

面向无人机视角下小目标检测的YOLOv8s改进模型

潘 玮, 韦 超, 钱春雨, 杨 哲

苏州大学 计算机科学与技术学院, 江苏 苏州 215006

摘 要:从无人机视角进行目标检测, 面临图像目标小、分布密集、类别不均衡等难点, 且由于无人机的硬件条件限制了模型的规模, 导致模型的准确率偏低。提出一种融合多种注意力机制的YOLOv8s改进模型, 在骨干网络中引入感受野注意力卷积和CBAM(concentration-based attention module)注意力机制改进卷积模块, 解决注意力权重参数在感受野特征中共享问题的同时, 在通道和空间维度加上注意力权重, 增强特征提取能力; 通过引入大型可分离卷积注意力思想, 改造空间金字塔池化层, 增加不同层级特征间的信息交融; 优化颈部结构, 增加具有丰富小目标语义信息的特征层; 使用inner-IoU损失函数的思想改进MPDIoU(minimum point distance based IoU)函数, 以inner-MPDIoU代替原损失函数, 提升对困难样本的学习能力。实验结果表明, 改进后的YOLOv8s模型在VisDrone数据集上mAP、P、R分别提升了16.1%、9.3%、14.9%, 性能超过YOLOv8m, 可以有效应用于无人机平台上的目标检测任务。

关键词:无人机; 小目标检测; YOLOv8s; 感受野注意力; 大型可分离卷积

文献标志码:A **中图分类号:**TP391.4 **doi:**10.3778/j.issn.1002-8331.2312-0043

Improved YOLOv8s Model for Small Object Detection from Perspective of Drones

PAN Wei, WEI Chao, QIAN Chunyu, YANG Zhe

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Abstract: Facing with the problems of small and densely distributed image targets, uneven class distribution, and model size limitation of hardware conditions, object detection from the perspective of drones has less precise results. A new improved model based on YOLOv8s with multiple attention mechanisms is proposed. To solve the problem of shared attention weight parameters in receptive field features and enhance feature extraction ability, receptive field attention convolution and CBAM (concentration based attention module) attention mechanism are introduced into the backbone, adding attention weight in channel and spatial dimensions. By introducing large separable kernel attention into feature pyramid pooling layers, information fusion between different levels of features is increased. The feature layers with rich semantic information of small targets are added to improve the neck structure. The inner-IoU loss function is used to improve the MPDIoU (minimum point distance based IoU) function and the inner-MPDIoU instead of the original loss function is used to enhance the learning ability for difficult samples. The experimental results show that the improved YOLOv8s model has improved mAP, P, and R by 16.1%, 9.3%, and 14.9% respectively on the VisDrone dataset, surpassing YOLOv8m in performance and can be effectively applied to unmanned aerial vehicle visual detection tasks.

Key words: unmanned aerial vehicle (UAV); small object detection; YOLOv8s; receptive field attention; large separable kernel attention

在无人机(unmanned aerial vehicle, UAV)平台上, 通过机载相机捕捉的实时画面进行目标检测, 在紧急救援、边境巡逻、测绘评估、环境检测等领域有着广泛的应用。从无人机视角进行目标检测, 常常拥有复杂的环境背景, 一般以俯视角拍摄, 目标尺寸通常很小且相对模

糊, 还会有巨大的角度变化。而且由于无人机的硬件条件限制了模型的规模, 导致检测效果并不理想。

目前, 以深度学习为基础的目标检测算法可以分为两类, 一类是基于回归的单阶段(one-stage)算法, 如CenterNet^[1]、RetinaNet^[2]、SSD系列^[3]、YOLO系列^[4-6]等;

基金项目:教育部产学研合作协同育人项目(220606363154256); 国家级大学生创新创业训练计划项目(202210285042Z)。

作者简介:潘玮(2000—), 女, 硕士研究生, 研究方向为目标检测; 韦超(2003—), 男, 研究方向为图像处理、目标检测; 钱春雨(2003—), 男, 研究方向为计算机视觉; 杨哲(1978—), 通信作者, 男, 博士, 副教授, CCF会员, 研究方向为机器视觉、人工智能、大数据等, E-mail: yangzhe@suda.edu.cn。

收稿日期:2023-12-04 **修回日期:**2024-02-05 **文章编号:**1002-8331(2024)09-0142-09

另一类是基于区域建议的两阶段(two-stage)算法,如R-CNN^[7]、Fast R-CNN、Faster R-CNN^[8]等。两阶段算法检测速度慢,不适用于无人机视角的实时检测;单阶段算法实时性好,但是精度低,尤其对小目标的检测方面需要优化。

目前,性能较好的YOLO系列模型已广泛应用于其他领域的检测任务中,如何使用YOLO进行无人机视角下的目标检测,既保证目标检测的精度,又能保持较小的模型规模,是目前的研究难点。陈卫彪等人^[9]在YOLOv5基础上采用深度可分离多头网络结构,降低了模型参数数量和体积。陈范凯等人^[10]引入上采样算子CARAFE(content-aware reassembly of features),加强高、低分辨率特征图之间的融合,使小目标检测效果有所提升。刘涛等人^[11]根据无人机视角图像分布密集的特点,重新聚类锚框,并融合通道注意力机制构建SEC2f模块,实现更精准地检测。Li等人^[12]在YOLOv8中引入Bi-PAN-FPN的思想提高不同尺度间的特征融合,并用GhostblockV2结构代替部分卷积模块,得到了精度上的大幅提升,但在所有的小目标类别中无法获得比其他模型更好的结果。Lou等人^[13]提出一种新的下采样方法,并改造YOLOv8的特征融合网络,提升对密集小目标的检测能力,但总体提升效果不是很高。Guo等人^[14]提出一种密集连接机制,设计新的网络结构模块C3D替换YOLOv8中的C2f以更全面地保留信息,但是提升了计算量。Wang等人^[15]针对无人机图像的小目标特点,在YOLOv8的Neck中嵌入小目标检测结构STC(small target detection structure),充分捕获全局信息和上下文信息,并引入全局注意力GAM减少采样过程中特征信息的丢失,得到较高的性能提升,但是导致了大幅增加的参数数量。Wang等人^[16]融合了BiFormer注意力机制优化骨干网络,并提出FFNB(focal FasterNet block)特征处理模块,增加了两个新的检测尺度,实现五个尺度进行检测,大大降低漏检率,但对于自行车等很小物体的检测精度仍然不高。

