计算机科学
COMPUTER SCIENCE

面向深度卷积神经网络的小目标检测算法综述

杜紫薇, 周恒, 李承阳, 李忠博, 谢永强, 董昱辰, 齐锦

引用本文

杜紫薇, 周恒, 李承阳, 李忠博, 谢永强, 董昱辰, 齐锦面向深度卷积神经网络的小目标检测算法综述[J]. 计算机科学, 2022, 49(12): 205-218.

DU Zi-wei, ZHOU Heng, LI Cheng-yang, LI Zhong-bo, XIE Yong-qiang, DONG Yu-chen, QI Jin. Small Object Detection Based on Deep Convolutional Neural Networks:A Review [J]. Computer Science, 2022, 49(12): 205-218.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220500260

面向深度卷积神经网络的小目标检测算法综述

杜紫薇[1] 周 恒[1,2] 李承阳[1,3] 李忠博[1] 谢永强[1] 董昱辰[1] 齐 锦[1]
1 军事科学院系统工程研究院 北京 100141
2 西安电子科技大学电子工程学院 西安 710071
3 北京大学信息科学与技术学院 北京 100871
(15271035@bjtu.edu.cn)

摘 要 小目标检测是计算机视觉领域最具挑战性的问题之一。相比大目标,小目标覆盖面积小,空间分辨率低,可用特征少,检测效果通常不理想。近年来,基于深度卷积神经网络的小目标检测算法蓬勃发展,并在卫星遥感、无人驾驶等领域取得了重大成就。文中对国内外现有的小目标检测算法进行了归类、分析和比较。首先介绍小目标检测的难点和常用的数据集;接着分别从骨干网络、金字塔结构、锚框设计、优化目标、增益组件5个方面系统地梳理了已有检测算法,为进一步改进小目标检测算法的性能提供了思路;然后对现有小目标检测算法进行全面总结,并比较分析了列举算法在常用数据集上的性能;最后介绍了小目标检测的应用前景,并对该领域未来的研究方向做出了展望。

关键词:深度学习;小目标检测;多尺度特征融合;无锚机制;注意力机制
中图法分类号 TP391

Small Object Detection Based on Deep Convolutional Neural Networks:A Review

DU Zi-wei[1],ZHOU Heng[1,2],LI Cheng-yang[1,3],LI Zhong-bo[1],XIE Yong-qiang[1],DONG Yu-chen[1] and QI Jin[1]
1 Institute of Systems Engineering,Academy of Military Sciences,Beijing 100141,China
2 School of Electronic Engineering,Xidian University,Xi'an 710071,China
3 School of Electronics Engineering and Computer Science,Peking University,Beijing 100871,China

Abstract Small object detection has long been one of the most challenging problems in computer vision. Since small objects have the characteristics of small coverage area,low resolution,and lack of feature information,their detection effect is not ideal compared to large-sized objects. In recent years,the small object detection algorithm based on deep convolutional neural networks has developed vigorously,and been successfully used in fields such as satellite remote sensing and driverless vehicles. This survey makes a taxonomy,analysis and comparison of existing algorithms. First,the difficulties of small object detection and common detection datasets are introduced. Second,the existing detection algorithms are systematically described from five aspects:backbone network,pyramid structure,anchor design,optimization of object and a bag of species,to provide ideas for further improving the performance of small object detection algorithms. Then,we briefly summarize the existing small object detection algorithms and analyze their performance of the listed algorithm on common dataset. Finally,the application and the future research direction in the field of small object detection has been prospected.

Keywords Deep learning,Small object detection,Multi-scale feature fusion,Anchor-free,Attention mechanism

1 引言

目标检测旨在从复杂的背景中精准定位每个目标的位置并判断目标的所属类别。近年来,深度卷积神经网络(Convolutional Neural Networks,CNN)推动了目标检测的发展并取得了长足进步[1-2],但对于小目标的检测一直是计算机视觉领域的难点之一。从表1可以看出,小目标由于占比面积小、分辨率低、可用特征少,相比大目标,其检测精度平均下降了30%左右。造成小目标检测性能下降的具体原因包括以下几点:

(1)特征信息少。相比大目标,小目标在图像中占有的像素较少,其表观特征不足,检测器很难对其进行有效识别。并且基于CNN的常用目标检测网络通过不断下采样来提取深度特征,在此过程中小目标的特征信息丢失严重。例如,在SSD网络中,8×8像素的区域在经过步长为8的卷积后被压缩为一个像素点[3]。

(2)正负样本不平衡。正负样本不平衡分为目标-背景不平衡和目标-目标不平衡。在目标检测中,与真实框重叠程度高于阈值的边界框被标记为正样本,低于阈值的边界框被标记为负样本。目标-背景不平衡表现为在设置的交并比(In-

到稿日期:2022-05-27 返修日期:2022-10-28
通信作者:李忠博(lzb05296@163.com)

tersection over Union, IoU)阈值下，背景实例的数目远大于小目标正样本的数目。目标-目标不平衡表现为在大多数检测数据集上进行训练时，大目标样本数量远多于小目标样本数量，模型更偏向于学习大目标特征，而忽略小目标。

(3)数据集不完备。当前目标检测领域的通用数据集PASCAL VOC[4]和MS COCO[5]拥有丰富的目标类别，但小目标的标记数量占比低。常用的小目标检测数据集Wider Face[6]和TinyPerson[7]只包含人脸和行人数据，虽然小目标数量占比高，但只对单个类别进行训练，模型泛化能力差。

(4)锚框设计难。常用的目标检测方法需要将预先设置的锚框(Anchor)密集平铺在特征层上。对于小目标而言，设计尺寸合适的锚框存在两个难点：1)数据集不同，锚框的各项超参数也不同；2)模型的检测效果对锚框的尺寸、长宽比、数目比较敏感，超参数的微小变化会影响模型的检测效果，模型鲁棒性低。

针对以上问题，本文系统地梳理了面向卷积神经网络的小目标检测算法，从骨干网络、金字塔结构、锚框设计、优化目标、增益组件5个方面详细阐述了各类方法的思想与特点。此外，本文在常用数据集上对现有的代表性算法进行了总结分析与性能评估，并对小目标检测的应用场景和未来研究方向进行了展望。

表1列出了在MS COCO数据集上性能排名前八的模型结果[8]。其中，AP^L和AP^S分别指大、小目标的检测精度。

表1 在MS COCO数据集上性能排名前八的模型结果

Table 1 Results of top-8 models on MS COCO

Model	AP^L	AP^S	性能下降/%
Noah CV Lab	0.720	0.407	31.30
mmdet	0.706	0.399	30.70
DeepAR	0.668	0.378	29.00
DetectoRS	0.669	0.377	29.20
KiwiDet2	0.685	0.362	32.30
360 AI	0.674	0.369	30.50
ZFTurbo	0.671	0.367	30.40
AD Lab	0.675	0.343	33.20

2 小目标检测数据集

数据集是目标检测领域用来衡量检测效果的基准，在大规模数据集下进行检测能更好地评估算法的优劣。除了通用数据集PASCAL VOC和MS COCO，本文还从目标的类型对小目标检测常用数据集进行划分，包括航空遥感图像、无人机图像、行人图像、人脸图像和交通标志图像等。

表2列出了常用的小目标检测数据集。航空遥感图像检测数据集包括DOTA[9]，RSOD[10]，AI-TOD[11]；无人机图像数据集包括VisDrone[12]；行人检测数据集包括CityPersons[13]和Tiny-Person；在人脸检测领域的数据集包括Wider Face；交通标志图像数据集包括T-T 100K[14]和Behrendt[15]。

表2 常用的小目标检测数据集

Table 2 Common small object detection datasets

目标类型	数据集	年份	类别数	图像数	实例数	尺寸	数据集划分/%		
							训练	验证	测试
通用图像	PASCAL VOC	2012	20	11530	27450	—	—	—	—
	MS COCO	2014	80	32800	328000	—	50	25	25
遥感图像	DOTA	2018	15	2806	188282	4000×4000	50	33	17
	RSOD	2014	4	976	5529	—	—	—	—
	AI-TOD	2021	8	28036	700621	800×800	40	10	50
无人机图像	VisDrone	2019	10	10209	2600000	—	75	18.5	6.5
行人图像	CityPersons	2017	—	5050	—	—	59.50	10	31.50
	TinyPerson	2020	—	1610	72651	—	50	—	50
人脸图像	Wider Face	2016	—	32203	393703	—	40	10	50
交通标志图像	Behrendt	2017	4	8334	13493	—	—	—	—
	T-T 100K	2016	—	100000	30000	2048×2048	36	18	46

当前研究将小目标的定义分为绝对尺寸和相对尺寸两种方式。1)基于绝对尺寸的定义：MS COCO数据集将尺寸小于32×32像素的目标定义为小目标。DOTA数据集和Wider Face数据集将小目标的像素限定在[10,50]之间。TinyPerson数据集对小目标进行进一步细分，将像素在[20,32]之间的目标定义为小目标，像素在[2,20]之间的目标定义为微小目标。2)基于相对尺寸的定义：按照小目标像素占图像的比值进行划分。SOD[16]数据集中，小目标相对面积的中位数分布在0.08%～0.58%之间。Wang等[17]将数据集中尺寸小于原图尺寸10%的目标设为小目标。

