

DOI:10.19651/j.cnki.emt.2517978

基于YOLOv8-ERD的无人机航拍小目标检测算法^{*}

蓝慧玲 刘琼

(上海电力大学数理学院 上海 201306)

摘要: 针对航拍图像中小目标检测存在的精度不足和检测困难等问题,提出了YOLOv8s的改进型小目标检测方法,即YOLOv8-ERD。首先,利用Efficient Neck方法对YOLOv8s的Neck部分进行改进,有效整合了高层语义特征与低层空间信息;接着,引入感受野注意力卷积RFACConv,通过增强感受野中关键特征的权重,进一步提升特征提取能力;随后,采用经过动态卷积DynamicConv改进的DyC2f模块替换C2f,不仅能减少计算量,还能提升模型性能;最后,增设一个小目标检测层,有效增强了模型对小目标的检测能力。实验结果表明,在Visdrone2019公开数据集上,与基准模型YOLOv8s相比,YOLOv8-ERD的mAP@0.5提高了5.0%,准确率P提高了4.0%,并且在与其他主流目标检测方法的比较中表现优异。

关键词: 无人机;小目标检测;YOLOv8s;感受野注意力;动态卷积

中图分类号: TN919.5 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Small target detection algorithm for UAV aerial images based on YOLOv8-ERD

Lan Huiling Liu Qiong

(College of Mathematics and Physics, Shanghai University of Electric Power, Shanghai 201306, China)

Abstract: Given the difficulty and low accuracy of small target detection in aerial images, YOLOv8s improved small target detection method, namely YOLOv8-ERD, was proposed. Initially, the YOLOv8s Neck network was optimized using the Efficient Neck feature fusion method to facilitate the efficient merging of high-level semantic information with low-level spatial information. Subsequently, the receptive-field attention convolutional operation RFACConv was incorporated to emphasize the significance of various features within the receptive field slider and to bolster the feature extraction capability. Additionally, the C2f module was replaced with the improved DyC2f module, which employs dynamic convolution DynamicConv, thereby not only reducing computational overhead but also enhancing model performance. Lastly, a small target detection layer was added to refine the recognition accuracy of diminutive targets. Experimental results show that on Visdrone2019 public data set, compared with the benchmark model YOLOv8s, the mAP@0.5 of YOLOv8-ERD has increased by 5.0%, and the accuracy P has increased by 4.0%, and it performs well in comparison with other mainstream target detection methods.

Keywords: unmanned aerial vehicle(UAV); small object detection; YOLOv8s; receptive-field attention convolutional operation; dynamic convolution

0 引言

随着无人机制造成本的逐渐降低以及飞行控制技术的不断进步,无人机在智慧城市管理、农业监测、电力巡检^[1]等领域的应用范围不断扩展。然而,由于无人机航拍视角的固有特性,航拍目标通常存在像素占比低、特征表征能力弱等不足,且在密集场景中容易受到目标间重叠效应的干

扰。无人机飞行高度的动态变化也会导致同一目标表现出显著的多尺度特征差异,从而增加检测难度。同时,复杂的光照条件和环境干扰会引发背景噪声,导致检测目标与背景特征的高度耦合。无人机检测过程中存在的这些问题对遥感检测算法,尤其是小目标检测的鲁棒性、实时性和泛化性能提出了更高的要求。

在遥感图像目标检测领域,深度学习技术以其出色的

收稿日期:2025-01-20

*基金项目:国家自然科学基金(11801356)项目资助

特征提取能力,逐步发展成为主流技术方法。现有方法主要分为两类:基于候选区域和基于回归的检测框架。前者以 R-CNN 及其改进模型^[2-4]为代表,通过生成候选区域实现高精度检测,但计算开销较大且速度较慢;后者如 SSD 和 YOLO 系列^[5-7],直接在特征图上进行目标定位,推理速度快,适合实时应用,但精度可能略低。此外,鉴于 Transformer^[8]模型在自然语言处理领域的优异表现,Carion 等^[9]将该其引入计算机视觉领域,提出了 DETR 模型,为目标检测提供了新的研究思路。

在遥感图像中,小目标的像素占比相对较低,目标的尺度变化多样且背景复杂。因此,将目前广泛使用的检测算法直接应用于无人机拍摄的场景,存在一定的局限性。近年来,针对无人机航拍目标检测的研究逐渐增多。Luo 等^[10]通过优化 YOLOv5 的网络模块提升了检测性能,但由于模型在浅层特征的空间细节保留方面存在不足,因此在小目标检测方面仍未能达到理想的效果。Zhou 等^[11]从数据增强的角度提出背景替换方法,缓解了无人机图像背景单一的问题,但未能显著提升小目标的检测精度。Du 等^[12]则从轻量化角度设计了稀疏卷积检测头,虽然减少了计算量,却牺牲了检测精度。为改善小目标检测效果,研究者们提出了多种改进方案。Fu 等^[13]提出的反卷积单发检测器(DSSD)通过反卷积融合深层与浅层特征,提升了小目标检测精度,但计算成本较高。Liu 等^[14]使用 Efficientlite 替代 YOLOv5 的主干网络,降低了模型的复杂度并提升了多旋翼无人机的检测精度,但该方法在通用性方面表现不足。Wang 等^[15]通过聚合-分发(gather-and-distribute, GD)机制将全局信息注入到各个层级,显著提高了多尺度信息的融合能力,但计算复杂度较高,限制了其在实时应用场景中的部署。闫钧华等^[16]提出 CC-YOLO 算法,通过引入注意力机制和执行多层次特征提取与融合,提升了检测小目标的精度和速度。但该算法使用的自建数据集可能限制其泛化能力。朱圣博等^[17]提出了多头感知注意力检测头(dynamic head, Dyhead),在特征、空间和输出通道 3 个维度上进行感知,以提高目标检测的准确性。然而,由于计算成本的考虑,研究中仅使用了一个 Dyhead,这限制了小目标检测精度的进一步提升。

