

基于 RSD-YOLO 的无人机航拍图像小目标检测

汤伟博¹, 方强^{2,3,4}, 李沛根^{2,3,4}, 艾龙金¹, 熊金红¹, 夏海廷¹

(1. 昆明理工大学民航与航空学院, 云南 昆明 650500; 2. 昆明理工大学建筑工程学院, 云南 昆明 650500;
3. 云南省绿色建造与智慧运维国际联合实验室, 云南 昆明 650500; 4. 云南省土木工程防灾重点实验室, 云南 昆明 650500)

摘要: 针对无人机(UAV)航拍图像存在的检测性能低、遮挡严重、小目标特征提取难度大及模型参数量大的问题, 提出了基于 YOLOv8s 的 RSD-YOLO 算法。首先, 设计了感受野注意力(RFA)模块 CSP-RFA 替代 C2f 模块, 以提升小目标特征提取能力, 有效应对传统卷积操作对位置变化不敏感的问题。其次, 对主干网络和特征融合网络进行了轻量化处理, 新增了大量特征图检测分支, 并提出了感受野金字塔网络(RFPN), 优化特征流动方向, 增强特征表达能力。再次, 检测头模块经过优化, 将多尺度特征集成至具有多级注意力机制的检测头中, 并替换了损失函数, 提升了模型对小目标的检测性能。最后, 在模型压缩方面, 采用层自适应幅度剪枝(LAMP)算法, 进一步减少了模型的参数量和大小。实验结果表明, 轻量化后的 RSD-YOLO 在公开数据集 VisDrone2019 上较基线模型有显著提升, 精度提高了 10.0 个百分点, mAP@0.5 提升 9.5 个百分点(增幅 24.1%), mAP@0.5:0.95 提高 6.9 个百分点(增幅 29.4%)。模型参数量从 11.12×10^6 减少至 4.05×10^6 (减少 63.6%), 计算量从 42.7 GFLOPs 降至 25.5 GFLOPs(减少 40%)。此外, 在仅检测遮挡小目标的新数据集上, RSD-YOLO 在精度、mAP@0.5、mAP@0.5:0.95 上分别提升了 9.1、16.1 和 10.7 个百分点。

关键词: 无人机; 小目标检测; YOLOv8s; 注意力机制; 特征融合; 遮挡目标检测

中图分类号: TP391.41

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070151

RSD-YOLO-Based Small Target Detection in UAV Aerial Images

TANG Weibo¹, FANG Qiang^{2,3,4}, LI Peigen^{2,3,4}, AI Longjin¹, XIONG Jinhong¹, XIA Haiting¹

(1. Faculty of Civil Aviation and Aeronautics, Kunming University of Science and Technology, Kunming 650500, Yunnan, China;

2. Faculty of Civil Engineering and Mechanics, Kunming University of Science and Technology, Kunming 650500, Yunnan, China;

3. International Joint Laboratory for Green Construction and Intelligent Maintenance of Yunnan Province, Kunming 650500, Yunnan, China;

4. Yunnan Key Laboratory of Disaster Reduction in Civil Engineering, Kunming 650500, Yunnan, China)

【Abstract】 The RSD-YOLO algorithm, based on YOLOv8s, is proposed to address the challenges of low detection performance, severe occlusion, difficulty of small target feature extraction, and large number of model parameters inherent in Unmanned Aerial Vehicle (UAV) aerial images. First, the Receptive Field Attention (RFA) module CSP-RFA is designed to replace the C2f module for enhancing the capability of small target feature extraction, effectively addressing the insensitivity of traditional convolutional operations to positional changes. Second, the backbone and feature fusion networks are made lightweight, a new large-size feature map detection branch is added, and a Receptive Field Pyramid Network (RFPN) is proposed to optimize the feature flow direction and improve feature representation. Third, the detection head module is optimized by integrating multi-scale features with a multi-level attention mechanism and the loss function is updated to improve the model's detection performance for small targets. Finally, in terms of model compression, Layer-Adaptive Magnitude-based Pruning (LAMP) algorithm is employed to further reduce the number of parameters and model size. The experimental results demonstrate that the lightweight RSD-YOLO model significantly outperforms the baseline model on the publicly available VisDrone2019 dataset, with a 10.0 percentage point increase in precision, a 9.5 percentage point increase in mAP@0.5 (equivalent to a 24.1% increase), and a 6.9 percentage point increase in mAP@0.5:0.95 (equivalent to a 29.4% increase). The number of model parameters is reduced from 11.12×10^6 to 4.05×10^6 , representing a 63.6% reduction, and the computational cost is reduced from 42.7 GFLOPs to 25.5 GFLOPs, showing a 40% reduction. Furthermore, for a newly filtered dataset focusing on small occluded targets, RSD-YOLO shows improvements of 9.1, 16.1, and 10.7 percentage points in terms of precision, mAP@0.5, and mAP@0.5:0.95, respectively.

【Key words】 Unmanned Aerial Vehicle (UAV); small target detection; YOLOv8s; attention mechanism; feature fusion; occluded target detection

基金项目: 国家自然科学基金(12262015)。

作者简介: 汤伟博(CCF 学生会员), 男, 硕士研究生, 主研方向为目标检测; 方强、李沛根, 博士研究生; 艾龙金、熊金红, 硕士研究生; 夏海廷(通信作者), 教授、博士。

收稿日期: 2024-07-19

修回日期: 2024-09-18

E-mail: haiting.xia@kust.edu.cn

0 引言

无人机(UAV)在智慧城市^[1-2]、人员搜救^[3]、路桥检测^[4]和高科技农业^[5-6]等民用领域通过其获取的图像信息发挥着重要作用。图像检测作为计算机视觉的核心任务,旨在精确定位并识别图像中的目标类别。在这些应用中,小目标的准确检测具有极其重要的现实意义。在智慧城市的管理中,精确识别和监控小目标可以显著提高城市安全性和运营效率;在人员搜救中,精准定位小型目标有助于加速救援行动并挽救生命;在路桥检测中,精准的小目标检测能够及时发现桥梁结构的细微损伤和潜在隐患,确保基础设施的安全与有效维护;在高科技农业中,对小型农业目标的监测能够优化作物管理和病虫害控制。小目标检测技术的进步不仅提升了这些领域的应用效果,还为未来的智能化发展提供了坚实的技术基础。因此,提高小目标检测的准确性,对于推动这些民用领域的技术进步和应用发展具有重要的现实意义。

然而,小目标检测是图像检测领域的一大难点。相较于大目标,小目标的检测性能大幅降低,其中 MS COCO 数据集上的性能分析^[7]显示,小目标的平均检测准确度不足 30.8%。在无人机航拍图像中,遮挡问题尤为突出。由于在航拍过程中,摄像机通常位于高空,拍摄范围广阔,但地面上的目标物体常常被树木、建筑物或其他物体遮挡,导致目标的特征无法被清晰捕捉。遮挡不仅影响了目标的显著性和辨识度,还造成目标表达弱、密集堆叠等问题^[8],进一步增加了检测难度。因此,面对无人机图像中的小目标检测问题,迫切需要开发更为专门的方法来克服这些独特的挑战。

为了提高无人机视角下的小目标检测性能,众多研究致力于目标检测算法的改进。VADDI 等^[9]结合了特征金字塔网络和 ResNet 或 MobileNet 主干、采用 Focal Loss 的实时端到端检测模型加快了检测速度。基于 SSD (Single Shot MultiBox Detector)算法,CHEN 等^[10]提出了一种扩展层特征的方法,以保留更多小尺度信息。CAI 等^[11]引入的特征金字塔引导注意网络 GANet 则通过融合不同尺度特征图之间的语义信息,进一步提升了检测精度。在航拍影像的应用中,JIANG 等^[12]提出的基于特征融合和空间注意力的单阶段检测器有效用于小目标的特征表示。GUO 等^[13]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。潘玮等^[14]提出

一种融合多种注意力机制的 YOLOv8s 改进模型,提升对困难样本的学习能力。柳进元等^[15]将 YOLOv5 算法用于城市洪水参照物模型训练,设计了道路内涝深度信息。杨秀娟等^[16]在 YOLOv5 的基础上引入 Dynamic Head 注意力并且设计了 C3-DCN 结构,有效提高了小目标检测的精度。

文献^[9]虽然使得检测速度更快但是对检测精度的提升并不大。文献^[10-12]中对多尺度特征融合进行了改进,在一定程度上缓解了无人机航拍视角下目标尺度不均衡的问题,但对小目标特征识别分类效果仍不好,且对被遮挡目标检测效果差。文献^[13-16]通过在网络中添加注意力机制,提升了网络对航拍小目标的识别率,但导致了大幅增加的参数量。

针对无人机航拍图像小目标检测中检测性能差、特征识别不清晰、模型参数量大、对遮挡目标效果不佳的问题,本文提出 RSD-YOLO 算法,提升了模型对小目标检测的性能。主要贡献如下:

- 1) 受到感受野注意力(RFA)机制的启发,本文采用 RFACONV 替换了主干网络中的传统卷积,并设计了一个即插即用的 CSP-RFA 模块。该模块突破了卷积核参数共享的限制,从而增强了模型对小目标特征的提取能力。

- 2) 本文提出了全新的感受野金字塔网络(RFPN)对网络结构进行了改进,将浅层特征图引入特征融合网络,并重新设计了检测头结构。同时,去除了主干网络中的最后一组下采样结构即 P5 检测头,这不仅提升了检测精度,也减少了模型的参数和计算量,提高了算法实时检测的能力。

- 3) 优化了检测头模块和边界框的损失函数,使其能够集成不同尺度的特征到一个配备多级注意力机制的检测头中。显著增强了小目标检测模块的表示效果,而且不增加额外的计算成本。