尽管上述方法在一定程度上提升了无人机目标检测的性能,但由于无人机视角图像数据集的复杂性与特殊性、无人机硬件平台的限制性,难以平衡检测性能和硬件资源消耗的关系。本文以YOLOv8s模型为基础,提出一种融合了多种注意力机制的改进模型,在无人机硬件资源受限的情况下,提升模型的检测性能。首先,引入由RFACnv改进后的RFCBAM模块替换骨干中的卷积,并改造主干中的C2f模块加强特征提取能力;其次,使用大型可分离卷积注意力改造金字塔池化层,并增加一个小目标检测层,加强特征融合能力;最后,使用inner-IoU改造后的inner-MPDIoU损失函数改善检测边框回归效果。

1 YOLOv8 模型

YOLOv8是ultralytics公司在2023年1月发布的最新YOLO模型,从小到大有v8n、v8s、v8m、v8l、v8x五个版本。随着模型大小的提高,模型精度不断提升,可以根据任务需要选择不同深度和宽度的网络模型进行目标检测、图像分类、实例分割、关键点检测任务。目前的YOLOv8模型仍在不断更新中,本文采用YOLOv8.0.202版本。

基于无人机设备的硬件限制,本文选取体积小、精度高的YOLOv8s模型,它主要由骨干网络(Backbone)、颈部(Neck)、检测头(Head)三部分组成。

Backbone部分主要进行特征提取,采用了Darknet-53框架,引入了全新的C2f(CSPLayer_2Conv)模块进行残差学习,借鉴了CSP和ELAN,采用更多的跳层连接和额外的Split操作,将梯度变化从头到尾集成到特征图中。Conv卷积模块和C2f模块串行叠加4次,每叠加一次称为一个stage。最后采用YOLOv5等架构采用的SPPF(spatial pyramid pooling fusion)模块,固定不同尺度特征图的向量大小。

Neck部分主要进行特征融合,同样使用C2f模块替换C3模块,采用PAN(path aggregation networks)-FPN(feature pyramid networks)的思想,实现自上向下和自下向上的特征金字塔。同时,YOLOv8去除了YOLOv5、YOLOv6中上采样前的 1×1 卷积,直接对Backbone不同阶段输出的特征进行上采样操作。

Head部分变化较大,由YOLOv5的anchor-based耦合头变成anchor-free的解耦头,不再有objectness分支,仅有解耦的分类分支和使用DFL(distribution focal loss)的回归分支。

作为最后的预测部分,Head部分包含了三个拥有不同尺寸特征图的检测头来对不同大小的目标物体进行检测输出。

2 多注意力机制融合的YOLOv8s改进模型

面对低空无人机数据集中目标所占像素少、分布密集、类别不均衡的问题,原始的YOLOv8s存在漏检误检、精度低的问题。针对低空无人机数据集特点,做出以下改进:

(1)使用RFACnv改造CBAM得到的RFCBAM模块替换骨干中的卷积,以及设计全新的模块C2f RFCBAM替换骨干中的C2f,增强网络骨干的特征提取能力。

(2)将LSKA(large separable kernel attention)融合到SPPF(spatial pyramid pooling fusion)中,提升不同特征层间的语义融合。

(3)根据无人机视角图像的小目标特点,改善颈部结构,增加具有丰富小目标语义信息的特征层,四个检测头辅助进行目标检测任务。

(4)用 inner-IoU 损失函数的思想改造 MPDIoU,以 inner-MPDIoU 作为模型边框回归的损失函数,提升对困难小样本的学习能力。改进后的模型网络结构如图1所示。

2.1 改进 C2f 模块

感受野注意力卷积(receptive-field attention convolutional operation, RFACnv)^[17]是在关注空间注意力特征的基础上,将注意力引向感受野空间特征,解决卷积核参数共享的问题。RFACnv 以参数和计算工作量的小幅增加为代价就可以替代标准卷积,并带来显著性能提升。在 CBAM^[18](concentration-based attention module)注意力机制中,已经从空间和通道两个维度赋予卷积注意力权重,将 CBAM 中的空间注意力机制的注意力引到感受野空间特征中,消除卷积参数共享的问题,得到 RFCBAM(receptive-field concentration-based attention module),以类似自注意力机制的方式获得长程信息依赖,提升卷积性能。

改进后的 RFCBAM 模块和原始 CBAM 模块的对比如图2所示,不再是单独执行通道和空间注意力,而是两者同时进行加权,并且为了减少计算开销,用 SE(squeeze-and-excitation)注意力替代 CBAM 中的通道空间注意力,去除了平均池化的分支,用分组卷积提取感受野空间特征,减少特征重叠。

用 RFCBAM 改造主干中的 C2f 模块的 Bottleneck,可以获得更好的特征信息,具体结构如图3所示。RFCBAM 改造的 C2f 模块主要用 RFCBAM_neck 替换原模块中的 Bottleneck。在 RFCBAM_neck 中,依然采用两个卷积模块,同样先经过一个 Conv 层,将第二个 Conv 使用 RFCBAM 替换。因为 RFCBAM 中就进行了残差信息连接的信息融合,在 RFCBAM_neck 中不再进行残差连接。YOLOv8s 模型的 Backbone 主要进行特征提取,将原始模型 Backbone 中的普通卷积 Conv 和 C2f 卷积全用 RFCBAM、C2f_RFCBAMF 替换,在原提取机制上既增加了感受野注意力改进后的空间特征,又增加了通道注意力,从空间和通道两个维度增强模型的特征提取能力。