尽管现有方法在一定程度上提升了无人机目标检测的性能,但在模型精度与资源消耗之间仍难以实现有效平衡。本文旨在改进 YOLOv8s 模型,在不显著增加计算成本的前提下,提升无人机场景下小目标检测的精度,从而优化模型整体性能。具体而言,首先通过优化 Neck 部分,采用改进的 Efficient Neck 替代基准模型中的 Neck,以更好地融合高层语义信息与低层空间特征;其次,引入感受野注意力卷积(receptive-field attention convolutional operation, RFACConv)替换 Backbone 和 Efficient Neck 中的传统卷积,强化感受野内不同特征的重要性,提升特征提取能力;接着,使用动态卷积(DynamicConv)改进的 DyC2f 模块替

代 C2f,在降低计算量的同时提升模型性能;最后,增加一个小目标检测层,进一步提高对小目标的检测精度。

1 YOLOv8 介绍

在 2023 年 1 月,Ultralytics 公司发布了 YOLOv8 目标检测算法。YOLOv8 相较于其前身 YOLOv5 和 YOLOv7,在检测精度和速度上都有显著提升,彰显了其技术的优越性。该算法的网络架构主要由 3 个核心部分构成:Backbone、Neck 和 Head,如图 1 所示。

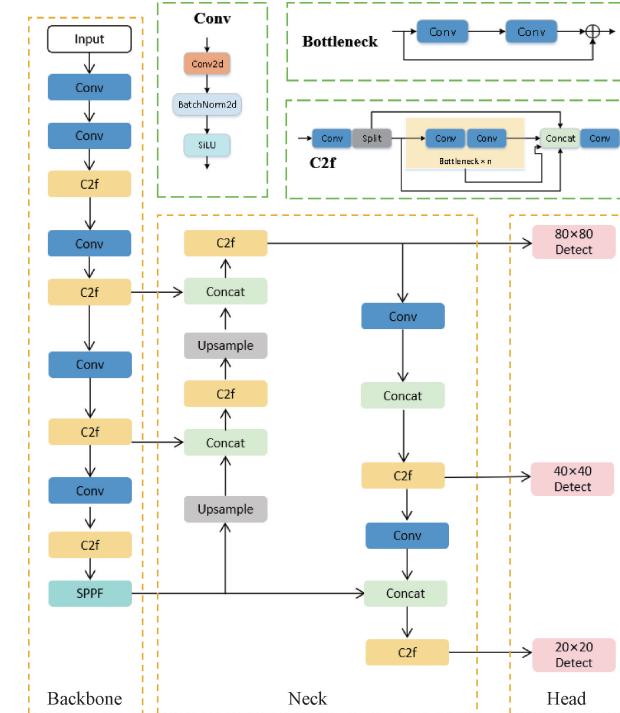


图 1 YOLOv8s 网络结构

Fig. 1 YOLOv8s network structure

YOLOv8 的主干网络基于 CSPDarknet53^[7],通过五次下采样提取多尺度特征。将 CSP 模块升级为 C2f 模块,利用梯度分流技术来强化特征信息流,同时确保模型保持轻巧。Conv 模块融合了卷积运算、批量归一化以及 SiLU 激活功能。主干网络最终通过 SPFF 模块将特征图池化为固定尺寸输出。

传统的特征金字塔网络(feature pyramid network, FPN)采用自上而下的方式传递深层语义信息,虽然增强了语义特征,但会损失部分目标定位信息。借鉴路径聚合网络(path aggregation network, PANet^[18])的设计,YOLOv8 在 Neck 部分设计采用了 PAN-FPN 结构,构建了一个结合上下路径的网络架构,实现了浅层位置信息与深层语义信息的融合,确保了特征多样性和完整性。

采用解耦设计的 YOLOv8 检测头,通过两个独立分支处理目标分类和边界框回归任务,不仅提高了检测精度,还加速了模型的收敛速度。

2 小目标检测方法 YOLOv8-ERD

针对 YOLOv8s 模型在遥感小目标检测中精度低、难度大的问题，并结合遥感小目标数据集中目标像素占比少、类别分布不均衡、目标尺度变化大等特点，本文提出了一种改进的检测网络模型 YOLOv8-ERD，改进的模型如图 2 所示。通过多尺度特征融合、动态卷积优化以及增强小目标检测能力等策略，显著提升了模型在复杂遥感场景下的检测性能与鲁棒性。具体改进如下：

1) 借鉴 DAMO-YOLO 和 BGF-YOLO 的设计思路, 优化了 Neck 结构, 采用改进的 Efficient Neck 特征融合技术, 实现了高层语义信息与低层空间信息的深度融合。

2)引入RFACconv,替代模型中的标准卷积Conv,解决了卷积核参数共享的局限性,从而提升了模型的特征提取能力。

3)利用 DynamicConv 技术对原始的 C2f 模块进行优化,构建了新的 DyC2f 模块,在降低内存占用的同时,高效提取丰富的语义信息,进一步提升了模型的整体性能。