3 小目标检测算法

针对小目标检测面临的挑战，本节从检测模型的主要结构，即骨干网络、金字塔结构、锚框设计、优化目标和增益组件这5个方面对现有的小目标检测算法进行全面的梳理和总结。如图1所示，骨干网络是特征提取的初级阶段。金字塔结构主要用于多尺度特征融合，通过加深特征提取并融合不同尺度的特征，使得特征具有高语义信息。锚框设计作为网络的输出阶段，利用金字塔结构特征层融合的信息来获得预测结果。优化目标围绕分类和定位两大任务展开，通过筛选合适的正负样本并且生成损失函数(Loss Function)来监督模型的训练。增益组件是模型中用较小的代价来换取较大增益的部件，在处理特殊目标检测问题时，增益组件起着非常重要的作用。

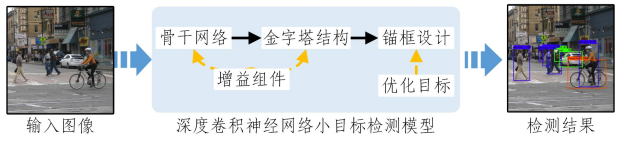


图1 小目标检测架构

Fig. 1 Architecture of small object detection

3.1 骨干网络

骨干网络对目标的特征提取起到了重要作用。骨干网络的选择是精度与速度的权衡。根据不同检测任务的需求，骨干网络分为以下两种：精度优先骨干网络和速度优先骨干网络。

图 2 给出了骨干网络的发展脉络。

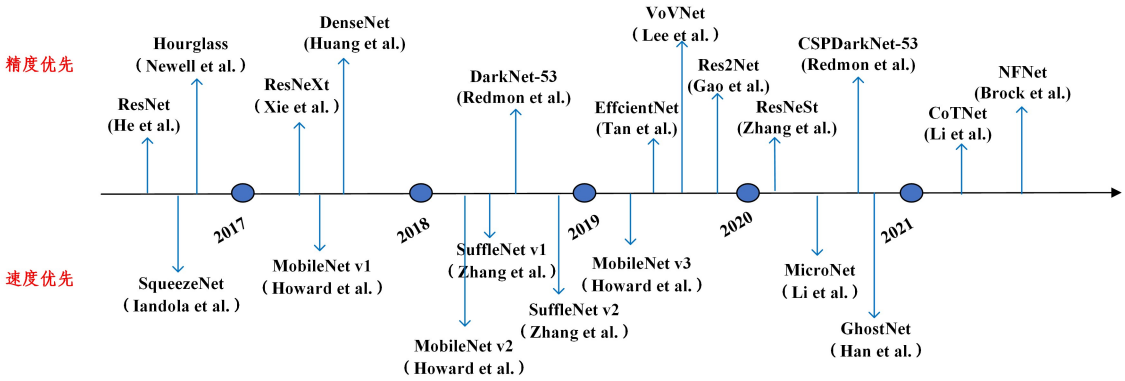


图 2 骨干网络发展脉络

Fig. 2 Timeline of backbone network

3.1.1 精度优先骨干网络

在深度网络中，随着网络层数的加深，各个特征通过非线性变换生成具有强大表征能力的深度特征。其中残差网络[18]（Residual Networks，ResNet）的提出有效突破了神经网络层数的限制，推动了更深层模型的发展。当前存在诸多 ResNet 的改进模型，如 ResNeXt[19]、ResNeSt[20]、DarkNet[21-23]系列、Res2Net[24]和 NFNet[25]等。Hourglass[26]（沙漏网络）使用残差模块作为基础单元。DenseNet[27]采用密集连接的方式，将每一层与其他层密集连接。VoVNet[28]对前者做出了改进，将每一层直接与最后一层拼接。EfficientNet[29]利用神经网络搜索（Neural Architecture Search，NAS）找到最优基准模型，并对模型的宽度、深度和图片分辨率进行复合缩放。CoTNet[30]将 Transformer[31]技术带入骨干网络的设计中，并利用输入的上下文信息指导注意力矩阵的生成。下文将对小目标检测中常用的 ResNeXt，Hourglass，DenseNet 和 DarkNet 模型进行详细介绍。

（1）ResNeXt。ResNeXt 在 ResNet 的基础上，学习了 Inception[32]中分组卷积的思想，通过堆叠多个相同的拓扑结构来加宽网络。例如，DetectoRS[33]将骨干网络 ResNet-50 替换为 ResNeXt-101 后，在 MSCOCO 数据集上小目标检测精度提升了 1.5%。

（2）Hourglass。沙漏网络由沙漏模块组成，单个沙漏模块通过对称的拓扑结构获取不同尺度下的特征信息，并利用跳层连接将丢失的细节信息带回模型。针对小目标漏检误检问题，Zhao 等[34]将 Hourglass 中由 2 个 3×3 卷积组成的基本残差块替换成 2 个 1×1 和 1 个 3×3 卷积模块，在不降低模型性能的情况下，计算参数量减少为了原来的 1/5。

（3）DenseNet。DenseNet 通过特征重用和旁路设置将若干由卷积层组成的密集连接块进行拼接。DenseNet 缓解了梯度消失问题，并且有效提升了特征信息的利用率。VoVNet 提出了 OSA（One-Shot Aggregation）结构，通过产生更大的感受野，捕获了多种特征并且解决了密集连接效率低的问题。针对小目标特征不明显的问题，Yuan 等[35]将 VoVNet 密集连接的思想带入了 YOLOv3 中，增强了网络对特征的重复利用，从而提升了模型的检测效果。

（4）DarkNet。DarkNet-53[22]共有 53 个卷积层，它借鉴了 ResNet 中快捷路径[36]（Shortcut）的思想，通过堆叠大量的残差块来增加网络深度。针对小目标检测精度低的问题，Wang 等[17]提出了 YOLOv3-CS，通过在遥感数据集上进行稀疏训练，统计骨干网络中不同尺度特征的重要性信息，来对相应卷积层的深度进行调整。在 RSOD 数据集上骨干网络重构后的模型的小目标检测精度提升了 4.48%。YOLOv4[23]设计了 CSP DarkNet-53 为骨干网络，引入 CSPNet[37]（Cross Stage Partial Networks）中梯度流截断的设计思想，避免了梯度信息重复学习的问题，并且减少了骨干网络中 20%的参数量。针对小目标特征信息损失的问题，Dou 等[38]提出了 YOLOv4-RF 算法。该算法用空洞卷积替换了 YOLOv4 中的池化金字塔，从而获取了更大的感受野，并在骨干网络的反馈机制上增加了特征金字塔，以保留更多小目标的特征信息。实验结果证明，相比 YOLOv4，该模型缩小了 138 MB，并且在 KITTY 数据集[39]上平均精度提升了 1.4%。

3.1.2 速度优先骨干网络

为了更加适应工程部署，计算量小和检测速度快的轻量化网络受到了更多学者的关注。早期的 SqueezeNet[40]通过 Fire 模块减少了参数并压缩了模型。近年来，MobileNet 系列[41-43]和 ShuffleNet 系列[44-45]模型从信息交流和缩减参数两个方向进行优化，是轻量化骨干网络中的研究热点。MicroNet[46]基于 MobileNet 系列进行改进，使用低秩矩阵降低网络节点之间的连通性，并利用非线性激活函数弥补精度损失。GhostNet[47]结合普通的卷积和简单的线性操作，以较小的代价增强特征。下文对常用的 MobileNet 和 ShuffleNet 系列模型进行了详细介绍。

（1）MobileNet 系列。MobileNetv1[41]将标准卷积分解成一个深度卷积（Depthwise Convolutions，DWConv）和一个 1×1 卷积（Pointwise Convolutions，PWConv）。DWConv 将不同的卷积核应用到每一个通道，PWConv 通过组合 DWConv 的输出，在保证模型准确率的同时缩短了计算时间并减少了参数数量。MobileNetv2[42]用紧凑的低维特征表达高维特征，

设计倒残差结构(Inverted Residuals)进行特征提取并映射降维。MobileNetv3[43]结合 NetAdapt[48]技术搜索最佳的神经网络结构,并引入 SE(Squeeze and Excitation)模块来加强网络的学习能力。Zhang 等[49]对 YOLOv3 进行了改进,用 MobileNetv3 替换 DarkNet-53 并引入空间注意力模块,以更有效地判断远距离小目标。

(2)ShuffleNet 系列。ShuffleNet v1[44]设计了逐点群卷积(Pointwise Group Convolution,PGConv)和通道混洗(Channel Shuffle)两个结构。PGConv 减小了计算复杂度,通道混洗用于抵消 PGConv 造成的通道信息流通不畅,加强通道内的信息表达能力。在此基础上,ShuffleNetv2[45]提出了 4 条轻量化网络设计准则:1)在同等输入输出通道下内存访问成本最小;2)慎用组卷积;3)网络碎片化不可取;4)减少元素级运算。ShuffleNetv2 引入通道分割(Channel Split)算子,将输入特征的通道拆分为两支,并借鉴 DenseNet 中的特征重用以获取更多有效信息。

从应用的角度出发,精度优先的骨干网络适用于大规模检测,它通常使用增加层数、拓宽结构的方式进一步提取特征,并且使用分组以及结构复用的手段来节省参数,从而提升检测效果,但同时也带来了大量的计算量以及高内存消耗。而速度优先的网络更适用于小型移动端检测,它趋于使用深度可分离卷积、通道分割等降低计算量的手段,以达到模型轻量化的目的。

