

文章编号:1671-6833(2025)04-0062-08

基于 FEW-YOLOv8 遥感图像目标检测算法

席阳丽¹, 屈丹^{2,3}, 王芳芳¹, 都力铭¹

(1. 郑州大学 网络空间安全学院, 河南 郑州 450001; 2. 战略支援部队信息工程大学 信息工程学院, 河南 郑州 450001; 3. 先进计算与智能工程(国家级)实验室, 河南 郑州 450001)

摘要: 针对遥感图像目标检测任务中进行特征提取时缺少小目标信息, 特征融合过程中部分信息丢失, 小目标特征信息不明显, 导致小目标检测精度不高的问题, 提出了一种基于 FEW-YOLOv8 模型的遥感图像目标检测算法。首先, 优化骨干网络架构, 使用 FasterNet 骨干网络, 更有效地提取了遥感图像中小目标的空间特征, 使得网络模型更专注于微小目标, 从而提升小目标检测精度。其次, 使用 EMA 注意力与 C2f 构建全新的 C2f_EMA 模块, 替换 Neck 结构中的 C2f 模块, 在融合特征前进行特征注意力加强操作, 使网络模型更突出特征信息中小目标部分, 有效解决特征融合过程中小目标特征丢失问题。最后, 采用带有动态非单调 FM 的 WIoUv3 作为边界框的损失函数, 提高了模型的边界框定位精度, 并且提升了对小目标的检测性能。实验结果显示: 在 NWPU VHR-10 数据集上经过优化的 YOLOv8 算法的 mAP50 相较于原始 YOLOv8 算法提高了 7.71 百分点, 在 HRSC2016 和 DOTA v1.0 上分别提高了 9.70 百分点和 12.32 百分点, 证明所提算法能够有效提升遥感图像中小目标的检测精度。

关键词: 遥感图像; YOLOv8; FasterNet 骨干网络; EMA 注意力机制; WIoU 损失函数

中图分类号: TP389.1

文献标志码: A

doi: 10.13705/j.issn.1671-6833.2025.04.007

遥感图像目标检测是指从遥感图像中识别出感兴趣的物体或目标, 并将其轮廓用矩形或多边形等形式表示出来的技术。遥感图像目标检测是地理信息技术领域中一个重要的研究方向, 在环境检测、地质勘查、植被检测、海洋生态监测、地表灾害监测等领域发挥着重要的作用。

遥感图像中的目标检测算法通常被划分为两阶段检测算法与单阶段检测算法两类。R-CNN 系列算法作为典型的两阶段检测方法, 在准确性上有明显的优势, 但这类算法存在训练复杂及检测速度较慢的问题。相比之下, 单阶段检测算法如 YOLO (you only look once)^[1] 系列算法和 SSD (single shot multibox detector)^[2] 算法, 则采用了直接回归的方式来确定目标的位置和分类信息。这种方法大大减少了计算资源的需求, 显著加快了检测速度, 不过在精度上有所妥协。YOLO 系列算法^[3] 在遥感图像目标检测领域, 常见的改进措施包括: 增强骨干网络的特征提取能力^[4]、引入注意力机制^[5]、采用密集连接

技术^[6]等。

改进目标检测算法可以更换骨干网络。一些骨干网络利用深度卷积 (DWconv)^[7] 或组卷积 (Gconv)^[8] 提取网络特征来解决特征冗余, 但会导致精度下降, 而 Chen 等^[9] 提出了一种简单的部分卷积 (Pconv), 并基于此构建了 FasterNet 网络有效解决了此类问题。此外, 融合注意力机制^[10] 可以突出目标特征权重来改进算法性能, 如 SE 注意力机制^[11]、CA 注意力机制^[12] 和 EMA 注意力机制^[13]。其中, EMA 是一种新型的高效多尺度注意力机制, 可以有效解决特征融合过程中小目标特征丢失问题。YOLOv8 算法的损失函数中, CIoU 损失函数关注中心坐标的回归, 而忽略了边界框长宽比的影响, 可能导致细长或压缩的物体难以准确检测, 而 WIoU 边界框损失函数对于小目标检测更为有效。综上所述, 本文提出了一种基于 FEW-YOLOv8 (FasterNet-EMA-WIoU-YOLOv8) 模型的遥感图像目标检测算法。主要工作如下: 使用 FasterNet 作为算法的骨干

收稿日期: 2024-10-15; 修订日期: 2024-12-06

基金项目: 国家自然科学基金资助项目 (62171470); 河南省中原科技创新领军人才项目 (234200510019); 河南省自然科学基金资助项目 (232300421240)

通信作者: 屈丹 (1974—), 女, 吉林九台人, 战略支援部队信息工程大学教授, 博士, 博士生导师, 主要从事语音识别、智能信息处理、机器学习等方面研究, E-mail: qudan_xd@163.com。

引用本文: 席阳丽, 屈丹, 王芳芳, 等. 基于 FEW-YOLOv8 遥感图像目标检测算法[J]. 郑州大学学报(工学版), 2025, 46(4): 62-69. (XI Y L, QU D, WANG F F, et al. Target detection algorithm based on FEW-YOLOv8 remote sensing images[J]. Journal of Zhengzhou University (Engineering Science), 2025, 46(4): 62-69.)

