化上下文模型[J].浙江大学学报(工学版),2020,54(3):529-539.
ZHENG Chenbin, ZHANG Yong, HU Hang, et al. Object detection enhanced context model[J]. Journal of Zhejiang University (Engineering Science), 2020, 54(3): 529-539.

[64] DIVVALA S K, HOIEM D, HAYS J H, et al. An empirical study of context in object detection[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2009: 1271-1278.

[65] TORRALBA A, SINHA P. Statistical context priming for object detection[C]// Proceedings of the Eighth IEEE International Conference on Computer Vision. New York: IEEE, 2001: 763-770.

[66] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32(9): 1627-1645.

[67] OUYANG W, WANG X, ZENG X, et al. Deepid-net: Deformable deep convolutional neural networks for object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2015: 2403-2412.

[68] CHEN Z, HUANG S, TAO D. Context refinement for object detection[C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 71-86.

[69] BARNEA E, BEN-SHAHAR O. Exploring the bounds of the utility of context for object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 7412-7420.

[70] CHEN Z M, JIN X, ZHAO B, et al. Hierarchical context embedding for region-based object detection[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2020: 633-648.

[71] 张瑞琰,姜秀杰,安军社,等.面向光学遥感目标的全局上下文检测模型设计[J].中国光学,2020,13(73): 138-149.
ZHANG Ruiyan, JIANG Xiujie, AN Junshe, et al. Design of global-contextual detection model for optical remote sensing targets [J]. Chinese Optics, 2020, 13(73): 138-149.

[72] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 2961-2969.

[73] ZHAO X, LIANG S, WEI Y. Pseudo mask augmented object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 4061-4070.

[74] ZHANG Z, QIAO S, XIE C, et al. Single-shot object detection with enriched semantics[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 5813-5821.

[75] CHEN Q, SONG Z, DONG J, et al. Contextualizing object detection and classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(1): 13-27.

[76] GUPTA S, HARIHARAN B, MALIK J. Exploring person context and local scene context for object detection[EB/OL].(2015-11-25)[2015-11-25].https://arxiv.org/abs/1511.08177.

[77] LIU Y, WANG R, SHAN S, et al. Structure inference net: Object detection using scene-level context and instance-level relationships[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 6985-6994.

[78] XU H, JIANG C H, LIANG X, et al. Reasoning-RCNN: Unifying adaptive global reasoning into large-scale object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 6419-6428.

[79] CHEN X, GUPTA A. Spatial memory for context reasoning in object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 4086-4096.

[80] HU H, GU J, ZHANG Z, et al. Relation networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 3588-3597.

[81] LIM J S, ASTRID M, YOON H J, et al. Small object detection using context and attention[EB/OL].(2019-12-13)[2019-12-16].

https://arxiv.org/abs/1912.06319.

[82] SHEN W, QIN P, ZENG J. An indoor crowd detection network framework based on feature aggregation module and hybrid attention selection module[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Los Alamitos:IEEE, 2019: 82-90.

[83] FU K, LI J, MA L, et al.Intrinsic relationship reasoning for small object detection[EB/OL].(2020-09-02)[2020-09-02].https://arxiv.org/abs/2009.00833.

[84] PATO L V, NEGRINHO R, AGUIAR P M Q. Seeing without looking: Contextual rescoring of object detections for ap maximization[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 14610-14618.

[85] HARIS M, SHAKHNAROVICH G, UKITA N. Task-driven super resolution: Object detection in low-resolution images[EB/OL].(2018-03-30)[2018-03-30].https://arxiv.org/abs/1803.11316.

[86] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al.Generative adversarial networks[EB/OL].(2014-06-10)[2014-06-10].https://arxiv.org/abs/1406.2661.

[87] RADFORD A, METZ L, CHINTALA S.Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL].(2015-11-19)[2016-01-07].https://arxiv.org/abs/1511.06434.