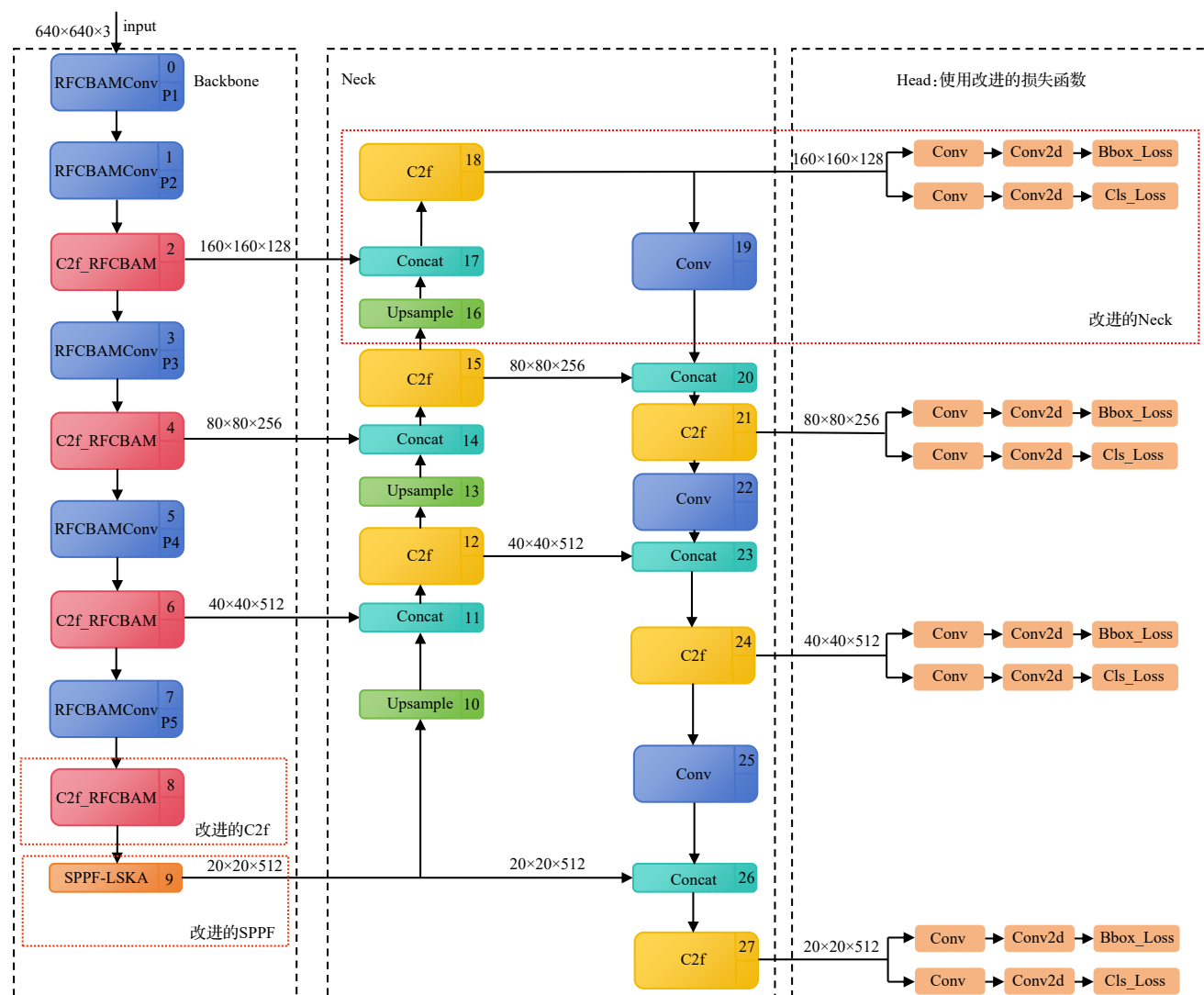


图1 改进的YOLOv8s模型结构

Fig.1 Structure of improved YOLOv8s network

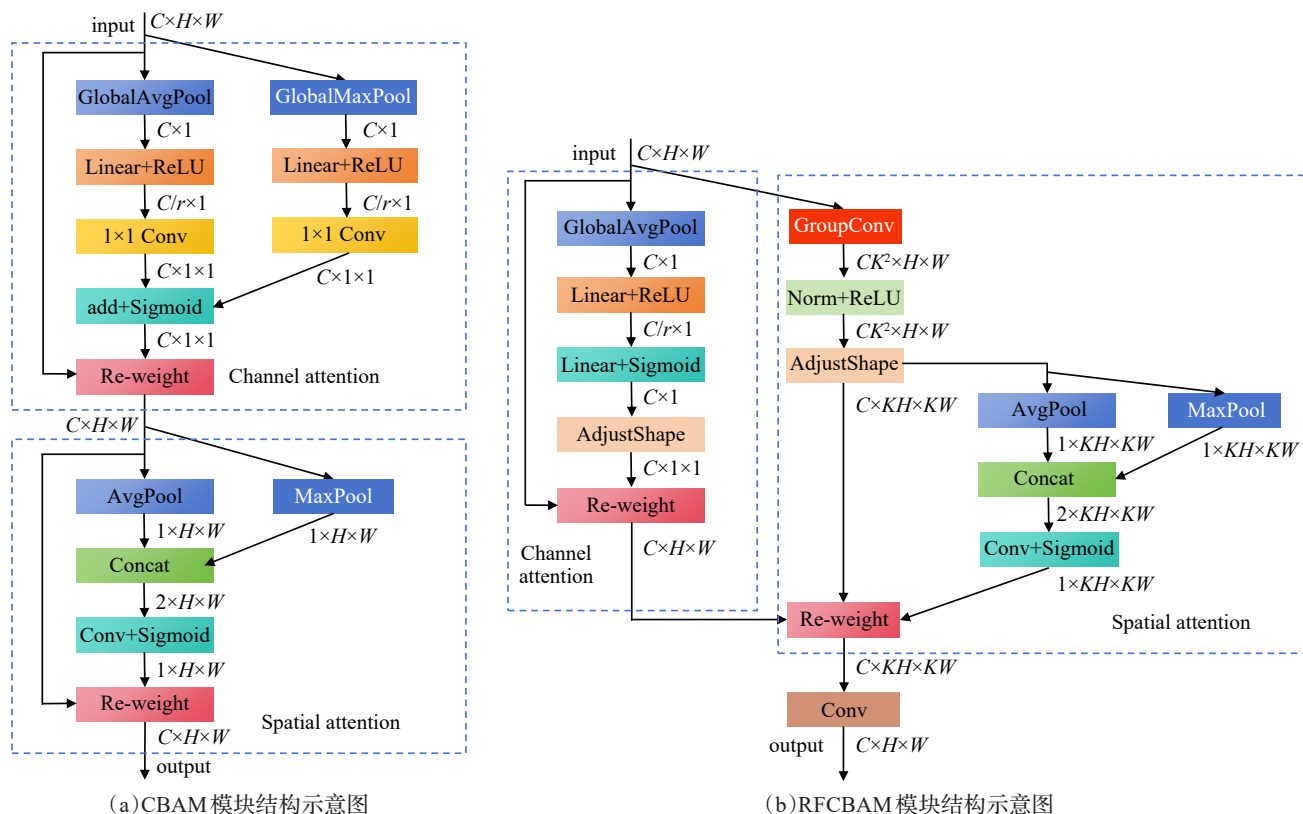


图2 CBAM 与 RFCBAM 的结构对比图

Fig.2 Structure comparison diagram of CBAM and RFCBAM

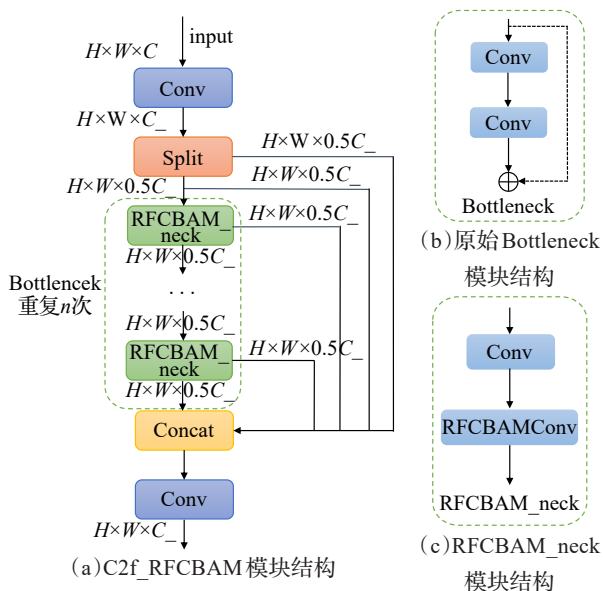


图3 C2f_RFCBAM 模块与子模块示意图

Fig.3 Structure of C2f_RFCBAM and sub modules

2.2 改进的 SPPF 模块

SPPF 将输入串行通过三个 5×5 大小的最大池化层,每层的输出都拼接起来,得到多尺度的特征。LSKA^[19] 中将 $k \times k$ 卷积核平均分为 $1 \times k$ 和 $k \times 1$ 的可分离卷积核,以级联方式处理输入特征,有效减少大卷积核带来的参数量大幅增长。

使用 LSKA 改进 SPPF 如图 4 所示,将原始 SPPF 模

块中经过多个池化层拼接后的结果先输入 11×11 的 LSKA 卷积模块,通过大型可分离卷积注意力获得长距离依赖,以更大感受野进行特征提取,然后再通过普通卷积进行特征融合,调整模型 Backbone 部分最终输出的特征向量大小。融合了 LSKA 注意力机制的 SPPF 层并没有增加很多参数,却在多尺度特征中得到了更加丰富的感受野信息,有助于模型进行更好地特征融合。

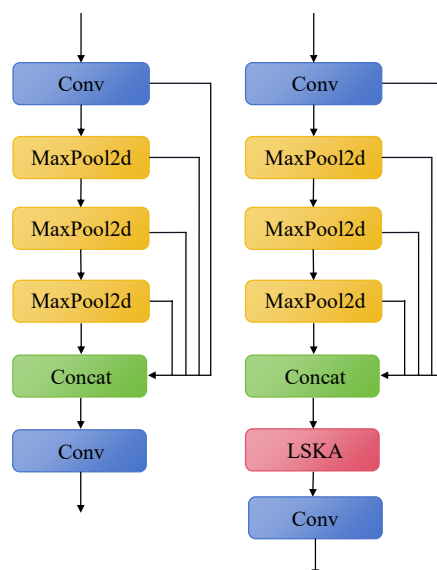


图4 SPPF 和 SPPF-LSKA 模块示意图

Fig.4 Structure of SPPF and SPPF-LSKA

2.3 改进的 Neck

在无人机目标检测任务中,成像距离远、速度快、低空飞行等特点会产生运动模糊,导致目标物体较小,难以捕捉特征,影响网络的训练效率。原始的YOLOv8模型有 80×80 、 40×40 、 20×20 三种保留不同尺度特征层信息的检测头,但模型本身经过了高倍数下采样后,小目标的特征已经不够明显,甚至融入背景中。