3.2 金字塔结构

在卷积神经网络中,深层特征分辨率低,拥有更大的感受野与更强的语义信息。浅层特征空间分辨率高,具有丰富的纹理信息和准确的空间信息。利用金字塔结构融合不同层次的特征信息,构建多尺度表达,是提升小目标检测性能的有效策略。金字塔结构主要包括图像金字塔和特征金字塔。

3.2.1 图像金字塔

图像金字塔通过采样来进行扩大或缩放操作以获得多分辨率图像集合,不同分辨率图像的输入使神经网络学会对不同尺度的目标进行检测。

由于简单地放大图像或小目标并不能提升检测效果,Singh 等[50]设计了尺度归一化图像金字塔策略(Scale Normalization for Image Pyramids,SNIP)。图像金字塔的每一层只对合适的尺度范围内的目标进行监督训练,即目标实例的损失梯度在对应分辨率图像上进行反向传播时会忽略尺度偏差大的目标。SNIP 充分发挥了图像金字塔多尺度训练的优势,但计算成本也较高。为此,Singh 等[51]提出了带有重采样的尺度归一化图像金字塔策略(Scale Normalization for Image Pyramids with Efficient Resampling,SNIPER)。该策略将金字塔每一层提取出的相同分辨率大小(512×512)的图像块作为训练的单元,模型不用处理图像上的每个像素点,只需要选择有合适的尺度的实例来参与训练。

针对深度卷积网络中小目标信息易丢失的问题,Liu 等[52]设计了图像金字塔引导网络(Image Pyramid Guidance,IPG),如图 3 所示。IPG 网络首先使用图像金字塔提取浅层子网络中的空间位置信息与纹理细节信息,然后在融合模块中将来自图像金字塔的浅层特征和来自特征金字塔的深层特征相融合,改善深层网络中特征错位与细节丢失的问题,从而增强小目标检测的效果。

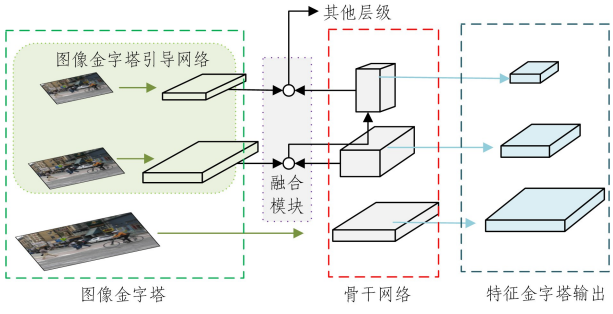


图 3 图像金字塔引导网络

Fig. 3 Image pyramid guidance network

图像金字塔通过对不同分辨率的图像进行训练,让网络适应不同尺度的目标检测,一定程度地提升了检测精度。但是,整个图像金字塔的输入会给训练带来内存与时间上的压力。

3.2.2 特征金字塔

特征金字塔网络[53](Feature Pyramid Network,FPN)通过对输入的图像进行多次上采样得到不同尺度的特征图,然后从上至下将高层提取的抽象语义信息与特征提取过程中低层轮廓纹理等的细节信息相融合,从而达到特征增强的目的。

由于 FPN 网络在网络加深的过程中会不断丢失浅层空间信息,Liu 等[54]提出了路径聚合网络(Path Aggregation Network,PANet)。如图 4(b)所示,PANet 在特征金字塔的底部连接出了一条自下而上的增强路径,用于缩短信息交融的传递途径,从而将具有细粒度特征的位置信息输入融合网络中,以提升特征金字塔架构的检测能力。BiFPN[55]在 PANet 基础上,移除只有一个输入的节点,减少参数计算量。如图 4(c)所示,在同一尺度下的输入特征层和输出特征层之间增加了一条额外的连接,以加强浅层特征与深层特征的融合。并且 BiFPN 给各层赋予了权重以进行自适应学习,让网络学习到不同层次的重要性。

为探讨反馈机制中浅层信息对检测的影响,Qiao 等[33]推出了 DetectoRS 检测网络。通过设计的递归特征金字塔(Recursive Feature Pyramid,RFP)将 FPN 输出的结果作为额外反馈连接到骨干网络中进行二次特征提取,增强了 FPN 的特征表示,如图 4(d)所示。作者还用可切换空间卷积(Switchable Atrous Convolution,SAC)替换标准的 3×3 卷积,并在其前后插入上下文模块,在节省参数的同时增强了模型中的全局信息。在 RFP 和 SAC 的共同作用下,DetectoRS 在 MS COCO 数据集上相比基准网络 HTC[56]的小目标检测精度提升了 6.4%。

为实现多尺度信息的充分融合,GiraffeDet[57]提出了两种特征连接方式。1)$\log_2 n$-连接。相比 DenseNet 中的密集连接,每个特征层需要接收前面所有层的特征信息,$\log_2 n$-连接中,第 l 层只需要接受 $\log_2 l+1$ 个先前层的特征信息,并且在反向传播中将最短的距离 1 提升到 $\log_2 l+1$,这有助于扩展到更深的网络。2)跨尺度连接。作者提出了一种同时考虑同级与领域特征的 Queen 融合,通过双线性插值

与最大值池化对各尺度的特征信息进行融合。GiraffeDet在MS COCO数据集上获得了SOTA性能，其中小目标检测性能提升了2.8%。多尺度特征融合是小目标检测的常用方法。其通过结合高层语义信息与低层细节信息，显著提升了小目标的检测性能。然而，特征金字塔网络的构建方式主要分为跨层连接和并行支路，在提升性能的同时也增加了额外的参数计算和存储空间。在下一阶段的工作中进行金字塔架构设计时需要探索提升了小目标检测效果的原因是设计的特征融合方式还是对特征的反复堆叠利用。

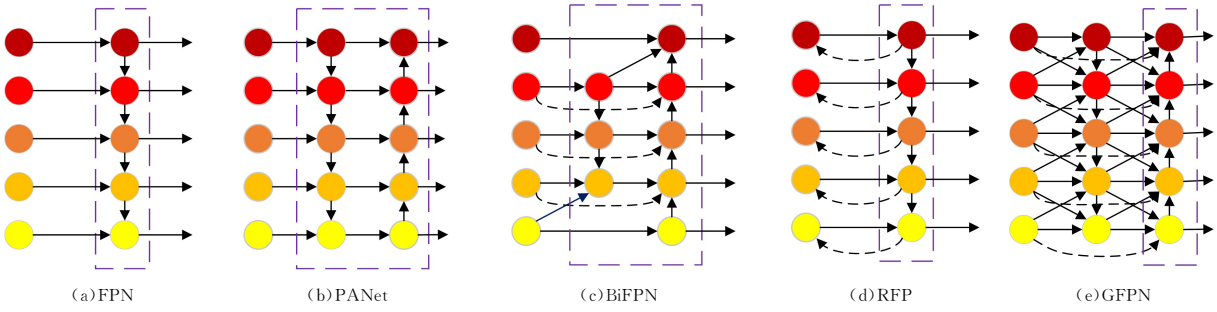


图4 5种特征金字塔网络架构

Fig. 4 Five ways of feature pyramid network

3.3 锚框设计

模型的输出是对大量预测框进行分类和回归的过程，根据生成预测框方式的不同分为有锚机制(Anchor-based)和无锚机制(Anchor-free)。

3.3.1 有锚机制

锚框起源于Faster RCNN[58]模型。相比传统检测方法中的遍历滑窗方式，锚框通过在特征图中对每个位置覆盖预先设定的不同大小和纵横比的先验框，来处理多尺度特征信息，极大程度地提高了检测速度。对于小目标检测而言，设计尺寸、数目适合的锚框和锚框匹配策略尤为重要。

将不同比例的锚框平铺在检测层中，通过计算各尺度锚框的平铺密度会发现，小锚框的密度相比大锚框而言过于稀疏，这造成了小目标的召回率低。针对平铺密度不平衡的问题，Zhang等[59]提出了锚框稠密策略，通过对不同尺寸的锚框进行“加密”处理，来保证不同尺寸的锚框在图像上密度统一，使小目标能够匹配到足够数量的锚框。Yang等[60]根据人脸检测的特性，对该策略进行拓展，设定高宽比为1的锚框检测正脸，高宽比为1.5的锚框检测侧脸。该设计覆盖了大部分人脸检测任务中的目标尺寸，有效提升了小目标检测性能。

相比直接通过超参数设定锚框尺寸，Zheng等[61]从经过调窗算法处理过的数据中提取先验信息，然后设计各特征层锚框尺寸，并且采用IoU优化的锚框补选增强策略，来降低小目标丢失的风险。该策略将迭代过程中淘汰的候选框进行随机补选，将其加入之后的迭代训练中，防止置信度阈值相邻的锚框被错误舍弃，从而提升小目标的召回率。

针对小目标检测中预定义的锚框与目标位置不匹配的问题，Zhou等[62]采用基于形状相似度的聚类算法生成一组锚框，通过IoU自适应阈值选择(Adaptive Threshold Selection，ATS)方法将生成的锚框划分为正样本和负样本，以保证训练时样本均衡。Chen等[63]提出了一种基于聚类的锚框匹配策略，将每个目标选取的与真实值最近的k个锚框作为正样本，保证网络训练时能够充分考虑各个不同尺寸大小的目标，以此提升小目标检测精度。