- 4) 从 VisDrone2019 数据集中提取被遮挡的检测目标,制作了一个只包含被遮挡目标的新数据集。并且通过实验对比,验证了 RSD-YOLO 算法对被遮挡小目标的检测性能。

- 5) 本文引入了层自适应幅度剪枝(LAMP)算法,通过通道剪枝对模型进行压缩,显著减少了模型的参数量和尺寸。在保持原有模型性能的基础上每秒帧数(FPS)增加了 42 帧,降低了对计算设备的要求。

这些改进措施综合提升了针对小目标检测的效率和准确性,为解决小目标检测的难题提供了有效的方法。

1 YOLOv8 概述

YOLOv8 网络是 Ultralytics 公司于 2023 年 1 月发布的一系列最新 YOLO 模型,包括从小到大的 5 个版本: v8n、v8s、v8m、v8l 和 v8x。随着模型尺寸的增大,其精度逐步提高。鉴于无人机硬件条件的限制,本研究选择了尺寸较小且精度较高的 YOLOv8s 模型作为基础。

YOLOv8 网络架构分为 3 个主要部分:主干网络(Backbone)、颈部(Neck)和检测头(Head)。主干网络主要负责特征提取,采用 Darknet-53 框架构建,包括 Conv、C2f 和 SPPF(Spatial Pyramid Pooling Fast)3 种模块。其中,Conv 和 C2f 模块按序叠加 4 次,每次叠加形成一个阶段(Stage),在这 4 个阶段之后,通过 SPPF 模块

固定不同特征图的向量大小,从而优化特征提取过程。

颈部采用特征金字塔网络(FPN),主要用于特征融合,以增强检测性能。此外,检测头部分相较于 YOLOv5 有显著变化,借鉴自 YOLOX 的设计,实现了耦合头向解耦头(Decoupled Head)的转变。通过分离分类头和回归头,网络能更有效地处理不同的特征尺度和语义信息,同时也更好地解决了分类和定位的关注点差异问题。

2 RSD-YOLO 算法

本文以 YOLOv8s 为基准模型,分析其不足与小目标检测的难点,在保证算法精度与速度的前提下,提出 RSD-YOLO 算法,如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同)。

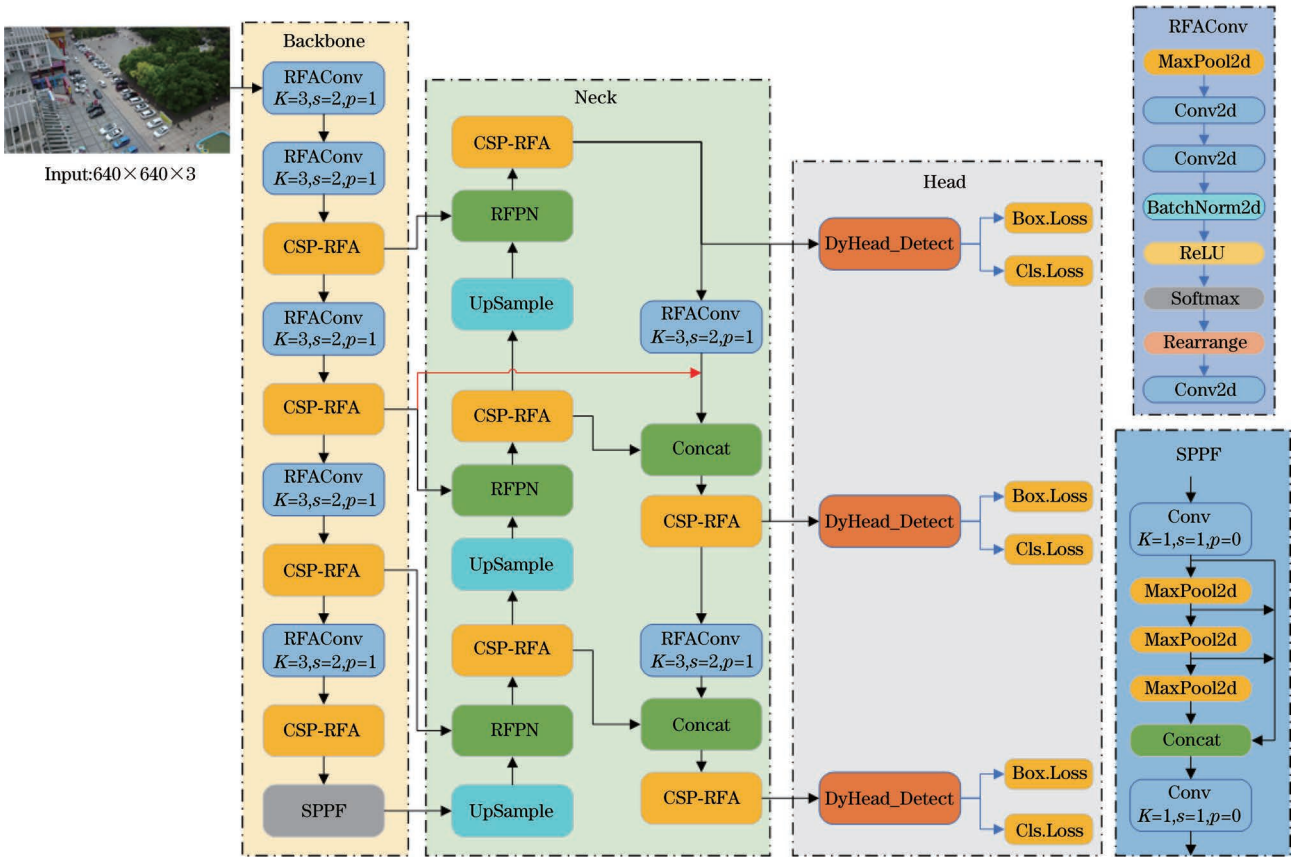


图 1 RSD-YOLO 算法网络结构图

Fig.1 Network structure diagram of RSD-YOLO algorithm

2.1 CSP-RFA 模块

感受野注意力卷积(RFACnv)旨在克服现有空间注意力机制中大规模卷积核参数共享的限制,将注意力集中于感受野内的空间特征。不同于传统卷积依赖固定大小卷积核,RFACnv 根据卷积核大小动态生成感受野内的空间特征,强调这些特征的重要性,并对它们进行优先排序。这种方法有效解决了传统卷积操作对于位置变化信息不敏感的问题,

并突破了参数共享的约束。RFACnv 采用组卷积(Group Conv)替代了传统方法中无参数的展开操作,以更有效地提取感受野特征并减少参数量。为了降低计算量并减少参数,RFACnv 使用平均池化(AvgPool)来整合每个感受野的全局信息。最终,通过 Softmax 函数强化特征权重的归一化处理,确保注意力机制能够精准地捕捉到每个感受野特征的重要性,从而显著提高了网络的性能。

$$F = \text{Softmax}(g^{1 \times 1}(\text{AvgPool}(X))) \times \text{ReLU}(\text{Norm}(g^{k \times k}(X))) = A_{rf} \times F_{rf} \quad (1)$$