网络,更有效地提取了遥感图像中小目标的空间特征;使用 EMA 注意力与 C2f 构建全新的 C2f_EMA 模块,在融合特征前进行特征注意力加强操作;采用具有动态非单调 FM 的 WIoUv3 作为边界框的损失函数,提高模型的边界框定位准确度,强化对小目标的定位能力。

1 基于 FEW-YOLOv8 遥感图像目标检测算法

1.1 改进的 FEW-YOLOv8 算法

YOLO 系列算法主要包括骨干网络 (Backbone)、颈部结构 (Neck) 和头部预测层 (Head) 3 个组件,YOLOv8 相较于 YOLOv5^[14] 和 YOLOv7^[15] 检测精度和速度都有所提升。骨干网络使用 CSP-Darknet^[16],由一系列卷积层和池化层构成,用于图像的特征提取;颈部设计了基于 PANet^[17] 的 PAN-FPN 结构,即在 PAN 结构中去掉了上采样后的卷积;在头部预测层中使用分布焦点损失 DFL^[18] 和 CIOU^[19],实现了检测精度的提升,加速了模型的收敛。但 YOLOv8 对于遥感图像的小目标检测还存在一些问题,如在检测过程中特征提取时缺少小目标信息、特征融合过程中部分信息丢失、小目标特征信息不明显等。

针对上述问题,本文基于 YOLOv8 设计了一种 FEW-YOLOv8 算法,对 YOLOv8 算法进行改进以提升小目标检测的性能,其网络结构如图 1 所示。

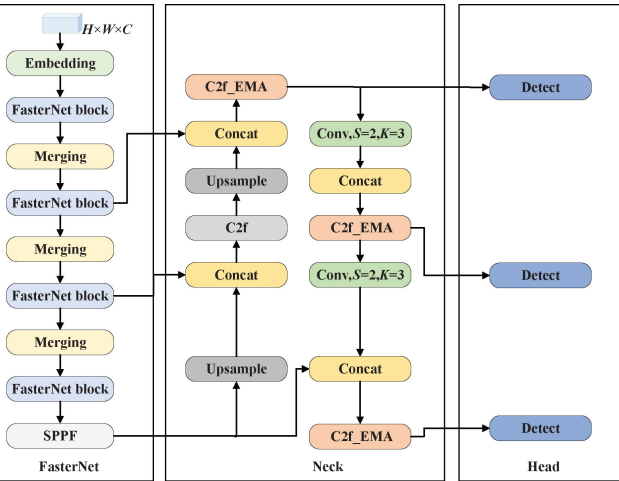


图 1 FEW-YOLOv8 的网络结构

Figure 1 Network structure of FEW-YOLOv8

首先,优化骨干网络架构,使用 FasterNet 骨干网络,更有效地提取了遥感图像中小目标的空间特征,使得网络模型更专注于微小目标,从而提升小目标检测精度。其次,使用 EMA 注意力与 C2f 构建全新的 C2f_EMA 模块,替换 Neck 结构中的 C2f 模块,

在融合特征前进行特征注意力加强操作,使网络模型更突出特征信息中小目标部分,有效解决特征融合过程中小目标特征丢失问题。最后,采用具有动态非单调 FM 的 WIoUv3 作为边界框损失函数,提高模型的边界框定位准确度,强化对小目标的定位能力。

1.2 FasterNet 骨干网络

FasterNet^[9] 网络结构如图 2(a) 所示,运行速度快,并且可以提升检测系统性能。FasterNet 网络分为 4 个模块,每个模块之前是一个卷积为 4×4 、步幅为 4 的嵌入层对图片的每一块处理,或是一个卷积为 2×2 、步幅为 2 的合并层将先前按块处理的图像以重叠的方式进行合并,这一合并过程实现了空间的下采样和增加了通道的数量。网络末端的 3 个层——全局平均池化层、 1×1 卷积层和全连接层则共同负责特征的转换与分类任务。

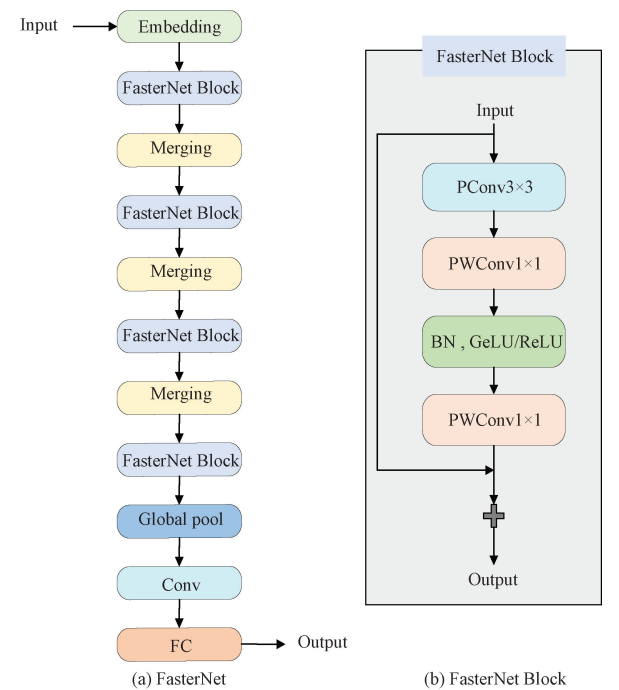


图 2 FasterNet 和 FasterNet Block 的网络结构
Figure 2 Network structure of FasterNet and FasterNet Block