3.3.2 无锚机制

对于小目标检测而言，基于锚框的目标检测算法中与锚框相关的超参数过多，并且对检测精度敏感。若目标的大小和设定的锚框大小相差过大，则不易检测；若增加锚框数量，检测速度则会变慢。近年来，无锚框算法的出现缓解了这个难题。

常规的无锚框算法分为基于关键点检测和基于密集检测。基于关键点的代表算法有Corner-Net[64]，ExtremeNet[65]和CenterNet[66]，通过检测特定关键点，并对关键点进行分组生成高质量检测框。基于密集点检测的无锚框算法有FSAF[67]，FCOS[68]，FoveaBox[69]和VFNet[70]，以上方法都利用特征金字塔网络进行多尺度检测，并将分类与回归分为两个子网络进行。表3列出了部分无锚框检测算法。

表3 部分无锚框检测模型的总结

Table 3 Summary of various anchor-free detection models

算法名称	发表时间	算法描述	局限性
CornerNet[64]	ECCV 2018	通过检测边界框的一对角点以及计算角点之间的距离对其进行分组	对边缘敏感；缺乏对目标全局信息的参考
ExtremeNet[65]	CVPR 2019	预测4个多峰值极值点与1个中心点，根据中心点对应的分数判断4个极值点是否为一组	效率低；无法处理共线且尺度相同的目标
CenterNet[66]	ICCV 2019	是对CornerNet的改进，通过引入中心点去除大量的错误目标框，尤其对小目标而言有较大作用	无法处理两个物体在真实框中的中心点重叠的情况
FSAF[67]	CVPR 2019	预测中心点到4个边界的距离并且让每个实例动态选择最适宜的特征层进行训练以优化网络	没有完成彻底的标签分配，与锚框分支结合才能达到理想效果
FCOS[68]	ICCV 2019	预测中心点到4个边界的距离并且利用center-ness分支剔除远离目标中心的错误边界框	容易得到过多的假阳性样本；中心度可解释性弱

（续表）

算法名称	发表时间	算法描述	局限性
FCOSv2[73]	TPAMI 2020	将 center-ness 与回归放在一起，将正样本的选取调整到符合边界约束目标的中心区域	实用性不高；模型速度慢
FoveaBox[69]	TIP 2019	预测坐标转化，将实例映射到对应尺寸的特征图上，引入超参数动态设置正样本范围	需要引入超参数动态设置正样本范围和金字塔网络不同层的尺度
VFNet[70]	CVPR 2021	使用融合目标存在置信度和定位精度的 IoU 感知分类评分机制对检测进行分级，并设计了一个准确捕获上下文信息的星形边界框	需要对 9 个特征采样点使用分离卷积输出缩放系数
HoughNet[71]	ECCV 2020	使用投票方法整合长短距离的视觉证据找到 map 的峰值，峰值表示目标实例的存在	投票机制不利于位置敏感的回归分支

除常规方法外，HoughNet[71]将投票机制引入无锚框目标检测任务中。模型将输入图片转化为存在视觉信息的分值图，通过分值叠加对包含目标的区域进行投票。在偏心视觉系统的启发下，与目标区域越近则投票的空间精度越高，最后累积投票的峰值表示目标实例的存在。

总而言之，有锚机制通过调整锚框的尺度范围和数量来提高小目标正样本的召回率，但难以平衡检测精度和计算成本的矛盾；无锚机制可以避免复杂的超参数设计，更易于训练，但由于缺乏目标区域内外部的上下文信息，因此容易获得大量无效边界框。Zhang 等[72]提出，有锚机制与无锚机制的本质区别是正负样本的定义方式，标签分配策略可以提升小目标检测的上限。未来可以从标签分配策略这一角度入手，探讨利用有锚机制与无锚机制来共同提升小目标检测效果。

3.4 优化目标

本节从交并比（IoU）、非极大抑制（NMS）算法和损失函数（Loss Function）3 方面对小目标检测的优化目标进行了总结，并分析了其优缺点。

3.4.1 交并比

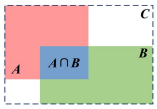
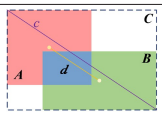
交并比是目标预测框（Bounding Box）和真实框（Ground Truth）交集和并集的比值。若 IoU 阈值过高，正样本质量高、数量少，则易导致正负样本不平衡。若 IoU 阈值较低，正样本数量增加，则会导致质量下降。

由于直接提高 IoU 阈值易导致模型过拟合以及正负样本不平衡，Cai 等[74]设计了级联检测器 Cascade R-CNN，通过分阶段逐步提高 IoU 阈值，将检测器的输出作为下一个检测器的输入进行训练，让每个阶段的检测器专注于检测此阈值范围内的候选框。级联方式为检测器筛选出高质量正样本，避免了过拟合问题。在此基础上，Li 等[75]对 Cascade R-CNN 检测器进行改进，采用特征共享的思路缩小了低 IoU 阈值下不同阶段之间的差距，缓解了低 IoU 阈值的候选框在进入第三阶段被当作负样本处理的情况，从而稳定地提升小目标检测性能。

除了通过级联获取高质量正样本，还可以通过优化交并比（IoU）函数来平衡正负样本。如表 4 所列，A 和 B 分别代表预测框和真实框。若预测框和真实框不重叠，则 $IoU=0$，模型训练中止。针对非交叉框无法回归这一难题，Rezatofighi 等[76]提出了 GIoU（Generalized IoU）。该策略考虑了无法优化的非重叠部分（见表 4 中的 C，为最小闭包区域，除交叠部分外的面积），保证当两框不相交时惩罚项和损失依然存在，模型始终会对回归产生监督的作用。

表 4 4 种 IoU 函数的对比

Table 4 Comparison of four IoU functions

	示意图	数学表达式	优势	局限
IoU		$IoU=\frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}$	具有尺度不变性、非负性、同一性	若 A 和 B 两框不相交，则无法进行训练；无法区分重合情况
GIoU		$GIoU=IoU-\frac{\lvert C\backslash(A\cup B)\rvert}{\lvert C\rvert}$	同时关注重叠区域和非重叠区域	当 A 和 B 两框重叠时，会退化为 IoU；当 $C-(A\cup B)\to 0$ 时，水平和垂直方向收敛速度变慢
DIoU		$DIoU=1-IoU+\frac{d^2}{c^2}$	最小化中心点距离 d，加快收敛速度	若多个候选框中心点重合，则无法区分哪个与真实框更相似
CIoU		$CIoU=DIoU+\alpha v$	考虑了预测框拟合真实框的长宽比	未解决难易样本不平衡问题

为提高回归的收敛速度，Zheng 等[77]引入两框的距离和尺度，提出了 DIoU（Distance IoU）。如表 4 所列，d 为两框中心点间的欧氏距离，c 为两框最小闭包区域的对角线长度。相比 IoU 和 GIoU，DIoU 在训练过程中不会发散并且两框不重叠时也能收敛，同步提升了准确率和收敛速度。作者还提出了回归定位的 3 个重点因素：重叠面积、中心点距离和长宽比。

CIoU（Complete IoU）作为 DIoU 的改进版本，它将两框的长宽比代入计算，增加了权重系数 α 和度量预测框与真实框长宽比相似性的参数 v，如式（1）和式（2）所示，它解决了当多个候选框中心点重合时难以找出最优预测框的问题。

$$\alpha=\frac{v}{(1-IoU)+v} \tag{1}$$

$$v=\frac{4}{\pi^2}\left(\arctan\frac{w^A}{h^A}-\arctan\frac{w^B}{h^B}\right)^2 \tag{2}$$

虽然 CIoU 会带来更高的精度收益，但反三角函数的计算延长了整体训练时间，其应用范围有限。

在不同的检测任务中，固定 IoU 阈值易导致样本不平衡，设置动态 IoU 阈值更适合不同尺度的目标检测任务。其中级联方法通过逐阶段提升 IoU 阈值筛选出高质量正样本，并通过多步预测进行渐进定位，但也存在小目标漏检的问题。通过优化交并比函数可以进一步准确评价两个目标框的重叠程度，加快预测框的筛选，但也会带来更复杂的计算。此外，对于 IoU 的选择需视具体情况而定，最优的 IoU 函数不一定在各类数据集上都能达到最优的训练效果。

3.4.2 非极大抑制算法

非极大抑制（Non-Maximum Suppesion，NMS）的作用是筛除冗余候选框，找到最佳物体检测位置。其算法的步骤如下：1）计算每一个候选框的置信度得分并按照得分降序排列；2）遍历其余的候选框，若该候选框和分值最高的候选框之间的 IoU 大于设定阈值，则删除该候选框；3）重复以上步骤，直到所有的候选框都被处理。

由于小目标占比面积小，因此在严格的 NMS 剔除机制中包含小目标的候选框容易被暴力剔除。Bodla 等[78]对 NMS 算法进行延伸，提出了可以对得分进行惩罚衰减的 Soft-NMS，式(3)为惩罚衰减公式。其中，预测框的重叠度越高，得分衰减得越多，这就极大地改善了小目标重叠问题。并且由于置信度永不为 0，因此避免了部分优良候选框被错误筛除。

$$s_i = s_i e^{-\frac{IoU(M,b_i)^2}{\sigma}}, b_i \notin \mathcal{D} \tag{3}$$

其中，s_i 为每个框的置信度，b_i 是第 i 个检测框，M 为得分最高的检测框。

Jiang 等[79]认为分类置信度和定位置信度之间存在偏差，直接对分类的结果做非极大抑制并不合理。为此，作者提出了 IoU-guided NMS，将分类置信度替换成 IoU 带入回归优化过程中，如式(4)所示。通过 IoU 引导非极大抑制过程，保留部分定位更精准的候选框，修正分类和定位之间的不平衡。

$$c^* = \arg\min_c crit(transform(box_{\det}, c), box_{gt}) \tag{4}$$

其中，$box_{\det}$ 和 box_{gt} 分别为检测边界框和 GT 框，transform 为边界框转化函数，$crit$ 表示 $\ln(IoU)$ 函数。