[88] SIXT L, WILD B, LANDGRAF T.Rendergan: Generating realistic labeled data[J].Frontiers in Robotics and AI, 2018, 5: 66.

[89] WANG X, SHRIVASTAVA A, GUPTA A. A-fast-RCNN: Hard positive generation via adversary for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 2606-2615.

[90] LI J, LIANG X, WEI Y, et al. Perceptual generative adversarial networks for small object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 1222-1230.

[91] BAI Y, ZHANG Y, DING M, 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, 2018: 206-221.

[92] 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]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 9725-9734.

[93] TYCHSEN-SMITH L, PETERSSON L. Denet: Scalable real-time object detection with directed sparse sampling[C]// Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 428-436.

[94] WANG X, CHEN K, HUANG Z, et al.Point linking network for object detection[EB/OL].(2017-06-12)[2017-06-13].https://arxiv.org/abs/1706.03646.

[95] LAW H, DENG J. Cornernet: Detecting objects as paired keypoints[C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 734-750.

[96] LAW H, TENG Y, RUSSAKOVSKY O, et al.Cornernet-lite: Efficient keypoint based object detection[EB/OL].(2017-06-12)[2017-06-13].https://arxiv.org/abs/1706.03646.

[97] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 6569-6578.

[98] ZHOU X, ZHUO J, KRAHENBUHL P. Bottom-up object detection by grouping extreme and center points[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 850-859.

[99] ZHOU X, WANG D, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-16) [2019-04-25]. https://arxiv. org/abs/1904.07850.

[100] YANG Z, LIU S, HU H, et al. Reppoints: Point set representation for object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 9657-9666.

[101] KONG T, SUN F, LIU H, et al.Foveabox: Beyound anchor-based object detection[J].IEEE Transactions on Image Processing, 2020, 29: 7389-7398.

[102] TIAN Z, SHEN C, CHEN H, et al. Fcos: Fully convolutional one-stage object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 9627-9636.

[103] ZHU C, HE Y, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]// Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 840-849.

[104] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 9759-9768.

[105] FU J, SUN X, WANG Z, et al. An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images[J].IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(2): 1331-1344.

[106] YAN J, ZHAO L, DIAO W, et al. AF-EMS detector: Improve the multi-scale detection performance of the anchor-free detector[J].Remote Sensing, 2021, 13(2): 160.

[107] ZHANG S, ZHU X, LEI Z, et al. Faceboxes: A CPU real-time face detector with high accuracy[C]//Proceedings of 2017 IEEE International Joint Conference on Biometrics (IJCB). New York: IEEE, 2017: 1-9.

[108] ZHANG S, ZHU X, LEI Z, et al. S3FD: Single shot scale-invariant face detector[C]// Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 192-201.

[109] EGGERT C, ZECHA D, BREHM S, et al. Improving small object proposals for company logo detection[C]// Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval. New York: Assoc Computing Machinery, 2017: 167-174.

[110] WANG J, CHEN K, YANG S, et al. Region proposal by guided anchoring[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 2965-2974.

[111] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]// Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. New York: IEEE, 2001: 1-9.

[112] LI A, YANG X, ZHANG C. Rethinking classification and localization for cascade R-CNN[EB/OL].(2019-07-27)[2019-07-27]. https://arxiv.org/abs/1907.11914.

[113] LIU W, LIAO S, HU W, et al. Learning efficient single-stage pedestrian detectors by asymptotic localization fitting[C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 618-634.

[114] YANG B, YAN J, LEI Z, et al. Craft objects from images[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 6043-6051.

[115] YANG F, CHOI W, LIN Y. Exploit all the layers: Fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. New York: IEEE, 2016: 2129-2137.

[116] GAO M, YU R, LI A, et al. Dynamic zoom-in network for fast object detection in large images[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 6926-6935.

