控制与决策
Control and Decision

基于弱特征增强的轻量化小目标检测方法

周葳楠, 吴治海, 张正道, 彭力, 谢林柏

引用本文:
周葳楠,吴治海,张正道,彭力,谢林柏. 基于弱特征增强的轻量化小目标检测方法[J]. 控制与决策, 2024, 39(2): 381−390.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2022.1432

基于弱特征增强的轻量化小目标检测方法

周葳楠[1,2], 吴治海[1,2†], 张正道[1,2], 彭 力[1,2], 谢林柏[1,2]

(1. 江南大学 物联网工程学院，江苏 无锡 214122；
2. 江南大学 物联网技术应用教育部工程研究中心，江苏 无锡 214122)

摘 要: 针对复杂背景下小目标特征经多次卷积被背景噪声淹没导致的检测精度低的问题,提出一种增强弱特征表达的一阶段轻量级小目标检测算法SA-YOLO. 首先,用改进的ShuffleNetv2网络构建骨干网络,通过嵌入SE注意力模块和Inception结构,提升网络在复杂背景下的特征提取能力,有效地抑制背景噪声,充分提取弱特征;其次,在颈部网络,采用新的特征融合模块,以含有弱特征较多的低层级特征块的空间位置信息对高层级特征进行权重调整,提高不同层级的特征融合利用率,减少小目标的特征损失;最后,在头部网络,用解耦的检测头替换原YOLO耦合的检测头, 解耦分类任务和回归任务, 提高弱特征的解码能力, 增强小目标检测的性能. 在公开数据集COCO2017上进行实验,结果表明,SA-YOLO参数量仅有1.14 M,小目标平均检测召回率AR_S达到31.6 %. 同时,将所提出算法与近几年主流算法进行对比,结果表明,所提出算法在小目标检测方面具有较强的竞争力.

关键词: 小目标检测；背景噪声；特征融合；特征增强；轻量级网络

中图分类号: TP183 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2022.1432

引用格式: 周葳楠, 吴治海, 张正道, 等. 基于弱特征增强的轻量化小目标检测方法 [J]. 控制与决策, 2024, 39(2): 381-390.

Lightweight small target detection method based on weak feature enhancement

ZHOU Wei-nan[1,2], *WU Zhi-hai*[1,2†], *ZHANG Zheng-dao*[1,2], *PENG Li*[1,2], *XIE Lin-bo*[1,2]

(1. School of Internet of Things Engineering，Jiangnan University，Wuxi 214122，China；2. Engineering Research Center of Internet of Things Technology Applications of MOE，Jiangnan University，Wuxi 214122，China)

Abstract: Aiming at the problem of low detection accuracy caused by small target features in complex backgrounds, which are submerged by background noise after multiple convolutions, a one-stage lightweight small target detection algorithm SA-YOLO is proposed to enhance weak feature expression. First, the improved ShuffleNetv2 network is used to build the backbone network, and by embedding the SE attention module and the Inception structure, the feature extraction ability of the network in complex backgrounds is improved, background noise is effectively suppressed, and weak features are fully extracted. Second, in the neck network, a new feature fusion module is adopted to adjust the weights of high-level features based on the spatial location information of low-level feature blocks containing more weak features, so as to improve the utilization rate of feature fusion at different levels and reduce the feature loss of small targets. Finally, it replaces the original YOLO-coupled detection head with the decoupled detection head, decouples the i classification task and the regression task, improves the decoding ability of weak features, and enhances the performance of small target detection. Experiments are carried out on the public dataset COCO2017, and the results show that the parameter size of the SA-YOLO is only 1.14 M, and the average detection recall rate AR_S of small targets reaches 31.6%. At the same time, the proposed algorithm is compared with the mainstream algorithms in recent years. The results show that the proposed algorithm has strong competitiveness in small target detection.

Keywords: small object detection；background noise；feature fusion；feature enhancement；lightweight network

收稿日期: 2022-08-08；**录用日期:** 2023-02-25.
基金项目: 国家自然科学基金项目(61876073).
责任编委: 巩敦卫.
†通讯作者. E-mail: wuzhihai@jiangnan.edu.cn.

0 引 言

目标检测[1]是计算机视觉中的一项重要任务,其目的是通过建立数学模型来识别一幅图像中特定的实例(如行人、车辆、动物等). 现有目标检测的研究对象大部分集中在简单背景下大、中尺度的目标,对于复杂背景下[2]的小目标(MS-COCO度量评价[3]中对于面积小于或等于32×32像素的物体定义为小目标)检测的研究较少.

针对小目标检测,基于深度学习的目标检测算法已成为主流. 这类算法可以分为一阶段检测框架和二阶段检测框架,其中一阶段检测算法将检测看作是一个回归任务,兼顾了精度与速度,代表性算法有SSD[4-6]、YOLO[7-14]等. Du等提出了一种基于YOLOv2[8]扩展感受野模型ERF-YOLO[15],通过获得更多的特征信息来提升在小目标场景下的检测性能. Deng等提出了一种带有额外高分辨特征层的扩展特征金字塔(FPN[16]),设计了一种特征提取模块FTT来增强小目标的检测性能[17]. Lim等[18]提出一种通过拼接多尺度特征,将来自不同层的附加特征作为上下文的小目标检测方法,该方法可以提升小目标检测在低分辨率场景下的准确率. Zhu等提出了一种将YOLOv5[11]与Transform相结合的小目标检测模型TPH-YOLOv5[19],先增加了一个更高分辨率的检测头,后用Transform Prediction Head替换原本YOLOv5的检测头,将其变为具有自注意力机制的检测头,提升了检测微小目标的能力,能够在无人机低空飞行的视角下检测小目标. Yang等[20]为了减少高分辨率图的计算成本并保留其丰富的特征,提出了一种先使用低分辨率图来计算目标的粗位置,再经过稀疏引导在高分辨率特征图上计算精确结果的小目标检测算法QueryDet. Zhao等[21]提出一种全局与局部图像特征自适应融合的一阶段小目标检测算法SODet,通过AFS模块对Transform和CNN提取的全局与局部特征进行融合,最后在4个尺度上进行分类和回归. 与上面所展现的小目标检测算法相似,能够增强小目标检测性能最常见最有效的方法是提高输入图像的分辨率和增加一路分辨率更大检测头,然而这两种方法都会增加不小的计算量. 同时,由于小目标的像素较少,模型无法提取到足够的特征,容易被背景噪声影响,甚至在卷积神经网络的前向传播中会丢失全部的信息. 因此如何在保证检测实时性的情况下,尽可能地滤除背景噪声并获取更多的小目标特征是检测性能提升的关键.