非极大抑制算法作为筛选优良候选框的重要手段，也是影响检测精度的重要因素。通过改进 NMS 算法修正与目标匹配的边界框位置，能够有效改善小目标检测中因分类和定位不平衡造成的漏检，从而提升模型召回率和最优候选框的选取精度，但同时，算法的训练效率也会受到影响。后续的小目标检测算法研究应着重于同时提高检测精度和检测效率。

3.4.3 损失函数

损失函数可以衡量模型输出预测值与真实值的差异程度，指导神经网络的学习过程。在模型训练过程中，损失函数通过反向传播不断修正权重参数来获取最优预测值。常见的损失函数有 L_1 Loss，L_2 Loss，IoU Loss 和 Focal Loss[80]等。

Focal Loss 通过调整难分样本在损失函数上的权重，来提高难分样本的准确度。针对 Focal Loss 难以应对检测后期难易样本不均衡的问题，Luo 等[81]设计了一个损失权重调整策略，将高于阈值但是没有匹配样本的目标设定为难拟合目标，并且通过网格搜索的方式获取最佳倍数系数，提升难拟合样本的损失权重。通过使损失函数倾向于训练难拟合的小目标，来有效提升小目标检测效率。

针对损失函数中大目标的损失占比远高于小目标的损失占比的问题，Liu 等[82]提出了一种反馈驱动损耗函数（Feedback-driven Loss Function），将损失比例信息作为反馈信号 $f(t)$ 带入损失函数的迭代计算中，如式(5)所示。其中目标尺寸越小，获得的增益越大。这种机制可以更有效地监督小目标且以更均衡的方式训练检测器。在 MS COCO 数据集上，该方法在几乎不增加计算量的情况下使小目标的检测精度提高了 3.5%。

$$f(t) = \begin{cases} \frac{1}{\pi\tau}\arccos t, & t < \omega \\ 2 - t, & t \geqslant \omega \end{cases} \tag{5}$$

其中，τ 为小目标损失占总损失的比值，$f(t)$ 为增益系数。

由于 IoU 对小目标的位置偏差敏感，因此 Wang 等[83]设计了一种基于 Wasserstein 距离的小目标检测评估方法。该方法通过将边界框建模为二维高斯分布，将边界框之间的相似性转化为两个高斯分布之间的分布距离，用来替代 IoU 阈值，如式(6)所示。实验结果表明，该方法能为微小目标保留更多的高质量正样本，并且该损失函数具有连续性，能够更好地指导网络的优化，表达式如式(7)所示。

$$W_2^2(N_a, N_b) = \left(\left[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \right]^{\mathrm{T}}, \left[cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \right]^{\mathrm{T}} \right) \tag{6}$$

$$\mathcal{L}_{NWD} = 1 - NWD(N_a, N_b) = 1 - \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{G} \right) \tag{7}$$

其中，(cx, cy, w, h) 分别代表边界框的中心坐标、宽度和高度。

损失函数指导模型获得更准确的定位和更高的识别精度，具有较强的可解释性。损失函数的优化方向分为调整小目标损失权重和设计新的度量指标，两者都能缓解小目标的类别不平衡问题。前者几乎不需要增加计算成本就能提高小目标检测性能，指导模型收敛方向；后者建立的新的度量模型对目标尺度不敏感，更容易保留高质量小目标。

3.5 增益组件

3.5.1 注意力机制

人类视觉系统在处理复杂的问题时，会优先注意关键部分[84]。小目标检测专注于小目标所在的区域，注意力机制[85-87]能够帮助神经网络在众多信息中聚焦于任务相关信息。

Hu 等[85]提出的 SE 模块通过挖掘特征通道之间的相互依赖关系，自适应校正各个通道的权重参数，让网络聚焦于更关键的特征信息。Woo 等[86]在空间维度上进行扩展，设计了卷积注意力模块（Convolutional Block Attention Module，CBAM）。它由通道注意力模块（Channle Attention Module，CAM）和空间注意力模块（Spatial Attention Module，SAM）串联而成，提高了网络对特征信息的提取能力和强化能力。如图 5 所示，CAM 结合了全局平均池化和全局最大池化，能够在减少信息的同时帮助网络捕获特征图中更具判别性的特征。SAM 通过学习特征间的空间关系，生成与原图相同空间尺寸的注意力图。CBAM 模块轻量且可嵌入各类检测网络中，提升了网络对判别性特征的提取能力。

在无人机小目标检测任务中，TPH-YOLOv5[88]将 CBAM 模块集成于 YOLOv5 中，帮助网络在区域覆盖范围大的图像中提取小目标所在区域的特征信息。在 VisDrone2021 挑战赛上，TPH-YOLOv5 的小目标检测精度相比基准模型提升了 7%。

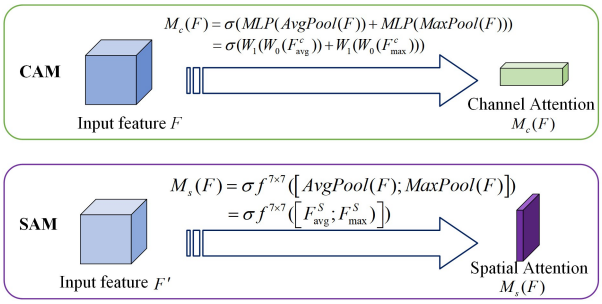


图 5 CBAM 示意图

Fig. 5 Diagram of CBAM

Wang 等[87]设计的一种高效通道注意力模块(Efficient Channel Attention,ECA)直接实现了局部跨通道交互,大大减小了降维操作给通道学习带来的复杂度。ECA 通过逐通道全局平均池化以及 k 个快速卷积来捕获相邻通道特征的依赖关系,学习有效的通道注意力信息。针对深层次小目标检测任务,Zhao 等[89]提出了以 ECA 为基础的特征融合策略,通过融合不同层次的特征信息来增强浅层语义信息的表达,从而提升小目标的检测精度。

针对小目标低分辨率和尺寸小的特性,Zhang 等[90]提出了多分辨率注意力检测器(Multi-Resolution Attention Extractor,MRAE)。它利用残差网络结构的特点,通过计算其他输出层与模板层的余弦相似度来获取注意力权重,然后将特征提取网络中前三层的特征信息进行加权融合来生成注意力图,以丰富小目标的特征表达。

目标检测领域中出现了众多轻量高效的注意力模块。其中通道注意力侧重于学习特征之间的相互关系,对不同的输入任务进行特征分配。空间注意力侧重于找到网络中最需要关注的区域。近年来,将自注意力(Self Attention)机制[31]应用于计算机视觉领域也是一大热门趋势,通过计算非局部模块特征图中空间点之间的相关矩阵,可以捕获远距离的依赖关系,从而更好地学习局部与全局特征信息。

3.5.2 上下文建模

上下文建模通过捕获目标区域周围的信息并且学习目标与背景信息之间的内在联系来辅助目标物体的检测。由于小目标本身包含的特征信息有限,因此其周边区域的上下文信息起着重要的作用。下文将从局部上下文、全局上下文和上下文交互 3 个角度介绍基于上下文建模的小目标检测方法。

(1)局部上下文。局部上下文指检测目标周边区域中的上下文信息。不同尺度的特征层上候选框的尺寸不同,感受野也不相同,各层提取的特征信息之间具有互补性。Zeng 等[91]设计了门控双向神经网络(Gated Bi-directional CNN,GBD-Net),该网络通过在目标窗口放大或缩小感受野来收集更多上下文信息,在不同分辨率和不同层次间进行特征交互传递。

为了探索局部上下文信息的作用机制,Li 等[92]提出了关注上下文卷积神经网络(Attentive Contexts CNN,AC-CNN),它由基于注意力的全局上下文(AGC)模块和多尺度局部上下文(MLC)模块组成。AGC 通过多个堆叠的长短期记忆(LSTM)层为输入图像递归生成注意力图,用于突出有效的全局上下文位置。MLC 利用多个尺度上每个候选框的特征信息来获取局部上下文信息。AC-CNN 将全局上下文信息和局部上下文信息结合,共同提升小目标检测效果。

(2)全局上下文。全局上下文指对整个图像的信息进行学习。Bell 等[93]设计了一种同时结合感受野内外部信息的目标检测器(Inside-Outside Net,ION),通过多尺度特征学习内部细节信息,并将循环神经网络(RNNs)拓展到二维,从 4 个方向对感受野外的上下文信息进行编码,实现对全局上下文信息的感知。

针对密集人脸检测,Song 等[94]指出局部上下文忽略了浅层特征与深层特征之间的联系,浅层特征图中与人脸相似的特征会被错误分类,需要从深层和更大的感受野中提取全局上下文信息并将其反馈给浅层以进行辅助检测。作者提出了全局上下文融合模块架构(Global Context Fusion,GCF),通过自上而下融合所有尺度的上下文信息进行预测。其中每个尺度从特征金字塔的当前层提取特征,并通过上下文模块与其他尺度被上采样到相同大小的特征进行融合,以提高人脸检测性能。

