

基于改进 YOLOv8 的遥感图像检测算法^{*}

宋树成 程换新

(青岛科技大学自动化与电子工程学院 青岛 266061)

摘要: 针对小目标在遥感图像中的局限性,如图像背景复杂、小目标分布密集、目标尺度多样等问题,本文提出了一种基于 YOLOv8n 的改进算法。首先,设计了一个多尺度空洞注意力模块,在主干网络中引入多尺度空洞注意力机制与 C2f 模块结合,以有效捕捉多尺度的语义信息并减少自注意力机制的冗余;其次,设计了一个残差快速卷积模块,减小模型计算量并提高特征提取能力;最后,使用 PIoU v2-Iou 损失函数代替 CIOU 损失函数,提升模型的检测精度。通过在 DOTA、RSOD 和 VisDrone2019 数据集上的实验结果显示,改进后 YOLOv8n 模型与原模型 YOLOv8n 相比, mAP 分别提升了 2.7%、3.3% 和 3.8%,计算量降低了 0.5 GFLOPs,验证了新算法的有效性。

关键词: 遥感图像; YOLOv8; 多尺度空洞注意力机制; 损失函数

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

Remote sensing image detection algorithm based on improved YOLOv8

Song Shucheng Cheng Huanxin

(College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: Aiming at the limitations of small targets in remote sensing images, such as complex image background, dense distribution of small targets, and diverse target scales, this paper proposes an improved algorithm based on YOLOv8n. Firstly, a multi-scale null attention module is designed to introduce a multi-scale null attention mechanism in the backbone network in combination with the C2f module to effectively capture multi-scale semantic information and reduce the redundancy of the self-attention mechanism; secondly, a residual fast convolution module is designed to reduce the model computation and improve the feature extraction capability; finally, the PIoU v2-Iou loss function is used instead of the CIOU loss function to improve the detection accuracy of the model. The experimental results on DOTA, RSOD and VisDrone2019 datasets show that the improved YOLOv8n model improves the mAP by 2.7%、3.3% and 3.8%, respectively, and reduces the computation by 0.5 GFLOPs compared with the original model YOLOv8n, which validates the effectiveness of the new algorithm.

Keywords: remote sensing images; YOLOv8; multi-scale null attention mechanisms; loss function

0 引言

当前目标检测技术是计算机视觉领域的重点发展对象,是其他复杂视觉任务的基础。目标检测作为遥感图像自动化分析的关键技术,被广泛应用于无人机^[1]、智能交通监控^[2]、航空航天^[3]等领域。遥感图像存在大量小目标,其中有背景噪声、信息量小等缺陷,导致对背景复杂、分辨率低、分布密集的小目标的检测往往效果不佳^[4],因此,对遥感技术的自动化、智能化、高效化具有迫切需要^[5]。

目前,基于深度学习的目标检测算法根据不同的检测方法可以分为双阶段检测算法和单阶段检测算法。双阶段

检测算法通过样本处理得到多个候选框,然后通过卷积神经网络对样本进行特征提取完成分类任务,最后通过后处理操作精确得到目标框,它以 R-CNN 系列 (R-CNN^[6]、Faster R-CNN^[7]、Masked R-CNN^[8]) 为代表。虽然两阶段算法在检测精度和目标定位方面已经足够成熟,但在速度方面依然不能满足实时性的要求。

近年来,为了统一目标的分类和定位,提出了以 YOLO^[9] 系列为为代表的单阶段检测算法,采用回归思维解决目标检测问题,从输入图像到获得边界框的位置和类别,一步到位。Guo 等^[10]提出了一种 PG-YOLO 检测算法,该算法通过设计了一组关注模块,学习网络模型中不同的合

并特征,对不同的合并层赋予不同的关注权重,从而获得更高的检测结果,但未考虑图像背景复杂情况,当目标分布密集时,该算法检测容易误检。张邵文等^[11]使用二维离散余弦变换对卷积块注意力模块(convolutional block attention module, CBAM)进行改进,提出一种加权多重感受野空间金字塔池化模块,并结合跨层特征融合的思想,提升模型对小尺度目标的感知能力。Zhou 等^[12]在结合上下文转换器模块的基础上,对 YOLOv5-s 进行了特征提取优化。Liu 等^[13]提出了 YOLO-Extract,该算法借鉴残差网络的思想,将协调关注融合到网络中,优化了模型对不同尺度目标的特征提取能力,但检测的目标特征种类较少,缺乏广泛适用性。Wang 等^[14]提出了基于 FasterNet 块的高效快速的特征处理模块 FFNB,实现浅层和深层特征的全面融合,在主干中引入动态稀疏注意力 BiFormer,提高对特征图的关注度,但对微小物体的检测效果较差,仍有优化空间。

虽然以上的研究发展已经有效的提高了遥感图像检测的准确性,但是小目标的漏检误检、目标定位不准确等问题仍然存在,为进一步优化检测效果,本文以 YOLOv8n^[15]为基线网络做出改进,主要贡献如下:

1) 设计了一个多尺度空洞注意力模块,把多尺度空洞注意力机制(multi-scale dilated attention, MSDA)^[16]引入到 C2f 模块,构成 C2f_MSDA 模块,提高对重要信息的关注,解决部分特征丢失的问题,从而提高检测性能。

2) 设计了一个残差快速卷积模块(residual_fasterconv, R-FConv),替换颈部网络中部分 C2f 模块,达到减小计算冗余、提高推理速度的效果,有效地减小了模型计算量。

3) 使用 PIoU v2-IoU^[17]损失函数替换原网络的 CIoU^[18]损失函数,提高检测框的定位能力,改善模型收敛速度和鲁棒性,最终更好的完成目标检测任务。

1 YOLOv8 网络

本文采用 YOLOv8 作为基线网络,是目前性能最优的模型,能够兼具速度和精度优势的单阶段检测算法。主要由输入模块(Input)、主干模块(Backbone)、颈部模块(Neck)及输出模块(Head)四部分组成。YOLOv8 网络结构图如图 1 所示。

YOLOv8 作为当前最主要的目标检测模型,与 YOLOv5^[19]相比,在 Backbone 和 Neck 部分,两者都使用了 CSP 梯度分流的思想,且都使用了 SPPF 模块,不同的是 YOLOv8 使用 C2f 结构替换了 C3 结构,并增加了额外的 split 操作,取消了分支中的卷积操作,可以在保证轻量化的同时获得更加丰富的梯度流信息,根据模型尺度来调整通道数,加强了特征信息并减少计算量,大幅提升了模型性能。Head 部分采用目前主流的解耦头结构,有效减少参数量和计算复杂度,增强模型的泛化能力和鲁棒性^[20]。最后由 Conv2d 计算出边界框损失。YOLOv8 还摒弃了之前

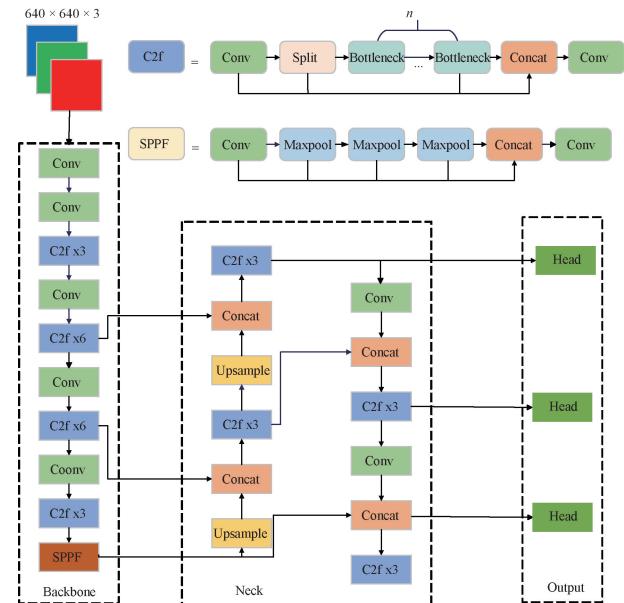


图 1 YOLOv8n 网络模型
Fig. 1 YOLOv8n network model

YOLO 系列中使用有锚节点来预测 Anchor 框的位置和大小的设计,转而使用无锚节点的检测方式,直接预测目标的中心点和宽高比例,减少 Anchor 框的数量,从而进一步提升模型的检测速度和精度。

YOLOv8 的核心思想在于将目标检测转化为回归问题。该算法将图像划分成 $s \times s$ 大小的网格,在每个网格中预测是否存在物体、物体的位置和大小,并输出一组包括中心坐标、宽度、高度的边界框,以及每个边界框所包含物体的置信度分数。鉴于同一目标可能被多个边界框检测到,为了消除重复检测结果,YOLOv8 使用非极大值^[21]抑制算法来筛选最终的检测结果,并进行类别预测。

2 改进 YOLOv8 算法

在原 YOLOv8n 模型基础上,使用 C2f_MSDA 模块替换原主干网络中部分的 C2f 模块,设计了 R_FConv 模块替换原网络中颈部的 CBSModule 结构,使用 PIoU v2 作为边界框损失函数。改进的网络模型如图 2 所示。

2.1 MSDA 模块

采用多头的设计的 MSDA 模块把特征图的通道分成 n 个不同的头部,在这 n 个不同的头部使用不一样的空洞率执行滑动窗口膨胀注意力(sliding window dilated attention, SWDA)。SWDA 可用公式描述为:

$$X = SWDA(Q, K, V, r) \quad (1)$$

其中, Q, K, V 分别表示查询矩阵、键矩阵和值矩阵, r 表示膨胀率, 3 个矩阵的每一行表示一个特征向量。在原始特征映射中,对于坐标 (i, j) 的位置, SWDA 稀疏选择键和值在以 (i, j) 为圆心, 大小为 $w \times w$ 的滑动窗口中进行自关注。这样可以更有效捕捉多尺度的语义信息,减少自注