针对以上问题,本文提出一种基于弱特征增强的轻量级小目标检测算法SA-YOLO. 主要工作为:

1) 提出一个基于ShuffleNetv2[22]的弱特征增强的骨干网络,该网络在具有轻量性的同时,通过加入SE[23]模块改变特征块各通道权重,将Inception[24]结构和通道混洗相结合,提取不同大小感受野的信息,加强骨干网络在复杂背景下的特征提取能力.

2) 针对小目标检测,设计一种结合通道注意力与空间注意力的特征融合模块来替代原始Concat,同时引入更底层的特征图来弥补前向传播中小目标的特征丢失.

3) 使用解耦的检测头代替原本的YOLO检测头,将回归任务和分类任务充分解耦,大幅度提升模型的收敛速度和小目标检测精度,并且在损失函数部分采用EIOU[25]优化小目标检测. 本文在公开的COCO2017数据集以及自建的工业现场下人体IFHB (industrial field human body)数据集上进行了充分的实验,验证了本文方法的有效性. 在输入分辨率为640的情况下,能够仅凭1.14 M的参数量达到31.6%的小目标平均召回率,同时,通过消融实验进一步验证了本文各个模块的合理性.

1 SA-YOLO目标检测算法

1.1 SA-YOLO网络整体结构

本文提出的小目标检测模型SA-YOLO主要在YOLOv5的基础上进行改进,YOLOv5通常由3部分组成:骨干网络由CSPDarknet53[10]组成,颈部网络由PANet[26]构成,头部网络是常规的YOLO检测头. YOLOv5能够输出3种尺度的预测图,对应大、中、小3种目标,这种多尺度的检测方法有效提高了目标检测的性能. 图1是SA-YOLO模型的整体结构图,可以看到改进的部分主要集中在骨干网络、颈部网络中的特征融合模块、检测头以及损失函数部分. 首先,用基于SAM (shuffle attention module)的骨干网络替换了原本的CSPDarknet53网络,该重构网络融入了SE注意力模块以及Inception结构,有效提高了骨干网络在复杂背景下的特征提取能力;然后,将融合了通道注意力与空间注意力的CAM (concat attention module)模块替代原本的Concat操作,这种新的特征融合方式可以调整不同尺度特征图的特征分布,有利于不同层级的特征融合表达;接着,引入底层特征层C2来弥补前向传播中小目标的特征丢失,再用解耦的YOLO检测头代替常规检测头,将回归任务和分类任务解耦,有效地提高了模型的收敛速度和精度;最后,在损失函数部分用EIOU替换原本的CIOU[27],有助于增强模型在复杂背景下小目标检测的性能. 以上

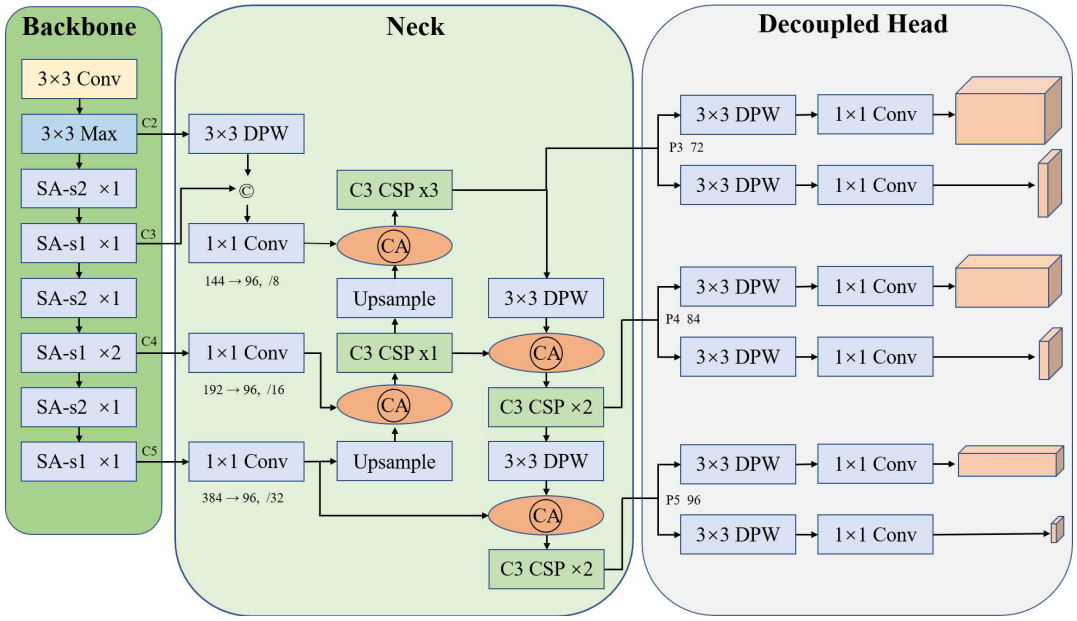


图 1　SA-YOLO模型整体结构可视化

改进均在保持模型轻量性的前提下进行,模型参数量保持在1 M左右,运算量保持在1.6 GFLOPs左右,能够部署到普通嵌入式设备上.

1.2　更强的骨干网络

实验对比发现,ShuffleNet比同等级的轻量级网络MobileNet[28]以及EfficientNet[29]在嵌入式移动设备上更健壮,因此本文选择ShuffleNetv2作为基础网络进行升级改进. 改进的ShuffleNetv2衍生出两种下采样步长的基础模块SAM,图2详细地描述了两类模块的内部结构. 步长为1的SAM模块延续了ShuffleNetv2的Channle Split和Channle Shuffle思想,Split操作将输入特征平均分成两部分,一部分进行深度可分离卷积计算,将计算结果与另一部分直接Concat连接,该操作在减少计算量的同时,还会保留部分浅层特征,尤其是可以减少小目标由于过度卷积造成的特征丢失. 最后在模块的尾部进行Shuffle操作,完成不同通道间的信息交换,但Shuffle之后会导致通道融合特征的丢失,因此在步长为2的模块中紧接着用1×1点卷积去整合不同通道的信息. 为了获取不同感受野大小的特征,增强模型的特征表达能力,还在步长为2的模块中融入了Inception结构,对应3×3和5×5两种规格的并行卷积核. 两个分支能够提取到同层次下的不同特征,并且避免了网络的过度碎片化,保证了模型的并行化以及运行速度. 需要注意的是,为了减少5×5卷积分支的计算量,在分支的第一个点卷积的时候,将通道缩放至输入的一半,3×3卷积分支的通道则保持与输入一致. 在对Inception的分支进行融合时,加入通道注意力模块SE,能够自适应调节所有特征块通道间的加权以增强模型的健壮性. SE模块中的激活函数依旧是经典的Relu和H-Sigmoid,其他深度可分离卷积中的激活函数是Silu.