(3)上下文交互。上下文交互指通过目标与目标之间、目标与背景之间的约束关系进行上下文语义信息的传递。Liu 等[95]提出了一种新型的结构化推理网络(Structure Inference Net,SIN)。作者将目标视为节点,将目标与目标之间的关系视为边,各个节点和边相连接构建成一个图网络。推理网络在对图像的场景信息进行特征提取的同时对节点之间的关系进行编码,通过不断更新迭代,将最后的二节点信息用于分类与回归任务。

针对小目标检测中语义信息弱且空间信息衰减的问题,Fu 等[96]模仿人类视觉机制,提出了上下文推理方法(Intrinsic Relationship Reasoning,IRRNet)来捕获目标物体之间的关系。IRRNet 通过构建基于初始区域特征的稀疏语义关系模块以及基于位置和形状信息的稀疏空间关系模块,将语义关系和空间关系生成无向图进行编码,找到小目标与大目标的相关性,从而推断小目标的位置。

基于上下文建模的方法通过挖掘图像中与目标相关联的信息,构建目标与周围环境的模型范式,有效提升了小目标检测性能。然而不是所有的上下文信息都有效,当周围环境缺少与目标关联性高的信息,并且没有其他易检测的物体可以辅助对小目标的检测时,这些冗余的信息噪声会影响模型的处理性能。

3.6 小目标检测总结

表 5 列出了从不同结构总结与分析的小目标检测算法的优缺点。从表中可以看出,在骨干网络、金字塔网络以及增益机制结构上的改进方案聚焦于特征的利用,通过对特征充分提取、融合包含不同分辨率信息的特征层、找到聚焦于任务的特征信息,来获得高精度的检测结果。在锚框设计与优化目标上的改进着力于优化模型的训练,加速检测模型的分类和回归。

表 5 不同结构改进方法的优缺点对比

Table 5 Comparision of advatages and disadvatages of improved methods in different structures

结构	分类	优点	缺点
骨干网络	精度优先	充分提取特征；高精度和高准确度	参数量多，计算量大
	速度优先	处理速度快；可在小型移动设备上搭载	准确度偏低
金字塔结构	图像金字塔	提供富含细节信息的多分辨率的图像	计算量大，占用内存多
	特征金字塔	融合浅层特征的细节信息和深层特征的语义信息	参数量多，计算量大
锚框设计	有锚机制	在密集的锚框上直接进行目标分类和回归，且有效提高了网络目标召回能力	尺度、数量等超参数需预定义； 大部分锚框对检测没帮助
	无锚机制	无需设置与锚框相关的各类参数； 回归不同尺度特征图的目标中心点和宽高，提高了推理速度	边界框回归难； 密集检测前景与背景极度不平衡； 语义模糊（目标中心点重叠）
优化目标	交并比 非极大抑制	通过设定阈值调整正负样本的质量和数量； 适用范围广	高阈值易导致小目标漏检误检； 需要针对性设计
	损失函数	可解释性强，易调整权重	函数不连续； 鲁棒性低
增益组件	注意力机制	帮助网络捕获与当前任务相关的判别性特征	增加参数量； 位置信息难捕捉
	上下文建模	利用不同对象之间的相互关系、对象和环境之间的关系作为条件辅助对新目标进行识别	有效上下文信息匮乏； 易造成无效信息冗余

4 小目标检测算法对比

本节对当前主流的小目标检测模型进行总结分析以及检测性能评估。其中，AP 代表平均精度，AP^{50} 和 AP^{75} 分别代表 IoU 阈值为 0.5 和 0.75 时的平均精度，AP^S，AP^M，AP^L 分别代表小、中、大 3 种不同尺度目标的平均精度。

表 6 列出了对文中列举的部分基于锚框的小目标检测算法的描述和局限性分析。表 7 将表 3 和表 6 中列举的具有代表性的小目标检测算法按照时间顺序进行排列，并在 MS COCO 数据集上对比了实验结果。

表 6 基于锚框的小目标检测模型总结

Table 6 Summary of various anchor-based small detection algorithms

算法名称	发表时间	算法描述	局限性
Faceboxes[59]	IJCB 2017	在不同尺度特征图上设置锚框检测物体，对锚框进行稠密化处理	不能处理过小的目标，会过滤像素小于 20 的目标框
YOLO v3[22]	CVPR 2018	骨干网络是使用残差跳层结构的全卷积网络，并且引用多尺度预测	识别物体位置精准性差，召回率低
SNIP[50]	CVPR 2018	对金字塔每一个分辨率进行训练，并且只对特定范围（即尺度相近）的目标回传损失	需要处理金字塔层图像上的每个尺度图片的像素计算，参数量大，计算慢
SNIPER[51]	NIPS 2018	训练时不需处理整张图片，仅处理图像金字塔上真实框旁边的图片块	只提升训练速度；模型必须依赖图像金字塔
Cascade R-CNN[74]	CVPR 2018	通过不断提升 IoU 阈值来保证每个检测器都有等效尺寸的正样本；逐步找到最接近目标的预测	多阶段级联检测调整参数量难； 在低 IoU 阈值上不理想
PANet[54]	CVPR 2018	在特征金字塔上连接出一条自下而上的增强路径，引入浅层特征进行分割；引入自适应特征池化融合特征	所有层级输入贡献视为相同
FSCascade[75]	BMVC 2019	采用特征共享的思路缩小了低 IoU 阈值下最后阶段与之前阶段之间的差距	预测中的全连接换成全卷积；可解释性弱
IPG-Net[52]	CVPR 2020	引入图像金字塔，为主干网络的每个阶段提供更多的空间信息	参数量大，计算慢
IRRNet[96]	CVPR 2020	通过构建语义模块和空间模块集成上下文信息	难以区分同类别下语义特征强的小目标；引入噪声风险大
EfficientDet[55]	CVPR 2020	提出混合缩放方法来缩放目标检测模型；给特征金字塔每个层级的输出赋予权重并且网络自适应进行权重更新	基准模型是用 NAS[97] 搜索出来的，对计算资源要求高
DetectoRS[33]	CVPR 2020	将 FPN 输出的结果作为额外反馈连接到骨干网络中进行二次特征提取，使用可切换空洞卷积自适应选择合适的感受野	可切换空洞卷积耗时严重且更适用于大目标检测
FDNet[82]	IVC 2021	将损失比例信息作为反馈信号带入损失函数的迭代计算中，有效监督小目标的训练	鲁棒性低

（1）从骨干网络对比来看，由于 ResNet 具有能够逐层深入表征以及易于优化的特点，被大部分算法模型作为基础骨干网络。其改进模型 ResNeXt-101 和 Res2Net 分别被作为有锚机制 DetectoRS 和无锚机制 VFNet 的骨干网络，其平均精度达到了 54.7% 和 55.1%，分别达到了两类算法的最佳效果。

（2）从纵向对比来看，随着时间的推移，小目标检测模型的性能有了明显的提升。早年 YOLO v3 算法中小目标检测精度不足大目标检测精度的一半，后期 VFNet 将小目标检测精度提升至 37.4%，性能提升了一倍。

（3）从时间维度进行分析，早期对小目标检测问题的探索聚焦于金字塔网络的构建、锚框设计等方面，此阶段

的代表算法有 SNIP, SNIPER, Faceboxes 等。其中，图像金字塔方法虽然检测精度较高，例如 SNIPER 在 2018 年检测精度达到了 46.1%，但其由于受到训练时的内存问题和推断时的速度问题的限制，因此难以推广。然而，随着特征金字塔网络的深入探索，小目标检测算法得到了大幅度提升，无锚机制方法成为了新的研究热点。其中多尺度特征学习方法 EfficientDet 和 DetectoRS 利用双向金字塔 BiFPN 和递归金字塔 RFP 达到了 52.2%和 54.7%的平均精度，远胜于同期的各类方法。并且，特征金字塔可以通过融合浅层细节信息和深层语义信息，来改善无锚机制算法的语义模糊问题，从而弥补有锚机制和无锚机制的检测差距。例如，FCOSv2 通过嵌入 BiFPN 网络并且利用中心度分支剔除低质量边界框从而提升回归预测，将小目标检测精度提升至 33.2%。

(4)从优化目标的角度进行分析，IoU 和 NMS 算法立足于保留高质量正样本并且加快正负样本的筛选。其中，FSCascade 对级联方法 CascadeR-CNN 进行改进，以减小低阈值下不同阶段之间的差距，使小目标检测精度提升了 1.4%。而损失函数的优化设计具有较强的解释性，可以对模型进行监督训练。FDNet 算法将小目标损失比例信息带入损失函数的训练中，在基准网络 IPG-Net101 上将小目标检测精度提升至 28.0%。

(5)从增益组件的角度进行分析，注意力机制和上下文建模是小目标检测的常用增益组件，能够插入骨干网络以及金字塔网络中，以较小的成本增强目标区域的特征，并有效地提升目标检测精度。

表 7 MS COCO 数据集上小目标检测模型的平均精度对比

Table 7 Comparison of average accuracy of small detection algorithms models on MS COCO dataset