式中： $g^{i \times i}$ 表示大小为 $i \times i$ 的组卷积； k 表示卷积核的大小；Norm 表示归一化； X 表示输入特征映

射； A_{rf} 为感受野注意图； F_{rf} 为变换后的感受野空间特征。 F 由 A_{rf} 与 F_{rf} 相乘得到。

具有 3×3 大小卷积核的 RFACConv 的整体结构如图 2 所示。

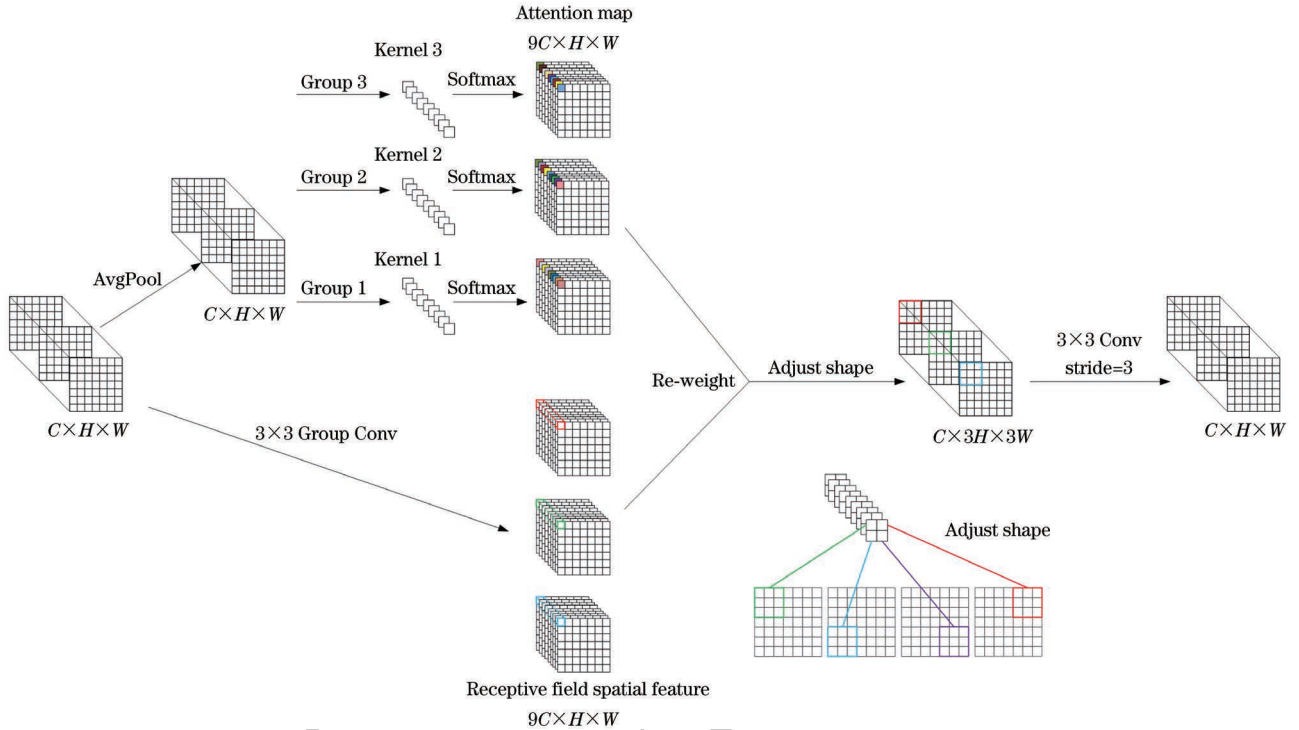


图 2 RFA 的详细结构

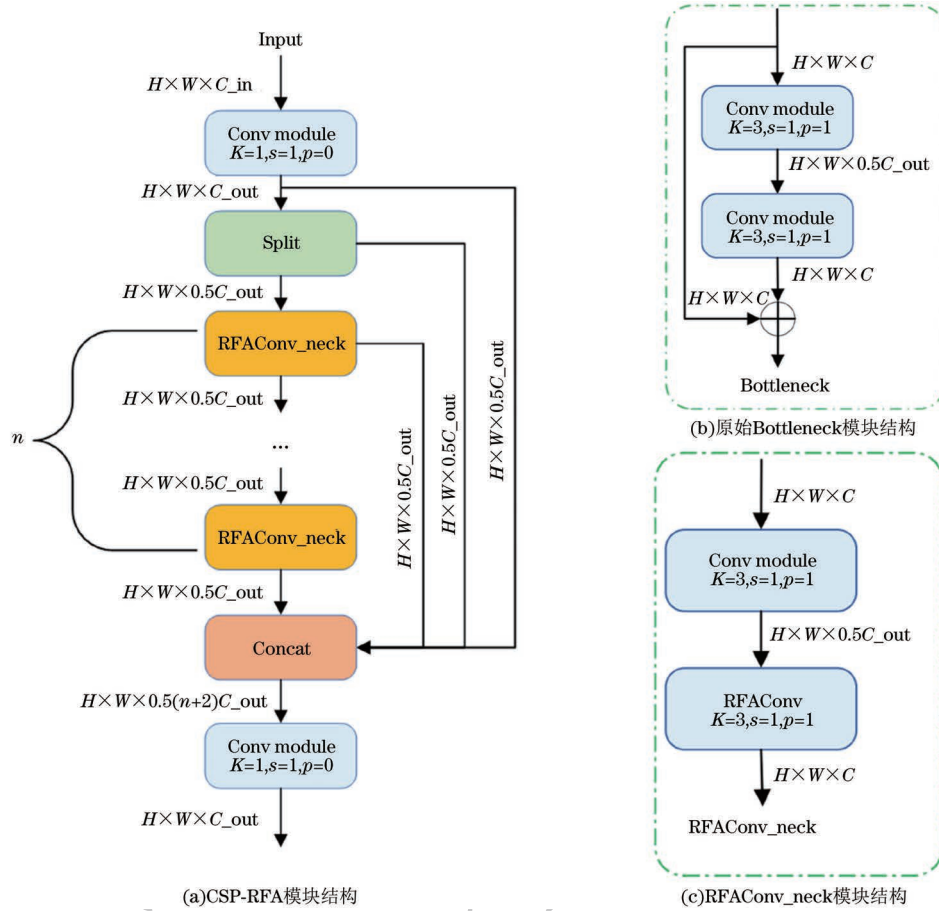
Fig.2 Detailed structure of RFA

本文受 RFACConv 的思想启发,改造了主干网络中 C2f 模块的 Bottleneck,提出了即插即用的 CSP-RFA 模块,其具体结构如图 3 所示。在 CSP-RFA 结构中,通过使用 RFACConv_neck 来替代传统的 Bottleneck 模块,实现了模型结构的创新。RFACConv_neck 维持了双卷积模块的配置,首先通过一个标准的卷积层进行特征处理,随后将第二个卷积层替换为 RFACConv,从而引入了残差信息的融合机制。值得注意的是,在 RFACConv_neck 的设计中,由于 RFACConv 本身已经融合了残差信息,因此取消了额外的残差连接操作。YOLOv8 模型主要依靠其主干网络部分进行特征提取,本研究通过将 YOLOv8 原有主干网络中的标准卷积层替换为 RFACConv,进一步将整个模型的 C2f 模块升级为 CSP-RFA。此举不仅保留了原有的特征提取机制,还通过动态调整基于卷积核大小的感受野,增加了对感受野内不同特征重要性的强调,并对感受野空间特征优先排序,有效解决了传统卷积操作对位置变动信息不敏感的问题,同时突破了传统卷积核参数共享的限制,提升模型对小目标特征的提取能力。

2.2 特征融合网络

在主干网络中,输入图像通过多个下采样和卷积模块的处理,采用自底向上的方法生成了多尺度特征图。深层的小尺寸特征图拥有更大的感受野和更高的通道数,携带丰富的语义信息,适合检测大尺寸目标;相反,浅层的大尺寸特征图虽然感受野较小、通道数较少,但它包含更多的空间信息,更适合小目标的检测^[17-18]。在无人机目标检测的场景中,由于目标尺寸多变、分布不均,图像中常见的多为小目标。为了降低计算成本和减少网络参数,本研究对模型的颈部进行了创新设计,移除了原网络中最后一组的下采样结构,即小尺寸特征图分支。

此外,为解决原版 YOLOv8 中 Concat 模块并不能很好地提取出浅层目标特征的问题,本文提出了全新的 RFPN,其主要功能是通过有效的特征融合策略来增强模型对复杂数据的理解能力。它能够将来自不同通道或尺度的特征信息整合在一起,从而显著提升模型的性能和准确度。该网络具备灵活的特征调整能力,能够适应多尺度数据的处理需求,同时减少过拟合风险,



(a) CSP-RFA模块结构

(c) RFACov_neck模块结构

图 3 CSP-RFA 模块示意图

Fig. 3 Schematic diagram of CSP-RFA module

增强模型在各种视觉任务中的表达能力和泛化能力。具体结构如图 4 所示,图中 BN2d 表示二维批量归一化。

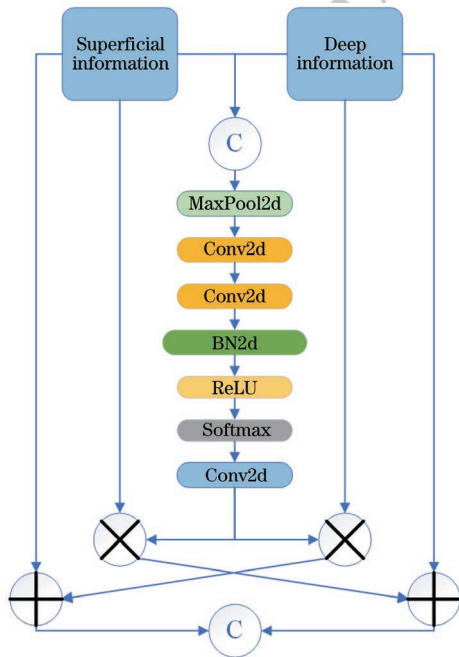


图 4 RFPN 结构图

Fig. 4 RFPN structure diagram

具体来说,本研究引入了一个专为小目标检测设计的分支并且去掉 YOLOv8 自带的 P5 模块,称为 SODH (Small Object Detection Head)。该分支通过将主干网络中进行了 4 倍下采样的大尺寸特征图与通过双线性插值上采样后富含语义信息的特征融合网络中的 4 倍特征图结合,提供给 SODH 作为输入,成为多尺度特征检测头的一部分。此外,输入到 P3 检测头的特征图采用了跳跃连接结构,它结合了主干网络中 8 倍下采样得到的特征图与特征融合网络中自上而下和自下而上处理得到的 8 倍特征图,通过通道维的拼接进行融合。与原网络相比,这种从主干网络获得的特征图的融合方式,进一步增强了特征的表达能力,使目标定位更为准确^[19]。改进网络结构如图 5 所示。

2.3 多级注意力检测头模块

YOLOv8 原有的检测机制采用了单一尺度的预测结构,仅从单一尺度的特征图进行目标预测,忽略了多尺度特征对于检测精度的潜在贡献。这限制了其在处理不同大小目标时的有效性。此外,其预测机制未充分考虑上下文信息,导致每个预测都是