[117] CHEN S, LI J, YAO C, et al. DuBox: No-prior box objection detection via residual dual scale detectors[EB/OL].(2019-04-15)[2019-04-16].https://arxiv.org/abs/1904.06883.

[118] DRENKOW N, BURLINA P, FENDLEY N, et al. Objectness-guided open set visual search and closed set detection[EB/OL].(2020-12-11)[2021-04-14].https://arxiv.org/abs/2012.06509.

[119] YANG F, FAN H, CHU P, et al. Clustered object detection in aerial images[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 8311-8320.

[120] ZHANG J, HUANG J, CHEN X, et al. How to fully exploit the abilities of aerial image detectors[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Los Alamitos:IEEE, 2019: 1-8.

[121] LI C, YANG T, ZHU S, et al. Density map guided object detection in aerial images[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Alamitos:IEEE, 2020: 190-191.

[122] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 2980-2988.

[123] ZHANG H, CHANG H, MA B, et al. Cascade retinanet: Maintaining consistency for single-stage object detection[EB/OL].(2019-07-16)[2019-07-16].https://arxiv.org/abs/1907.06881.

[124] SUN S, YIN Y, WANG X, et al. Multiple receptive fields and small-object-focusing weakly-supervised segmentation network for fast object detection[EB/OL].(2019-04-19)[2019-05-22].https://arxiv.org/abs/1904.12619.

[125] YOO J, LEE H, CHUNG I, et al. Density-based object detection: Learning bounding boxes without ground truth assignment

[EB/OL].(2019-11-28)[2020-10-04].https://arxiv.org/abs/1911.12721.

[126] BONDI E, JAIN R, AGGRAWAL P, et al. Birdsai: A dataset for detection and tracking in aerial thermal infrared videos[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Los Alamitos:IEEE, 2020: 1747-1756.

[127] BRAUN M, KREBS S, FLOHR F, et al. The eurocity persons dataset: A novel benchmark for object detection[EB/OL].(2018-05-18)[2018-06-05].https://arxiv.org/abs/1805.07193.

[128] ZHANG S, XIE Y, WAN J, et al. Widerperson: A diverse dataset for dense pedestrian detection in the wild[J]. IEEE Transactions on Multimedia, 2019, 22(2): 380-393.

[129] NEUMANN L, KARG M, ZHANG S, et al. Nightowls: A pedestrians at night dataset[C]//Proceedings of Asian Conference on Computer Vision. Cham: Springer, 2018: 691-705.

[130] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 3213-3223.

[131] ZHU Z, LIANG D, ZHANG S, et al. Traffic-sign detection and classification in the wild[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 2110-2118.

[132] DOLLÁR P, WOJEK C, SCHIELE B, et al. Pedestrian detection: A benchmark[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2009: 304-311.

[133] WANG L, SHI J, SONG G, et al. Object detection combining recognition and segmentation[C]//Proceedings of Asian Conference on Computer Vision. Berlin, Heidelberg: Springer, 2007: 189-199.

[134] WANG C Y, BOCHKOVSKIY A, LIAO H Y M.Scaled-YOLOv4: Scaling cross stage partial network[EB/OL].(2020-11-16)[2021-02-22].https://arxiv.org/abs/2011.08036.

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

[136] BAI Y, ZHANG Y, DING M, et al. Finding tiny faces in the wild with generative adversarial network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 21-30.

[137] REDMON J, FARHADI A. Yolov3: An incremental improvement[EB/OL].(2018-04-08)[2018-04-08].https://arxiv.org/abs/1804.02767.

[138] ZHU X, HU H, LIN S, et al. Deformable convnets v2: More deformable, better results[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 9308-9316.

[139] LI Y, CHEN Y, WANG N, et al. Scale-aware trident networks for object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 6054-6063.

[140] TAN Z, NIE X, QIAN Q, et al. Learning to rank proposals for object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 8273-8281.

[141] SONG G, LIU Y, WANG X. Revisiting the sibling head in object detector[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. new york: IEEE, 2020: 11563-11572.

