

引用格式: 杨雪锋, 刘佳尧, 周昌华. MITD-YOLO: 改进 YOLOv8n 的海上红外目标检测方法 [J]. 中国舰船研究, 2026, 21(2): 424-434.

YANG X F, LIU J Y, ZHOU C H. MITD-YOLO: an improved YOLOv8n-based method for maritime infrared target detection[J]. Chinese Journal of Ship Research, 2026, 21(2): 424-434 (in Chinese).

MITD-YOLO: 改进 YOLOv8n 的海上红外目标检测方法



扫码阅读全文

杨雪锋^{*1,2}, 刘佳尧¹, 周昌华¹

1 重庆交通大学 航运与船舶工程学院, 重庆 400074

2 交通安全应急信息技术国家工程实验室, 北京 100011

摘要: [目的] 海上红外图像背景复杂、目标尺寸变化大、海浪杂波干扰严重, 容易导致目标漏检和误检。为提高红外图像中的目标检测准确率, 提出一种基于 YOLOv8n 的海上红外目标检测方法——面向海上红外目标检测任务的 YOLO(MITD-YOLO)。[方法] 首先引入多样化分支模块 (DBB) 和多尺度卷积 (EMSCConv), 利用多个不同尺度的卷积使模型能够更好地捕捉复杂特征。然后, 采用三重注意力机制 (triple attention) 实现空间和通道维度的特征交互, 强化关键特征提取。最后, 使用 Powerful-IoUv2 (PIoUv2) 对原模型的损失函数进行改进, 以解决锚框扩展问题, 提高检测精度并增强模型的鲁棒性。[结果] 实验结果表明, 改进的 MITD-YOLO 模型对海上红外图像目标的检测效果有所提升: 准确率提升 2.3%, 召回率提升 1.7%; 平均准确率达 88.9%, 帧率 (FPS) 达到 132.8, 优于原模型。[结论] 该方法可提高海上红外目标检测效果, 为海上安全监控和船舶导航等领域提供更可靠的目标检测技术, 助力智能海洋系统发展。

关键词: 目标检测; 红外目标检测; 多尺度卷积; 三重注意力机制

中图分类号: U675.3

文献标志码: A

DOI: 10.19693/j.issn.1673-3185.04311

0 引言

近年来, 红外成像技术因其能够在完全黑暗或恶劣天气条件下探测目标, 受到越来越多研究人员的关注。在海上环境中, 红外成像技术可以穿透海雾, 减少海面反射干扰, 提高对远距离小目标的探测能力, 已广泛应用于船舶航行环境感知、海事监管、搜救巡逻、海上军事行动等领域。然而, 由于海上红外图像背景复杂、目标尺寸变化大、海浪杂波干扰严重等, 海上红外图像目标检测的准确率还不高。

现阶段, 红外目标的检测方法主要分为背景差法、优化方法和深度学习方法 3 类。背景差法^[1-2]的主要思路是对图像背景建模, 然后从原始图像中减去背景, 灰度值为 0 的区域为背景, 非 0 区域为目标。这种方法假设背景可预测, 在静态背景且无明显光照变时具有较好的效果, 但难

以适应动态背景或目标运动场景。海上红外图像背景中的海浪杂波难以用该方法建模。基于优化的红外目标检测方法^[3]是将目标检测问题转化为稀疏低秩矩阵的优化问题, 通过找到原始图像的最佳稀疏表示, 使重构图像与原始图像差异最小化, 同时确保目标特征被准确表示。该方法可以通过模型优化提升性能, 但低秩背景的假设使得其对背景均匀性的要求很高, 海面杂波会降低目标检测的准确率。基于深度学习的方法主要通过构建深度网络模型, 利用大量数据训练模型, 挖掘图像特征, 再利用训练好的网络模型检测目标。该方法具有较高准确率和快速检测能力, 是目前红外图像目标检测领域的重要发展方向。