因此改造neck增加一个具有丰富小目标语义信息的特征层SODH(small object detection head)。SODH通过与backbone中 160×160 尺度的特征向量进行残差连接,在当前融合全局信息的较深语义中加上浅层的特征信息,以 160×160 的大尺度可以更好地保留小目标的相关特征信息,扩大模型分辨率,增强对小目标的感知能力,四个检测头共同辅助模型做出最后的检测任务。

2.4 改进的损失函数

边框回归的目的是通过对检测器输出的检测窗口进行微调去接近真实检测窗口。IoU(intersection over union)自提出后就成为了检测领域预测框损失评价的主流标准,其公式如式(1)所示:

$$IoU = \frac{|B^{\text{pred}} \cap B^{\text{gt}}|}{|B^{\text{pred}} \cup B^{\text{gt}}|} \quad (1)$$

基于IoU的边框回归损失函数不断迭代发展,如GIoU、DIoU、CIoU、EIoU、SIoU等,YOLOv8s中采用了CIoU,计算公式如式(2)所示:

$$CIoU = IoU - \left(\frac{\rho^2(B^{\text{pred}}, B^{\text{gt}})}{c^2} + \alpha v \right) \quad (2)$$

其中, $v = \frac{4}{\pi} (\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w^{\text{pred}}}{h^{\text{pred}}})$, 用来衡量长宽比, $\alpha = \frac{v}{(1-IoU)+v}$, α 作为平衡参数, $\rho^2(B^{\text{pred}}, B^{\text{gt}})$ 表示预测框与真实框中心点之间的欧氏距离, CIoU在IoU基础上加上对中心点距离和长宽比的权衡,检测效果得到有效提升。

这些改进的损失函数仍一直是通过加入新损失项来加速收敛,并没有意识到IoU自身的限制。inner-IoU^[20]提出以辅助边框来计算IoU,提升泛化能力,具体计算过程如式(3)~(7)所示,用尺度因子比ratio来控制辅助边界框的大小。

$$b_l = x_c - \frac{w \times ratio}{2}, b_r = x_c + \frac{w \times ratio}{2} \quad (3)$$

$$b_t = y_c - \frac{h \times ratio}{2}, b_b = y_c + \frac{h \times ratio}{2} \quad (4)$$