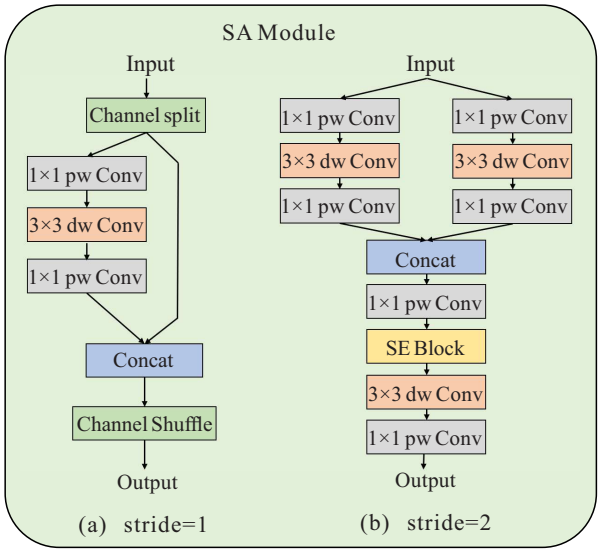


图 2　SAM模块结构可视化

1.3　颈部特征融合优化与底层特征引入

在目标检测算法中,输入图像从RGB三维数字信息经过一层一层的卷积,特征不断被编码,越来越抽象,浅层特征蕴含着丰富的空间位置信息,深层特征更多的是语义信息. 颈部网络往往采用特征金字塔来融合浅层特征和深层特征,从FPN开始,自顶向下将高层特征上采样和低层特征做Concat,向下传播语义信息来增强各层级的语义表达,便于头部网络的分类解码. FPN进一步演化至PAN,PAN是在FPN之后再增加一个自底向上的传播定位信息的通道,对FPN进行空间位置信息的补充,头部网络

能够对输入特征做到更充分的解码. 在高层级和低层级的特征融合中,两者的特征分布并不趋同,无法做到特征融合利用的最大化. 对此,本文提出一个结合通道注意力和空间注意力的CAM模块. 如图3所示,该模块有两个输入端 F_L 和 F_H,一个输出端 F_R,其中 F_L 表示低层特征块,F_H 表示高层特征块,且 $F_L, F_H \in \mathbb{R}^{C\times H\times W}$,$F_R$ 表示输出特征块,且 $F_R \in \mathbb{R}^{2C\times H\times W}$. 先是对 F_L 特征块进行垂直方向的压缩,得到空间注意力特征块 $F_S \in \mathbb{R}^{1\times H\times W}$,然后把经过卷积和激活函数的输出加1做到与残差连接类似的效果,最后再分别与两个层级的特征块在水平方向上点对点相乘来调整融合特征块的空间分布,做到融合统一,具体实现如下所示:

$$F_M = (M_S(F_L)\otimes F_L)©(M_S(F_L)\otimes F_H). \quad (1)$$

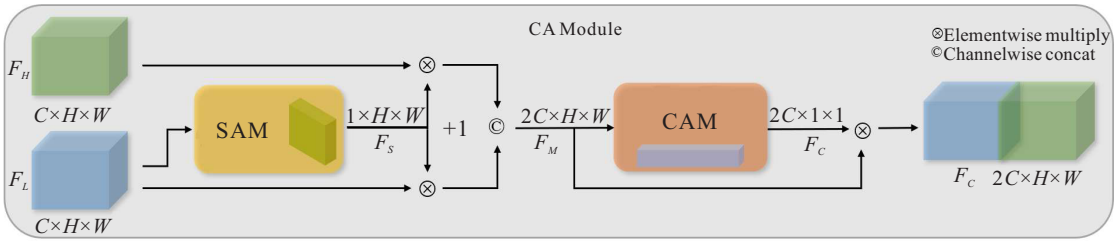


图 3 CAM模块结构可视化

对Concat之后的特征块,本文采取通道注意力的方式,将得到的 $F_M \in \mathbb{R}^{2C\times H\times W}$ 特征块在水平方向进行压缩,得到通道注意力输出的一维特征向量 $F_C \in \mathbb{R}^{2C\times 1\times 1}$,然后在垂直方向上将经过卷积和激活函数的输出与 F_M 来点对点相乘调整特征块各通道的权重,最后输出 F_R,即

$$F_R = M_C(F_M)\otimes F_M. \quad (2)$$

其中:$\otimes$ 表示点对点相乘,© 表示Concat相连操作.

空间注意力 M_S 和通道注意力 M_C 的表达式分别为

$$M_S(F) = \sigma(f^{3\times 3}(\text{AvgPool}(F)©\text{MaxPool}(F))), \quad (3)$$

$$M_C(F) = \sigma(\text{MLP}(\text{AvgPool}(F)+\text{MaxPool}(F))). \quad (4)$$

其中:σ 代表Sigmod激活函数,$f^{3\times 3}$ 表示卷积核大小为3×3的卷积操作,AvgPool和MaxPool分别表示全局平均池化和全局最大池化,MLP代表带有一个隐藏层的多层感知器.

小目标因为其所占像素较少,它的特征容易在多次卷积操作中被背景噪声淹没. 为了弥补这种小目标的特征丢失,本文引入图1中底层C2特征块,将其通过一个深度可分离卷积降维后与C3进行Concat,最后通过一个1×1卷积输入到颈部网络.