FasterNet 网络的核心组件是 FasterNet Block,其结构如图 2(b) 所示,包括一个 PConv^[20] 层和两个 PWConv 层,其中间层拥有扩展的通道数,并且加入了 BN 层和 GeLU/ReLU 层。卷积层的主要职责在于捕捉图像中的局部特征;批归一化层的作用是加快训练速度并提高网络的稳定性;激活层通过引入非线性元素,增强了网络的表现力,保持了特征的多样性。批归一化层和激活层可能会略微减缓整体计算速度,在激活层中,对于较小的变体 FasterNet 选

择了 GeLU,而对于较大的变体则使用了 ReLU,这样的选择兼顾了运行效率和有效性。

传统卷积通过对所有输入通道进行全面扩张,导致了内存访问次数的增加;而深度可分离卷积(DWConv)不对通道进行扩张,可能会导致精度下降。为了解决这些问题,可以采用部分卷积(PConv)。作为 FasterNet 的主要运算单元,PConv 的结构如图 3 所示,它能够同时减少计算的冗余性和内存访问次数,仅对部分通道执行卷积操作,而其他通道则不进行处理。

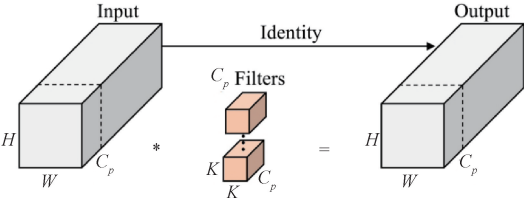


图 3 PConv 示意图

Figure 3 Schematic diagram of PConv

PConv 的计算量为 $H \times W \times K^2 \times C_p^2$, 对于 $r = C_p/C = 1/4$, 常规卷积计算量是其 16 倍。对于内存访问量,PConv 也有较大优势,其访问量为 $H \times W \times K \times 2C_p + K^2 \times C_p^2 \approx H \times W \times K \times 2C_p$, 其中, H 、 W 分别为通道的高和宽。

为了充分而高效地利用来自所有通道的信息,在 PConv 的后面添加一个 PWConv,这两个结合起来在输入上的有效感受野看起来像一个 T-shaped Conv,与普通卷积相比更关注于中心区域。通过计算 Frobenius norm 可以证明这种 T 型感受野进而评估每个位置的重要性。如果一个位置的 Frobenius norm 比其他位置都大,那么它往往更重要。一个普通卷积 $F \in \mathbf{R}^{K^2 \times C}$ 位置 i 处的 Frobenius norm 通过式(1)计算得到:

$$\|F\| = \sqrt{\sum_{j=1}^C |f_{ij}|^2}, i = 1, 2, \dots, K^2. \quad (1)$$

已知一个显著的位置是具有最大 Frobenius norm 的位置。图 4 是研究 ResNet18 中的每个卷积核,找到每个卷积核的显著位置,然后画出一个显著位置的直方图,得出中心位置作为最显著位置的频率是最高的。中心位置的权重比周围都要大,这与 T 形计算集中于中心处是一致的。

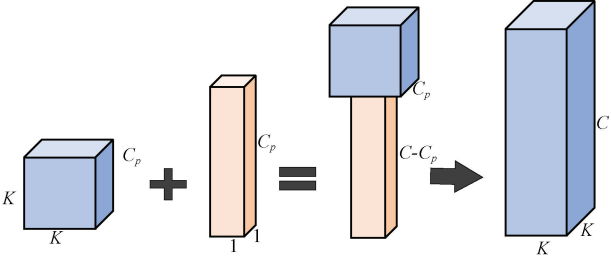


图 4 T-shaped Conv 示意图

Figure 4 Schematic diagram of T-shaped Conv

将 YOLOv8 中的骨干网络换为引入 PConv 的 FasterNet 网络,减少了算法的参数数量和计算量,更有效地提取了遥感图像中小目标的空间特征,最终在保证遥感图像小目标检测精度的同时大幅度提升了检测速度。

1.3 多尺度注意力机制模块

EMA^[13] 高效多尺度注意力机制结构如图 5 所示,其中, G 代表分组; X Avgpool 为在一维水平方向的全局平均池化; Y Avgpool 为在一维垂直方向的全局平均池化。该机制将输入特征映射 $X \in \mathbf{R}^{C \times H \times W}$, 分割成 G 个独立的子集即 $X = [X_0, X_i, \dots, X_{G-1}]$, $X_i \in \mathbf{R}^{(C//G) \times H \times W}$, 通过使用 3 个并行子网结构来得到分组的特征图。在第一个 1×1 分支中,输入的张量被重塑成特定的形式 $(C // G) \times H \times W$, 随后组内的两个通道注意力图通过乘法操作被合并,接着采用 Avgpool (全局平均池化) 对上层输入进行编