4)加入一个 160×160 尺寸的小目标检测模块,增强模型对小型物体识别的能力。

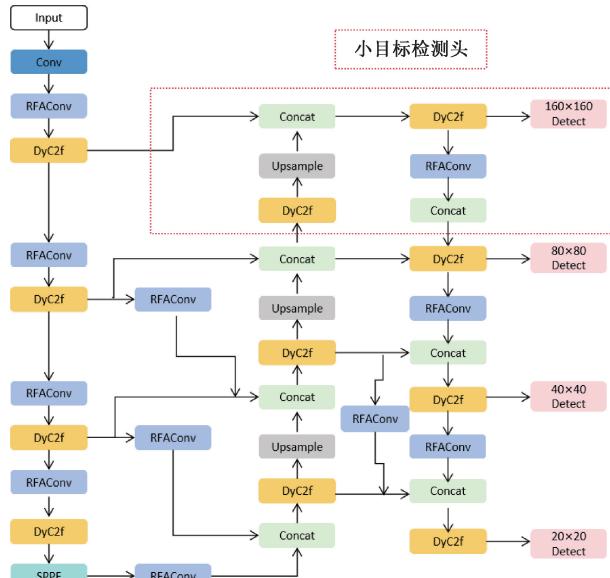


图 2 YOLOv8-ERD 结构

Fig. 2 YOLOv8-ERD structure

2. 1 Efficient Neck

在目标检测领域，Neck 结构负责整合高级语义信息与低级空间特征。尽管 YOLOv8 所采用的 PAN-FPN 架构在一定程度上增强了对小目标的检测能力，但在特征传递过程中，信息损失的问题依然存在。特别是低层次的细节特征在经历多次上采样与下采样后逐渐减弱，这导致了小目标定位精度的降低。此外，PAN-FPN 在融合不同层级特征时采用的方式较为单一，这在背景复杂或目标密集的

场景中容易导致漏检或误检的问题。针对这些问题，本文对 PAN-FPN 进行了优化，通过强化多路径融合机制，实现了更为有效不同的层级特征整合。本文借鉴了 DAMO-YOLO^[19] 和 BGF-YOLO^[20] 中 Neck 网络的设计理念，引入额外的卷积操作以深化特征提取，并增设跳跃连接，使得模型能够在 Neck 部分充分融合高层语义信息与低层空间信息，优化后的 Efficient Neck 网络结构如图 3 所示。

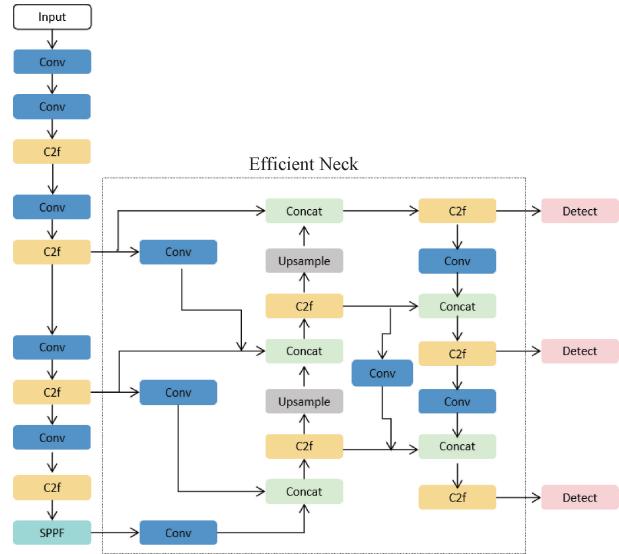


图 3 Efficient Neck 网络结构

Fig. 3 Efficient Neck network structure

图 3 中 Efficient Neck 的特征融合机制通过多阶段处理实现高效信息整合。首先, Backbone 提取的特征在输入 Neck 模块后, 并未直接进行上采样或特征拼接, 而是通过 1×1 和 3×3 卷积核的组合进一步提取特征的局部和全局上下文信息, 以实现全局特征的深度聚合。其次, 在 PAN-FPN 结构的跨层连接中, SPPF 层通过多尺度池化操作生成的特征与 Backbone 中 C2f 模块输出的 40×40 尺度特征分别经过卷积增强后, 在通道维度上进行特征融合。与此同时, 40×40 尺度的特征还会与经过卷积优化的 80×80 浅层细节特征直接融合, 从而实现了抽象语义特征与细节空间特征的深度融合, 显著减少了信息在层级传递中的衰减问题。最后, 通过从多个尺度的特征图中提取信息并进行融合, 再送入后续处理模块。与传统的 FPN 结构相比, Efficient Neck 的设计模式能够显著提升多尺度特征的融合效果, 尤其在目标检测任务中, 其对小目标检测的鲁棒性提升尤为明显。

2.2 RFACConv

在标准卷积操作中，特征提取过程采用统一的卷积核，意味着每个滑动窗口内相同位置的特征均共享同一组参数。这种设计虽然简化了计算过程，但卷积核参数固定，无法根据输入特征动态调整，使得模型对图像特征响应不灵活，难以捕捉关键信息。此外，标准卷积对特征空间位置变

化不敏感, 难以区分不同位置特征的重要性, 影响了模型提取细节特征的能力, 特别是在小目标检测中容易导致特征信息丢失。针对这一问题, 本文引入感受野注意力卷积 RFACConv^[21], 结合了空间注意力和传统卷积技术, 突出了感受野滑块内各种特征的重要性。此外, 它优先处理感受野内的空间特征, 使网络能够更有效地处理图像的局部区域, 从而提升特征提取的精确度。图 4 展示了具有 3×3 卷积核的 RFACConv 结构。