通过式(3)、(4)可以对检测框的中心点进行一个变换得到辅助检测框的角顶点。对模型输出的预测框和真实框都做相应变换,用 b^{gt} 、 b^{pred} 分别表示真实框、预测框的计算结果。

$$inter = (\min(b_r^{\text{gt}}, b_r) - \max(b_l^{\text{gt}}, b_l)) \times (\min(b_b^{\text{gt}}, b_b) - \max(b_t^{\text{gt}}, b_t)) \quad (5)$$

$$union = (w^{\text{gt}} \times h^{\text{gt}}) \times (ratio)^2 + (w \times h) \times (ratio)^2 - inter \quad (6)$$

$$IoU^{\text{inner}} = \frac{inter}{union} \quad (7)$$

所以,如式(5)-(7)所示,inner-IoU实际上计算的是辅助边框之间的IoU。 $ratio \in [0.5, 1.5]$, $ratio < 1$ 时,辅助边框比实际边框小,回归的有效范围小于IoU损失,但梯度绝对值比IoU损失所得的梯度大,可以加速高IoU样本的收敛。 $ratio > 1$ 时,辅助边框比实际框大,扩大回归的有效范围,有益于低IoU样本的回归。

MPDIoU^[21]是一种直接最小化预测框和真实框对应左上角、右下角间点距离的改进算法,对于边界框重叠与不重叠的情况都能很好处理,提高收敛速度。

$$MPDIoU = IoU - \frac{\rho^2(P_1^{\text{pred}}, P_1^{\text{gt}})}{w^2 + h^2} - \frac{\rho^2(P_2^{\text{pred}}, P_2^{\text{gt}})}{w^2 + h^2} \quad (8)$$

其中, P_1^{pred} 、 P_2^{pred} 、 P_1^{gt} 、 P_2^{gt} 分别指预测框与真实框的左上角、右下角的点, $\rho^2(P_1^{\text{pred}}, P_1^{\text{gt}})$ 计算相应点之间的距离。使用inner-IoU的思想改造MPDIoU,替换其中的IoU计算部分可以很好提升检测效果。

$$MPDIoU^{\text{inner}} = IoU^{\text{inner}} - \frac{\rho^2(P_1^{\text{pred}}, P_1^{\text{gt}})}{w^2 + h^2} - \frac{\rho^2(P_2^{\text{pred}}, P_2^{\text{gt}})}{w^2 + h^2} \quad (9)$$

3 实验设计与结果分析

3.1 数据集与实验环境

实验使用由天津大学机器学习与数据挖掘实验室所整理公开的VisDrone2019数据集^[22],数据集共包含10类航拍检测目标,包括行人、人、自行车、汽车、面包车、卡车、三轮车、遮阳三轮车、公共汽车、摩托车,训练集、验证集、测试集分别由6471张、548张、1610张图片组成。数据集中,像素面积小于 16×16 像素的极小目标占比为12.05%,像素面积小于 32×32 的小目标占比为44.7%,人和行人的类别中小目标占比更是分别高达77.45%、64.59%。训练集中共有353550个目标,其中被部分遮挡的目标个数为142873,被严重遮挡的目标个数为33804,近50%的目标都被遮挡^[23]。

为了验证改进有效性,还使用了SHWD(safety helmet (hardhat) wearing detect dataset)数据集和CARPK数据集。SHWD数据集有戴头盔和不戴头盔的头两个类别,由5944张训练集和1637张验证集组成;CARPK数据集^[24]是无人机在40m低空拍摄的单类别汽车检测数据集,包含从不同停车场拍摄的近90000辆汽车,由989张训练集和459张测试集组成。

实验使用的GPU为NVIDIA A100-PCIE-40GB,使用了6核CPU,pytorch框架版本为2.0.0+cu117,python版本为3.9.0,cuda版本为11.7。具体的实验环境参数设置如表1所示。

表1 训练参数设置

Table 1 Traning parameters setting

参数	设置	参数	设置
epochs	300	close_mosaic	10
patience	50	warmupepochs	3.0
batch	16	lrf	0.01
imgsz	640	lr0	0.01
workers	4	momentum	0.937
optimizer	SGD	weight_decay	0.000 5

实验中所有模型的输入图像尺寸都为640×640,在同样的实验环境下使用相同训练参数都训练300 epochs,当模型近50轮训练中并没有得到性能提升时,将触发早停机制提前结束训练。所有的YOLO模型在两个数据集上均采用表1中的设置进行训练。

3.2 消融实验

对YOLOv8s模型进行改进,并对每个改进的结果进行统计分析,结果如表2所示。其中,A、B、C、D分别表示在YOLOv8s模型的基础上各自加上LSKA、RFCBAM、SODH、inner-MPDIoU改进后的实验结果,可以看出A、B在仅增加少量计算量的情况下就能够将注意力机制融入原始模型有效提升各项指标,mAP@0.5分别提升0.016、0.018;C虽然增加了较大计算量,但是模型整体大小下降0.8 MB,性能提升明显,P、R、mAP@0.5分别提升了0.028、0.028、0.033,有助于在实际无人机硬件上的部署;D替换检测框回归的损失函数,在不改变模型大小和计算量的前提下,mAP@0.5也可以有效提升0.005。E、F、G将不同的改进方法分别加入模型中,都能够原改进基础上再次得到优化,将这四个改进同时加入模型的实验结果H,虽然P比G下降了0.003,但R和mAP@0.5分别提升了0.13和0.08,整体上替换损失函数还是提升了模型性能。与原始模型相比,H模型大小仅增加1.7 MB,但P、R、mAP@0.5分别提升了0.047、0.051、0.063,可以更好实现无人机视角图像的目标检测任务。

对于inner-MPDIoU损失函数中的ratio,取不同值的消融实验结果如表3所示。

表3 ratio 消融实验

Table 3 Ablation experiments of ratio

ratio	P	R	mAP@0.5	mAP@0.5:0.95
0.50	0.554	0.423	0.449	0.272
0.70	0.542	0.424	0.441	0.268
1.00	0.562	0.422	0.450	0.274
1.15	0.551	0.433	0.450	0.275
1.25	0.544	0.437	0.452	0.276
1.40	0.552	0.438	0.455	0.276
1.50	0.543	0.431	0.447	0.273

当ratio=1时,inner-MPDIoU损失函数实质就是MPDIoU损失函数,实验结果表明,对于无人机视角的图像都是小目标,标注框略有偏移,IoU就较低,当ratio>1时,辅助边框比实际框大,有益于低IoU的样本回归。所以在ratio>1时的实验效果基本优于ratio≤1时。当ratio=1.4时,实验整体取得一个最佳的结果,但在ratio>1时,效果也并不是ratio越大越好,在ratio=1.25和ratio=1.5时,实验效果有所波动,ratio的具体取值需要针对实验数据集进行调整。

3.3 对比实验

3.3.1 改进效果对比实验

为了更直观地感受模型的改进效果,将原模型和改进的模型放在同样的实验平台进行测试,效果对比如图5所示。

可以看出改进模型和原模型收敛的速度差不多,但是改进模型整体性能更优,mAP曲线更为平滑,且触发了YOLOv8s的早停机制,近50轮没有优化就提前结束了模型训练,节省计算资源。