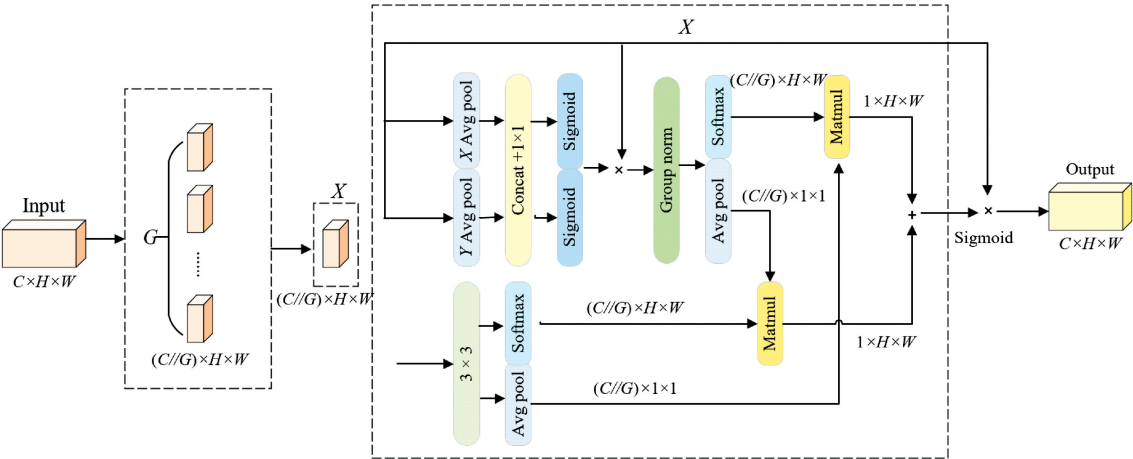


图 5 EMA 网络结构

Figure 5 EMA network architecture

码,得到第一个输出为 $\mathbf{R}_1^{1 \times (C//G)} \times \mathbf{R}_3^{(C//G) \times H \times W}$, 最后通过 2D 全局池化和矩阵点积运算操作,将这些并行处理的结果结合起来形成第一个空间注意力图,其公式为 $Z_c = \frac{1}{HW} \sum_j^H \sum_i^W x_c(i,j)$ 。在第二个分支中,采用的是 3×3 的卷积进行捕捉,接着类比于第一个分支的方法进行编码,经过系列调整后得到第二个输出为 $\mathbf{R}_3^{1 \times (C//G)} \times \mathbf{R}_1^{(C//G) \times H \times W}$, 进而得到了第二个分支的空间注意力图。然后,将每组内的输出特征映射与第一、第二个空间注意力图结合,并通过 Sigmoid 激活函数进行处理,以捕捉像素级别的配对关系并强调全局上下文信息,使得 EMA 的输出保持与输入相同的尺寸。

本文将 EMA 与 C2f 模块相结合,构成 C2f_EMA 模块,以提高模块提取小目标特征的能力。EMA 在通道方向上对跨通道特征交互进行建模,可以捕获所有通道之间的依赖关系并减少计算量,提高特征融合的精确度。设计的 C2f_EMA 模块在小目标、模糊目标识别任务上具有更强大的应对能力,在 YOLOv8 中使用 C2f_EMA 模块使算法在融合特征前进行特征注意力加强操作,使网络模型更突出特征信息中小目标部分,有效解决特征融合过程中小目标特征丢失问题。

1.4 WIoU 损失函数

WIoU^[21]使用了一种非单调的聚焦机制,该机制通过生成动态的梯度增益因子来改进锚框的质量评估过程。这一方法采用离群度来替代交并比(IoU)作为评估锚框质量的标准,有助于缓解传统 IoU 损失函数中存在的问题,即该函数基于尺寸差异、长宽比等因素对低质量样本施加过大的惩罚,从而影响模型的泛化能力。通过这种方式,可以更有效地处理不同质量的样本,提高模型的整体性能。WIoU 公式如下所示:

$$L_{\text{WIoUv1}} = R_{\text{WIoU}} L_{\text{IoU}}; \quad (2)$$

$$R_{\text{WIoU}} = \exp \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}; \quad (3)$$

$$L_{\text{IoU}} = 1 - \text{IoU}; \quad (4)$$

$$\text{IoU} = \frac{A \cap B}{A \cup B}. \quad (5)$$

式中: (x,y) 为锚框的中心点坐标; (x_{gt},y_{gt}) 为目标框的坐标; W_g 和 H_g 为最小包围框的尺寸; * 代表将 W_g 和 H_g 从计算图中分离; R_{WIoU} 用于放大普通质量锚框的 L_{IoU} , L_{IoU} 用于降低高质量锚框的 R_{WIoU} , 借此可以根据需要增加或降低对中心点距离的关注度。

WIoUv3 在 WIoUv1 的基础上引入了一个非单

调聚焦系数 r 。这样, WIoUv3 能够根据训练过程动态调整梯度增益。具体来说,对于那些与目标框匹配良好的锚框(即离群度较低的锚框),赋予较小的梯度增益,从而使边界框回归更加关注于一般质量的锚框;而对于匹配较差的锚框(即离群度较高的锚框),则赋予较大的梯度增益,以此来避免低质量锚框产生过大的负面影响。具体公式如下所示:

$$L_{\text{WIoUv3}} = r L_{\text{WIoUv1}}; \quad (6)$$

$$r = \frac{\beta}{\delta \alpha^{\beta-\delta}}; \quad (7)$$

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

$$m = 1 - \sqrt[n]{0.05}. \quad (9)$$

式中: α 和 δ 均为超参数,设定为 1.9 和 3.0,此时对于异常框抑制效果更加显著; β 为离群度,可以表示锚框的质量,与锚框质量负相关; L_{IoU}^* 为梯度增益,可以根据每个检测目标的情况来动态计算和调整; \bar{L}_{IoU} 为动量 m 的滑动平均值,可以动态调节最大的梯度增益; t 为 epoch 值; n 为 batches 值。