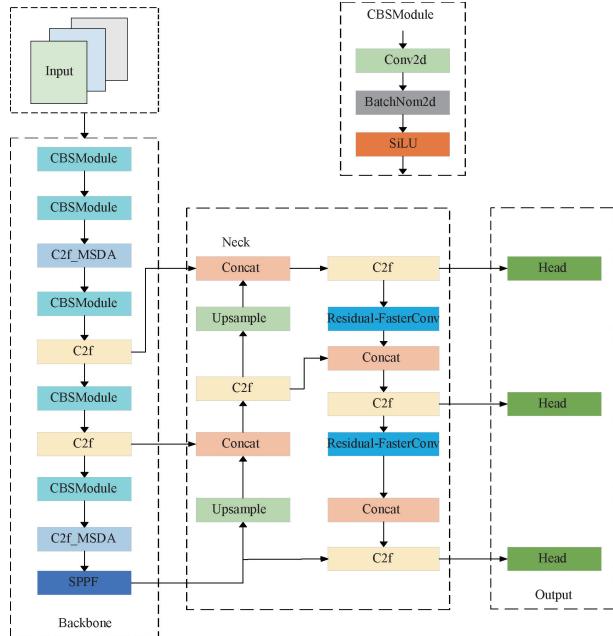


图 2 改进 YOLOv8n 结构

Fig. 2 Improvement of YOLOv8n structure

意力机制的冗余,降低计算成本,其结构如图 3 所示。

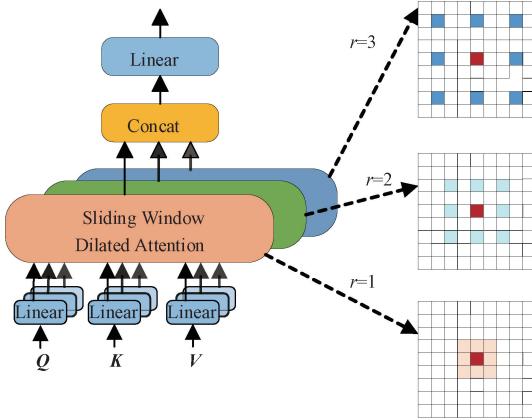


图 3 MSDA 模块

Fig. 3 MSDA module

给定一个特征映射 X , 通过线性投影获得相应的查询、键和值, 将特征映射的通道分割成不同的头部, 把不同的头部的输出连接在一起, 然后通过一个线性层进行特征聚合, 对于每个头部, 都会有一个独立的膨胀率 r_i , 在默认情况下, 使用 3×3 的内核大小, 膨胀率 $r = 1, 2$ 和 3 , 不同头部的参与接受野的大小分别为 $3 \times 3, 5 \times 5$ 和 7×7 。在红色查询补丁周围窗口的彩色补丁之间进行自注意操作, 在不同的头部使用不同的扩张率, 从特征图中获取切片 Q_i, K_i 和 V_i , 执行 SWDA, 得到输出 h_i , 可用数学表达式表示:

$$h_i = \text{SWDA}(Q_i, K_i, V_i, r_i), 1 \leq i \leq n \quad (2)$$

$$X = \text{Linear}(\text{Concat}[h_1, \dots, h_n]) \quad (3)$$

把 MSDA 模块与原网络中 C2f 模块相结合, 可以让模型更准确的捕获目标信息, 达到提高模型检测精度^[22]的要求, 如图 4 所示。

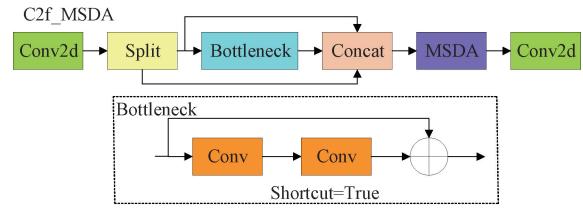


图 4 C2f_MSDA 模块

Fig. 4 C2f_MSDA module

2.2 Residual_FasterConv 模块

为了更好地实现模型轻量化和检测性能之间的平衡, 本文设计了残差快速卷积模块, 用于 YOLOv8 模型颈部网络中。部分卷积(partial convolution, PConv)^[23]利用了特征映射中的冗余性, 系统地对部分输入通道应用传统卷积, 而其余的通道保持不变来提取空间特征, 这种方法在提高推理速度的同时减小了计算量。PConv 的 FLOP 为:

$$h \times w \times k^2 \times c_p^2 \quad (4)$$

对于输入 $X \in R^{c \times h \times w}$ 、输出 $Y \in R^{c \times h \times w}$ 的常规卷积, FLOPs 为:

$$h \times w \times k^2 \times c^2 \quad (5)$$

其中, c 和 c_p 为道编号, h 和 w 是输入数据的高度和宽度。由于 PConv 把第一个或最后一个连续的 c_p 通道作为整个特征映射的代表进行计算, 一般性情况下, 通常认为输入特征映射和输出特征映射的通道数相同。比例比率 $R = c_p = 1$, 与 Conv 相比, PConv 的 FLOPs 降低了 16 倍。PConv 的工作原理如图 5 所示。