算法名称	骨干网络	输入图像尺寸	AP/%	AP^{50}/%	AP^{75}/%	AP^{S}/%	AP^{M}/%	AP^{L}/%
YOLO v3[22]	DarkNet-53	320×320	33.0	57.9	34.4	18.3	35.4	41.9
SNIP[50]	ResNet-50	多尺度	43.6	65.2	48.8	26.4	46.5	55.8
SNIPER[51]	ResNet-101	多尺度	46.1	67.0	51.6	29.6	48.9	58.1
CascadeR-CNN[74]	ResNet-101	—	42.8	62.1	46.3	23.7	45.5	55.2
PANet[54]	ResNeXt-101	多尺度	45.0	65.0	48.6	25.4	48.6	59.1
CornerNet[64]	Hourglass-104	511×511	40.5	56.5	43.1	19.4	42.7	53.9
ExtremeNet[65]	Hourglass-104	511×511	43.7	60.5	47.0	24.1	46.9	57.6
FSCascade[75]	ResNet-101	—	43.2	62.3	47.5	25.1	46.2	54.8
CenterNet[66]	Hourglass-104	512×512	44.9	62.4	48.1	25.6	47.4	57.4
FSAF[67]	ResNeXt-101	1300×800	44.6	65.2	48.6	29.7	47.1	54.6
FoveaBox[69]	ResNet-101	1300×800	44.2	65.4	47.8	28.8	46.7	53.7
IPG RCNN[52]	IPG-Net101	1333×800	45.7	64.3	49.9	26.6	48.6	58.3
IRRNet[96]	ResNet-101	—	39.7	62.0	43.2	22.9	42.4	50.2
FCOS[68]	ResNeXt-101	1300×800	44.7	64.1	48.4	27.6	47.5	55.6
HoughNet[71]	Hourglass-104	512×512	46.4	65.1	50.7	29.1	48.5	58.1
FCOSv2[73]	ResNeXt-101	1300×800	50.4	68.9	55.0	33.2	53.0	62.7
EfficientDet[55]	EfficientNet-B6	多尺度	52.2	71.4	56.3	34.8	55.5	64.6
DetectoRS[33]	ResNeXt-101	多尺度	54.7	**73.5**	**60.1**	**37.4**	57.3	66.4
FDNet[82]	IPG-Net101	—	46.6	65.4	50.8	28.0	49.6	59.1
VFNet[70]	Res2Net-101-DCN	1300×800	**55.1**	73.0	**60.1**	**37.4**	**58.2**	**67.0**

5 小目标检测应用前景

在许多领域中，小目标检测算法面临着不同类型的检测挑战，如采集图像分辨率低、复杂场景下目标遮挡严重等问题，因此小目标检测算法有广泛的应用前景。下文将军用无人机检测、智能视频监控和自动驾驶这 3 类应用作为代表进行简要介绍。

(1)军用无人机检测

无人机在现代军事战场上不可或缺，在军事侦察、军事打击等任务中能有效辅助我方军事人员进行战场态势感知。由于无人机进行远距离的高空作业，检测的目标常有不清晰、不准确的问题，因此小目标检测技术对于军事任务意义重大。针对小目标在军事中的智能感知技术研究，Hui[98]把骨干网络 ResNet 上的 ROI 池化层舍去，保留小目标特征信息。通过锚点和锚框一一对应来减少特征提取的计算量，并且设计多尺度特征层来提升小目标的检测。此外，Liu 等[99]首次将 CenterNet 和自适应特征激活模块相结合，从而有效抑制冗余特征的表达，并且加入了 GC-Block(Global Context Block)注意力模块来增强关键点语义信息。该方法在小目标数据集中的性能有着不错的提升。

(2)智能视频监控

视频监控被广泛应用于商场、火车站等公共场所，是保障城市安全的重要工具。由于监控中常出现人口密集现象，所采集图片中的小目标呈现数量多、被遮挡、模糊等特点，具有一定的检测难度。为此，Huang 等[100]提出了一种新颖的 NMS 方法，用于对冗余候选框进行有效删减，并且从成对模型的对应关系中相互学习到更多的特征信息，从而提升行人检测任务的准确性。针对人群中的密集人脸检测，Li 等[101]采用平衡数据锚点采样策略来平衡不同尺度人脸样本的占比数量，并且将密集上下文模块嵌入预测层中以增强信息的传递。

(3)自动驾驶

自动驾驶是未来车辆技术发展的趋势，对行车环境中交通标志的精准检测是衡量自动驾驶技术的一项重要指标。为了检测小而密集的交通标志，Ge 等[102]将 VGG16 作为 R-FCN[103]的骨干网络，提出了减小特征图缩放倍数的策略。该策略聚焦于小尺寸范畴的物体，并采用特征拼层的方式组合 VGG16 同组的卷积层来丰富小目标的特征信息。与上述

方法关注骨干网络的设计不同，Tang 等[104]设计了一种具有特征聚合的集成特征金字塔。它将集成操作引入金字塔层级以解决 RoI 不平衡问题，并且特征聚合结构增强了特征图的表征能力，从而能更精准地检测交通标志。

6 未来展望

虽然当前小目标检测已经取得显著的成果，但由于小目标检测场景具有复杂性与广泛性，因此其仍存在一些关键问题需要解决，未来工作可以从以下 5 个角度展开研究。

(1)小目标数据集的扩展与设计。由于当前通用数据集中的小目标数量占比低，并且现有的小目标数据集只针对单一类别或单一场景下的常见类别，因此训练出的模型可迁移性差。为了更好地评估小目标检测算法的性能，建立小目标检测专用的大规模数据集十分必要。此外，由于小目标对位置偏移比较敏感，MS COCO 数据集上关于精度的评价指标并不适用于小目标与极小目标，因此为该小目标数据集设计新的评价指标具有重要意义。

(2)标签分配方法。小目标检测任务存在严重的正负样本不平衡问题，划分正确的正负样本至关重要。标签分配指检测器在训练阶段给特征图的每个位置区分正负样本的过程。当前存在的基于先验信息的静态匹配策略和自适应动态匹配策略[72,105]，都有助于帮助模型找到合适的正负样本，从而加速训练收敛并提升模型的检测精度。

(3)弱监督小目标检测方法。由于图像中小目标标注的困难性和费时性，在训练数据有限的情况下，利用全监督学习方法难以训练出用于小目标检测的通用模型。弱监督学习指使用少量的全注释图像来检测大量的非全注释图像，利用对象类标注图像能够高效地训练小目标检测器，有助于对小目标进行分类；反之使用边界框训练检测器能增强目标的定位能力。通过巧妙平衡两种弱监督信息，可以进一步提升小目标检测性能。

(4)结合图像超分辨重建方法。小目标的特征分辨率低，难以检测。目前存在诸多基于生成对抗网络的超分辨率重建[106-108]方法。此类方法通过对原图像进行多次上采样后获得超分图像，从而恢复更多的纹理细节和像素信息，并且可以弥补深层网络中小目标的细节损失，从特征增强方面提升小目标的检测效果。

(5)模型的可解释性研究。当前小目标检测模型对数据集有较强的依赖性，建立的模型往往对该数据集拟合程度高。而模型中间涉及的大量映射关系以卷积的形式呈现，没有展现其本身的规律。模型的可解释性是通过规则与学习结合的方式，建立高精度、可解释、可通用且不依赖大量标注数据的目标检测模型。

结束语 本文回顾了近年来基于深度卷积神经网络的小目标检测研究进展，并进行了全面归类和分析。从骨干网络、金字塔结构、锚框设计、优化目标、增益组件 5 个方面详细阐述了现有小目标检测算法。同时，对各个方面的代表算法进行了对比分析，并结合实验结果概括了该技术的优势和局限。最后，详细阐述了军用无人机检测、智能视频监控和自动驾驶这 3 个领域的应用前景，以及小目标检测未来的研究方向。

参 考 文 献

[1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 580-587.

[2] SUN F W, LI C Y, XIE Y Q, et al. Review of Deep Learning Applied to Occluded Object Detection[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(6): 1243-1259.

[3] 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. 2017: 192-201.

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

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

[6] 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. 2016: 5525-5533.

[7] 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. 2020: 1257-1265.

[8] Detection Leaderboard [EB/OL]. https://cocodataset.org/#detection-leaderboard.

[9] 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. 2018: 3974-3983.

[10] CHENG G, HAN J, ZHOU P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98: 119-132.

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

[12] CAO Y, HE Z, WANG L, et al. VisDrone-DET2021: The Vision Meets Drone Object detection Challenge Results[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 2847-2854.

[13] 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. 2017: 3213-3221.

[14] 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. 2016: 2110-2118.

[15] 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). 2017: 1370-1377.