1.4 头部网络及损失函数的优化

在目标检测任务中,如图4所示,YOLOv5构建了一个端到端的模型,同时解决分类和回归问题,但也造成了两者的耦合. SA-YOLO则采用了分类和回归解耦的头部解码网络,对由颈部PAN输入的特征块经过两路深度可分离卷积处理,接着分别输入到分类和回归的检测头里,完成目标检测任务. 实验表明,解耦的检测头能增加模型的收敛速度和检测精度.

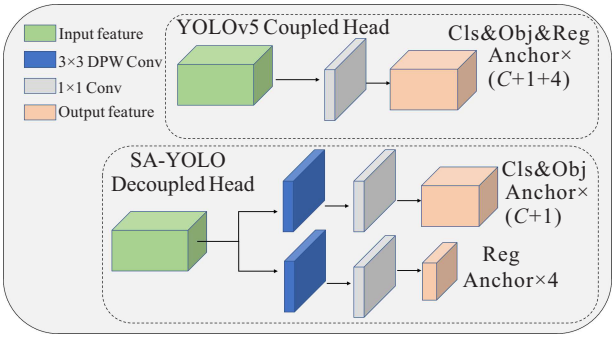


图 4 耦合头和解耦头的结构可视化

在损失函数部分,YOLOv5采用的是CIOU,该损失函数在DIOU的基础上加入了预测框和真值框的长宽比惩罚项,但长宽比描述的是相对值,不能反映出预测框与真值框的实际误差. 于是SA-YOLO引入EIOU来解决这个问题,EIOU的改进是将CIOU的长宽比惩罚项拆开,分别计算预测框和真值框的宽高损失,EIOU损失函数包含3个部分:交并比损失、中心点损失和宽高损失. 其中前两部分与CIOU一致,但是EIOU的宽高损失是直接计算目标框与真值框的实际误差,使得收敛速度更快. EIOU如下所示:

$$L_{\text{EIOU}} = L_{\text{IOU}} + L_{\text{dis}} + L_{\text{asp}} = 1 - \frac{B\bigcap B^{gt}}{B\bigcup B^{gt}} + \frac{\rho^2(b,b^{gt})}{c^2} + \frac{\rho^2(w,w^{gt})}{c_w^2} + \frac{\rho^2(h,h^{gt})}{c_h^2}. \quad (5)$$

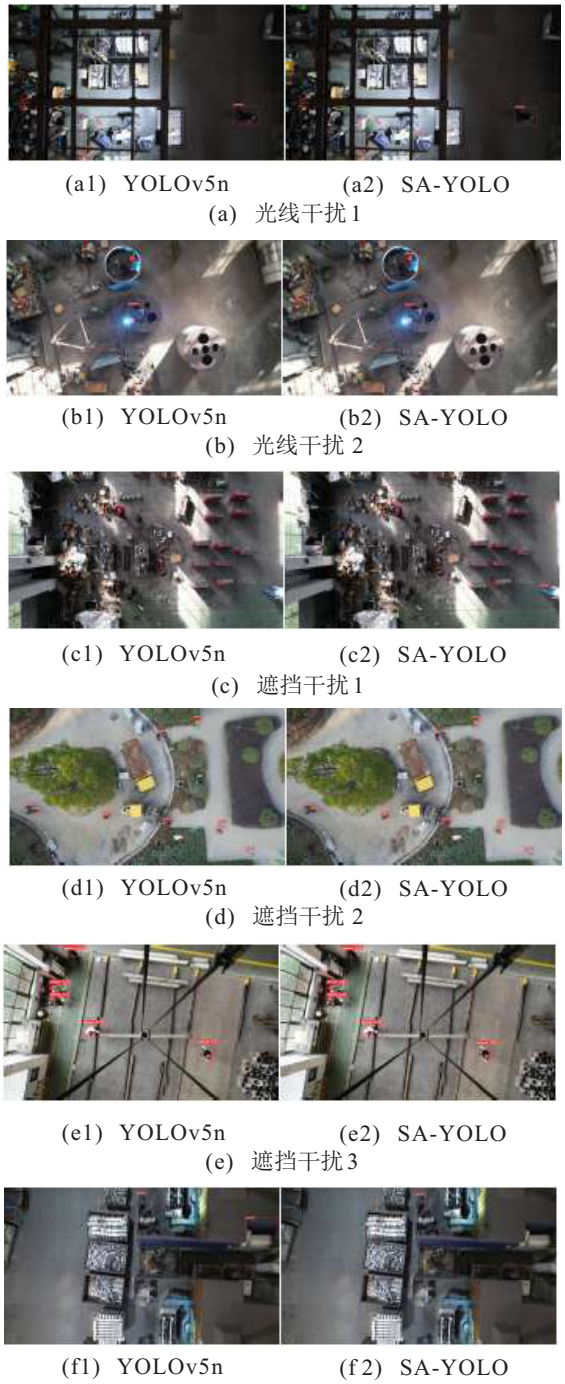
其中:$\rho(\cdot)$ 表示欧式距离;b, w, h 分别表示预测框的中心点和宽高;b^{gt}, w^{gt}, h^{gt} 分别表示真值框的中心点与宽高;c_w 与 c_h 分别表示预测框和真值宽合并后最

小外接矩形的宽与高.

2　实验结果与分析

2.1　实验平台和数据集介绍

本文所使用的CPU是Intel(R)Core(TM)i7-8700; GPU为NIVIDA GeForce GTX1070,显存8 G;操作系统为Ubuntu16.04. 本文算法在自建的数据集IFHB以及公共数据集COCO2017上进行实验验证,以表明SA-YOLO的有效性. IFHB数据集由无人机俯视拍照近百家工厂中的工人人体图片组成,总共6 000张图片,按常规比例7∶1∶2,将数据集划分成训练集、验证集和测试集. IFHB数据集中有近一半图片中工人未按照规定佩戴安全帽,还存在部分图片包含遮挡、光线变化或相似物干扰等复杂背景. 由于拍摄高度较高,导致图片中人体所占像素较少,人体符合小目标的定义,再结合俯视视角,人体形状特征进一步减少,给人体检测带来较大影响. 图5展现了IFHB数据集上复杂背景下的人体图片,图5(d)是室外工地施工图,其他是室内加工厂的图片. COCO2017是由微软发布的一个大型的、丰富的公开数据集,具有80个物体类别,提供超过33万张图片,其中包含众多场景下的小目标.