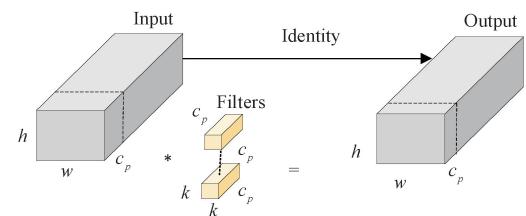


图 5 PConv 的工作原理

Fig. 5 Working Principle of PConv

本文通过 PConv 和 Conv 设计出残差快速卷积模块, 在每个 R_FConv 模块中, 3×3 PConv 层占据中心位置, 融合了点式卷积的概念, 最大程度地利用所有信道信息, 同时在 PConv 后面插入两个 1×1 Conv 层, 在 3×3 PConv 与 1×1 Conv 之间建立残差连接, 确保丰富的特征提取, 把 BatchNormalization 放在 Conv 之后, 同时以 RELU 激活函数作为补充, 优化模型推理速度。把 R_FConv 模块整合到 YOLOv8 中, 相较于原始的 Bottleneck, RF_Bottleneck 去掉了残差网络结构, 有效的替换了原颈部网络中 C2f 的 Bottleneck, 如图 6 所示。

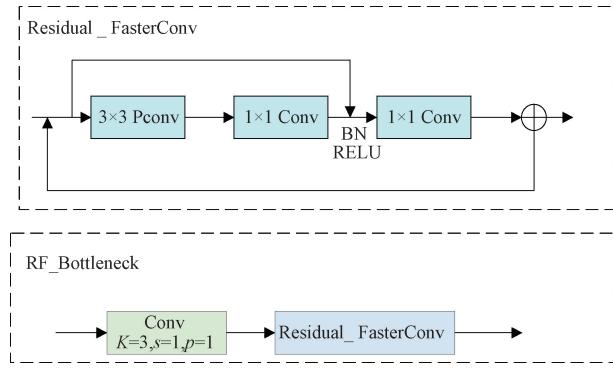


图 6 R-FConv 模块和 RF_Bottleneck 模块

Fig. 6 R-FConv module and RF_Bottleneck module

2.3 损失函数

在进目标检测任务时, 边界框回归损失函数对于模型至关重要, 它通过学习预测边界框的位置, 使得模型能够尽可能地接近真实边界框, 从而提供更准确的定位和区域信息。YOLOv8 使用的损失函数为 CIOU, 计算公式如下:

$$L_{\text{CIOU}} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{D^2} + \alpha v \quad (6)$$

$$v = \frac{4}{\pi^2} (\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h})^2 \quad (7)$$

$$\alpha = \frac{v}{1 - IoU + v} \quad (8)$$

其中, ω^{gt}, h^{gt} 和 ω, h 分别代表目标框的宽高和预测框的宽高, $\rho^2(b, b^{gt})$ 代表预测框与目标框中心点距离, IoU 代表目标框与预测框交并比, D 表示预测框与目标框之间的最小外接矩形的对角线距离, α 代表可调节的超参数, v 表示长宽比的模板匹配项。

CIOU 存在一定局限性, 它并没有直接捕捉锚框和目标框之间的形状差异, 从而导致不理想的收敛行为。为解决上述缺陷, 本文引入 PIoU v2 损失函数, 其中一个适应目标尺寸的惩罚因子 P 定义如下:

$$P = (\frac{d\omega_1}{\omega_{gt}} + \frac{d\omega_2}{\omega_{gt}} + \frac{dh_1}{h_{gt}} + \frac{dh_2}{h_{gt}}) / 4 \quad (9)$$

其中, $d\omega_1, d\omega_2, dh_1$ 和 dh_2 代表预测框和目标框边界之间距离并取其绝对值, ω_{gt} 和 h_{gt} 代表目标框的宽高。 P 只取决于目标框的尺寸, 与锚框和目标框的最小外部框的尺寸无关, 锚框扩大不会改变 P 。PIoU 损失函数计算公式如下:

$$f(x) = 1 - e^{-x^2} \quad (10)$$

$$PIoU = IoU - f(P), -1 \leqslant P \leqslant 1 \quad (11)$$

$$L_{\text{PIoU}} = 1 - P \text{IoU} = L_{\text{IoU}} + f(P), 0 \leqslant L_{\text{PIoU}} \leqslant 2 \quad (12)$$

为了增强了对中高质量锚框的关注能力, 添加一个由超参数控制的非单调注意力函数与 PIoU 相结合, 得到 PIoU v2 损失函数, 其计算公式如下:

$$q = e^{-p}, q \in (0, 1] \quad (13)$$

$$u(x) = 3x \cdot e^{-x^2} \quad (14)$$

$$L_{\text{PIoU_v2}} = u(\lambda q) \cdot L_{\text{PIoU}} = 3 \cdot (\lambda q) \cdot e^{-(\lambda q)^2} \cdot L_{\text{PIoU}} \quad (15)$$

其中, $u(\lambda q)$ 表示注意函数, λ 是控制注意力函数行为的超参数。本文用惩罚因子 q 代替 P , 用于测量锚框的质量, 当 $q = 1$, 意味着 $P = 0$, 表示锚框与目标框完全对齐, P 增加, q 逐渐减小, 代表较低质量的锚框。PIoU v2 损失函数损失不仅融合了非单调注意力函数, 还充分利用了注意力机制的潜力, 只需要一个超参数便简化了调整过程。