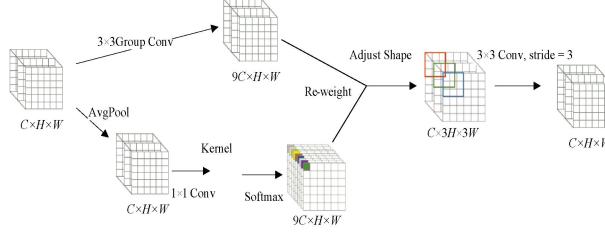


图 4 RFACConv 详细结构

Fig. 4 The detailed structure of RFACConv

图 4 中 RFACConv 先利用全局平均池化 (global average pooling, AvgPool) 对输入特征图 X 进行空间维度上的特征压缩, 以提取每个感受野的全局上下文信息。随后, 通过 1×1 分组卷积对特征通道数进行调整, 在降低计算复杂度的同时, 促进不同通道间的特征交互, 从而提取更具判别性的特征表示。接下来, 对经过 1×1 卷积处理的全局特征应用 Softmax 函数, 生成特征重要性权重矩阵, 用于量化每个特征的贡献度, 同时抑制冗余信息的干扰。与此同时, 对输入特征图 X 进行 $k \times k$ 卷积操作并通过 ReLU 和归一化处理, 实现特征的标准化。最终, 将 Softmax 生成的权重矩阵与标准化后的特征图进行点乘操作, 得到融合后的输出特征图。这一过程通过高效聚合全局信息、促进特征间相互作用、合理加权特征重要性以及标准化处理, 显著增强了卷积神经网络在特征表达和信息整合方面的能力。

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

式中: F 是注意力空间特征图 A_{rf} 与经过变换的感受野空间特征图 F_{rf} 相乘所得。 $g^{i \times i}$ 代表尺寸为 $i \times i$ 的分组卷积操作, k 指的是卷积核尺寸, Norm 代表归一化操作, X 是输入的特征图。

RFACConv 通过引入注意力机制, 增强了感受野内特征的重要性, 解决了传统卷积对位置变化敏感度不足的问题。它允许特征根据重要性动态调整权重, 实现非共享特性。与传统卷积相比, RFACConv 在增加少量参数和计算开销的情况下, 显著提升了模型性能。

2.3 DyC2f

由于 YOLOv8 算法中频繁的下采样操作, 使得 C2f 模块在处理小目标特征时表现欠佳, 细节信息在传递过程中逐渐丢失, 且其固定的卷积核结构和参数共享机制限制了模型对动态特征的捕捉能力, 尤其在复杂背景或多尺度目

标场景中表现不足。此外, C2f 模块的计算量较大, 在处理高分辨率图像时浮点运算次数 (FLOPs) 显著增加, 影响了模型的实时性。针对这些问题, 本文提出了 DyC2f 模块, 通过引入动态卷积技术, 实现了卷积核参数的动态调整, 在增强特征提取能力的同时, 显著降低了计算开销, 有效提升了模型对小目标的检测性能与计算效率。

动态卷积 DynamicConv^[22] 技术能够在不显著增加 FLOPs 的情况下, 将额外的参数引入网络中。这种设计使得模型在保持计算效率的同时, 拥有了更多的参数来“记忆”复杂的知识, 从而提高了模型的性能, 具体结构如图 5 所示。

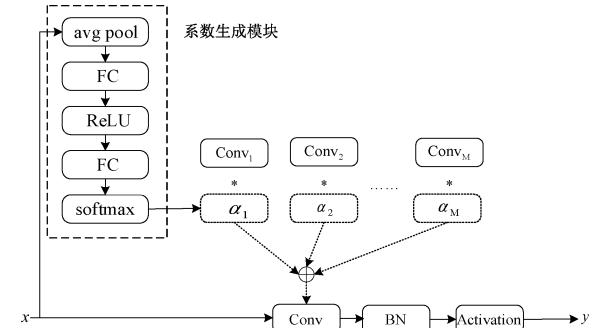


图 5 DynamicConv 结构

Fig. 5 DynamicConv structure

DynamicConv 利用系数生成模块 (如 MLP 模块) 根据输入特征动态生成权重系数, 随后将这些系数应用于不同的卷积核, 实现卷积核的动态选择与加权。具有 M 个动态专家的动态卷积操作可以写成:

$$Y = X * W' \quad (2)$$

$$W' = \sum_{i=1}^M \alpha_i W_i \quad (3)$$

式中: $W_i \in R^{C_{out} \times C_{in} \times H \times W}$ 是第 i 个动态卷积核的参数, α_i 是对应的动态超参数, 随不同的输入而动态地生成。典型的生成策略是使用 MLP 模块根据输入生成, 即: 对于输入 X , 应用全局平均池化将信息融合为向量, 然后使用具有 Softmax 激活的两层 MLP 模块动态生成系数:

$$\alpha = \text{Softmax}(\text{MLP}(\text{Pool}(X))) \quad (4)$$

式中: $\alpha \in R^M$ 。与传统卷积相比, DynamicConv 只引入了动态生成系数 α , 在使用更多参数学习模型的同时最大限度地减少了 FLOPs 的增加。