采用带有动态非单调 FM 的 WIoUv3 作为边界框的损失函数,能够在训练过程中消除许多不利影响,可以根据实际情况来动态地调整梯度增益,从而提高模型边界框定位的准确性。

2 实验结果与分析

2.1 实验环境和参数设置

本文的实验环境配置环境如下:使用配备 24 GB 显存的 NVIDIA GeForce RTX 4090 GPU,并运行在 Ubuntu 20.04 操作系统上。编程语言采用 Python 3.8, CUDA 版本为 11.3。模型训练共迭代了 100 轮,初始学习率设定为 0.01,批量大小为 16,采用线性学习率衰减策略,优化算法采用随机梯度下降(SGD)。

2.2 实验数据集

一共选择了 3 个具有代表性的公开遥感图像的数据集进行实验,使得实验结果具有科学性和普遍性。3 个数据集分别为 NWPU VHR-10、HRSC2016 和 DOTA v1.0。

NWPU VHR-10 数据集中有 650 幅图像包含目标,选择 520 幅图像作为训练集,130 幅图像作为测试集。HRSC2016 数据集包含 1 070 张图像和 2 976 个实例,使用旋转矩形框进行标注,专门用于舰船检测。DOTA v1.0 数据集有 2 806 张航空遥感图像,将其分割成 21 046 张子图像,在其中选用 15 749 张图像作为训练集,5 297 张作为测试集。

2.3 公共评价指标

本文实验结果使用目标检测常用的评估指标:准确率 P 、召回率 R 、参数量 $Parameters$ 、计算量和平均精度均值 $mAP^{[22]}$ 。采用 $mAP50$ 和 $mAP50-95$ 表示不同 IoU 阈值下的平均精度均值。计算公式如下所示:

$$P = \frac{TP}{TP + FP}; \tag{10}$$

$$R = \frac{TP}{TP + FN}; \tag{11}$$

$$AP = \int_0^1 P(R) dR; \tag{12}$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N}. \tag{13}$$

式中: TP 为真正例,指的是实际存在并且被模型正确识别为目标类别的样本数量; FP 为假正例,指的是实际上不属于某个类别但被模型错误地识别为目标类别的样本数量; FN 为假反例,指的是实际存在某个类别但模型未能正确识别出来的样本数量; N 为类别数; AP 为平均精度; mAP 为各个 AP 的加权平均值, mAP 考虑了所有类别,是一个综合指标,能够全面反映模型在多个类别上的表现,可以有效评估模型性能。

2.4 消融实验

为验证各个模块的性能,以 YOLOv8 原始模型为基准方法,采用 NWPU VHR-10 数据集设计一系列消融实验进行对比验证,使用 R 、 $mAP50$ 、和 $mAP50-95$ 为定量评价指标,实验结果如表 1 所示。由表 1 中数据可得,单独使用 FasterNet 网络替换原

始骨干网络时,实验结果和基准方法相比, R 提升了 2.02 百分点, $mAP50$ 和 $mAP50-95$ 分别提升了 5.42 百分点和 2.86 百分点。完整模型与没有采用 FasterNet 网络相比, $mAP50$ 提升了 1.32 百分点。说明 FasterNet 网络可以有效解决遥感图像小目标检测精度不高的问题。

单独使用 C2f_EMA 改进颈部结构时,实验结果与基准方法相比, R 提升了 4.44 百分点, $mAP50$ 和 $mAP50-95$ 分别提升了 6.33 百分点和 4.40 百分点。C2f_EMA 模块结合 FasterNet 网络和 WIoUv3 损失函数相比没有采用 C2f_EMA 模块的模型,其 $mAP50$ 和 $mAP50-95$ 都有所上升。说明加入 C2f_EMA 模块可以优化系统对遥感图像小目标的检测性能。

采用 WIoUv3 作为回归损失函数时,相比基准方法其 $mAP50$ 和 $mAP50-95$ 分别提高了 1.59 百分点和 0.74 百分点;对于完整模型来说,改进损失函数 $mAP50$ 和 $mAP50-95$ 分别提高了 1.41 百分点和 2.40 百分点。由此可知,WIoUv3 可以提高回归框精度,提高模型鲁棒性。

当更换为 FasterNet 骨干网络,引入 C2f_EMA 和 WIoUv3 时,其召回率、 $mAP50$ 和 $mAP50-95$ 分别比基准方法高出 5.87 百分点、7.71 百分点和 5.24 百分点,对于小目标检测的性能较好,有效改善遥感图像中小目标检测精度较低的问题,所提出的 FEW-YOLOv8 网络检测效果良好。

2.5 对比实验

为了验证本文算法的有效性,在保证参数设置不变的情况下,分别对 NWPU VHR-10、HRSC2016 和 DOTA v1.0 数据集进行实验,实验结果如表 2 所示。