3 实验分析

3.1 实验环境

训练环境为 Windows11 操作系统、Intel Core i9-13900H、32 G 内存, 采用 NVIDIA GeForce RTX 4070 显存 8 GB 的 GPU 网络进行实验, 并使用 Python 3.8、Pytorch 1.11.0 版本和 torchvision 0.12.0 作为实验性深度学习框架。训练时初始学习率 0.01 batch-size 为 8, epochs 为 300。

3.2 数据集

为了验证本文算法的有效性, 在 DIOR 遥感数据集^[24]、RSOD 遥感数据集^[25]和 VisDrone2019 数据集^[26]上进行了实验。DIOR 数据集是一个光学遥感图像目标检测的大型数据集, 该数据集有 23 463 张图像, 共有 20 种类别。RSOD 数据集是一个公开的目标检测数据集, 该数据集包括 946 张图像, 共有飞机、操场、立交桥和油箱 4 种类别。VisDrone2019 数据集是天津大学实验室团队在各种复杂场景和光照条件下进行拍摄的, 该数据集包括 8 629 张图像, 共有 10 类标签。3 组数据集训练中将数据按 7:2:1 的比例划分成训练集、验证集、测试集。在 RSOD 数据集中, 有 40 张有关操场图像没有标签, 所以训练过程中仅使用标签完整的操场图像。

3.3 实验评估指标

为了直观的体现改进后模型的性能, 实验评估指标选用准确率(precision, P)、召回率(recall, R)、平均精度的均值(mean average precision, mAP)以及浮点运算次数(giga floating-point operations per second, GFLOPs)。P、R、mAP 具体计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (16)$$

$$R = \frac{TP}{TP + FN} \quad (17)$$

$$mAP = \frac{1}{k} \sum_{i=1}^k AP_i \quad (18)$$

其中, TP 表示真正例, 即模型正确地预测为正例的样本数, FP 表示假正例, 即模型错误地将负例预测为正例的样本数, FN 表示假负例, 即模型错误地将正例预测为负例的样本数, AP_i 表示准确率和召回率曲线下方的面积, k

表示所有的类别数。

3.4 消融实验

为更好地验证各个模块对原网络的改进效果,在 DIOR

数据集上设计了 4 组消融实验,在训练过程中分别评估 C2f_MSDA 模块、R_FConv 模块以及 PIoU v2 损失函数对原 YOLOv8n 模型的改进效果,具体的测试数据如表 1 所示。

表 1 DIOR 数据集的消融实验数据

Table 1 Data from ablation experiments on the DIOR dataset

算法	P/%	R/%	mAP/%	GFLOPs
YOLOv8n	90.1	80.3	86.5	8.9
YOLOv8n+C2f_MSDA	89.5	80.8	87.2	9.0
YOLOv8n+R_FConv	90.3	90.1	88.3	8.5
YOLOv8n+PIoU v2	89.2	82	87.7	8.9
YOLOv8n+C2f_MSDA+R_FasterConv+PIoU v2	90.6	81.5	89.2	8.4

通过实验数据可知,原网络模型的 P、R、mAP、GFLOPs 分别为 89%、80.3%、86.5%、8.9。在主干网络部分加入部分 C2f_MSDA 模块后,R 和 mAP 分别提高了 0.5% 和 0.7%,P 降低了 0.6%,由此可知 C2f_MSDA 模块使改进后的模型可以更好捕获语义信息,提高对标注目标的注意;通过加入 R_FConv 模块,使得 P、R 和 mAP 分别提高了 0.2%、0.8% 和 1.8%,计算量也有明显减小,R_FConv 模块实现了模型的轻量化^[27];损失函数的选择方面,选用 PIoU v2 损失函数,与原模型各项指标相比,P 下降 0.9%,R 和 mAP 提高了 1.7% 和 1.2%,虽然准确率略有降低,但是召回率和平均精度都有显著提升,改善模型收敛速度和鲁棒性;最后同时添加 C2f_MSDA 模块、R_FConv 模块以及替换 PIoU v2 损失函数后,P、R 和 mAP 分别提高了 0.5%、1.2% 和 2.7%,计算量减少了 0.5 GFLOPs,各项性能指标均有不同程度的提升。

3.5 对比实验

考虑到更好的验证本文方法的有效性,设置相同的训练参数,在 DIOR 数据集上与其他经典方法如 YOLOv3、YOLOv5、YOLOv6、YOLOv8n 以及改进方法 DS-YOLOv8^[28]进行了对比试验,结果如表 2 所示。

表 2 DIOR 数据集的对比实验数据

Table 2 Comparative experimental data from the

DIOR dataset

模型	P/%	R/%	mAP/%	GFLOPs
YOLOv3	81.7	79.0	84.3	12.8
YOLOv5	88.1	78.0	84.7	7.8
YOLOv6	87.7	77.3	83.6	13.1
YOLOv8n	89.0	79.3	85.5	8.9
DS-YOLOv8	90.1	82.9	88.1	10.2
改进 YOLOv8n	90.6	81.5	89.2	8.4