为了验证改进模型的检测性能,将改进模型与YOLOv3s、YOLOv5s、YOLOv6s等具有代表性的网络以及其他针对无人机视角的YOLO改进模型进行对比,实验结果如表4所示。YOLOv3s总体上实现了最佳的性能,但是模型参数量和计算量太大,不太适用于无人机的硬件设备,且对小目标的检测结果AP,效果不如改进模型,而无人机视角下的图像大多为小目标。更轻量的模型YOLOv5s参数小,但是在检测性能各项指标上不如其他模型,满足不了无人机视觉检测任务对性能的要求。

表2 消融实验

Table 2 Ablation experiments

方法	LSKA	RFCBAM	SODH	inner_MPDIoU	模型大小/MB	GFLOPs	P	R	mAP@0.5
YOLOv8s	×	×	×	×	22.5	28.5	0.505	0.381	0.392
A	√	×	×	×	24.7	29.3	0.521	0.398	0.408
B	×	√	×	×	23.2	29.8	0.538	0.395	0.410
C	×	×	√	×	21.7	36.7	0.533	0.409	0.425
D	×	×	×	√	22.5	28.5	0.501	0.387	0.397
E	√	√	×	×	25.1	30.3	0.529	0.392	0.408
F	√	×	√	×	23.9	37.5	0.533	0.425	0.441
G	√	√	√	×	24.2	38.5	0.555	0.425	0.447
H	√	√	√	√	24.2	38.5	0.552	0.438	0.455

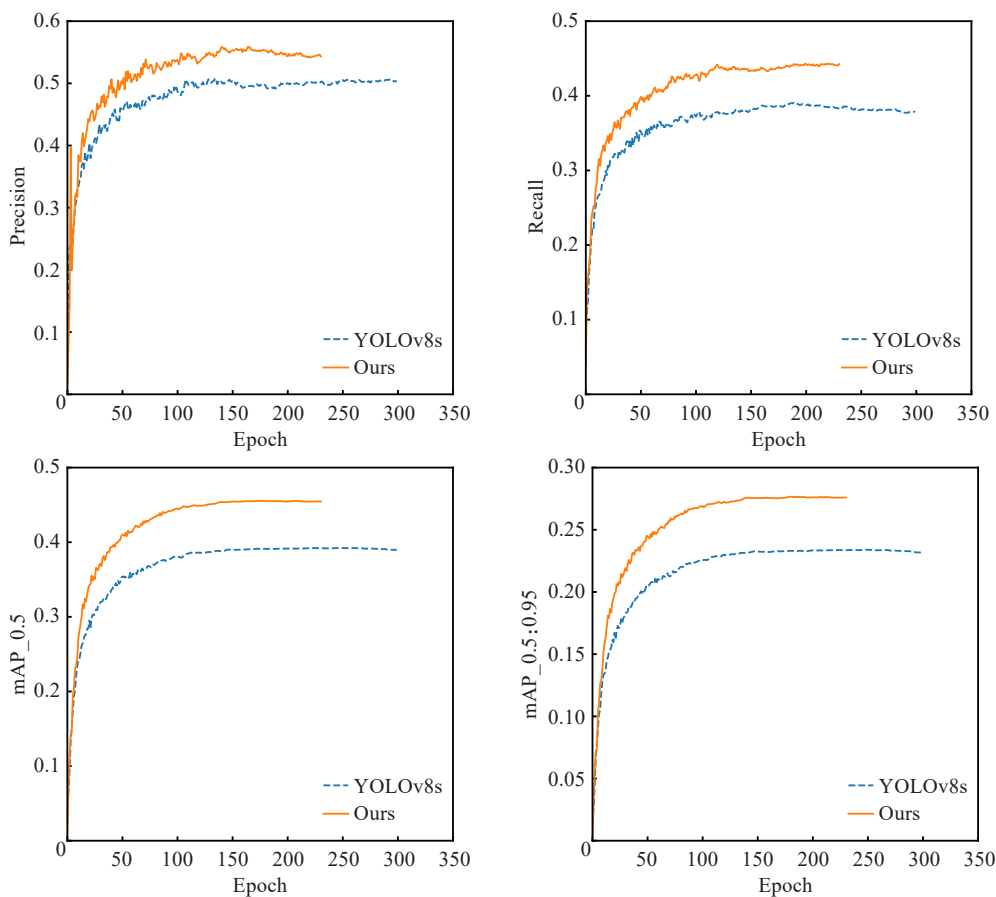


图5 改进效果对比

Fig.5 Comparison of improvement effects

表4 算法对比实验

Table 4 Algorithm comparative experiments

方法	模型大小/MB	GFLOPs	P	R	AP _s	AP _m	AP _l	mAP	FPS
YOLOv5s	5.3	7.2	0.446	0.326	0.093	0.275	0.379	0.331	109.7
YOLOv8s	22.5	28.5	0.505	0.381	0.129	0.330	0.381	0.392	100.3
YOLOv6s	32.9	44.0	0.487	0.369	0.112	0.324	0.407	0.377	150.2
YOLOv8m	52.1	78.7	0.532	0.420	0.154	0.360	0.425	0.427	82.4
YOLOv8l	87.6	165.4	0.545	0.437	0.160	0.378	0.458	0.441	63.0
YOLOv3s	207.8	282.3	0.580	0.438	0.163	0.395	0.457	0.453	98.8
DSM-YOLOv5 ^[9]	11.0	—	0.510	0.370	—	—	—	0.368	—
DC-YOLOv8 ^[13]	—	—	0.518	0.394	—	—	—	0.415	—
Ours	24.2	38.5	0.552	0.438	0.173	0.366	0.410	0.455	60.2

本文的改进模型在YOLOv8s的基础上仅增加1.8 MB的模型大小,mAP、P、R分别提升了16.1%、9.3%、14.9%,特别是对于像素面积小于 32×32 的小目标检测的性能比原模型提升了0.044,成为所有对比模型中的最佳。改进后的模型大小和计算量比YOLOv8m的一半还少,但除了对大目标的检测略低于YOLOv8m,其余性能全面超过YOLOv8m,mAP提升了0.028。即使和参数量是本文改进模型400%的YOLOv8l相比,也提升了P、R和对小目标的检测精度AP_s,在无人机视角以小目标为主的情况下,mAP总体仍可以提升3.2%。虽然对比之下,改进模型的FPS有所下降,但仍在60之上,可以较好适用于无人机的实时场景检测。目前小型无人机基本都满足30 FPS的拍摄,追求更流畅画质的