表 1 消融实验

Table 1 Ablation experiment

YOLOv8	FasterNet	C2f_EMA	WIoUv3	$R/\%$	$mAP50/\%$	$mAP50-95/\%$	参数量/ 10^6	计算量/GFLOPs
✓				75.96	79.34	47.06	3.76	11.2
✓	✓			77.98	84.76	49.92	23.97	61.6
✓		✓		80.40	85.67	51.46	23.97	61.7
✓			✓	74.54	80.93	47.80	3.76	11.2
✓	✓		✓	76.28	84.70	49.91	23.97	61.6
✓		✓	✓	79.92	85.72	50.04	23.97	61.7
✓	✓	✓		80.06	85.64	49.90	11.19	29.5
✓	✓	✓	✓	81.83	87.05	52.30	11.19	29.5

表 2 不同算法在 NWPU VHR-10、HRSC2016 和 DOTA v 1.0 数据集上的性能比较

Table 2 Performance comparison of different algorithms with NWPU VHR-10,HRSC2016 and DOTA v1.0 datasets

算法	NWPU VHR-10			HRSC2016			DOTA v1.0		
	$P/\%$	$R/\%$	$mAP50/\%$	$P/\%$	$R/\%$	$mAP50/\%$	$P/\%$	$R/\%$	$mAP50/\%$
YOLOv8	81.47	75.96	79.34	84.46	62.99	75.64	69.01	54.24	56.95
YOLOv8+FasterNet	86.88	77.98	84.76	77.27	64.06	74.56	69.32	59.31	61.87
YOLOv8+FasterNet+WIoUv3	85.99	76.28	84.70	77.93	68.60	77.50	69.24	58.84	61.68
FEW-YOLOv8	89.16	81.83	87.05	91.08	76.33	85.34	73.52	66.41	69.27

可以看出,对于不同的改进算法,FEW-YOLOv8 的检测效果在 3 个数据集上均比基准方法效果好,其中, $mAP50$ 在 NWPU VHR-10、HRSC2016、DOTA v1.0 数据集上分别提升了 7.71 百分点、9.7 百分点和 12.32 百分点。所以 FEW-YOLOv8 对小目标的检测性能的提升有效且具有普遍性。其中,FEW-YOLOv8 在 NWPU VHR-10 数据集上的召回率曲线如图 6 所示。

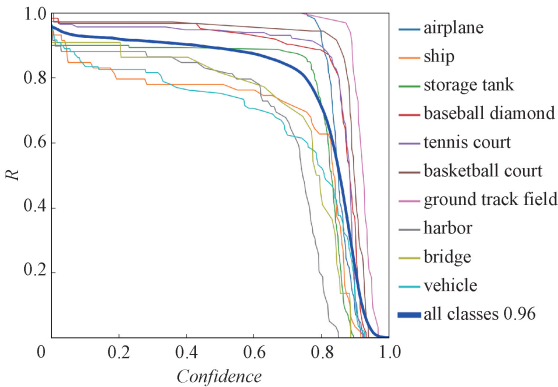


图 6 NWPU VHR-10 数据集中每个类别的召回率曲线
Figure 6 Recall curves for each category in the NWPU VHR-10 dataset

如表 3 所示,在各骨干网络改进的实验中,

表 3 各骨干网络改进与 FasterNet 的对比

Table 3 Comparison between the improvement of each backbone network and FasterNet

骨干网络	$R/\%$	$mAP50/\%$	$mAP50-95/\%$	参数量/ 10^6	计算量/GFLOPs
YOLOv8	75.96	79.34	47.06	3.76	11.2
YOLOv8+GhostNetv1	67.95	74.68	42.85	2.73	7.7
YOLOv8+GhostNetv2	60.04	61.20	33.42	3.79	8.0
YOLOv8+PP-LCNet	38.61	47.13	24.69	1.88	5.7
YOLOv8+MobileNetv3	58.30	58.21	31.51	2.36	5.8
YOLOv8+MobileViT	57.62	60.09	32.69	1.18	5.4
YOLOv8+EfficientNet	56.58	59.41	33.14	2.27	7.0
YOLOv8+SwinTransformer	74.39	80.18	46.59	5.13	455.2
YOLOv8+FasterNet	77.98	84.76	49.92	23.97	61.6

GhostNetv1 和 GhostNetv2^[23] 使用更少的参数生产更多的特征图,以此替换骨干网络的改进效果并不理想,虽模型参数量有所减少,但检测精度均有较大幅度降低。由 PP-LCNet^[24]、MoblieNetv3^[25]、MobileViT^[26]、EfficientNet^[27]、SwinTransformer 替换骨干网络的改进,检测精确度和召回率均有不同幅度的下降,检测性能相比基准方法并未提升。而由 FasterNet 网络替换骨干网络的改进表现优秀,更有效地提取了遥感图像中小目标的空间特征,从而提升小目标检测精度。