DyC2f 模块的核心设计思路是将 YOLOv8 中 C2f 模块的 Bottleneck 层第 2 个标准卷积替换为 DynamicConv, 从而在保留原有结构优势的同时, 进一步提升特征提取能力并优化计算效率。如图 6 所示, 改进后的 Bottleneck 模块首先通过 3×3 卷积进行初步特征提取, 随后利用 DynamicConv 动态调整卷积核参数, 进一步挖掘特征的上下文信息并显著降低计算量, 最后通过残差连接将输出特征与输入特征融合, 有效缓解了梯度消失问题并增强了特

征的表达能力。图 7 是 DyC2f 模块的整体结构,其通过动态卷积技术在不显著增加计算开销的前提下,显著提升了模型性能。基于这一优势,本文在 YOLOv8-ERD 模型中将所有 C2f 模块替换为 DyC2f 模块,进一步优化了模型的计算效率与检测精度。



图 6 改进后的 Bottleneck 模块
Fig. 6 Improved Bottleneck module

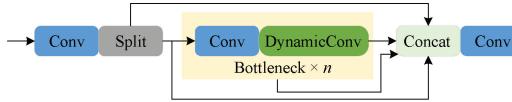


图 7 DyC2f 模块
Fig. 7 DyC2f module

2.4 小目标检测层

原始 YOLOv8s 模型有 3 个检测头,尺寸分别为 20×20 、 40×40 和 80×80 。虽然能满足大多数通用目标检测任务的需求,但在处理小目标占比较高的无人机航拍场景时,尤其是针对 VisDrone2019 数据集,其表现存在明显不足。由于多次下采样操作,小目标的细节特征在传递过程中逐渐丢失,导致模型对小目标的检测能力显著下降,容易出现漏检或误检现象。

为了解决上述问题,本文引入了一个 160×160 的小目标检测头,以增强模型对图像中小目标的检测能力,具体可见图 2。首先,利用 DyC2f 模块和上采样技术对 Efficient Neck 层输出的 80×80 尺度特征进行提取;随后,将这些特征与 Backbone 中 160×160 尺度的浅层特征进行通道维度上的特征融合,从而获得同时包含高层抽象语义特征和低层空间细节特征的特征层。接下来,这些融合后的特征被送入 DyC2f 模块进行进一步优化,以提升对小目标的检测能力。通过这一优化的特征提取流程,模型不仅能够保留丰富的语义信息,还提升了对小目标位置的敏感性。新设计的小目标检测模块使得小目标的特征信息能够传递到其他尺度的特征层,并能从深层网络中提取出细节特征,进而显著降低小目标检测中的误报率和漏检率。与传统的单尺度检测头相比,多尺度检测头能更好地适应不同尺寸目标的检测需求,尤其适用于小目标检测任务。

3 实验

3.1 数据集与实验环境

本文采用天津大学机器学习与数据挖掘实验室公开的 VisDrone2019 数据集^[23]。该数据集包含 10 类航拍目标,例如行人、汽车等,分为训练集(6 471 张)、验证集(548 张)和测试集(1 610 张)。数据集中小目标占比较高,其中像素面积小于 16×16 的目标占比 12.05%,小于 32×32 的目标占比 44.7%。训练集中共包含 353 550 个目标,其中近半

数目标存在遮挡问题^[24]。

实验环境基于 Ubuntu 20.04 系统,硬件配置包括:12 vCPU Intel(R) Xeon(R) Platinum 8255C 处理器,24 GB 显存的 NVIDIA RTX 3090 显卡。软件环境为 Python 3.8、PyTorch 1.11.0 和 CUDA 11.3。训练过程中未使用任何预训练模型,所有实验均采用默认超参数设置。

3.2 评价指标

实验性能评估主要采用以下指标:准确率(P)、召回率(R)、平均精度平均值(mAP)以及浮点操作次数(GFLOPs)。

准确率 P 衡量了模型对正样本进行正确分类的性能,具体计算公式为:

$$P = \frac{TP}{TP + FP} \quad (5)$$

召回率 R 展示了模型在全面检测目标方面的性能,其计算公式为:

$$R = \frac{TP}{TP + FN} \quad (6)$$

mAP 是评估目标检测模型性能的关键指标,其综合了各类别的检测效果。mAP 值越高,说明模型在检测小目标方面表现越好。mAP 的计算公式为:

$$mAP = \frac{\sum_i^N AP_i}{N} \quad (7)$$

其中, AP_i 表示单个类别的平均精度, N 表示数据集中目标类别的总数。

GFLOPs 是衡量模型计算复杂度的指标,其值越小,表明模型所需的计算资源越少,计算效率越高。

3.3 对比实验

为了验证改进模型在检测性能方面的提升,本文对 YOLOv8-ERD 和原始 YOLOv8s 在保持一致的参数配置下,各自进行了 200 个周期的训练,并在测试集上进行了相应的评估。图 8 展示了 YOLOv8-ERD 与 YOLOv8s 在训练过程中 mAP 随训练轮数的变化曲线。从图 8 中可以看出,经过 200 轮训练后,模型性能趋于稳定,且改进后的 YOLOv8-ERD 在 mAP 上显著优于基准模型。

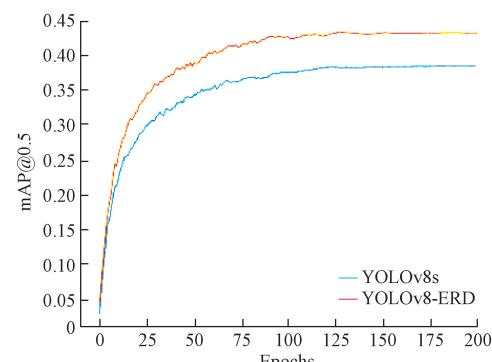


图 8 YOLOv8-ERD 和 YOLOv8s mAP@0.5 对比
Fig. 8 Comparison of YOLOv8-ERD and YOLOv8s mAP@0.5