[16] CHEN C, LIU M Y, TUZEL O, et al. R-CNN for small object

detection[C]// Asian Conference on Computer Vision. 2016: 214-230.
[17] WANG J J, WEI J, MEI S H, et al. Improved YOLOv3 for Small Object Detection in Remote Sensing Images[J]. Computer Engineering and Applications, 2021, 57(20): 133-141.
[18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
[19] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1492-1500.
[20] ZHANG H, WU C, ZHANG Z, et al. Resnest: Split-attention networks[J]. arXiv: 2004. 08955, 2020.
[21] 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. 2016: 779-788.
[22] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv: 1804. 02767, 2018.
[23] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv: 2004. 10934, 2020.
[24] GAO S H, CHENG M M, ZHAO K, et al. Res2net: A new multi-scale backbone architecture[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(2): 652-662.
[25] BROCK A, DE S, SMITH S L, et al. High-performance large-scale image recognition without normalization[C]// International Conference on Machine Learning. 2021: 1059-1071.
[26] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]// European Conference on Computer Vision. 2016: 483-499.
[27] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.
[28] LEE Y, HWANG J W, LEE S, et al. An energy and GPU-computation efficient backbone network for real-time object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 752-760.
[29] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]// International Conference on Machine Learning. 2019: 6105-6114.
[30] LI Y, YAO T, PAN Y, et al. Contextual transformer networks for visual recognition[J]. arXiv: 2107. 12292, 2021.
[31] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv: 1706. 03762, 2017.
[32] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2818-2826.
[33] QIAO S, CHEN L C, YUILLE A L. DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution[C]// Computer Vision and Pattern Recognition. 2020.
[34] ZHAO W Q, KONG Z X, ZHAO Z B, et al. Small target detection based on a combination of feature pyramid and CornerNet [J]. CAAI Transactions on Intelligent Systems, 2021, 16(1): 108-116.
[35] YUAN L, LIU Z Y, ZHU M C, et al. Improved YOLOv3 Remote Sensing Target Detection Based on Improved Dense Connection and Distributional Ranking Loss[J]. Computer Science, 2021, 48(9): 168-173.
[36] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Highway networks[J]. arXiv: 1505. 00387, 2015.
[37] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 390-391.
[38] KOU Y C, HOU J, ZENG L M, et al. Road small target detection network based on feedback mechanism and hole convolution [J/OL]. Computer Engineering, 2022: 1-10. [2022-11-21]. DOI: 10. 19678/j. issn. 1000-3428. 0063575.
[39] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 3354-3361.
[40] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0. 5 MB model size[J]. arXiv: 1602. 07360, 2016.
[41] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv: 1704. 04861, 2017.
[42] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4510-4520.
[43] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 1314-1324.
[44] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 6848-6856.
[45] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018: 116-131.
[46] LI Y, CHEN Y, DAI X, et al. MicroNet: Towards image recognition with extremely low FLOPs[J]. arXiv: 2011. 12289, 2020.
[47] HAN K, WANG Y, TIAN Q, et al. Ghostnet: More features from cheap operations[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1580-1589.
[48] YANG T J, HOWARD A, CHEN B, et al. Netadapt: Platform-aware neural network adaptation for mobile applications[C]//

Proceedings of the European Conference on Computer Vision (ECCV). 2018:285-300.

[49] ZHANG X,LI N,ZHANG R. An improved lightweight network MobileNetv3 Based YOLOv3 for pedestrian detection[C]// 2021 IEEE International Conference on Consumer Electronics and Computer Engineering(ICCECE). IEEE,2021:114-118.

[50] 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. 2018:3578-3587.

[51] SINGH B,NAJIBI M,DAVIS L S. Sniper:Efficient multi-scale training[J]. Advances in Neural Information Processing Systems,2018,31:9310-9320.

[52] LIU Z,GAO G,SUN L,et al. IPG-net:Image pyramid guidance network for small object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020:1026-1027.

[53] 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. 2017:2117-2125.

[54] LIU S,QI L,QIN H,et al. Path aggregation network for instance segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:8759-8768.

[55] 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. 2020:10781-10790.

[56] CHEN K,PANG J,WANG J,et al. Hybrid task cascade for instance segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019:4974-4983.

[57] JIANG Y,TAN Z,WANG J,et al. GiraffeDet:A Heavy-Neck Paradigm for Object Detection[J]. arXiv:2202.04256,2022.

[58] REN S,HE K,GIRSHICK R,et al. Faster r-cnn:Towards real-time object detection with region proposal networks[J]. arXiv:1506.01497,2015.

[59] ZHANG S,WANG X,LEI Z,et al. Faceboxes:A CPU real-time and accurate unconstrained face detector[J]. Neurocomputing,2019,364:297-309.

[60] YANG S P,LIU H Z,WANG X Q. Small Size Face Detection Based on Feature Map Fusion[J]. Computer Science,2020,47(6):126-132.

[61] ZHEN X K,NIU Y,LI J. Research on Remote Sensing Image Target Detection Based on Improved SSD Algori-thm[J]. Laser Journal,2022,43(7):106-112.

[62] ZHOU H,YAN F L,CHU N,et al. Approach to Improve Detection Model for Small Object in Complex Scenes[J]. Computer Engineering and Applications,2022,58(11):187-192.

[63] CHEN Q,WANG Y,YANG T,et al. You only look one-level feature[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:13039-13048.

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

[65] 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. 2019:850-859.

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

[67] 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. 2019:840-849.

[68] 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. 2019:9627-9636.

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

[70] ZHANG H,WANG Y,DAYOUB F,et al. Varifocalnet:An iou-aware dense object detector[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:8514-8523.

[71] SAMET N,HICSONMEZ S,AKBAS E. HoughNet:Integrating near and long-range evidence for visual detection[J]. arXiv:2104.06773,2021.

[72] 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. 2020:9759-9768.

[73] TIAN Z,SHEN C,CHEN H,et al. Fcos:A simple and strong anchor-free object detector[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,44(4):1922-1933.

[74] CAI Z,VASCONCELOS N. Cascade R-CNN:high quality object detection and instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,43(5):1483-1498.

[75] LI A,YANG X,ZHANG C. Rethinking Classification and Localization for Cascade R-CNN[C]// British Machine Vision Conference. 2019.

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

[77] ZHENG Z,WANG P,LIU W,et al. Distance-IoU loss:Faster and better learning for bounding box regression[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020:12993-13000.

[78] BODLA N,SINGH B,CHELLAPPA R,et al. Soft-NMS-improving object detection with one line of code[C]// Proceedings of the IEEE international Conference on Computer Vision. 2017:5561-5569.

[79] JIANG B,LUO R,MAO J,et al. Acquisition of localization confidence for accurate object detection[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018:784-799.

[80] 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. 2017:2980-2988.

[81] LUO Y T, JIANG P F, DUAN, et al. Small Object Detection Oriented Improved-RetinaNet Model and Its Application[J]. Computer Science, 2021, 48(10): 233-238.
[82] LIU G, HAN J, RONG W. Feedback-driven loss function for small object detection[J]. Image and Vision Computing, 2021, 111: 104197.
[83] WANG J, XU C, YANG W, et al. A Normalized Gaussian Wasserstein Distance for Tiny Object Detection[J]. arXiv: 2110.13389, 2021.
[84] WANG W G, SHEN J B, JIA Y D. Review of Visual Attention Detection[J]. Journal of Software, 2019, 30(2): 416-439.
[85] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.
[86] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018: 3-19.
[87] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[J]. arXiv. 1910.03151, 2020.
[88] ZHU X, LYU S, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 2778-2788.
[89] ZHAO P F, XIE L B, PENG L. Deep Small Object Detection Algorithm Integrating Attention Mechanism[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(4): 927-937.
[90] ZHANG F, JIAO L, LI L, et al. Multiresolution attention extractor for small object detection[J]. arXiv: 2006.05941, 2020.
[91] ZENG X, OUYANG W, YAN J, et al. Crafting GBD-Net for Object Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40: 2109-2123.
[92] LI J, WEI Y, LIANG X, et al. Attentive contexts for object detection[J]. IEEE Transactions on Multimedia, 2016, 19(5): 944-954.
[93] 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. 2016: 2874-2883.
[94] SONG L, YANG J F, SHANG Q Z, et al. Dense Face Network: A Dense Face Detector Based on Global Context and Visual Attention Mechanism[J]. Machine Intelligence Research, 2022, 19(3): 247-256.
[95] 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. 2018: 6985-6994.
[96] FU K, LI J, MA L, et al. Intrinsic Relationship Reasoning for Small Object Detection[J]. arXiv: 2009.00833, 2020.
[97] GHIASI G, LIN T Y, LE Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7036-7045.
[98] HUI G B. Research on intelligent perception technology of military small target[C]// The 9th China Conference On Command and Control. 2021: 101-106.
[99] LIU X, HUANG J, YANG T, et al. Improved small object detection for UAV acquisition based on CenterNet[J]. Computer Engineering and Applications, 2022, 58(14): 96-104.
[100] HUANG X, GE Z, JIE Z, et al. Nms by representative region: Towards crowded pedestrian detection by proposal pairing[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10750-10759.
[101] LI Z, TANG X, HAN J, et al. Pyramidbox++: High performance detector for finding tiny face[J]. arXiv: 1904.00386, 2019.
[102] GE Y Y, XU Y J, ZHAO S, et al. Detection of small and dense traffic signs in self-driving scenarios[J]. CAAI Transactions on Intelligent Systems, 2018, 13(3): 366-372.
[103] DAI J, LI Y, HE K, et al. R-fcn: Object detection via region-based fully convolutional networks[J]. arXiv: 1605.06409, 2016.
[104] TANG Q, CAO G, JO K H. Integrated feature pyramid network with feature aggregation for traffic sign detection[J]. IEEE Access, 2021, 9: 117784-117794.
[105] GE Z, LIU S, LI Z, et al. Ota: Optimal transport assignment for object detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 303-312.
[106] 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). 2018: 206-221.
[107] 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. 2017: 1222-1230.
[108] 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. 2019: 9725-9734.

DU Zi-wei, born in 1997, postgraduate. Her main research interests include small object detection and deep learning.

LI Zhong-bo, born in 1983, Ph.D, senior engineer. His main research interests include machine learning, multimedia technology and cloud computing, etc.

(责任编辑:何杨)