2.6 实验结果

为了直观地显示本文算法的检测性能,选取 HRSC2016 数据集中小目标数量多且特征不明显的港口的部分图像进行可视化实验,如图 7 所示,对于同一组图像,将原始 YOLOv8 检测结果和 FEW-YOLOv8 的检测结果进行了直观对照。其中,蓝色箭头所标出的是原 YOLOv8 误检部分,黄色箭头标出的是原 YOLOv8 漏检部分,通过图 7 中可以明显观察到 FEW-YOLOv8 能有效纠正这些错误,提高检测精度。由此可见,本文提出的 FEW-YOLOv8 算法能有效地提升遥感图像小目标的检测性能。

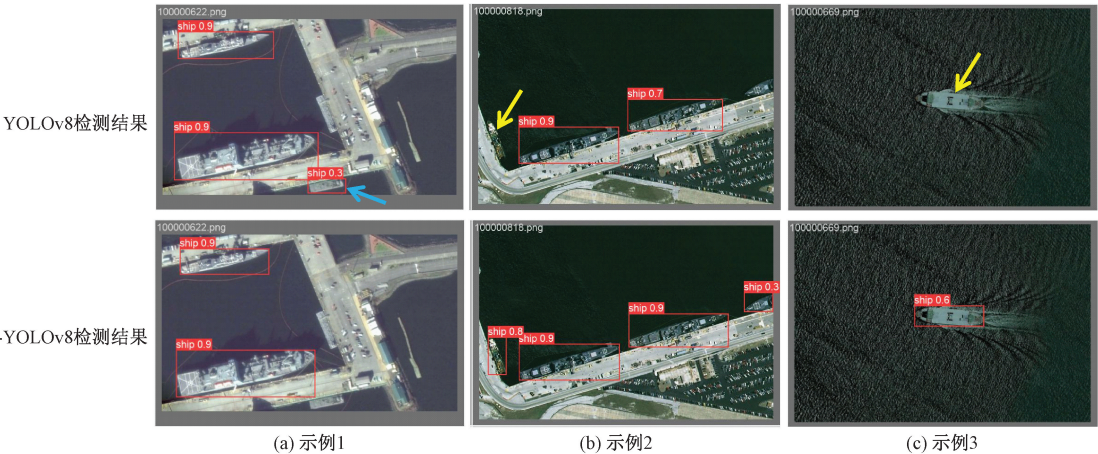


图 7 YOLOv8 与 FEW-YOLOv8 在 HRSC2016 数据集中部分目标检测结果对比

Figure 7 Comparison of partial detection results of YOLOv8 and FEW-YOLOv8 in the HRSC2016 dataset

3 结论

本文提出了基于 YOLOv8 的 FEW-YOLOv8 算法,可以有效提高遥感图像目标检测性能。首先,优化骨干网络架构,使用 FasterNet 骨干网络,能更有效地提取了遥感图像中小目标的空间特征,使得网络模型更专注于微小目标,从而提升小目标检测精度。其次,使用 EMA 注意力与 C2f 构建全新的 C2f_EMA 模块,替换 Neck 结构中的 C2f 模块,在融合特征前进行特征注意力加强操作,使网络模型更突出特征信息中小目标部分,可有效解决特征融合过程中小目标特征丢失问题。最后,采用具有动态非单调 FM 的 WIoUv3 作为边界框损失函数,提高模型的边界框定位准确度,强化对小目标的定位能力。通过实验结果表明,3 个模块的改进对遥感图像小目标检测的改进都是有效的,且将 3 个改进模块组合在一起检测效果较好;与 YOLOv8 等其他具有代表性的改进算法进行比较,FEW-YOLOv8 算法在检测精度和检测速度两方面可以兼顾,对于遥感图像小目标精度有较高的提升。

参考文献:

- [1] 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 (CVPR). Piscataway: IEEE, 2016: 779-788.
- [2] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [3] GAO F, CAI C X, JIA R H, et al. Improved YOLOX for pedestrian detection in crowded scenes[J]. Journal of Real-Time Image Processing, 2023, 20(2): 1-13.
- [4] XU D Q, WU Y Q. FE-YOLO: a feature enhancement network for remote sensing target detection[J]. Remote Sensing, 2021, 13(7): 1311.
- [5] CHEN L Q, SHI W X, DENG D X. Improved YOLOv3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images[J]. Remote Sensing, 2021, 13(4): 660.
- [6] XU D Q, WU Y Q. Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection[J]. Sensors, 2020, 20(15): 4276.
- [7] POUDEL R P K, BONDE U, LIWICKI S, et al. ContextNet: exploring context and detail for semantic segmentation in real-time[EB/OL]. (2018-11-05)[2024-10-12]. <http://arxiv.org/abs/1805.04554>.
- [8] CHEN J T, LEI B W, SONG Q Y, et al. A hierarchical graph network for 3D object detection on point clouds[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 389-398.
- [9] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 12021-12031.
- [10] ZHOU T, HUANG B, LI R R, et al. An attention-based deep learning model for citywide traffic flow forecasting[J]. International Journal of Digital Earth, 2022, 15(1): 323-344.
- [11] GAO R X, WANG T F. Motion deblurring algorithm for wind power inspection images based on Ghostnet and SE attention mechanism[J]. IET Image Processing, 2023, 17(1): 291-300.
- [12] LI G B, SHI G L, JIAO J. YOLOv5-KCB: a new method for individual pig detection using optimized K-means, CA attention mechanism and a bi-directional feature pyramid network[J]. Sensors, 2023, 23(11): 5242.
- [13] OUYANG D L, HE S, ZHANG G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2023: 1-5.
- [14] YUAN Z, FANG W, ZHAO Y M, et al. Research of insect recognition based on improved YOLOv5[J]. Journal on Artificial Intelligence, 2021, 3(4): 145-152.
- [15] 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[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 7464-7475.
- [16] YIN L L. Analysis recognition of ghost pepper and cili-padi using mask RCNN and YOLO[J]. Przegląd Elektrotechniczny, 2023, 1(8): 94-99.
- [17] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8759-8768.
- [18] 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.
- [19] WANG P J, BAYRAM B, SERTEL E. A comprehensive review on deep learning based remote sensing image super-resolution methods[J]. Earth Science Reviews, 2022, 232: 104110.
- [20] PENG X C, CHEN Y Z, CAI X W, et al. An improved YOLOv7-based model for real-time meter reading with PConv and attention mechanisms[J]. Sensors, 2024, 24