通过实验结果可知,YOLOv8n 网络模型的 P、R、mAP 指标与前几代 YOLO 模型各项指标相比都有提升,而本文

所改进后的 YOLOv8n 网络模型在原网络模型的基础上,P、R、mAP、GFLOPs 等各项指标均有提升,与 DS-YOLOv8 相比,R 略有下降。本文所改进后的模型不仅实现了网络的轻量化,还提高了模型对图像目标的检测精度,达到模型优化的目的,证明了本文改进方法的有效性。

同时为了证明本文方法的适用性,保持训练的各项参数不变,数据集由 DIOR 数据集换成 RSOD 数据集,通过实验可得各项性能指标的结果,如表 3 所示。

表 3 RSOD 数据集的对比实验数据

Table 3 Comparative experimental data for the RSOD dataset

模型	P/%	R/%	mAP/%	GFLOPs
YOLOv8n	92.5	82.8	89.8	8.9
改进 YOLOv8n	93.3	87.4	93.1	8.4

由表 3 分析可得,改进后网络模型的各项性能指标都有提升,其中 P、R 和 mAP 分别提高了 0.8%、4.6% 和 3.3%,计算量进一步降低,由 8.9 降为 8.4。此外,由 RSOD 数据集换成 VisDrone2019 数据集,通过实验可得各项性能指标的结果,如表 4 所示。

表 4 VisDrone2019 数据集的对比实验数据

Table 4 Comparative experimental data for the

VisDrone2019 dataset

模型	P/%	R/%	mAP/%	GFLOPs
YOLOv3	40.8	32.7	34.1	12.8
YOLOv5	43.9	34.2	33.1	7.8
YOLOv6	42.8	34.5	29.8	13.1
YOLOv8n	44.5	35.9	33.8	8.9
改进 YOLOv8n	48.2	38.7	37.6	8.4

由表 4 分析可得,本文改进的 YOLOv8n 网络模型与原网络模型相比,P、R、mAP、GFLOPs 等各项指标均有提升,与前几代 YOLO 模型相比,P、R 和 mAP 均有提升,与

YOLOv5 相比计算量略大, 验证了该改进方法的广泛适用性。

为了直观的对比, 将 YOLOv8n 算法与改进后

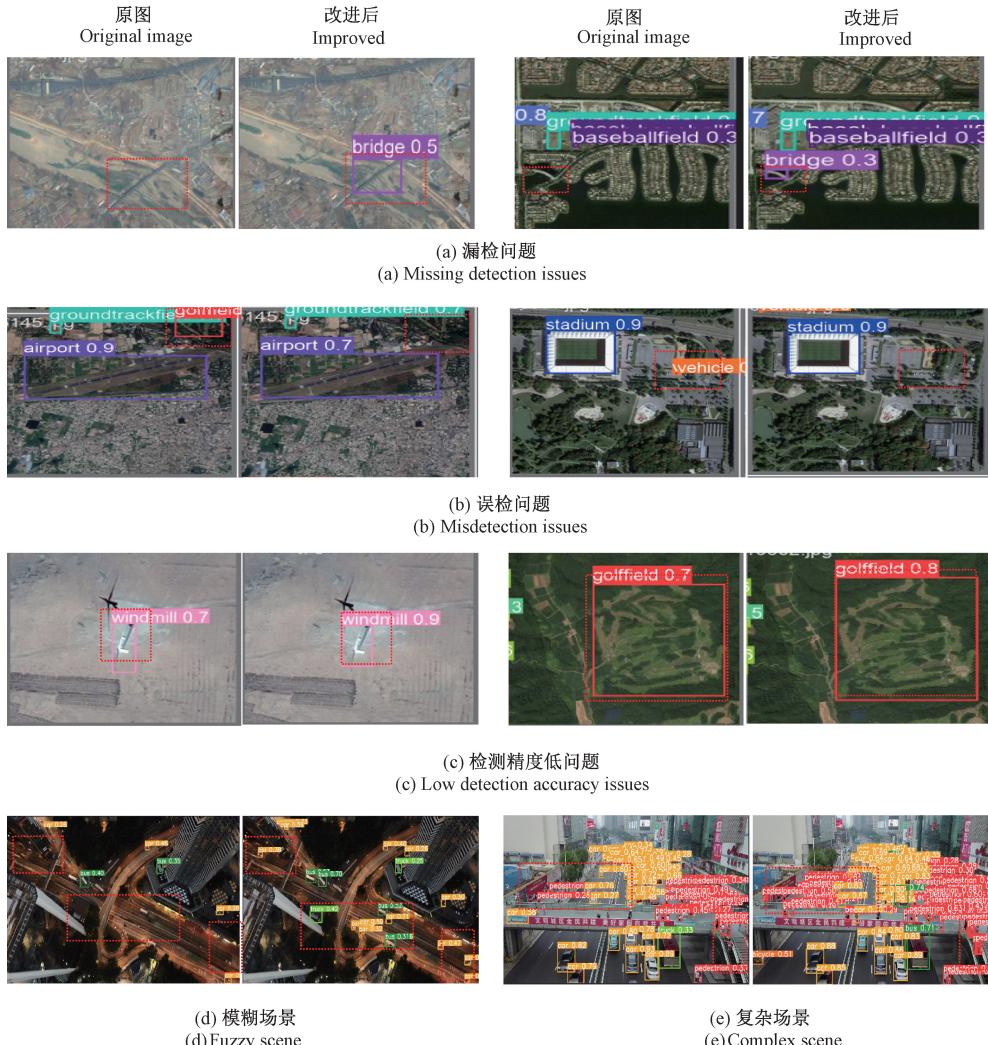
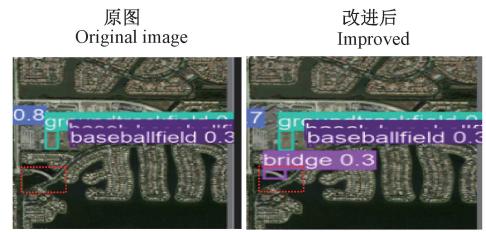