[142] ZHU X, SU W, LU L, et al.Deformable DETR: Deformable transformers for end-to-end object detection[EB/OL].(2020-10-08)[2021-03-18].https://arxiv.org/abs/2010.04159.

[143] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 10781-10790.

[144] YANG S, LUO P, LOY C C, et al. From facial parts responses to face detection: A deep learning approach[C]// Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2015: 3676-3684.

[145] ZHANG K, ZHANG Z, LI Z, et al.Joint face detection and alignment using multitask cascaded convolutional networks[J].IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

[146] ZHU C, ZHENG Y, LUU K, et al. CMS-RCNN: Contextual multi-scale region-based cnn for unconstrained face detection [C]//Deep learning for biometrics. Cham: Springer, 2017: 57-79.

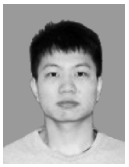
[147] HU P, RAMANAN D. Finding tiny faces[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 951-959.

[148] NAJIBI M, SAMANGOUEI P, CHELLAPPA R, et al. SSH: Single stage headless face detector[C]// Proceedings of the

IEEE International Conference on Computer Vision. New York: IEEE, 2017: 4875-4884.
[149] WANG Y, JI X, ZHOU Z, et al. Detecting faces using region-based fully convolutional networks[EB/OL].(2017-09-14)[2017-09-18].https://arxiv.org/abs/1709.05256.
[150] WANG J, YUAN Y, YU G. Face attention network: An effective face detector for the occluded faces[EB/OL].(2017-11-20)[2017-11-22].https://arxiv.org/abs/1711.07246.
[151] LIU Y, SHI M, ZHAO Q, et al. Point in, box out: Beyond counting persons in crowds[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 6469-6478.
[152] ZHANG C, XU X, TU D. Face detection using improved faster RCNN[EB/OL].(2018-02-06)[2018-02-06].https://arxiv.org/abs/1802.02142.
[153] SAM D B, PERI S V, SUNDARARAMAN M N, et al. Locate, size and count: Accurately resolving people in dense crowds via detection[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020. DOI: 10.1109/TPAMI.2020.2974830.
[154] WANG Y, HOU J, HOU X, et al. A Self-training approach for point-supervised object detection and counting in crowds[J].IEEE Transactions on Image Processing, 2021, 30: 2876-2887.
[155] GONG Y, YU X, DING Y, et al. Effective fusion factor in FPN for tiny object detection[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. New York: IEEE, 2021: 1160-1168.
[156] WAN J, DING W, ZHU H, et al. An efficient small traffic sign detection method based on YOLOv3[J]. Journal of Signal Processing Systems, 2020: DOI: 10.1007/S11265-020-01614-2.
[157] PANG J, CHEN K, SHI J, et al. Libra R-CNN: Towards balanced learning for object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 821-830.
[158] LU X, LI B, YUE Y, et al. Grid R-CNN[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2019: 7363-7372.
[159] ZHANG X, WAN F, LIU C, et al. Learning to match anchors for visual object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021. DOI: 10.1109/TPAMI.2021.3050494.
[160] SONG S, QUE Z, HOU J, et al. An efficient convolutional neural network for small traffic sign detection[J].Journal of Systems Architecture, 2019, 97: 269-277.

作者简介:

高新波(1972-),男,教授,博士生导师,研究方向:人工智能、机器学习、计算机视觉和模式识别等,E-mail: gaoxb@cqupt.edu.cn。



莫梦竟成(1997-),男,硕士研究生,研究方向:航拍图像目标检测。

汪海涛(1997-),男,硕士研究生,研究方向:行人重识别。

冷佳旭(1989-),**通信作者**,男,博士,研究方向:目标检测、人脸超分、行人重识别和视频异常检测等,E-mail: lengjx@cqupt.edu.cn。

(编辑:刘彦东)