(a1) YOLOv5n　(a2) SA-YOLO
(a) 光线干扰1
(b1) YOLOv5n　(b2) SA-YOLO
(b) 光线干扰2
(c1) YOLOv5n　(c2) SA-YOLO
(c) 遮挡干扰1
(d1) YOLOv5n　(d2) SA-YOLO
(d) 遮挡干扰2
(e1) YOLOv5n　(e2) SA-YOLO
(e) 遮挡干扰3
(f1) YOLOv5n　(f2) SA-YOLO
(f) 相似干扰

图5　IFHB检测效果可视化对比

2.2　实验细节

输入图片分辨率为416×416,模型在COCO2017数据集上从零开始训练,训练重要超参数如下:初始学习率lr0=0.01,最后一轮学习率衰减比例lrf=0.01,训练batch size=128,训练总epochs=300,前3个epochs采用warmup,优化算法采用SGD,动量因子为0.937;在数据增强方面,翻转系数fliplr=0.5,mosaic为1.0. 在IFHB数据集上的训练依靠在COCO2017上的结果权重进行微调,超参数因为数据集较小,故关闭数据增强,同时减少学习率和训练次数.

2.3　算法性能评价指标

本文使用的评价指标为平均准确度均值(mean average precision, mAP)、运行时间T、小目标平均召回率均值AR_S、模型参数量(parameters, Params)以及浮点运算数(floating point operations, FLOPs). mAP由准确率Precision和召回率Recall计算得出,其计算公式为

$$AP=\int_0^1 P(R)\mathrm{d}R, \tag{6}$$

$$mAP=\frac{1}{C}\sum_{i=1}^{c}AP_i=\frac{1}{C}\sum_{i=1}^{c}\int_0^1 P(R)\mathrm{d}R. \tag{7}$$

以召回率R为横坐标,正确率P为纵坐标,对形成的PR曲线进行积分得到AP值,然后累加各类别AP值,最后除以总类别数C,得到mAP. Params为模型中的权重和偏置总参数量,FLOPs为模型计算量指标,衡量模型计算复杂度.

2.4　消融实验及分析

为了验证改进点的有效性,本部分在COCO2017数据集和IFHB数据集上进行了不同模型的实验验证. 表1详细地展现了加入各组件后的模型参数量、计算量、在入门级移动端CPU(Cortex-A53)和个人笔记本电脑的CPU(I5-6300HQ)上的实验性能,以及在COCO2017数据集和IFHB数据集上的评价指

标. 图5展现的是YOLOv5n和SA-YOLO在IFHB上检测的可视化效果. 可以看到,在图5(a)～图5(c)中,由于环境的复杂性,存在光线突变以及目标缺失的情况,YOLOv5n存在漏检;在图5(f)中,YOLOv5n存在误检,将环境中的干扰物识别成人体;在图5(d)和图5(e)中,YOLOv5n和SA-YOLO都成功识别出目标,但YOLOv5n的识别置信度普遍低于SA-YOLO的结果. 双方在面对图5(a)中严重被遮挡的目标时,都存在漏检的情况,原因在于小目标被遮挡后,可用特征更少,已无法支持模型进行正确识别. 实验对比结果表明,SA-YOLO在复杂背景的环境中比YOLOv5n更加鲁棒,能够检测出更多的目标.

表 1 SA-YOLO消融实验性能对比

Methods	Backbone	Params/M	FLOPs/G	T/ms		COCO2017			IFHB
				Cortex-A53	I5-6300HQ	mAP/%	AP_{50}/%	AR_S/%	AP_{50}/%
YOLOv5n	Darknet53	1.872	1.91	157	52	25.4	41.3	19.4	80.3
SA-YOLO	SAM	1.019(↓0.853)	1.496(↓0.414)	134	44	26.3(↑0.9)	42.4(↑**1.1**)	20.8(↑**1.4**)	81.5
SA-YOLO+ C2 Layer	SAM	1.02	1.529	135	44	26.6(↑0.3)	42.7(↑0.3)	21.7(↑0.9)	81.9
SA-YOLO+CAM	SAM	1.094	1.554	135	45	27.0(↑0.4)	43.1(↑0.4)	22.3(↑0.6)	82.4
SA-YOLO+Decoupled Head	SAM	1.143	1.649	141	48	28.1(↑**1.1**)	44.0(↑0.9)	23.2(↑0.9)	83.2
SA-YOLO+EIOU	SAM	1.143	1.649	141	48	28.0(↓0.1)	44.1(↑0.1)	23.4(↑0.2)	83.5

2.4.1 骨干网络消融实验分析

SA-YOLO的骨干网络是在ShuffleNetv2提出的4个构建高效轻量级网络标准的指导下,基于ShuffleNetv2原网络,融入Inception结构和SE注意力模块重新构建的一个高效的轻量级网络. 该骨干网络在步长为2时引入的3×3和5×5卷积能够提取到不同规格大小的特征,完全可以替代原骨干网络CSPDarknet53后的SPP层. 为了验证SAM的优越性,本文在输入分辨率为416的前提下,对原YOLOv5模型和替换了SAM骨干网络的模型进行了实验验证. 由表1的前2行可知,SA-YOLO在参数量下降45.57%、计算量下降21.68%的同时,运行时间在Cortex-A53和I5-6300HQ上分别减少了14.7%和15.4%,在COCO2017数据集上,mAP提升0.9%,AP_{50}提升1.1%,小目标指标AR_S提升1.4%,在IHFB数据集上AP_{50}提升1.2%. 由此表明了YOLOv5的网络具有一定的冗余,结构并不高效,SA-YOLO的参数对mAP指标的贡献率是原YOLOv5模型的1.9倍. 表2是YOLOv5和SA-YOLO的骨干网络各层具体的参数显示. 可以看出,SA-YOLO的骨干网络更加紧凑,比YOLOv5层数少2层,总参数量减少433 678个,只有原来的58.55%. 优秀的网络结构不仅带来的是精度的提升,还更加方便嵌入式移动端的部署,用户能够选择成本更低的板卡.