图 7 改进前后检测效果对比

Fig. 7 Comparison of detection results before and after improvement

图 7 为原 YOLOv8n 算法检测效果图和改进后的检测效果图。由图 7(a)的检测结果可知, 若待检测的目标分布隐蔽或目标尺度较小, 原算法因为特征提取不充分而造成桥梁漏检, 改进后的算法加入了 C2f_MSDA 模块, 可以更好的提取目标的关键特征, 有效的解决了这一问题。由图 7(b)的检测结果可知, 若待检测目标分布集中且背景复杂, 原始 YOLOv8n 算法的关键特征提取不够准确, 误将树林的阴影识别为高尔夫球场, 误将墙壁检测为车辆, 改进后的算法有更好的特征提取能力, 有效地解决了对于车辆这种小目标的误检。由图 7(c)的检测结果如图所示, 若待检测目标相对单一或者周边目标模糊, 原模型检测精度较低, 改进后的算法使用 PIoU v2 损失函数代替 CIOU 损失函数, 提高检测框的定位能力, 比原算法检测精度更高。

YOLOv8n 算法在 DIOR 数据集和 VisDrone2019 数据集的部分图像进行检测效果展示, 对比结果用红色虚线框标注, 如图 7 所示。



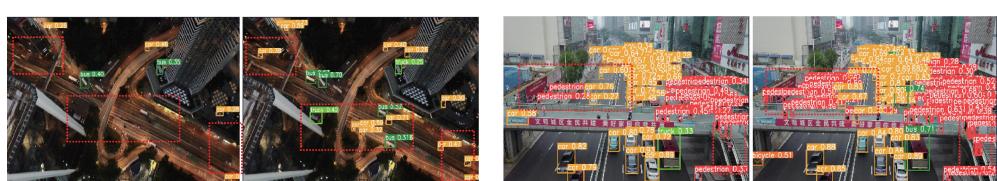
(a) 漏检问题
(a) Missing detection issues



(b) 误检问题
(b) Misdetection issues



(c) 检测精度低问题
(c) Low detection accuracy issues



(d) 模糊场景
(d) Fuzzy scene

(e) 复杂场景
(e) Complex scene

图 7(d)是模糊场景下的检测图, 由于光照的原因而拍摄到的模糊图像, 改进算法仍可以较全面的检测出目标。图 7(e)是复杂场景下的效果图, 在小样本目标众多的情况下, 改进后的算法能更准确的检测到微小目标。综合来看, 改进后的算法在一定程度上解决了目标漏检、误检的问题, 并在检测精度上有进一步提升。

4 结 论

为了解决遥感图像因背景复杂、待检测目标隐蔽、目标尺度变化大而导致检测效果不理想的问题, 本文提出了一种基于改进 YOLOv8 的遥感图像检测算法。该算法首先通过在骨干网络进行模块的替换, 把部分 C2f 模块替换为 C2f_MSDA, 弥补了采样过程中语义信息的丢失问题,

有效提高了算法对小目标的检测效果;然后,该模型通过加入 R_FConv 模块,既减小了模型计算量又提高特征提取能力;最后引入 IoU v2 损失函数,大大加快了模型收敛速度,提升目标定位精度。经实验对比,在 DIOR 数据集上改进后网络比原网络的准确率、召回率、平均精度的均值分别提高 0.5%、1.2% 和 2.7%,在 RSOD 数据集上分别提高 0.8%、4.6% 和 3.3%,在 VisDrone2019 数据集上分别提高 3.7%、2.8% 和 3.8%,证明改进方法有效。在接下来的研究中,准备收集更多恶劣天气下的遥感图像进行训练,以增强模型的适应能力。