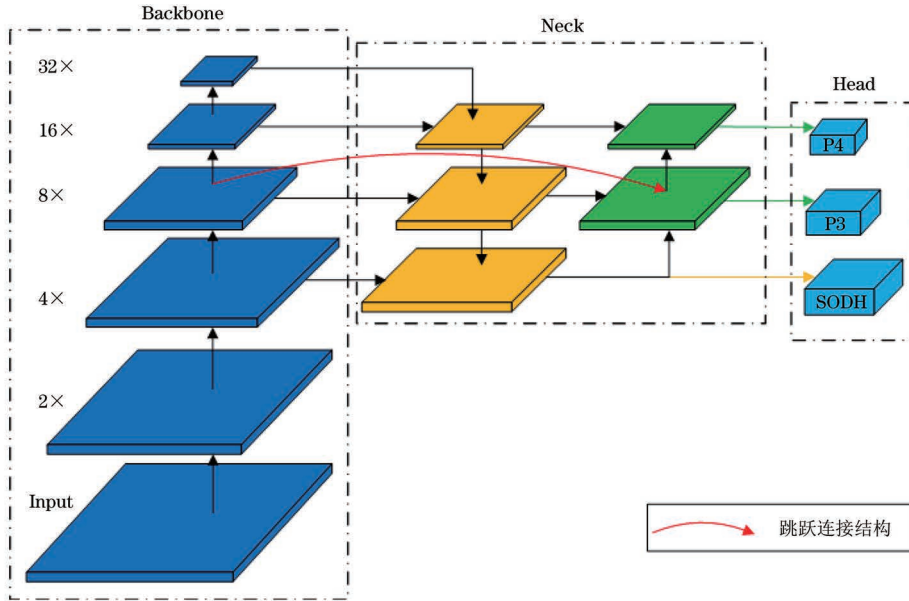


图 5 改进网络结构图

Fig. 5 Structural diagram of the improved network

孤立进行的,缺乏整体的视野。参数数量的限制也减弱了其在探索特征空间结构方面的能力。总的来说,YOLOv8的这种设计,特别是单尺度预测结构和缺乏动态调整能力,在多尺度目标检测任务中存在明显短板。

受到文献[20]的启发,本研究通过将注意力机制整合入目标检测框架,利用特征层间的自注意力机制进行有效交互,赋予了模型对尺度、空间和任务的感知能力,从而显著增强了目标检测的表现力,且未增加额外的计算成本。如图 6 所示,首先,通过主干网络和颈部网络处理图像,得到 3 种不同尺度的特征图,并将它们融合成一个统一

的特征表示 $F \in \mathbb{R}^{L \times S \times C}$ 。然后,将层级特征 L 通过尺度感知注意力模块处理,以融合不同尺度的特征并加强特征学习表达,使得特征图对前景对象的尺度差异展现出更高的敏感度;接着,空间特征 S 通过空间感知注意力模块来提升特征图中不同位置的学习权重,使特征图更具稀疏性,从而集中聚焦于前景对象的空间位置判别;后续,将通道特征 C 通过任务感知注意力模块处理以提升特征表达能力。最后,这些特征经过原有的检测头处理,输出最终结果。通过利用自注意力机制在尺度感知注意力、空间感知注意力与任务感知注意力三者上实现了统一。

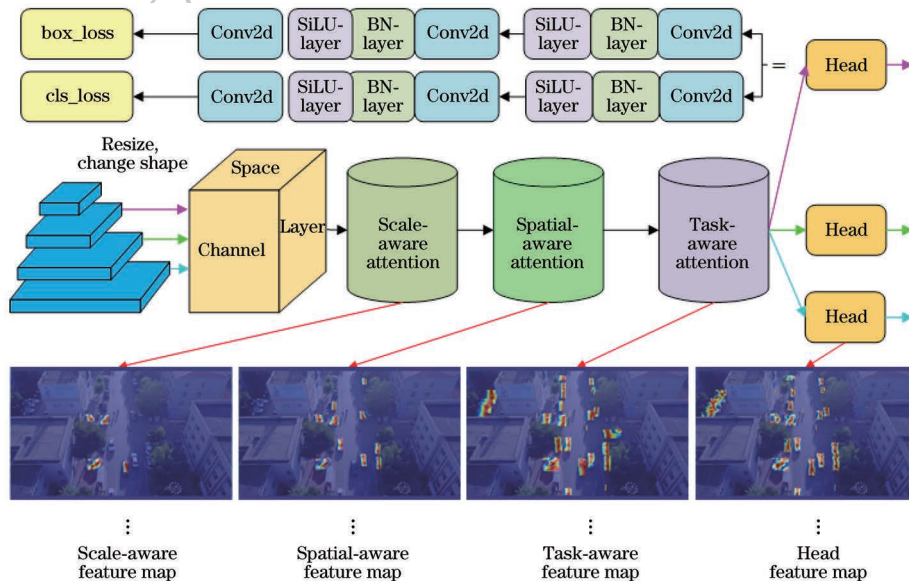


图 6 改进的检测头网络结构图

Fig. 6 Network structure diagram for the improved detection head

该注意力的计算公式为:

$$W(\mathbf{F}) = \pi_C(\pi_S(\pi_L(\mathbf{F}) \cdot \mathbf{F}) \cdot \mathbf{F}) \cdot \mathbf{F} \quad (2)$$

$$\pi_L(\mathbf{F}) \cdot \mathbf{F} = \partial\left(f\left(\frac{1}{SC} \sum_{S,C} \mathbf{F}\right)\right) \mathbf{F} \quad (3)$$

$$\pi_S(\mathbf{F}) \cdot \mathbf{F} = \frac{1}{L} \sum_{l=1}^L \sum_{k=1}^K W_{l,k} \cdot \mathbf{F}(l; P_k + \Delta P_k; c) \cdot \Delta m_k \quad (4)$$

$$\pi_C(\mathbf{F}) \cdot \mathbf{F} = \max(\alpha^1(\mathbf{F}) \cdot F_C + \beta^1(\mathbf{F}), \alpha^2(\mathbf{F}) \cdot F_C + \beta^2(\mathbf{F})) \quad (5)$$

式中: $\partial(\cdot)$ 为 Hard-sigmoid 函数; $f(\cdot)$ 为线性函数, 用 1×1 卷积实现; $S = H \times W$; \mathbf{F} 代表一个输入的三维张量 $L \times S \times C$; $\frac{1}{SC} \sum_{S,C} \mathbf{F}$ 为平均池化操作; K 为稀疏采样位置的数量; Δm_k 是一个关于位置 P_k 的可自学习的重要性度量因子; $W_{l,k} \cdot \mathbf{F}(l; P_k + \Delta P_k; c)$ 为可变形卷积操作; ΔP_k 为学习偏差; $W_{l,k}$ 为可变形卷积运行的参数; $[\alpha^1, \beta^1, \alpha^2, \beta^2]^T$ 为超函数, 用于学习控制激活阈值; $\pi_C(\cdot)$ 、 $\pi_S(\cdot)$ 和 $\pi_L(\cdot)$ 分别为作用于 C 维度的任务感知注意力模块、作用于 S 维度的空间感知注意力模块和作用于 L 维度的尺度感知注意力模块。

2.4 改进的损失函数模块

YOLOv8 在边界框预测中采用完全交并比 (CIoU) 损失函数, 有效解决了传统交并比 (IoU) 损失函数在处理不规则形状目标时可能遇到的梯度消失问题。尽管如此, CIoU 在定义边界框长宽比时的模糊性, 导致其损失计算无法根据样本质量进行自适应调整, 特别是在处理几何度量时, 如距离和长宽比, 这种处理方式可能过度惩罚低质量样本, 影响模型的泛化能力。为提高模型对高质量样本的敏感度, 减轻在锚框与目标框重合度较高时几何度量梯度的惩罚, 本研究引入了全新的 Wise-IoU v3 损失函数^[21]作为替代。与 CIoU 不同, Wise-IoU v3 避免了长宽比计算, 基于 Wise-IoU v1 的注意力机制边界框损失, 引入了一种动态非单调的聚焦机制, 并通过“离群度”概念来评估锚框的质量, 从而更加精准地优化模型的预测性能。基于注意力的边界框损失 Wise-IoU v1 计算式如下:

$$L_{\text{WIoUv1}} = R_{\text{WIoU}} L_{\text{IoU}} \quad (6)$$

式中: L_{WIoUv1} 为边界框损失 Wise-IoU v1; L_{IoU} 为边界框损失 IoU; R_{WIoU} 为距离注意力,

$$R_{\text{WIoU}} = \exp\left(\frac{(x - x_{\text{gt}})^2 + (y - y_{\text{gt}})^2}{W_g^2 + H_g^2}\right) \quad (7)$$

式中: X, Y 分别为预测框中心点的横、纵坐标; $X_{\text{gt}},$

Y_{gt} 分别为真实框中心点的横、纵坐标; W_g, H_g 分别为预测框与真实框最小外接矩形的宽和高。

接着用离群度来描述锚框的质量, 定义式如下:

$$\beta = \frac{L_{\text{IoU}}^*}{\bar{L}_{\text{IoU}}} \in [0, +\infty) \quad (8)$$

式中: \bar{L}_{IoU} 为滑动平均值; $*$ 表示从计算图中分离操作, 使其变为不带梯度的常量。

最后, 利用离群度 β 构造非单调聚焦系数 r , 得到边界框损失 Wise-IoU v3, 计算式如下:

$$L_{\text{WIoUv3}} = r L_{\text{WIoUv1}}, \quad r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (9)$$

式中: L_{WIoUv3} 为边界框损失 Wise-IoU v3; α, δ 为超参数, 本研究中分别设置为 1.9 和 3.0。

2.5 模型压缩

经过 2.2 的轻量化设计后, 模型在参数量和大小方面有所改善, 但仍然存在参数冗余和模型过大的问题。为了在尽量保证模型对航拍小目标检测效果的同时, 减少参数冗余和计算量, 并降低其在嵌入式设备部署的难度, 本文进一步采用通道剪枝技术, 对模型进行压缩, 以最大限度降低其对算力资源的消耗。