表 1 展示了改进后的模型 YOLOv8-ERD 与 YOLOv8s 在测试数据集上的性能对比。观察表 1 可知, 经过优化的 YOLOv8-ERD 在 mAP@0.5 和 P 这两个评价指标上均超越了 YOLOv8s。

表 1 指标对比

Table 1 Index comparison

评估指标	YOLOv8s	YOLOv8-ERD
P/%	43.5	47.5
mAP@0.5	31.0	36.0

为进一步验证本文提出的改进算法 YOLOv8-ERD 的优越性, 在 VisDrone2019 数据集上进行了对比实验, 结果如表 2 所示。实验中对比了多种模型, 包括基于候选区域的 Faster-RCNN、基于回归的 SSD 和 RetinaNet、YOLO 系列模型, 以及最新的深度学习方法如 VanillaNet 和 RT-DETR-L。实验结果表明, 基于回归的模型计算量更小, 更适合无人机检测任务; 其中, YOLO 系列在保持较低 GFLOPs 的同时, 检测精度表现优异。本文提出的 YOLOv8-ERD 在 mAP@0.5 上达到 36.0%, 相比 Faster-RCNN、SSD、RetinaNet 和 RT-DETR-L 分别提升了 2.9%、12.1%、13.6% 和 12.5%, 相比 YOLOv5s、YOLOX-s、YOLOv7-tiny 和 YOLOv8s 分别提升了 5.2%、2.1%、3.3% 和 5.0%。综合分析, 本文方法在无人机航拍

图像的小目标检测精度和计算效率方面具有显著优势。

表 2 对比实验

Table 2 Comparative experiments

方法	mAP@0.5	mAP@0.5:0.95	GFLOPs
Faster-RCNN	33.1	17.0	206.7
SSD	23.9	11.7	87.9
RetinaNet	22.4	11.8	93.7
VanillaNet	33.8	19.1	92.8
RT-DETR-L	23.5	12.6	100.9
YOLOv5s	30.8	16.7	16.5
YOLOX-s	33.9	19.9	26.8
YOLOv7-tiny	32.7	16.1	13.1
YOLOv8s	31.0	17.6	28.5
YOLOv8-ERD	36.0	20.2	33.0

3.4 消融实验

为了检验本研究提出的优化模块对模型检测性能的增强作用, 进行了消融实验来评估不同模块对模型检测性能的影响。实验选取 YOLOv8s 作为基准模型, 并逐步集成优化模块。首先, 实施了 Efficient Neck 结构, 随后加入了改进的卷积操作 RFACConv, 接着引入了 DyC2f 模型, 最终加入了针对小目标的检测头。在一致的实验条件下, 对各个模块对检测效果的影响进行了分析, 具体实验结果如表 3 所示。

表 3 消融实验结果

Table 3 Ablation results

方法	P/%	R/%	mAP@0.5	mAP@0.5:0.95	GFLOPs
YOLOv8s	43.5	32.8	31.0	17.6	28.5
+ Efficient Neck	45.0	32.7	31.7	18.0	29.3
+ RFACConv	43.8	34.2	32.3	18.6	29.1
+ DyC2f	43.3	33.1	31.6	18.1	23.8
+ 小目标检测头	45.9	35.4	34.8	19.7	37.0
+ Efficient Neck+RFACConv	45.2	34.5	32.9	18.7	30.4
+ Efficient Neck+RFACConv+DyC2f	45.6	34.3	33.4	18.9	25.7
Efficient Neck+RFACConv+DyC2f+小目标检测头	47.5	36.7	36.0	20.2	33.0

根据表 3 的数据显示, 经过优化模块的整合, 小目标检测任务的平均精度均值 mAP 普遍得到了提升。在 YOLOv8s 模型中, 将原有的 Neck 结构替换成 Efficient Neck 后, 精确度提高了 1.5%, mAP@0.5 和 mAP@0.5:0.95 指标分别增长了 0.7% 和 0.4%, 这表明 Efficient Neck 在特征融合方面表现更为优异; 引入 RFACConv 模块, 取代基准模型中的传统卷积后, 召回率提高了 1.4%, mAP@0.5 和 mAP@0.5:0.95 指标分别提升了 1.3% 和 1.0%; 在 YOLOv8s 模型中应用 DyC2f 后, mAP@0.5 提高了 0.6%, mAP@0.5:0.95 提高了 0.5%, 同时模型的计

算量减少了 4.7 GFLOPs; 引入小目标检测头, 模型 mAP@0.5 提升了 3.8%, mAP@0.5:0.95 提高了 2.1%。同时采用 Efficient Neck 和 RFACConv 替代 YOLOv8s 的 Neck 和模型中的传统卷积 Conv 后, mAP@0.5 和 mAP@0.5:0.95 分别提高了 1.2% 和 0.7%; 继续采用动态卷积 DynamicConv 与网络中的 C2f 进行融合, 可以在确保准确率的同时有效降低模型的计算量, 模型 mAP 增加了 0.5%, 计算量减少了 4.7 GFLOPs, 效果显著; 最后将 Efficient Neck, RFACConv, DyC2f 和小目标检测头共同引入基准模型, 与基准模型相比, mAP@0.5 提升了 5.0%,

mAP@0.5:0.95 提高了 2.6%，计算量仅升高 4.5 GFLOPs。

图 9 展示了本文改进方法与基线模型在测试集上针对各个类别以及整体的平均精度均值 mAP 和精确率-召回率 PR 曲线。图 9(a)是 YOLOv8s 模型的 PR 曲线, 图 9(b)为本研究提出的 YOLOv8-ERD 模型的 PR 曲线。观察图 9 可知, 改进后的 YOLOv8-ERD 模型在数据集的每个类别上均实现了不同程度的性能提升, 从而彰显了模型卓越的泛化性能。