升级款可以实现50 FPS甚至60 FPS,但高帧数的画质会大大占用存储内存,非商业需要的拍摄基本采用50帧以下。针对无人机平台硬件空间小、算力低的情况,本文的改进模型不仅模型占用硬件空间小,满足实时画面的帧率,而且提升了检测性能,可以更好实现检测任务。

3.3.2 通用性对比实验

尽管本文实验围绕YOLOv8s,其模型大小和性能更适配于无人机的硬件条件,但为了验证模型改进的有效性,不仅适用于YOLOv8s,对其他规模的YOLOv8模型做实验验证如表5所示。实验证明,在YOLOv8n、YOLOv8s、YOLOv8m上应用本文改进方法,各方面指标都能有效提升。但考虑到无人机的实际硬件情况,仍应以YOLOv8s为宜。

表5 YOLO 验证实验

Table 5 Validation experiments on YOLO

方法	P	R	mAP@0.5	mAP@0.5:0.95
YOLOv8n	0.448	0.327	0.335	0.195
YOLOv8n-ours	0.471	0.373	0.377	0.225
YOLOv8s	0.505	0.381	0.392	0.234
YOLOv8s-ours	0.547	0.441	0.457	0.277
YOLOv8m	0.532	0.420	0.427	0.260
YOLOv8m-ours	0.573	0.466	0.490	0.303

在CARPK数据集上,用改进方法分别在YOLOv8n、YOLOv8s、YOLOv8m上进行实验,验证改进对无人机视角下目标检测的普适性,实验结果如表6所示。

表6 在CARPK数据集上验证实验

Table 6 Validation experiments on CARPK

方法	P	R	mAP@0.5	mAP@0.5:0.95
YOLOv8n	0.960	0.899	0.959	0.653
YOLOv8n-ours	0.961	0.920	0.967	0.694
YOLOv8s	0.966	0.916	0.966	0.695
YOLOv8s-ours	0.972	0.922	0.970	0.715
YOLOv8m	0.965	0.922	0.970	0.687
YOLOv8m-ours	0.972	0.908	0.954	0.718

实验证明,改进方法不仅仅是针对VisDrone数据集,在其他无人机视角下的数据集上,改进算法也同样有效。将改进方法应用于YOLOv8n和YOLOv8s上,都能够全面实现模型性能提升。但值得注意的是,由于CARPK数据集是小规模的单类别数据集,较为简单,在太复杂的网络结构上如YOLOv8m上已经出现了性能退化,YOLOv8m的P、mAP@0.5:0.95已经低于YOLOv8s,改进的方法用于YOLOv8m使网络更复杂,使得R、mAP@0.5出现下降,但是还是提升了P、mAP@0.5:0.95。

在SHWD数据集上,采用改进方法在YOLOv8n、YOLOv8s、YOLOv8m进行实验,验证改进对小目标的普适性,实验结果如表7所示。

表7 在SHWD数据集上验证实验

Table 7 Validation experiments on SHWD

方法	P	R	mAP@0.5	mAP@0.5:0.95
YOLOv8n	0.914	0.827	0.913	0.597
YOLOv8n-ours	0.924	0.885	0.938	0.610
YOLOv8s	0.923	0.905	0.941	0.610
YOLOv8s-ours	0.939	0.910	0.956	0.629
YOLOv8m	0.939	0.909	0.951	0.623
YOLOv8m-ours	0.950	0.914	0.961	0.634

实验证明,本文对模型的改进方法不仅仅是针对无人机视觉的特定种类目标检测任务,对于小目标数据集都能够实现有效提升。在SHWD数据集上,尽管原始模型的精度已经很高,本文改进方法仍然能在较高水平的基础上将n、s、m三种规格的原始模型的P、R、mAP@0.5、mAP@0.5:0.95进行提升,如改进后YOLOv8n分别提升了1.1%、7%、2.7%、2.2%,使得YOLOv8n的性能接近YOLOv8s。

3.4 检测结果可视化

图6、图7分别展示了YOLOv8s原始模型和本文的改进模型在同一实验条件和参数下训练后,对无人机视角下密集目标和稀疏目标的检测效果。左边是原始模型的检测效果,右边是改进后模型的效果。

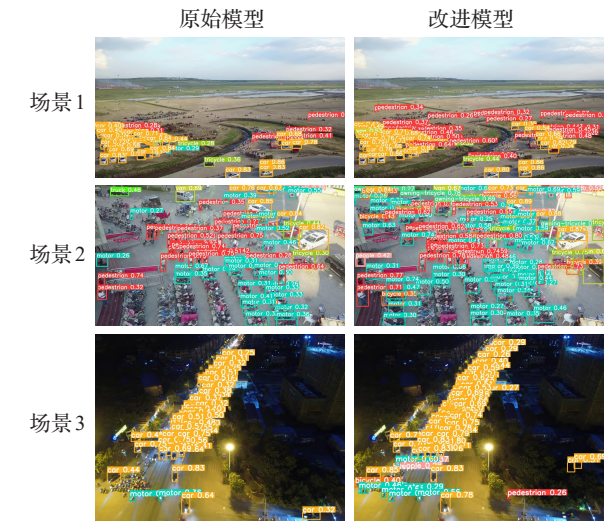


图6 密集目标检测效果对比

Fig.6 Comparison of dense object detection results

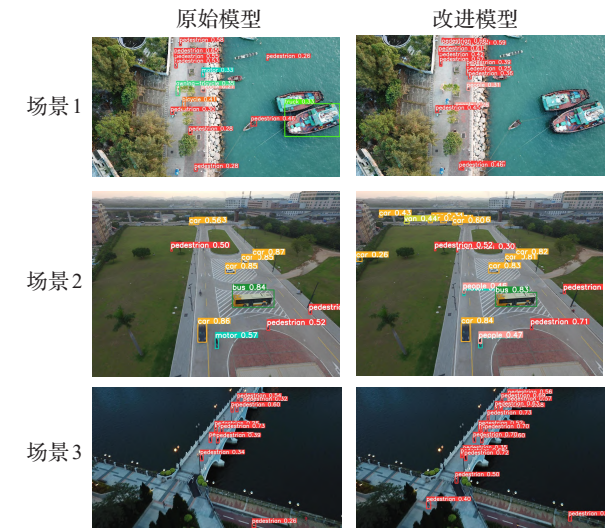


图7 稀疏目标检测效果对比

Fig.7 Comparison of sparse object detection results

从图6可以看出,在密集小目标场景中,改进模型可以更准确地检测出更多的小目标,如场景1草原中远处的行人、场景2密集小目标中的电动车、场景3夜晚道路上密集的电动车和路边昏暗光线下的行人与院子里的汽车,有效减少了漏检情况。