参考文献

- [1] OSCO L P, JUNIOR J M, RAMOS A P M, et al. A review on deep learning in UAV remote sensing[J]. International Journal of Applied Earth Observation and Geoinformation, 2021, 102: 102456.
- [2] YU Y T, GU T N, GUAN H Y, et al. Vehicle detection from high-resolution remote sensing imagery using convolutional capsule networks [J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(12): 1894-1898.
- [3] ZHANG B, WU Y F, ZHAO B Y, et al. Progress and challenges in intelligent remote sensing satellite systems[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 1814-1822.
- [4] CHENG G, XIE X X, HAN J W, et al. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 3735-3756.
- [5] 付涵,范湘涛,严珍珍,等.基于深度学习的遥感图像目标检测技术研究进展[J].遥感技术与应用,2022,37(2): 290-305.
- [6] FU H, FAN X T, YAN ZH ZH, et al. Advances in deep learning-based target detection technology for remote sensing images [J]. Remote Sensing Technology and Applications, 2022, 37(2): 290-305.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158.
- [8] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN [C]. IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [9] 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(CVPR), Las Vegas: IEEE, 2016: 779-788.
- [10] GUO W, LI W H, LI ZH H, et al. A slimmer network with polymorphic and group attention modules for more efficient object detection in aerial images[J]. Remote Sensing, 2020, 12(22): 3750.
- [11] 张绍文,史卫亚,张世强,等.基于加权感受野和跨层融合的遥感小目标检测[J].电子测量技术,2023,46(18):129-138.
- [12] ZHANG SH W, SHI W Y, ZHANG SH Q, et al. Remote sensing small target detection based on weighted receptive field and cross-layer fusion [J]. Electronic Measurement Technology, 2023, 46(18): 129-138.
- [13] ZHOU Q K, ZHANG W, LI R ZH, et al. Improved YOLOv5-S object detection method for optical remote sensing images based on contextual transformer[J]. Journal of Electronic Imaging, 2022, 31(4): 043049.
- [14] LIU ZH G, GAO Y, DU Q Q, et al. YOLO-extract: Improved YOLOv5 for aircraft object detection in remote sensing images[J]. IEEE Access, 2023, 11: 1742-1751.
- [15] WANG G, CHEN Y F, 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.
- [16] 王德伟,刘小芳.基于改进 YOLOv8n 的带钢表面缺陷检测方法[J].国外电子测量技术,2024,43(7): 158-169.
- [17] WANG D W, LIU X F. Steel strip surface defect detection method based on improved YOLOv8n[J]. Foreign Electronic Measurement Technology, 2024, 43(7): 158-169.
- [18] JIAO J Y, TANG Y M, LIN K Y, et al. Dilateformer: Multi-scale dilated transformer for visual recognition [J]. IEEE Transactions on Multimedia, 2023, 25: 8906-8919.
- [19] LIU C, WANG K G, LI Q, et al. Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism [J]. Neural Networks, 2024, 170: 276-284.

- [18] ZHENG ZH H, WANG P, REN D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2021, 52(8): 8574-8586.
- [19] 赵鸿亮, 郭佑民, 王建鑫, 等. 基于改进 YOLOv5 的轨道交通障碍物检测算法[J]. 电子测量技术, 2024, 47(1): 130-135.
- ZHAO H L, GUO Y M, WANG J X, et al. Obstacle detection algorithm for rail transit based on improved YOLOv5[J]. Electronic Measurement Technology, 2024, 47(1): 130-135.
- [20] 王泽宇, 徐慧英, 朱信忠, 等. 基于 YOLOv8 改进的密集行人检测算法: MER-YOLO[J]. 计算机工程与科学, 2024, 46(6): 1050-1062.
- WANG Z Y, XU H Y, ZHU X ZH, et al. An improved dense pedestrian detection algorithm based on YOLOv8: MER-YOLO[J]. Computer Engineering and Science, 2024, 46(6): 1050-1062.
- [21] NEUBECK A, GOOL L V. Efficient non-maximum suppression [C]. 18th International Conference on Pattern Recognition(ICPR'06), 2006: 850-855.
- [22] 曾勇杰, 范必双, 杨涯文, 等. 改进 YOLOv8 算法在风机叶片缺陷检测上的应用[J]. 电子测量与仪器学报, 2024, 38(8): 26-35.
- ZENG Y J, FAN B SH, YANG Y W, et al. Application of improved YOLOv8 algorithm on wind turbine blade defect detection[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (8): 26-35.
- [23] CHEN J R, KAO SH H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [24] LI K, WANG G, CHENG G, et al. Object detection in optical remote sensing images: A survey and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.
- [25] LONG Y, GONG Y P, XIAO ZH F, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55 (5): 2486-2498.
- [26] 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, 2019: 213-226.
- [27] 张国鹏, 周金治, 马光岑, 等. 改进 YOLOv8 的轻量化安全帽佩戴检测算法[J]. 电子测量技术, 2024, 47(17): 147-154.
- ZHANG G Q, ZHOU J ZH, MA G C, et al. Improved lightweight helmet wear detection algorithm for YOLOv8[J]. Electronic Measurement Technology, 2024, 47(17): 147-154.
- [28] SHEN L Y, LANG B H, SONG ZH X. DS-YOLOv8-Based object detection method for remote sensing images [J]. IEEE Access, 2023, 11: 125122-125137.

作者简介

宋树成, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 3177790996@qq.com

程换新(通信作者), 博士, 教授, 主要研究方向为人工智能、先进控制、机器视觉。

E-mail: 15762612129@163.com