3.5 可视化分析

为验证 YOLOv8-ERD 算法的鲁棒性, 使用 Visdrone2019 测试集中的白天和夜间场景图像进行实验, 涵盖背景复杂、目标密集和多尺度目标的情况。如图 10 所示, 图 10(a)展示了在日间及夜间条件下, 背景复杂图像的检测效果, 相较于 YOLOv8, YOLOv8-ERD 在不同光照条件下能更有效地区分背景与目标, 并检测到更多小目标。图 10(b)展示了目标密集图像的检测效果, YOLOv8-ERD 在检测精度和漏检率方面均优于 YOLOv8, 表现出更强的鲁棒性。图 10(c)展示了多尺度目标图像的检测效果, YOLOv8-ERD 在多尺度目标检测中表现优异, 错检和漏检现象显著减少。综合以上分析, 可以得出结论, YOLOv8-ERD 在这些特定的使用场景中展现出了更加卓越的检测能力。它在精确度和可靠性方面都表现得更为出色, 因此, YOLOv8-ERD 特别适合那些对检测精度和系统稳定性有着较高要求的应用场合。

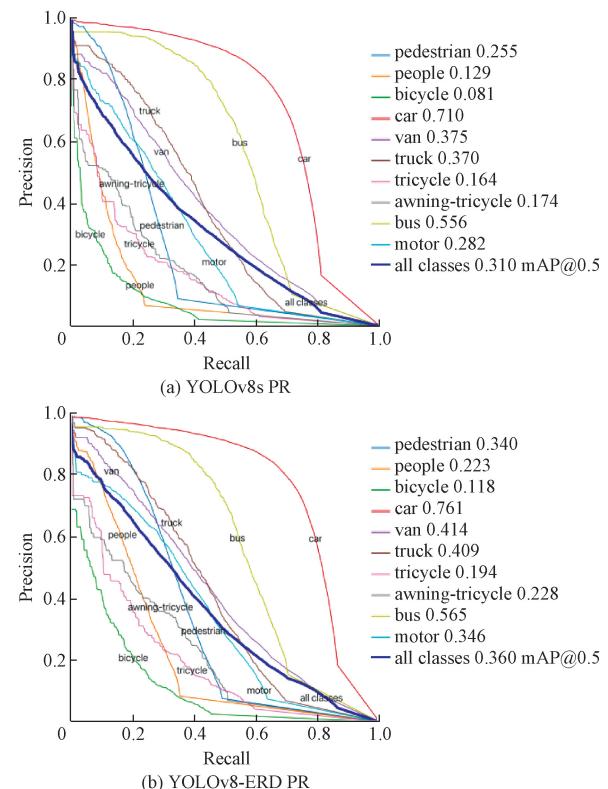


图 9 YOLOv8s 与 YOLOv8-ERD PR 曲线对比图

Fig. 9 Comparison of YOLOv8s and YOLOv8-ERD PR curves

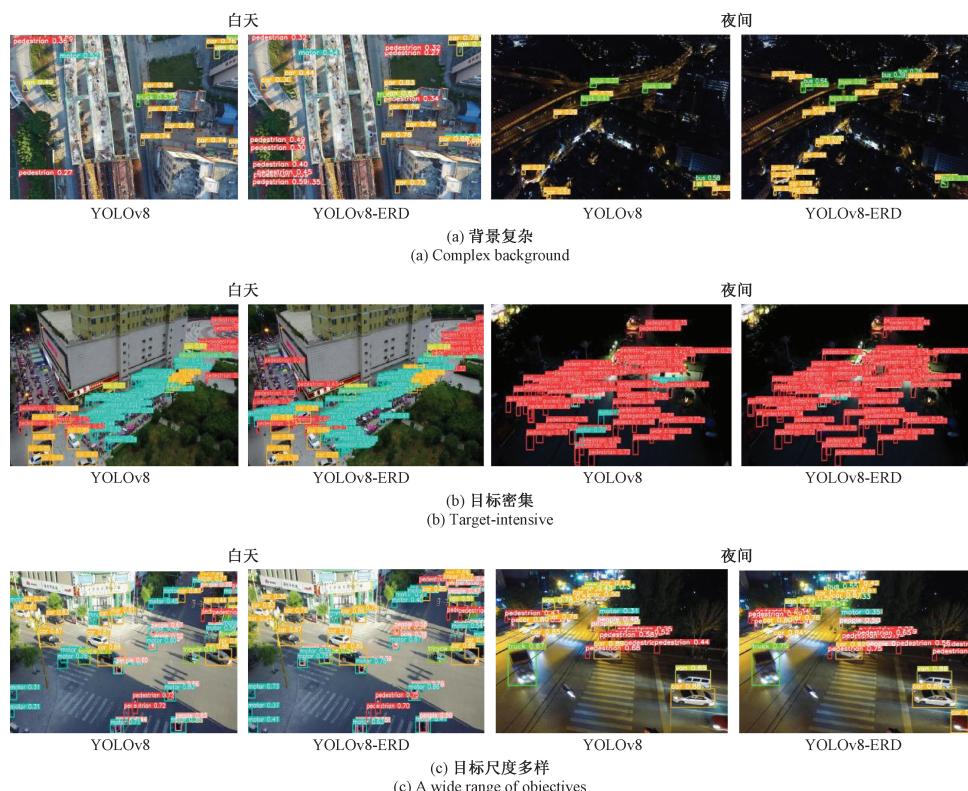


图 10 检测效果对比图

Fig. 10 Comparison of detection results

4 结 论

针对航拍小目标检测任务所面临的检测难度大和精度低等挑战,本文提出了YOLOv8-ERD小目标检测模型。该模型的Neck使用Efficient Neck特征融合方法,实现高层语义信息和低层空间信息充分融合;引入RFACConv模块替换传统卷积,提升特征提取能力;使用DyC2f模块减少计算量并提升模型准确度;最后,增加小目标检测层,提升小目标识别精度。在Visdrone2019数据集上,准确率P、召回率R、mAP@0.5、mAP@0.5:0.95分别提升4.0%、3.9%、5.0%、2.6%,在对比的8个主流目标检测模型中,检测精度最高,GFLOPs最小,证明了本文方法的有效性。

由于改进了Neck网络,增加了小目标检测层,因此本文的模型相较于基准模型计算量略有增加。虽然引入了轻量化模块DyC2f,但是模型计算量减少有限,后续的工作致力于保证模型精度的同时进一步对模型做轻量化处理。

参考文献

- [1] 黄郑,卞尹蕾,王红星,等.无先验地图条件下电力管廊无人机自主巡检方法研究[J].电子测量技术,2022,45(19):30-35.
- HUANG ZH, BIAN Y L, WANG H X, et al. Research on autonomous inspection method for power pipe gallery UAV without prior map conditions [J]. Electronic Measurement Technology, 2022, 45(19): 30-35.
- [2] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [3] REN SH Q, HE K M, GIRSHICK R, et al. Faster RCNN: Towards real time object detection with region proposal networks[J]. International Journal for Light and Electron Optics, 2017(39): 1137-1149.
- [4] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision, 2016: 21-37.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [7] REDMON J, FARHADI A. YOLOv3: An incremental improvement[C]. Computer Vision and Pattern Recognition, 2018: 1-6.
- [8] VASWANI A, SHAZEEB N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [9] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision, 2020: 213-229.
- [10] LUO X D, WU Y Q, WANG F Y. Target detection method of UAV aerial imagery based on improved YOLOv5[J]. Remote Sensing, 2022, 14(19): 5063.
- [11] ZHOU H L, MA A T, NIU Y F, et al. Small-object detection for UAV-based images using a distance metric method[J]. Drones, 2022, 6(10): 308.