LAMP 是一种深度学习模型剪枝算法, 通过选择性地去除神经网络中不重要的参数来减少模型的复杂性和计算成本。其核心思想是根据每一层参数的重要性 (即权重的绝对值大小) 来自适应地进行剪枝, 而不是对整个网络进行统一的剪枝。具体来说, LAMP 首先为每一层计算独特的重要性阈值, 这些阈值基于层内权重的绝对值大小确定, 并根据预设的剪枝率选择一个阈值。接下来, LAMP 确定剪枝权重, 将绝对值小于或等于阈值的权重标记为要剪枝的权重, 并将其置为零, 保留其余的权重。最后, 剪枝后的模型通常需要进行微调训练, 以恢复或提升剪枝前的性能。其公式如下:

$$s(w, W) = \frac{(W[u])^2}{\sum_{u \geq v} (W[u])^2} \quad (10)$$

式中: W 表示卷积神经网络中的卷积核; $W[u]$ 表示卷积核索引 u 对应的权重, 当 $u \leq v$ 成立时, $|W[u]| \leq |W[v]|$ 。

3 实验与结果分析

3.1 数据集与实验环境

为验证 RSD-YOLO 改进算法模型更具有普适性, 使用 VisDrone2019^[22] 和 DOTA^[23] 两个数据集作为验证对象开展实验。

VisDrone2019 数据集共包含 10 类航拍检测目标,包括:行人、静止人员、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 6 471 张、548 张、1 610 张图片组成。数据集中面积小于 16×16 像素的极小目标占比为 12.05%,面积小于 32×32 像素的小目标占比为 44.7%,静止人员和行人的类别中小目标占比更是分别高达 77.45%和 64.59%,与本文面向航拍图像的小目标检测任务契合。

DOTA 数据集是一个专为航空图像目标检测设计的开放数据集,旨在提升在高分辨率和大场景下检测小型目标的技术。DOTA 数据集提供了来自多个地区和场景的高分辨率图像,涵盖各种天气条件和拍摄角度。数据集包含 15 种目标类别,并提供详细的标注信息。由于该数据集图片尺寸过大,直接输入会导致计算过慢,本文对 DOTA 数据集进行裁剪得到 $1\,024 \times 1\,024$ 像素的图片,同时设置 200 像素重叠,防止目标被裁剪为两部分,处理之后训练集、验证集、测试集分别由 15 749 张、5 297 张、10 833 张图片组成。

采用以下指标评价模型:精度(Precision)表示所有被识别出的部位中识别正确部位所占的比例,反映模型正确分类的能力;召回率(Recall)是被正确识别的部位数量与测试集中所有部位数量的比值,反映模型全面检测目标的能力;平均精度值(mAP)由预测框与标注框的 IoU 计算,并根据 IoU 阈值不同分为 $mAP@0.5$ 和 $mAP@0.5:0.95$,反映算法生成的检测框对目标的贴合程度。此外,模型的复杂度也是本文实验的关键指标,参数量(Params)用于衡量模型内存资源消耗。同时,还采用每秒十亿次浮点运算数(GFLOPs)作为网络计算量的评估指标。

实验平台使用的 GPU 为 NVIDIA GeForce RTX 4090。实验环境为 Python 3.8,PyTorch 2.1.2,CUDA 12.1。为了公平对比算法的性能,所有方法

采用相同的超参数设置进行训练、验证。具体实验环境参数设置如表 1 所示。

表 1 训练参数设置

Table 1 Training parameter settings

参数编号	参数	设置
1	epochs	300
2	patience	50
3	batch	8
4	imgsz	640
5	workers	8
6	optimizer	SGD
7	close_mosaic	10
8	lr0	0.01
9	lrf	0.01
10	warmup_epochs	3.0
11	momentum	0.937
12	weight_decay	0.0005

3.2 实验结果分析

3.2.1 注意力模块和特征融合网络的实验比较

为了准确地展示 RFACnv 注意力模块在性能上的优势,本文选择在 VisDrone2019 数据集上与其他流行的注意力模块在相同的位置将普通卷积 Conv 和 C2f 模块中的 Bottleneck 改进替换进行了比较实验。本文的实验通过确保所有其他条件如网络结构、优化策略和超参数等都与表 1 保持一致来保证比较的公正性。

添加不同注意力方法的结果比较如表 2 所示(最好的数据用粗体表示,下同)。表 2 中,YOLOv8s 为本文的基线模型,+表示在 YOLOv8 的基础上进行相应模块改进替换。经过测试,本文提出的使用 RFACnv 将普通卷积 Conv 和 C2f 模块中的 Bottleneck 改进替换的方法在各项评估指标上都展现了卓越的性能,在模型计算量仅增加 1.3 GFLOPs 的情况下,将精度提高了 3.1 百分点,将 $mAP@0.5$ 提高了 1.3 百分点,超过了其他常用网络。

表 2 添加不同注意力方法的结果比较

Table 2 Comparison of results of adding different attention methods

实验方法	Precision/%	Recall/%	$mAP@0.5$ /%	$mAP@0.5:0.95$ /%	GFLOPs	Params/ 10^6
YOLOv8s	49.0	39.0	39.4	23.4	28.5	11.12
+CBAM-Attention ^[24]	50.9	38.3	39.5	23.6	29.1	11.58
+RevCol-net ^[25]	47.9	36.3	37.0	21.4	21.2	8.20
+SCConv ^[26]	51.5	38.5	39.8	23.9	33.5	14.91
+DySnakeConv ^[27]	50.6	39.1	39.7	23.8	33.8	14.42
+ODConv ^[28]	49.1	37.3	38.3	22.4	22.6	9.47
+Ours	52.1	38.9	40.7	24.2	29.8	11.35

3.2.2 网络结构的实验比较

同时对修改前后的网络结构的精度、召回率、平均精度值、GFLOPs 和参数量进行了对比,结果如表 3 所示。表 3 中, YOLOv8s 为本文的基线模型, X-YOLOv8s 为未剪去小尺寸特征图分支但新增了小目标检测分支 SODH 的网络结构, XS-YOLOv8s 为改进后的网络结构。数据集使用 VisDrone2019, 训练参数与表 1 保持一致。实验结果表明 X-YOLOv8 比基线模型提高了 5.5 百分点

的精度、5.7 百分点的 $mAP@0.5$, 然而该模型计算量增加了 8.4 GFLOPs。而改进后的网络结构 XS-YOLOv8s 在精度、召回率、平均精度值都达到最高的同时, 相比于 X-YOLOv8 模型计算量减少了 2.6 GFLOPs, 参数量下降了 3.23×10^6 。分析认为原网络结构中的小尺寸特征图检测分支感受野过大, 对小目标数据集的检测效果有限, 且深层特征图带来较多额外计算量, 影响算法的参数量和计算量。

表 3 网络结构对比

Table 3 Comparison of network structure

实验方法	Precision/%	Recall/%	$mAP@0.5$ /%	$mAP@0.5:0.95$ /%	GFLOPs	Params/ 10^6
YOLOv8s	49.0	39.0	39.4	23.4	28.5	11.12
X-YOLOv8s	54.5	43.5	45.1	27.6	36.9	10.64
XS-YOLOv8s	54.5	43.9	45.4	27.7	34.3	7.41

3.2.3 检测头模块的实验比较

YOLOv8 检测头串联的数量决定着模型检测的性能。本文对 YOLOv8 的检测头串联数量对于模型性能的影响进行了深入的探究和实验研究, 结果如表 4 所示。由表 4 可见, 尽管理论上使用 4 个改进检测头的配置模型提升了 5.4 百分点的精度和 2.4 百分点的 $mAP@0.5:0.95$, 带来了更强的小目标检测能力, 但是其增加了 1.2 GFLOPs

的计算量和 1.22×10^6 的参数量。而相较于基线模型, 使用 1 个检测头时模型提升了 2.5 百分点的精度和 1.4 百分点的 $mAP@0.5:0.95$, 减少了 0.4 GFLOPs 的计算量和 0.27×10^6 的参数量。这一发现为本文后续模型选择和优化提供了重要的参考依据, 也证明了增加模型复杂度不一定能带来更好的性能, 有时简化和精细调整可能是更佳的选择。

表 4 不同检测头数量的结果比较

Table 4 Comparison of results with different numbers of detection heads

实验方法	Precision/%	Recall/%	$mAP@0.5$ /%	$mAP@0.5:0.95$ /%	GFLOPs	Params/ 10^6
YOLOv8s	49.0	39.0	39.4	23.4	28.5	11.12
1 个改进检测头	51.5	40.0	41.1	24.8	28.1	10.85
2 个改进检测头	53.8	39.9	41.8	25.4	28.6	11.35
3 个改进检测头	53.5	40.8	42.2	25.7	29.2	11.84
4 个改进检测头	54.4	40.9	42.5	25.8	29.7	12.34

3.2.4 对遮挡目标实验

VisDrone2019 数据集的标注文件通常是以文本格式提供的, 每个文件对应一张图像, 包含每个目标的遮挡等级。遮挡等级表示目标的遮挡情况, 通常用 0(无遮挡)、1(部分遮挡)、2(严重遮挡)来表示。在航拍过程中, 摄像机通常位于高空, 拍摄范围广阔, 但地面上的目标物体常常被树木、建筑物或其他物体遮挡, 这使得遮挡问题尤为突出。本文通过遍历所有标签, 统计出训练集中共有 353 550 个目标, 其中被部分遮挡的目标个数为 142 873, 被严重遮挡的目标个数为 33 804, 再将所有未被遮挡目标信息删除, 剩余 6 253 张图片。将其中 5 210 张作为

训练集, 1 043 张作为验证集来验证算法对被遮挡目标的检测效果。

遮挡小目标检测效果比较如表 5 所示。本文提出的 RSD-YOLO 算法对被遮挡目标的精度、召回率、 $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别为 35.9%、48.8%、26.5% 和 14.60%, 较基线算法 YOLOv8s 分别提高了 9.1、16.1、10.7、6.46 百分点, 其中 $mAP@0.5$ 增幅为 67.7%, $mAP@0.5:0.95$ 增幅接近 80%, 参数量从 11.12×10^6 下降到 8.26×10^6 , 减少了 25.6%, 在所有模型当中效果最优。经过实验验证, 本文提出的 RSD-YOLO 算法对遮挡目标的检测有较好效果。