基于深度学习的目标检测方法主要分为双阶段方法和单阶段方法。双阶段方法通常先生成候选区域, 然后对每个候选区域进行分类和回归。常用的双阶段方法包括区域卷积神经网络特征

收稿日期: 2024-12-13 修回日期: 2025-04-07 网络首发时间: 2025-07-03 13:35

基金项目: 交通安全应急信息技术国家工程实验室开放基金项目 (YW170301-05); 重庆市自然科学基金资助项目 (cstc2019jcyj-msxmX0729)

作者简介: 杨雪锋, 男, 1987 年生, 博士, 讲师。研究方向: 水上智能运输系统, 船舶航行环境感知, 航迹规划。

E-mail: yangxuefeng536@cqjtu.edu.cn

*通信作者: 杨雪锋

(regions with CNN features, R-CNN)^[4]、掩码区域卷积神经网络(mask R-CNN)等。Gupta^[5]采用 Mask R-CNN 算法进行行人检测,展示了该算法在高质量图像分割和检测中的优势。Ou 等^[6]提出一种改进的 Faster R-CNN 模型,用于变电站电气设备的红外图像检测,通过优化特征提取网络和增加锚框比例,提高了检测精度和速度。单阶段方法将目标检测问题直接转化为回归问题,常用的单阶段算法有 YOLO 系列算法和单镜头多盒检测器(single shot multibox detector, SSD)算法。Zhao 等^[7]提出一种改进的 SSD 算法,通过使用跨步卷积层替换最大池化层,增强浅层特征信息,并引入残差单元和 MSRA 函数初始化权重,显著提升了红外小车目标检测的精度。陈德海等^[8]基于 YOLOv3 DarkNet-53 主干网络结构,增添一个浅层检测尺度,并添加 8 倍上采样使其与最后一层检测尺度进行信息融合,有效提升了小目标的检测效果。Wang 等^[9]设计细节增强模块,结合空间注意力机制以增强图像纹理和细节,并通过修改 YOLOv4 模型,引入 Alpha-IoU 损失和加权非极大值抑制(non-maximum suppression, NMS),显著提升了检测精度。张炳焱等^[10]运用基于 FasterNet 思想改进的 FasterNeXt 模块替换 YOLO 模型中的 C3 模块,同时添加神经注意力模块(neural attention module, NAM)注意力机制,在不影响准确性的条件下提高了运行速度。Hu 等^[11]提出一种基于 YOLOv7-tiny 的改进算法,根据纵横比分配锚框和使用困难样本挖掘损失函数,显著提升了无人机图像中小目标的检测精度。张瑶等^[12]提出一种改进的 YOLOv8 水面小目标检测算法,基于 BiFormer 双层路由注意力机制构建的 C2fBF 模块,并添加小目标检测头,使用多视角距离交并比(multi-perspective distance intersection over union, MPD IoU)替换损失函数,有效解决了水面目标检测中的噪声干扰和小目标漏检问题。

从上述分析可以看出, YOLO 系列模型在目标检测领域的改进主要围绕网络结构优化、多尺度特征融合、注意力机制、损失函数改进、锚框优化策略 5 个维度展开,具体改进方法需与应用场景目标特征相匹配。海上红外图像存在背景复杂、目标尺寸变化大、海浪杂波干扰严重等问题,导致现有方法对海上红外目标检测的准确率不高,存在误检和漏检现象,需结合海上红外目标特征对目标模型进行优化。因此,本文拟利用 YOLOv8 模型进行海上红外目标检测实验,结合检测结果和红外图像特征,设计一种基于 YOLOv8n 的海上红外目标检测方法 MITD-YOLO,并通过

实验验证所提改进方法在提高海上红外目标检测准确率方面的效果。

1 理论方法

海上红外图像通常具有背景复杂、对比度较低、船舶种类繁多且尺度差异大、遮挡重叠等问题。这些特性使得原始的 YOLOv8n 模型在处理海上船舶红外数据集时,存在漏检、误检、检测精度低等问题。针对这些问题,本文以 YOLOv8n 模型为基础,进行了以下改进:

1) 使用 C2f 动态分支模块 C2f_DBB^[13] 替换 C2f 模块。通过并行分支的多样化卷积操作,在保持计算成本相当的情况下提高模型的检测精度。

2) 采用 EMSConv 对检测头进行改进。通过精简的多尺度卷积捕捉不同尺度的红外目标特征,并减少参数数量,提升检测精度和速度。

3) 引入三重注意力机制(triplet attention)^[14]。在处理复杂场景和小目标时,该机制能增强特征表示,分离有用特征和噪声,从而提高模型的鲁棒性和可靠性。

4) 将损失函数替换为 PIoUv2^[15]。该函数能更准确地评估检测框的重叠情况,处理密集和重叠目标的检测问题,对不同形状和大小的目标具有更好的适应性,减少误检和漏检。

改进后的网络模型如图 1 所示,图中, Contact 表示拼接矩阵, Upsample 表示上采样, k 表示卷积核的尺寸, s 表示卷积的步幅, p 表示填充的像素数。

1.1 改进 C2f 模块

在 YOLOv8n 模型中,特征提取和融合部分采用 C2f 模块。该模块中的 Bottleneck 卷积使用固定卷积核,缺乏灵活性,导致网络感受野受限,仅能捕捉目标的局部信息。由于海上红外目标尺度变化大,固定卷积核无法有效捕捉不同尺度下目标的特征,且不能根据目标尺度进行自适应调整,因此检测效果不理想。针对该问题,本文构建多样化分支模块 C2f_DBB,以替换 YOLOv8n 模型中的 C2f 模块,其网络结构如图 2 所示。

DBB 采用多分支拓扑结构,包括多尺度卷积(conv)、顺序的 1×1 和 $k \times k$ 卷积、平均池化(average pooling)、批标准化(batch norm)和分支相加,其网络结构如图 3 所示。