表 2 YOLOv5n和SA-YOLO骨干网络参数

Index	YOLOv5n			SA-YOLO Backbone		
	Layer	Output	Params	Layer	Output	Params
1	Conv	16	1 760	Conv	48	1392
2	Conv	32	4 672	MaxPool	48	0
3	C3	32	4 800	SA-s2	96	23 706
4	Conv	64	18 560	SA-s1	96	6 540
5	C3 ×2	64	29 184	SA-s2	192	89 460
6	Conv	128	73 984	SA-s1 ×2	192	49 200
7	C3 ×3	128	156 928	SA-s2	384	347 112
8	Conv	256	295 424	SA-s1	384	95 280
9	C3	256	296 448			
10	SPPF	256	164 608			
Total			1 046 368			612 690

2.4.2 底层特征引入与CAM消融实验分析

为了更进一步增强小目标的检测性能,SA-YOLO引入了底层特征C2来丰富小目标的特征表达. 由表1第3行可以看出,引入底层特征后,mAP增加了0.3%,尤其是小目标指标AR_S和AP_{50}在COCO2017和IFHB上分别增加了0.9%和0.4%,表明了该策略的有效性.

特征融合是目标检测中一个重要的环节,能够融合高低不同层级的特征信息,传统的Concat只是单纯地将不同层级的特征在通道层面相连,实验发现,高层和低层的特征分布并不趋同,给特征的融合带来了负面影响. 本文提出的融合注意力的CAM模块,能够在颈部网络的自顶向下和自底向上的特征融合过程中自适应地调整高低层级特征块的空间和通道权重. 表1的实验结果表明了CAM模块的有效性,经过注意力网络对特征块权重的调整后,模型对融合后的特征块利用率更高. 对小目标来说,低层有更多的有效特征,经过CAM特征融合后,可以更有效地保留小目标的特征. 由表1中第4行可知,mAP提升了0.4%,

小目标指标 AR_S 和 AP_{50} 在 COCO2017 和 IFHB 上分别提升了 0.6 % 和 0.5 %, 参数量只增加了 0.074 M, 计算量增加了 0.025 G, 这些参数的增加是完全值得的.

2.4.3　解耦检测头与损失函数改进消融实验分析

传统的 YOLO 检测头可以同时执行分类任务和回归任务, 但造成一定程度上的耦合, 分类和回归的解码不够彻底, SA-YOLO 专门引入了解耦的检测头, 一路专门负责分类任务, 另一路负责回归任务. 引入解耦头不可避免地增加一定的参数量和计算量, 本文采用深度可分离卷积代替普通卷积能够消除掉一些负面影响. 由表 1 可知, 更换解耦头增加了 0.049 M 参数和 0.095 G 计算量, 运行时间在 Cortex-A53 和 I5-6300HQ 上分别增加了 4.4 % 和 6.7 %, 同时 mAP 提升 1.1 %, 小目标检测指标 AR_S 和 AP_{50} 在 COCO2017 和 IFHB 上分别提升了 0.9 % 和 0.8 %. 图 6 是 SA-YOLO 搭配原 YOLO 耦合的检测头和解耦的检测头在 COCO2017 上训练了 300 个 epochs 的 mAP 和 AP_{50} 的可视化图, 其中蓝线是 Decouple Head, 红线是 Coupled Head, 从图 6 中可以看到解耦的检测头能够大幅度增加收敛速度和精度.

损失函数部分的 EIOU 的宽高损失更能表达出预测框与真值框的宽高之差, 对模型的梯度下降有促进作用. 实验结果如表 1 所示, 替换掉 EIOU 后, AP_{50} 和 AR_S 分别小幅增加 0.1 % 和 0.2 %. 虽然增加的精度有限, 但改换损失函数并不影响模型的整体结构, 在实际部署时不会增加任何的参数和计算量.

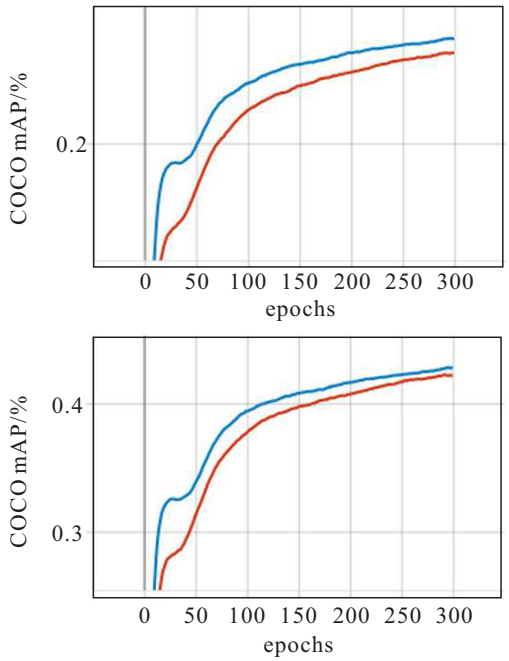


图 6　解耦头和耦合头 AP 值对比

2.5　对比实验及分析

为了更进一步验证算法的有效性, 本文还将 SA-YOLO 与近几年性能优异的轻量级目标检测算法在 COCO2017 数据集和 IFHB 数据集上进行了实验对比, 结果如表 3 所示. 本文提出的小目标检测模型 SA-YOLO 的小目标性能指标 AR_S 在同一分辨率下均优于其他检测算法, 且参数量和计算量进一步减少.

表 3　COCO2017 & IFHB 轻量级模型性能对比

Model	Size	Params/M	FLOPs/G	COCO2017		IFHB		year
				mAP/%	AP_{50}/%	AR_S/%	AP_{50}/%	
Model	416	8.86	5.62	16.6	33.1	–	–	2018
YOLOv4-Tiny[10]	416	6.06	6.96	22	42.1	19.1	78.4	2020
PP-YOLO-Tiny[30]	320	1.08	0.58	20.6	–	–	–	2020
PP-YOLO-Tiny[30]	416	1.08	1.02	22.7	–	–	–	2020
YOLOv5n(6.0)[11]	416	1.87	1.91	25.4	41.3	19.4	80.3	2021
YOLOv5n(6.0)[11]	640	1.87	4.5	28.4	46	29	–	2021
YOLOX-nano[12]	416	0.91	1.08	25.8	41.4	15.7	80.5	2021
YOLOX-tiny[12]	416	5.06	6.45	32.8	50.3	22.6	82.9	2021
PicoDet-S[13]	320	0.99	0.73	27.1	41.4	14.3	–	2021
PicoDet-S[13]	416	0.99	1.24	30.5	45.5	19.5	–	2021
PicoDet-M[13]	416	2.15	2.5	34.4	50.2	23.2	–	2021
YOLOv6-n[33]	416	4.3	4.7	30.8	47.2	21.1	82.4	2022
YOLOv6-n[33]	640	4.3	11.1	35	53	34.2	–	2022
YOLOv7-tiny-SiLU[14]	640	6.2	13.8	38.7	56.7	18.8	–	2022
SODet[21]	608	–	–	–	47.9	–	–	2022
SODet[21]	800	–	–	–	49.4	–	–	2022
SA-YOLO	320	1.14	0.98	24.5	39.8	18	–	2022
SA-YOLO	416	1.14	1.65	28	44.1	23.4	83.5	2022
SA-YOLO	640	1.14	3.9	31.9	50	31.6	–	2022