表 5 遮挡小目标检测效果比较

Table 5 Comparison of detection effect of occluded small targets

实验方法	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5+0.95/%	Params/10 ⁶
YOLOv8s	26.8	32.7	15.8	8.14	11.12
YOLOv5	25.9	30.8	14.7	7.21	2.50
YOLOv6	23.5	27.9	12.4	6.03	4.23
YOLOv8l	30.6	37.6	19.0	10.30	43.61
RLSM-YOLO ^[14]	35.5	48.2	26.2	14.40	11.85
Ours	35.9	48.8	26.5	14.60	8.26

3.2.5 消融实验比较

为验证所提出的 CSP-RFA 模块、改进网络结构、多级注意力检测头模块以及改进损失函数的有效性,本文进行消融实验对比不同模块的组合对小

目标检测算法性能的影响。YOLOv8 按照网络规模从小到大有 n、s、m、l、x 5 个版本,综合考虑网络性能与计算开销,以开源的 YOLOv8s 为基线模型,训练完毕后在验证集上验证,结果如表 6 所示。

表 6 消融实验结果

Table 6 Results of ablation experiment

实验方法	CSP-RFA	SODH	DyHead	Wise-IoU	Precision/%	Recall/%	mAP@0.5/%	GFLOPs	Params/10 ⁶
YOLOv8s					49.0	39.0	39.4	28.5	11.12
A	✓				52.2	39.0	40.7	29.8	11.35
B		✓			54.5	43.9	45.4	34.3	7.41
C			✓		52.0	39.9	41.1	28.1	10.85
D				✓	50.5	38.7	39.6	28.5	11.12
E	✓	✓			54.9	43.4	45.5	36.6	7.79
F	✓		✓		53.6	40.8	42.5	29.4	11.08
G	✓			✓	52.7	40.2	41.1	29.8	11.36
H	✓	✓	✓		57.5	46.6	49.1	42.7	8.27
I	✓	✓	✓	✓	59.4	46.6	49.6	42.7	8.27

对比基线模型 YOLOv8s,添加 CSP-RFA 模块的模型 A 在实验结果中的精度上有明显提升。其在计算量增加 1.3 GFLOPs 和参数量增加 0.23×10^6 的情况下精度和 mAP@0.5 分别提升了 3.2 个百分点和 1.3 个百分点。改进网络结构的模型 B 实验结果有明显提升,在计算量仅增加 5.8 GFLOPs 的情况下精度、召回率和 mAP@0.5 分别提升了 5.5、4.9、6.0 百分点。此外,参数量也由 11.12×10^6 降至 7.41×10^6 。结果表明,针对小目标设计的小目标检测头能够有效地检测到更多的小目标。同时,简化深层网络结构后,参数量明显减少,从而显著提高了模型的实用性和效率。模型 C 引入了多级注意力检测头,使得精度、召回率和 mAP@0.5 分别提升了 3.0、0.9、1.7 百分点,同时参数量减少了 0.27×10^6 ,计算量下降了 0.4 GFLOPs。此项改进显示了结构优化通过注意力机制与任务需求的结合,提高了特征表示的丰富性和模型的学习能力。相较于基线模型,模型 D 的精度和 mAP@0.5 分别提升了 1.5 个百分点和 0.2 百分点,可知替换 Wise-

IoU v3 损失函数,在不改变模型大小和计算量的前提下,也可以有效提升模型的检测效果,并且使得模型更快收敛。

模型 E 在模型 A 的基础上进一步改进了网络结构,通过再次优化,显著提升了性能。尽管计算量较模型 A 增加了 6.8 GFLOPs,精度、召回率和 mAP@0.5 分别提升了 2.7、4.4、4.8 百分点,同时参数量也减少了 3.56×10^6 ,进一步验证了优化后小目标检测头和简化深层网络结构的有效性。模型 F 在模型 A 的基础上引进了 DyHead,相较基线模型精度、召回率和 mAP@0.5 分别提升了 4.6、1.8、3.1 百分点。模型 G 在模型 A 的基础上引入了 Wise-IoU v3,相较基线模型精度、召回率和 mAP@0.5 分别提升了 3.7、1.2、1.7 百分点。模型 H 在模型 E 的基础上引入了 DyHead,相较基线模型精度、召回率和 mAP@0.5 分别提升了 8.5、7.6、9.7 百分点。

最终,将这 4 个改进结合至模型 I,相较于 YOLOv8s 基线模型,其精度提高了 10.4 百分点,

召回率提高了 7.6 百分点, mAP@0.5 提高了 10.2 百分点, 其中 mAP@0.5 的提升幅度达到了 25.9%。此外, 参数量减少了 2.85×10^6 , 下降幅度为 25.6%。这一切证明了模型在效率和性能方面取得了显著的进步。

3.2.6 不同剪枝率下的性能对比

为寻找最优剪枝率, 本文对改进后的模型进行不同比率的剪枝, 剪枝完成后通过对模型进行微调(finetune)来恢复精度。不同剪枝率下模型的

性能如表 7 所示。从表中数据分析得知随着剪枝率的增大, 模型的参数量和计算量都呈下降趋势。当剪枝率为 40% 时模型参数量下降 51%, 模型大小下降了 49%, 而检测性能并未明显降低。此外, 剪枝前后网络各层通道数变化如图 7 所示, 分析后不难发现通道数大幅减少发生在卷积模块, 表明卷积模块中网络冗余较高, 并且通过剪枝去掉这些冗余参数并不会对网络整体性能产生严重影响。

表 7 不同剪枝率下模型的性能

Table 7 Performance of the model under different pruning rates

剪枝率/%	Precision/%	mAP@0.5/%	mAP@0.5:0.95/%	GFLOPs	Params/10 ⁶	Model size/MB
0 (base)	58.2	49.4	30.3	42.7	8.26	16.20
20	59.9	49.7	30.8	34.1	4.86	9.78
34	58.8	49.1	30.4	28.5	4.35	8.80
40	59.0	48.9	30.3	25.5	4.05	8.21
50	58.0	47.5	29.1	21.3	3.64	7.45
67	43.1	25.4	14.5	14.2	2.80	5.88

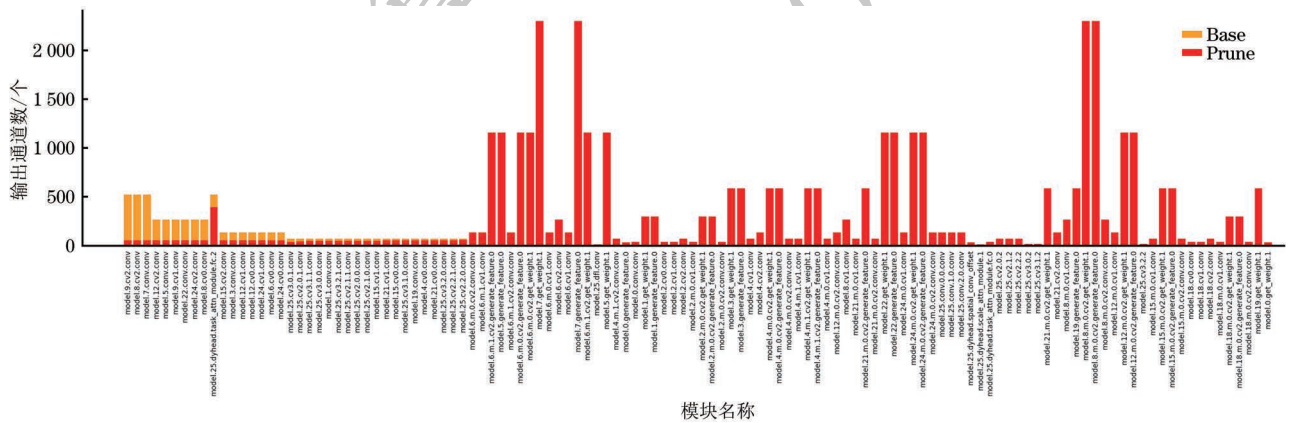


图 7 40%剪枝率下模型中各模块输出通道数的变化

Fig.7 Variation in the number of output channels of each module in the model at 40% pruning rate

3.3 对比实验

为证明本文算法的优势, 将本文算法与其他目标检测算法进行训练和测试, 其中包括目前较先进的小目标检测算法、YOLO 系列算法以及

YOLOv8s 基线网络。使用与表 1 相同的设置并对齐训练参数, 在 VisDrone2019 和 DOTA 数据集上进行实验, 对比 mAP@0.5、mAP@0.5:0.95、FPS、Params 结果如表 8 和表 9 所示。

表 8 VisDrone2019 数据集各模型对比实验结果

Table 8 Comparison experiment results of various models on the VisDrone2019 dataset

实验方法	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	FPS/(帧·s ⁻¹)	GFLOPs	Params/10 ⁶
RLSM-YOLO ^[14]	52.2	43.8	45.5	27.7	60	38.5	11.86
YOLOv5 ^[15]	41.9	32.7	32.0	18.5	126	7.1	2.50
DC-YOLO ^[29]	52.7	40.1	41.5	24.7	—	—	—
YOLOv8s (base)	49.0	39.0	39.4	23.4	256	28.5	11.12
YOLOv8l	54.5	42.3	44.2	27.4	97	207.6	436.10
Mamba-YOLO-B ^[30]	53.9	41.0	42.6	25.7	181	49.6	21.79
RT-DETR-r18 ^[31]	61.6	46.6	47.6	29.3	172	57.0	19.88
Ours	59.4	46.6	49.6	30.5	126	42.7	8.26
Ours (finetune)	59.0	45.8	48.9	30.3	171	25.5	4.05