同样地,在稀疏小目标场景中,改进模型能有效减少漏检、误检情况。如图7所示,场景1中,原始模型错误地将船识别为行人和卡车,将人识别为摩托。场景2中,原始模型只能识别出motor,改进模型不仅能识别出行驶中的motor还识别出了motor上的人,正确地识别出更多的人和远方的车。场景3中,改进模型有效避免了原模型将路灯误检为行人的情况,准确检测出所有的

行人。本文的改进模型有效地改善了YOLOv8s对无人机视角下小目标的检测效果。

4 结束语

本文提出一种高效的YOLOv8s改进模型对无人机视角的图像进行目标检测任务。通过引入RFCBAM改进backbone中的Conv和C2f,增强网络的特征提取能力;在SPPF中增加LSKA来提升不同特征层间的语义融合;增加具有丰富小目标语义信息的特征层来改善颈部结构更好地处理小目标;用inner-IOU改造后的inner-MPDIoU作为模型边框回归的损失函数,提升对困难小样本的学习能力。在YOLOv8s基础上改进的模型大小和计算量小于YOLOv8m的一半,性能基本全面超过YOLOv8m,较小的模型较高的精度可以满足无人机设备的硬件条件。接下来将继续研究模型性能的提升,并对模型进行剪枝轻量化,优化对无人机视角密集小目标的检测能力。

参考文献:

- [1] SAEED Z, YOUSAF M H, AHMED R, et al. On-board small-scale object detection for unmanned aerial vehicles (UAVs)[J]. Drones, 2023, 7(5): 310.
- [2] BISIO I, HALEEM H, GARIBOTTO C, et al. Performance evaluation and analysis of drone-based vehicle detection techniques from deep learning perspective[J]. IEEE Internet of Things Journal, 2021, 9(13): 10920-10935.
- [3] HOSHINO W, SEO J, YAMAZAKI Y. A study for detecting disaster victims using multi-copter drone with a thermographic camera and image object recognition by SSD[C]//2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), 2021: 162-167.
- [4] SORBELLI F B, PALAZZETTI L, PINOTTI C M. YOLO-based detection of halyomorpha halys in orchards using RGB cameras and drones[J]. Computers and Electronics in Agriculture, 2023, 213: 108228.
- [5] SINGHA S, AYDIN B. Automated drone detection using YOLOv4[J]. Drones, 2021, 5(3): 95.
- [6] LI Y, YUAN H, WANG Y, et al. GGT-YOLO: a novel object detection algorithm for drone-based maritime cruising[J]. Drones, 2022, 6(11): 335.
- [7] PIRASTEH S, RASHIDI P, RASTIVEIS H, et al. Developing an algorithm for buildings extraction and determining changes from airborne LiDAR, and comparing with R-CNN method from drone images[J]. Remote Sensing, 2019, 11(11): 1272.
- [8] SEO D M, WOO H J, KIM M S, et al. Identification of asbestos slates in buildings based on faster region-based convolutional neural network (Faster R-CNN) and drone-based aerial imagery[J]. Drones, 2022, 6(8): 194.
- [9] 陈卫彪, 贾小军, 朱响斌, 等. 基于DSM-YOLO v5的无人机航拍图像目标检测[J]. 计算机工程与应用, 2023, 59(18): 226-233.
- CHEN W B, JIA X J, ZHU X B, et al. Target detection for UAV image based on DSM-YOLO v5[J]. Computer Engineering and Applications, 2023, 59(18): 226-233.
- [10] 陈范凯, 李士心. 改进Yolov5的无人机目标检测算法[J]. 计算机工程与应用, 2023, 59(18): 218-225.
- CHEN F K, LI S X. UAV target detection algorithm with improved Yolov5[J]. Computer Engineering and Applications, 2023, 59(18): 218-225.
- [11] 刘涛, 丁雪妍, 张冰冰, 等. 改进YOLOv5的遥感图像检测方法[J]. 计算机工程与应用, 2023, 59(10): 253-261.
- LIU T, DING X Y, ZHANG B B, et al. Improved YOLOv5 for remote sensing image detection[J]. Computer Engineering and Applications, 2023, 59(10): 253-261.
- [12] LI Y, FAN Q, HUANG H, et al. A modified YOLOv8 detection network for UAV aerial image recognition[J]. Drones, 2023, 7(5): 304.
- [13] LOU H, DUAN X, GUO J, et al. DC-YOLOv8: small-size object detection algorithm based on camera sensor[J]. Electronics, 2023, 12(10): 2323.
- [14] GUO J, LOU H, CHEN H, et al. A new detection algorithm for alien intrusion on highway[J]. Scientific Reports, 2023, 13(1): 10667.
- [15] WANG F, WANG H, QIN Z, et al. UAV target detection algorithm based on improved YOLOv8[J]. IEEE Access, 2023, 11: 116534-116544.
- [16] WANG G, CHEN Y, AN P, et al. UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.
- [17] ZHANG X, LIU C, YANG D, et al. RFACConv: innovating spatital attention and standard convolutional operation[J]. arXiv:2304.03198, 2023.
- [18] 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.
- [19] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN[J]. Expert Systems with Applications, 2024, 236: 121352.
- [20] ZHANG H, XU C, ZHANG S. Inner-IOU: more effective intersection over union loss with auxiliary bounding box[J]. arXiv:2311.02877, 2023.
- [21] MA S L, XU Y. MPDIoU: a loss for efficient and accurate bounding box regression[J]. arXiv:2307.07662, 2023.
- [22] DU D, ZHU P, WEN L, et al. VisDrone-DET2019: the vision meets drone object detection in image challenge results[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.
- [23] 张智, 易华挥, 郑锦. 聚焦小目标的航拍图像目标检测算法[J]. 电子学报, 2023, 51(4): 944-955.
- ZHANG Z, YI H H, ZHENG J. Focusing on small objects detector in aerial images[J]. Acta Electronica Sinica, 2023, 51(4): 944-955.
- [24] HSIEH M R, LIN Y L, HSU W H. Drone-based object counting by spatially regularized regional proposal network [C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 4145-4153.