在COCO2017数据集上用320分辨率进行验证, PP-YOLO-Tiny[30] 的mAP只能达到20.6%, PicoDet-S[13] 的AR_S只能达到14.3%, 但SA-YOLO的mAP和AR_S可以分别达到24.5%和18.0%.

416分辨率是一个较合适的值, 在嵌入式平台上可以兼顾到计算量和精度. SA-YOLO由Pytorch[31] 框架转换至更适合移动端的框架ncnn[32] 后, 在CPU (Cortex-A53)上进行一次416分辨率的图片推理只需要141ms, 运行时间比YOLOv5n减少10%, 且精度可以大致保持一致. 在此分辨率下, 在COCO2017数据集上验证得出: SA-YOLO在mAP指标上比YOLOX-nano[12] 和YOLOv5n分别高2.2%和2.6%, 尤其是小目标指标AR_S, 分别高7.7%和4.0%. 在与PicoDet-S和YOLOv6-n[33] 的对比中, SA-YOLO在mAP指标上分别低2.5%和2.8%, 但在AR_S指标上分别高3.9%和2.3%. SA-YOLO还可以与PicoDet-M在AR_S指标上相媲美, 但SA-YOLO的参数量大小只有PicoDet-M的53.02%, 计算量只有其66%. 同时, SA-YOLO的平均准确度均值mAP能与YOLOv5n在输入分辨率为640时不相上下, 表明本文的改进是有效的, 能够在减少参数量和计算量的前提下, 做到比原YOLOv5更好的效果. 在IFHB数据集上, SA-YOLO在AP_{50}指标上比YOLOv6n、YOLOX-nano、YOLOv5n分别高1.1%、3.0%、3.2%.

在640的输入图像分辨率时, 在COCO2017数据集上, SA-YOLO算法的AR_S达到了31.6%, 而且仅只有3.90 G的计算量. 在与当前最新最优秀的目标检测算法PicoDet、YOLOv6和YOLOv7[14] 的比较中发现, 现有的研究在追求mAP指标提升时, 几乎都忽略了对小目标的性能提升. 在他们提出的轻量级目标检测模型下, 小目标检测的评价指标AR_S都偏低, 但只有小目标的检测性能有所提升, 才能成为一个成熟鲁棒的目标检测模型. 在与文献[20]的小目标检测算法SODet的比较中, SA-YOLO可以得到更优的精度和性价比. 以上结果表明: SA-YOLO算法具有较高的小目标检测精度, 同时具有较快的检测速度, 较好地权衡了精度和速度, 在一众轻量级目标检测算法中具有较强的竞争力.

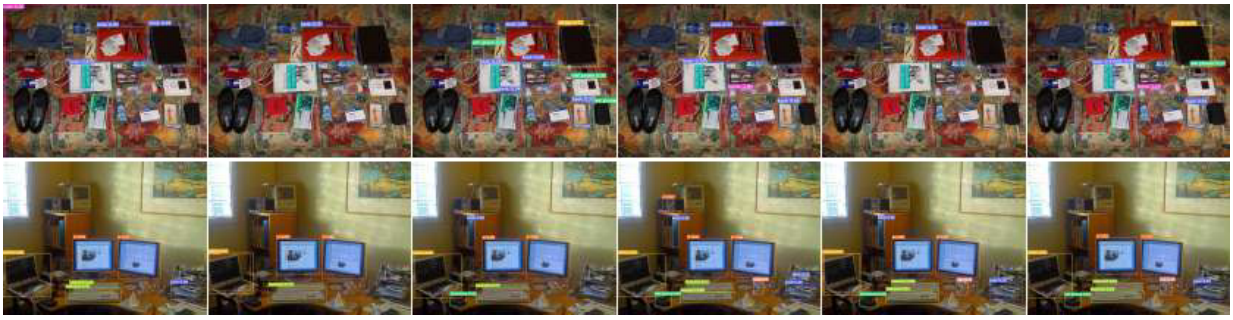
图7展现了目前比较优秀的轻量级目标检测模型在COCO2017数据集上的检测可视化效果, 可以明显看到, 目标检测的性能和输入分辨率呈现正相关, 对小目标更是如此. 图6(a)中YOLOv5n、YOLOv6n在640分辨率下还存在部分小目标未检测到的现象, 而SA-YOLO基本可都检测到. 图7(b)的背景属于开

(a1) YOLOv6n 416 (a2) YOLOv5n 416 (a3) SA-YOLO 416 (a4) YOLOv6n 640 (a5) YOLOv5n 640 (a6) SA-YOLO 640

(a) 遮挡干扰1

(b1) YOLOv6n 416 (b2) YOLOv5n 416 (b3) SA-YOLO 416 (b4) YOLOv6n 640 (b5) YOLOv5n 640 (b6) SA-YOLO 640

(b) 遮挡干扰2



(c1) YOLOv6n 416 (c2) YOLOv5n 416 (c3) SA-YOLO 416 (c4) YOLOv6n 640 (c5) YOLOv5n 640 (c6) SA-YOLO 640

(c) 相似干扰

图 7 COCO2017检测效果可视化对比

放简单的背景,但存在一些目标遮挡的干扰,YOLOv6n和SA-YOLO都能够检测出全部小目标,YOLOv5n还未能检测出遮挡严重的小目标. 图7(c)是常规的办公环境,背景比较杂乱,对目标检测存在一定的干扰,可以看到SA-YOLO依旧保持着较优的检测性能,能够在复杂背景下检测出更多的小目标.