表 9 DOTA 数据集各模型对比实验结果

Table 9 Comparison experiment results of various models on the DOTA dataset

实验方法	Precision/%	Recall/%	mAP@0.5/%	mAP@0.5:0.95/%	FPS/(帧·s ⁻¹)	GFLOPs	Params/10 ⁶
YOLOv5	50.3	48.2	46.9	33.7	666	7.1	2.50
YOLOv8s	60.1	53.7	56.1	40.6	500	28.5	11.13
YOLOv8m	67.7	58.5	62.6	45.3	285	78.7	25.84
Mamba-YOLO-B ^[30]	76.7	68.5	73.0	50.6	113	49.6	21.80
RT-DETR-r18 ^[31]	76.4	67.5	69.4	44.6	270	57.0	19.89
Ours	76.6	68.7	73.0	50.4	128	42.7	8.26
Ours (finetune)	76.0	68.4	72.2	50.0	173	25.5	4.05

在 VisDrone2019 数据集上进行各个模型的测试,对比实验结果如表 8 所示。RSD-YOLO 在召回率、mAP@0.5、mAP@0.5:0.95 上相较其他模型具有更好的性能表现。在验证集上召回率、mAP@0.5、mAP@0.5:0.95 分别达到了 46.6%、49.6%、30.5%,相比基线模型 mAP@0.5 提升了 10.2 个百分点,提升幅度为 25.9%,mAP@0.5:0.95 提升了 7.1 百分点,实现了 30.3% 的增幅,参数量减少了 2.85×10^6 。RT-DETR-r18 虽然在精度上超越了改进算法,达到 61.6%,但其计算量多出了 14.3 GFLOPs。在本文使用 LAMP 剪枝对模型进行压缩后,在模型精度基本不变的情况下减少了 40% 的模型计算量和 51% 的参数量,此外还使得模型 FPS 提升了 45 帧/s。检测速度方面,尽管与其他算法相比,改进算法的 FPS 有所下降,但考虑到小型无人机通常只需要满足 30 帧/s 的拍摄要求,而本文的算法 FPS 仍高于 100 帧/s,因此十分适用于无人机的实时场景检测。

RSD-YOLO 算法在 DOTA 数据集上同样有明显的优势,mAP@0.5 和 mAP@0.5:0.95 分别达到了 73.0% 和 50.4%,相较于基线模型分别提升了 16.9 个百分点和 9.8 个百分点。经过 LAMP 剪枝对模型进行压缩后,模型性能基本不变。对比 YOLOv8m、Mamba-YOLO-B、RT-DETR-r18,本文提出的 RSD-YOLO 算法在模型更小、参数量更少的情况下达到了更好的检测效果。而在 DOTA 数据集和 VisDrone2019 数据集上,算法表现出不同 FPS,这是由于在运行过程中的不稳定性导致的波动现象,实际检测速度并无显著差异。通过多次运行求取均值,FPS 结果基本一致。并且 FPS 计算时小数的取舍也会对 FPS 结果产生影响。

综上所述,本文的算法在航拍目标检测的精度上优于其他算法,同时实时性也得到了保证。这表明本文算法在密集小目标检测任务上能够很好地平衡检测精度与推理速度,展示了较高的实用价值。

3.4 检测结果可视化

改进算法与 YOLOv5 和 YOLOv8s 的指标对比如图 8 所示。为更直观地验证本文算法在实际场景中的表现,选取 VisDrone2019 验证集中具有代表性的场景,包括存在遮挡、密集堆叠、低亮度、复杂场景 4 种情形。与 YOLOv5s、YOLOv8s 基线模型以及 RLSM-YOLO 算法进行可视化对比,结果如图 9 所示。统计场景中所有目标与检测目标的数量,综合检测框的准确程度来看,本文算法误检率与漏检率显著改善,小目标的识别表现更优秀,表明本文算法在适应不同场景和应对目标遮挡等问题上更具鲁棒性和实用性。

4 结束语

针对无人机航拍图像小目标检测中检测性能差、特征识别不清晰、模型参数量大、对遮挡目标效果不佳的问题,本文提出了一种改进算法 RSD-YOLO。该算法首先通过 CSP-RFA 模块替换 C2f 模块增强了算法对小目标特征的提取能力;其次将浅层特征图引入特征融合网络,提出一个全新的 RFPN,并重新设计了检测头结构,提升了算法对小目标的定位和分类能力,并减少了算法的参数和计算量;然后,检测头模块通过集成不同尺度的特征,显著增强了小目标检测的表示效果,且未增加额外的计算成本;最后,替换了损失函数,加速了算法的收敛过程,并增强了对难以识别的小样本的学习能力。在公开数据集 VisDrone2019 上的实验结果显示,RSD-YOLO 算法的精度、mAP@0.5、mAP@0.5:0.95 分别为 59.4%、49.6%、30.5%,较 YOLOv8s 分别提高了 10.4、10.2、7.1 百分点,参数量从 11.12×10^6 下降到 8.26×10^6 ,减少了 25.6%。同时在 DOTA 数据集上 mAP@0.5、mAP@0.5:0.95 相较于基线模型分别提升了 16.9、9.8 个百分点,取得了良好的效果。对本模型使用 LAMP 剪枝进行模型压缩后,在模型性能基本不