- [12] DU B W, HUANG Y CH, CHEN J X, et al. Adaptive sparse convolutional networks with global context enhancement for faster object detection on drone images [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 13435-13444.
- [13] FU CH Y, LIU W, RANGA A, et al. DSSD: Deconvolutional single shot detector [J]. ArXiv preprint arXiv:1701.06659, 2017.
- [14] LIU B L, LUO H. An improved YOLOv5 for multi-rotor UAV detection[J]. Electronics, 2022, 11(15): 2330.
- [15] WANG CH CH, HE W, NIE Y, et al. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism [J]. Advances in Neural Information Processing Systems, 2023, 36: 51094-51112.
- [16] 闫钧华,张琨,施天俊,等.融合多层级特征的遥感图像地面弱小目标检测[J].仪器仪表学报,2022,43(3): 221-229.
- YAN J H, ZHANG K, SHI T J, et al. Detection of weak ground targets in remote sensing images by fusing multi-level features [J]. Journal of Instrumentation, 2022, 43(3): 221-229.
- [17] 朱圣博,魏利胜,高港,等.基于改进YOLOv8s的光学遥感小型船舶检测算法[J].电子测量与仪器学报,2024,38(10):48-57.
- ZHU SH B, WEI L SH, GAO G, et al. Optical remote sensing small vessel detection algorithm based on improved YOLOv8s [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(10): 48-57.
- [18] LIU SH, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [19] XU X Z, JIANG Y Q, CHEN W H, et al. DAMO-YOLO:A report on real-time object detection design[J].

- ArXiv preprint arXiv:2211.15444, 2022.
- [20] KANG M, TING CH M, TING F F, et al. BGF-YOLO: Enhanced YOLOv8 with multiscale attentional feature fusion for brain tumor detection [C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2024: 35-45.
- [21] ZHANG X, LIU C, YANG D, et al. RFACConv: Innovating spatial attention and standard convolutional operation[J]. ArXiv preprint arXiv:2304.03198, 2023.
- [22] HAN K, WANG Y H, GUO J Y, et al. ParameterNet: Parameters are all you need for large-scale visual pretraining of mobile networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 15751-15761.
- [23] DU D W, ZHU P F, WEN L Y, et al. VisDrone-
- DET2019: The vision meets drone object detection in image challenge results[C]. IEEE/CVF International Conference on Computer Vision Workshops, 2019: 213-226.
- [24] 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 ZH, YI H H, ZHENG J. Aerial image target detection algorithm focusing on small objects[J]. Journal of Electronics, 2023, 51(4): 944-955.

作者简介

蓝慧玲,硕士研究生,主要研究方向为计算机视觉、小目标检测。

E-mail:lanhuiling01@163.com

刘琼(通信作者),博士,副教授,主要研究方向为大数据分析。

E-mail:sky200547@126.com