3 结 论

本文针对复杂背景下小目标特征经多次卷积被背景噪声淹没导致的检测精度低的问题,提出了一种基于弱特征增强的轻量级小目标检测算法SA-YOLO. 首先,采用改进的ShuffleNetv2网络构建骨干网络,通过Inception结构下的两路并行卷积提取不同规格的特征和嵌入SE注意力模块,去动态调整两卷积分支合并后的特征块权重,提升网络在复杂背景下的特征提取能力,能够有效地抑制背景噪声,充分地提取小目标特征;其次,在颈部网络,采用新的特征融合模块,以低层级的空间位置信息对高层级特征进行权重调整,提高了不同层级的特征融合利用率,对特征进行更有效的深层次融合,增强网络的特征表达;然后,在头部网络,用解耦的检测头替换了原YOLO耦合的检测头,解耦了分类任务和回归任务,提高了模型的收敛速度和检测精度;最后,在损失函数部分,引入EIOU解决长宽损失定义不明确的问题,在不影响推理速度的前提下,小幅提高了算法的检测性能. 将SA-YOLO与近几年主流算法进行了充分的对比,结果表明本文算法具有较强的竞争力,较好地平衡了精度与速度.

参考文献(References)

[1] Masita K L, Hasan A, Shongwe T. Deep learning in object detection: a review[C]. 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD). Durban, 2020: 1-11.

[2] 王红梅, 王晓鸽, 王晓燕. 基于深度学习的复杂背景下目标检测[J]. 控制与决策, 2022, 37(12): 3115-3121. (Wang H M, Wang X G, Wang X Y. Target detection under complex background based on deep learning [J]. Control and Decision, 2022, 37(12): 3115-3121.)

[3] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context[C]. European Conference on Computer Vision. Cham: Springer, 2014: 740-755.

[4] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector[C]. European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[5] Lu X, Ji J, Xing Z, et al.Attention and feature fusion SSD for remote sensing object detection[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-9.

[6] Womg A, Shafiee M J, Li F, et al. Tiny SSD: A tiny single-shot detection deep convolutional neural network for real-time embedded object detection[C]. The 15th Conference on Computer and Robot Vision. Toronto, 2018: 95-101.

[7] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 779-788.

[8] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 6517-6525.

[9] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J/OL]. 2018, arXiv: 1804.02767.

[10] Wang C Y, Bochkovskiy A, Liao H Y M. Scaled-YOLOv4: Scaling cross stage partial network[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 13024-13033.

[11] Glenn-jocher. YOLOv5[EB/OL]. (2022-04-16)[2022-07-18]. https://github.com/ultralytics/yolov5.

[12] Ge Z, Liu S T, Wang F, et al. YOLOX: Exceeding YOLO series in 2021[J/OL]. 2021, arXiv: 2107.08430.

[13] Yu G H, Chang Q Y, Lv W Y, et al. PP-PicoDet: A better real-time object detector on mobile devices[J/OL]. 2021, arXiv: 2111.00902.

[14] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J/OL]. 2022, arXiv: 2207.02696.

[15] Du Z, Yin J, Yang J. Expanding receptive field yolo for small object detection[J]. Journal of Physics: Conference Series, 2019, 1314(1): 012202.

[16] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 936-944.

[17] Deng C F, Wang M M, Liu L, et al. Extended feature pyramid network for small object detection[J]. IEEE Transactions on Multimedia, 2022, 24: 1968-1979.

[18] Lim J S, Astrid M, Yoon H J, et al. Small object detection using context and attention[C]. 2021 International Conference on Artificial Intelligence in Information and Communication. Jeju Island, 2021: 181-186.

[19] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]. 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, 2021: 2778-2788.

[20] Yang C, Huang Z H, Wang N Y. QueryDet: Cascaded sparse query for accelerating high-resolution small object detection[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 13658-13667.

[21] 赵亮, 刘世鹏. 全局与局部图像特征自适应融合的小目标检测算法[J]. 控制与决策, 2023, 38(4): 935-943. (Zhao L, Liu S P. Small object detection algorithm based on adaptive fusion of global and local image features[J]. Control and Decision, 2023, 38(4): 935-943.)

[22] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design[C]. Proceedings of the 15th European Conference on Computer Vision. Munich, 2018: 122-138.

[23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 7132-7141.

[24] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 2818-2826.

[25] Peng H Y, Yu S Q. A systematic IoU-related method: Beyond simplified regression for better localization[J]. IEEE Transactions on Image Processing, 2021, 30: 5032-5044.

[26] Yang J F, Fu X Y, Hu Y W, et al. PanNet: A deep network architecture for pan-sharpening[C]. 2017 IEEE International Conference on Computer Vision. Venice, 2017: 1753-1761.

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

[28] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4510-4520.

[29] Tan M X, Le Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[J/OL]. 2019, arXiv: 1905.11946.

[30] Long X, Deng K P, Wang G Z, et al. PP-YOLO: An effective and efficient implementation of object detector[J/OL]. 2020, arXiv: 2007.12099.

[31] Paszke A, Gross S, Massa F, et al. PyTorch: An imperative style, high-performance deep learning library[J/OL]. 2019, arXiv: 1912.01703.

[32] Tencent.ncnn[EB/OL]. (2022-10-10)[2022-10-10]. https://github.com/Tencent/ncnn.

[33] Li C Y, Li L L, Jiang H L, et al. YOLOv6: A single-stage object detection framework for industrial applications[J/OL]. 2022, arXiv: 2209.02976.

作者简介

周葳楠(1999−), 男, 硕士生, 从事目标检测的研究, E-mail: 643567780@qq.com;

吴治海(1982−), 男, 副教授, 博士, 从事多智能体系统的实时、可靠与安全协同控制等研究, E-mail: wuzhihai@jiangnan.edu.cn;

张正道(1976−), 男, 副教授, 博士, 从事信息物理系统安全性、系统状态监测与故障诊断等研究, E-mail: wxzzd@jiangnan.edu.cn;

彭力(1967−), 男, 教授, 博士生导师, 从事物联网、图像处理等研究, E-mail: pengli@jiangnan.edu.cn;

谢林柏(1973−), 男, 教授, 博士生导师, 从事网格化控制系统、故障诊断与容错控制等研究, E-mail: xielinbo@jiangnan.edu.cn.