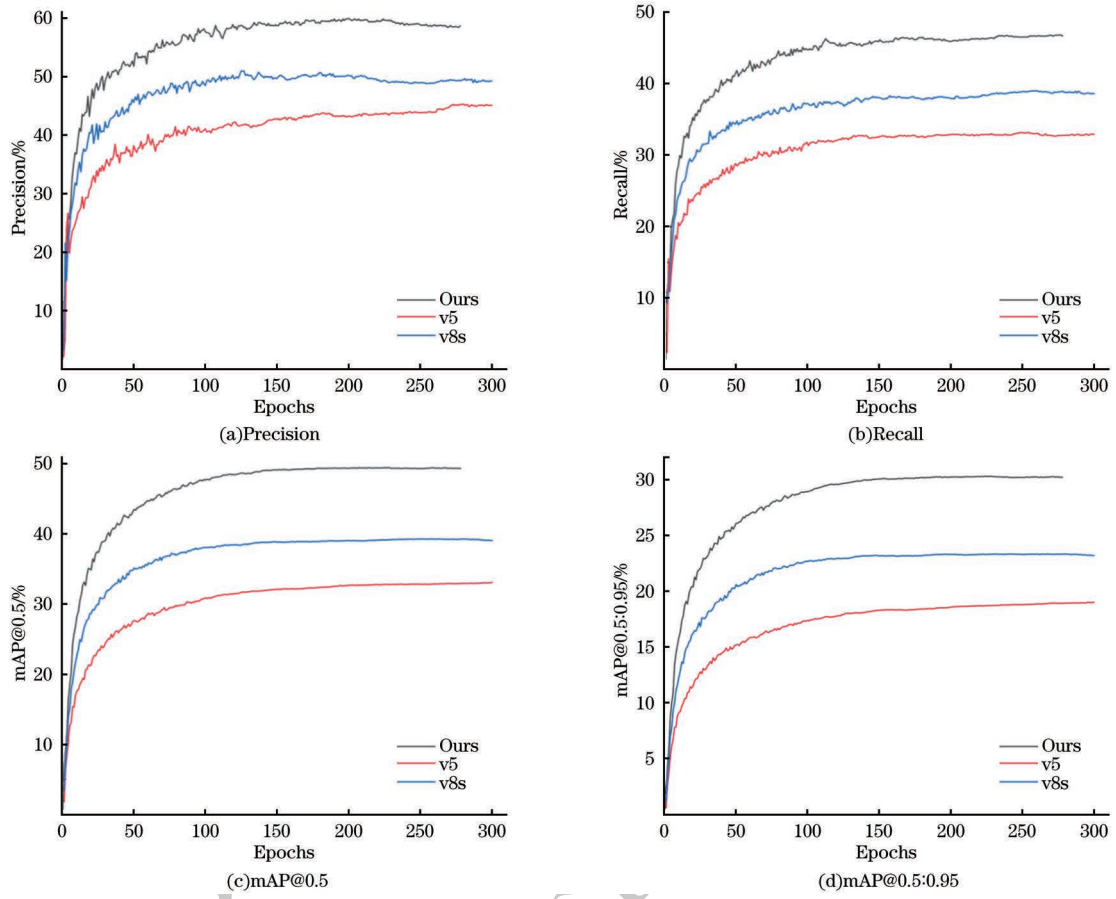


图 8 改进算法指标对比图

Fig. 8 Comparison of improved algorithm metrics

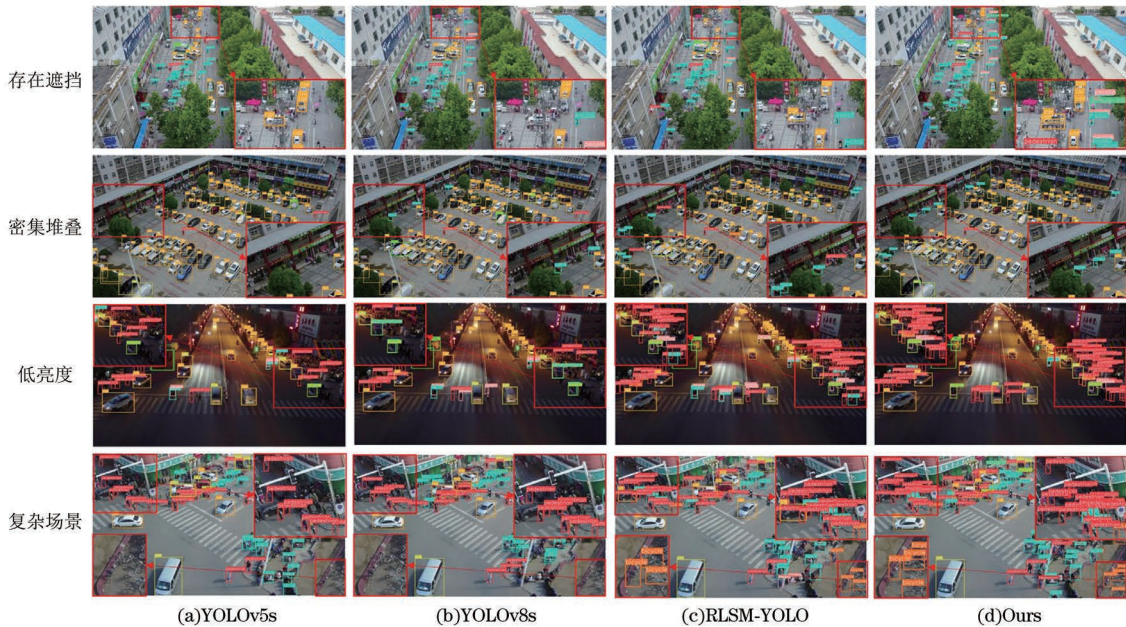


图 9 检测效果可视化对比

Fig. 9 Visualization comparison of detection results

变的情况下减少了 40% 的模型计算量和 51% 的参数数量,提高了 FPS。通过将遮挡的检测目标进行梳理,在 VisDrone2019 数据集的基础上重新制作了一个只有被遮挡目标的数据集,并且通过实验得到

本文算法在精度、召回率、mAP@0.5 上较基线模型分别提高了 9.1、16.1、10.7 个百分点。本文提出的算法对被遮挡目标有良好的检测效果。

与其他目标检测算法相比,本文提出的算法在

精度和速度方面达到了良好的平衡,具有明显的实用价值。未来工作将继续聚焦于提升算法性能,并对算法进行进一步的剪枝和轻量化处理,优化对无人机视角下密集小目标的检测效果。

参考文献

- [1] 谢竞, 邓月明, 王润民. 改进 YOLOv8s 的交通标识检测算法[J]. 计算机工程, 2024, 50(11): 338-349.
XIE J, DENG Y M, WANG R M. Improved YOLOv8s traffic sign detection algorithm[J]. Computer Engineering, 2024, 50(11):338-349. (in Chinese)
- [2] 孙亭, 杨洁, 李家璇, 等. 面向弱光交通场景的 YOLOv7 道路标志检测算法优化[J]. 计算机工程, 2025, 51(3): 342-351.
SUN T, YANG J, LI J X, et al. YOLOv7 algorithm enhancement for road sign detection in low-light traffic scenes [J]. Computer Engineering, 2025, 51(3): 342-351. (in Chinese)
- [3] DONG J, OTA K, DONG M. UAV-based real-time survivor detection system in post-disaster search and rescue operations [J]. IEEE Journal on Miniaturization for Air and Space Systems, 2021, 2(4): 209-219.
- [4] 缙永涛, 李文博, 段学锋. 基于无人机的路桥病害检测系统[J]. 测绘通报, 2022(12): 141-146, 169.
GOU Y T, LI W B, DUAN X F. Road and bridge disease detection system based on UAV[J]. Bulletin of Surveying and Mapping, 2022(12): 141-146, 169. (in Chinese)
- [5] 杨普, 赵远洋, 李一鸣, 等. 基于多源信息融合的农业空地一体化研究综述[J]. 农业机械学报, 2021, 52(S1): 185-196.
YANG P, ZHAO Y Y, LI Y M, et al. Review of research on integration of agricultural air-ground integration based on multi-source information fusion [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(S1): 185-196. (in Chinese)
- [6] 王林, 赵莉, 王无为. 高动态场景下无人机空对空目标检测方法研究[J]. 计算机工程, 2024, 50(12): 265-275.
WANG L, ZHAO L, WANG W W. Research on air-to-air target detection method for unmanned aerial vehicles in high dynamic scenarios [J]. Computer Engineering, 2024, 50(12): 265-275. (in Chinese)
- [7] COCO. Detection leaderboard [EB/OL]. [2024-06-19]. <https://cocodataset.org/#detectionleaderboard>.
- [8] DOMOZI Z, STOJCSICS D, BENHAMIDA A, et al. Real time object detection for aerial search and rescue missions for missing persons [C] // Proceedings of the 2020 IEEE 15th International Conference of System of Systems Engineering (SoSE). Washington D. C., USA: IEEE Press, 2020: 519-524.
- [9] VADDI S, KUMAR C, JANNESARI A. Efficient object detection model for real-time UAV applications [EB/OL]. [2024-06-19]. <https://arxiv.org/abs/1906.00786>.
- [10] WU C, ZHAO B J, TANG L B, et al. Small vehicles detection based on UAV[J]. The Journal of Engineering, 2019(21): 7894-7897.
- [11] CAI Y, DU D, ZHANG L, et al. Guided attention network for object detection and counting on drones[C] // Proceedings of the 28th ACM International Conference on Multimedia. New York, USA: ACM Press, 2020: 709-717.
- [12] JIANG D, SUN B, SU S, et al. FASSD: a feature fusion and spatial attention-based single shot detector for small object detection[J]. Electronics, 2020, 9(9): 1536-1555.
- [13] GUO J, LOU H, CHEN H, et al. A new detection algorithm for alien intrusion on highway [J]. Scientific Reports, 2023, 13(1): 10667.
- [14] 潘玮, 韦超, 钱春雨, 等. 面向无人机视角下小目标检测的 YOLOv8s 改进模型[J]. 计算机工程与应用, 2024, 60(9): 142-150.
PAN W, WEI C, QIAN C Y, et al. Improved YOLOv8s model for small object detection from perspective of drones [J]. Computer Engineering and Applications, 2024, 60(9): 142-150. (in Chinese)
- [15] 柳进元, 张明锋. 基于 YOLOv5 算法的图像水深自动提取[J]. 福建师范大学学报(自然科学版), 2023, 39(1): 86-92.
LIU J Y, ZHANG M F. Automatic image water depth extraction based on YOLOv5 algorithm [J]. Journal of Fujian Normal University (Natural Science Edition), 2023, 39(1): 86-92. (in Chinese)
- [16] 杨秀娟, 曾智勇. 基于 YOLOv5 的无人机航拍改进目标检测算法 Dy-YOLO[J]. 福建师范大学学报(自然科学版), 2024, 40(1): 76-86.
YANG X J, ZENG Z Y. Dy-YOLO: Improved UAV image target detection algorithm base on YOLOv5 [J]. Journal of Fujian Normal University (Natural Sciences Edition), 2024, 40(1): 76-86. (in Chinese)
- [17] 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 (CVPR'17). Washington D. C., USA: IEEE Press, 2017: 2117-2125.
- [18] ZHU L, GENG X, LI Z, et al. Improving YOLOv5 with attention mechanism for detecting boulders from planetary images[J]. Remote Sensing, 2021, 13(18): 3776.
- [19] 于傲泽, 魏维伟, 王平, 等. 基于分块复合注意力的无人机小目标检测算法[J]. 航空学报, 2024, 45(14): 629148.
YU A Z, WEI W W, WANG P, et al. Small target detection algorithm for UAV based on patch-wise co-attention [J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(14): 629148. (in Chinese)
- [20] DAI X Y, CHEN Y P, XIAO B. Dynamic head: unifying object detection heads with attentions[C] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 7369-7378.
- [21] TONG Z, CHEN Y, XU Z, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism [EB/OL]. [2024-06-19]. <https://arxiv.org/abs/2301.10051>.
- [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. Washington D. C., USA: IEEE Press, 2019: 213-226.
- [23] XIAG 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. Washington D. C., USA: IEEE Press, 2018: 3974-3983.
- [24] WOO S, PARK J, LEE J-Y, et al. CBAM: convolutional block attention module [EB/OL]. [2024-06-19]. <https://arxiv.org/abs/1807.06521>.
- [25] CAI Y X, ZHOU Y Z, HAN Q, et al. Reversible column networks [EB/OL]. [2024-06-19]. <https://arxiv.org/abs/2212.11696>.
- [26] LI J F, WEN Y, HE L H. SConv: spatial and channel reconstruction convolution for feature redundancy [C] // Proceedings of 2023 IEEE/CVF Conference on Computer

- Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 6153-6162.
- [27] QI Y L, HE Y T, QI X M, et al. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation [EB/OL]. [2024-06-19]. <https://arxiv.org/abs/2307.08388>.
- [28] LI Y, XIN Y, LI X, et al. Correction: omni-dimensional dynamic convolution feature coordinate attention network for pneumonia classification[J]. Visual Computing for Industry, Biomedicine, and Art, 2024, 7(1): 19.
- [29] 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.
- [30] CAO S Y, CHEN M X, SONG Y C, et al. MambaRefine-YOLO: a dual-modality small object detector for UAV imagery[EB/OL]. [2024-06-19]. <https://arxiv.org/abs/2511.19134>.
- [31] ZHAO Y A, LÜ W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection [EB/OL]. [2024-06-19]. <https://arxiv.org/abs/2304.08069>.

文字编辑 陆燕菲
栏目编辑 宋 圆

计算机工程
www.ecice06.com