(11): 3549.

[21] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[EB/OL]. (2023-04-08)[2024-10-12]. <https://arxiv.org/abs/2301.10051>.

[22] 贾云飞, 郑红木, 刘闪亮. 基于 YOLOv5s 的金属制品表面缺陷的轻量化算法研究[J]. 郑州大学学报(工学版), 2022, 43(5): 31-38.

JIA Y F, ZHENG H M, LIU S L. Lightweight surface defect detection method of metal products based on YOLOv5s[J]. Journal of Zhengzhou University (Engineering Science), 2022, 43(5): 31-38.

[23] 刘庆华, 杨欣仪, 接浩, 等. 基于融合 GhostNetV2 的 YOLO v7 水稻籽粒检测[J]. 农业机械学报, 2023, 54(12): 253-260, 299.

LIU Q H, YANG X Y, JIE H, et al. Rice grain detection based on YOLO v7 fusing of GhostNetV2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(12): 253-260, 299.

[24] 胡瑛, 刘狄昆, 刘拯, 等. 基于改进 YOLOv5 的复杂场景下交通标志识别方法[J]. 湖南工程学院学报(自然科学版), 2024, 34(2): 31-38.

HU Y, LIU D K, LIU Z, et al. Traffic sign recognition method in complex scenes based on improved YOLOv5[J]. Journal of Hunan Institute of Engineering (Natural Science Edition), 2024, 34(2): 31-38.

[25] 刘磊. YOLOv4 交通信号灯检测[J]. 电子世界, 2021(15): 92-94.

LIU L. YOLOv4 traffic light detection[J]. Electronics World, 2021(15): 92-94.

[26] 邓翔宇, 裴浩媛, 盛迎. 基于网络融合的改进 MobileViT 人脸表情识别[J]. 计算机工程与科学, 2024, 46(6): 1072-1080.

DENG X Y, PEI H Y, SHENG Y. Facial expression recognition based on network fusion to improve MobileViT[J]. Computer Engineering & Science, 2024, 46(6): 1072-1080.

[27] 胡施威, 邓建新, 王浩宇, 等. 基于改进 EfficientNetB0 模型的葡萄叶部病害识别方法[J]. 现代电子技术, 2024, 47(15): 73-80.

HU S W, DENG J X, WANG H Y, et al. Grape leaf disease identification method based on improved EfficientNetB0 model[J]. Modern Electronics Technique, 2024, 47(15): 73-80.

Target Detection Algorithm Based on FEW-YOLOv8 Remote Sensing Images

XI Yangli¹, QU Dan^{2,3}, WANG Fangfang¹, DU Liming¹

(1. School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450001, China; 2. School of Information System Engineering, Strategic Support Force Information Engineering University, Zhengzhou 450001, China; 3. Laboratory for Advanced Computing and Intelligent Engineering, Zhengzhou 450001, China)

Abstract: Aiming at the problems of lack of small target information during feature extraction, partial loss of information during feature fusion, and inconspicuous small target feature information in remote sensing image target detection task, which lead to the low accuracy of small target detection, an algorithm for remote sensing image target detection based on FEW-YOLOv8 model was proposed. Firstly, the backbone network architecture was optimized to use the FasterNet backbone network, which extracted the spatial features of small targets in remote sensing images more efficiently, making the network model more focused on tiny targets, thus improving the small target detection accuracy. Secondly, the new C2f_EMA module was constructed using EMA attention and C2f to replace the C2f module in Neck network, and the feature attention enhancement operation was performed before fusing the features, so that the network model highlighted the small-target part of the feature information more, which effectively solved the problem of small-target feature loss in the process of feature fusion. Finally, WIoUv3, which had a dynamic non-monotonic FM, was used as the bounding box loss function to improve the accuracy of the model's bounding box localization and strengthen the localization ability of small targets. The experimental results on NWPU VHR-10, HR-SC2016 and DOTA v1.0 datasets showed that the test *mAP*50 of the improved YOLOv8 algorithm was 7.71, 9.70 and 12.32 percentage points higher than that of the original YOLOv8 algorithm, respectively, which proved that the proposed algorithm could effectively improve the detection accuracy of small targets in remote sensing images.

Keywords: remote sensing images; YOLOv8; FasterNet backbone network; EMA attention mechanism; WIoU loss function