在 DBB 中, 1×1 卷积快速聚合通道信息, $K \times K$ 卷积扩展感受野,使得这些分支具有不同的感受野和不同复杂程度的路径,结合不同尺度的

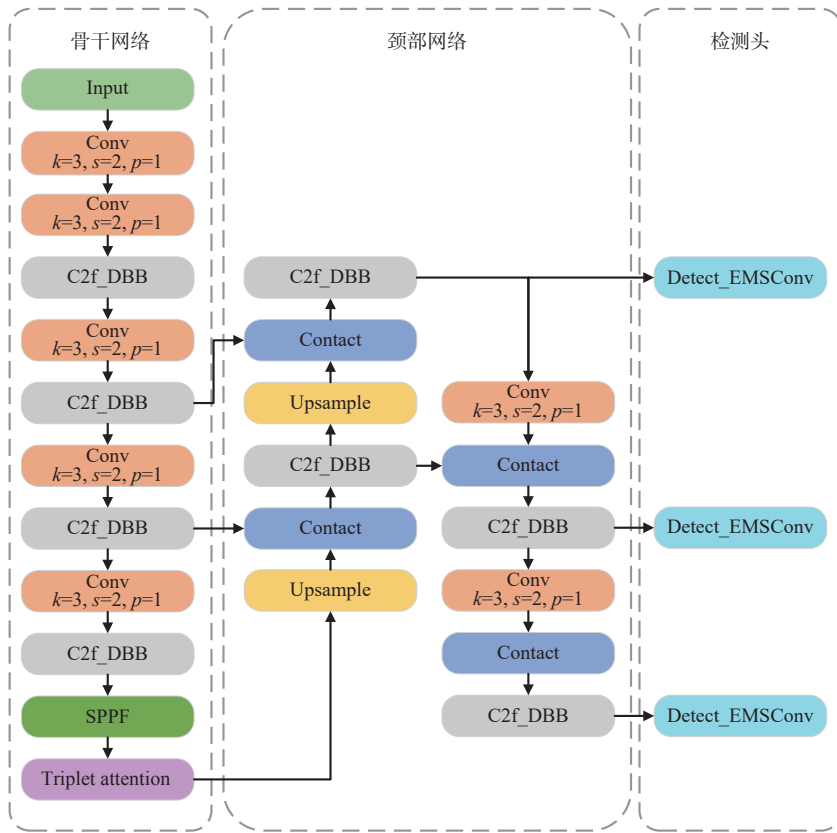


图1 MITD-YOLOv8 网络结构图

Fig. 1 Network structure of MITD-YOLOv8

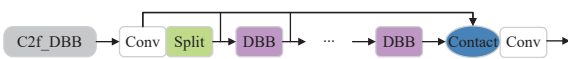


图2 C2f_DBB 网络结构图

Fig. 2 Network structure of C2f_DBB

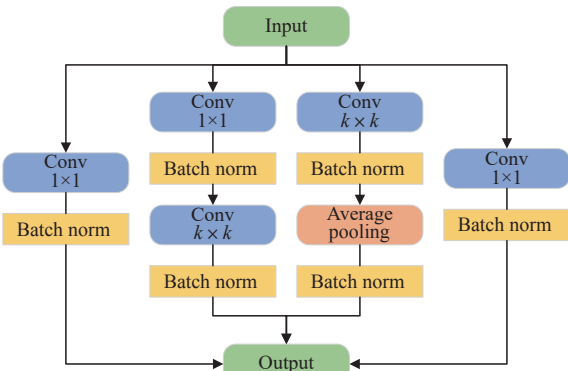


图3 DBB 网络结构图

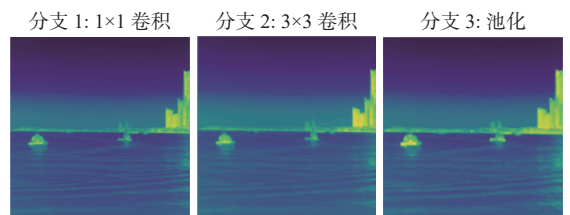
Fig. 3 Network structure of DBB

卷积,能够丰富特征空间,使模型能够捕获从局部细节到更广泛区域的特征,改善在复杂海洋背景中的识别效果。在DBB中,平均池化通常与卷积操作结合,作为下采样的手段,帮助模型在保留关键信息的同时减少冗余特征。在海上红外图像目标检测中,池化操作有助于减少海面背景噪声的影响。同时,DBB可以在运算时等效地转换为单一卷积。因此,可以在训练阶段用DBB替换常规卷积层来构建C2f_DBB模块,训练完成再将其转换回原始的卷积结构,从而降低运算成本。

DBB在每次前向传递之前不会推导出参数,而会在训练结束后一次性转换模型,然后只保存并使用转换后的模型。尽管DBB和常规卷积层在推理时的结构相同,但应用DBB可以提高模型的特征表达能力以及模型的鲁棒性,具有很高的灵活性和扩展性。C2f卷积与DBB卷积特征提取对比图如图4所示。



(a) C2f 卷积特征提取图



(b) DBB 卷积特征提取图

图4 C2f卷积与DBB卷积特征提取对比图

Fig. 4 Feature extraction comparison between C2f convolution and DBB convolution

1.2 改进检测头

YOLOv8n 检测头中采用了 2 个不同分支的

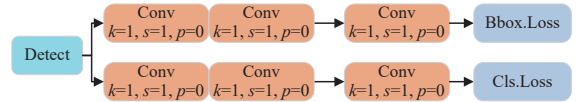
卷积层模块 cv2 和 cv3 用来表示分类和回归, 其中 cv2 主要用于回归输出预测物体的边界框信息, cv3 主要用于分类输出判断每个物体的类别。cv2 和 cv3 每个分支有 2 个 3×3 卷积和 1 个 1×1 卷积, 在训练过程中, 需要遍历 3 个检测头, 因此 YOLOv8n 模型中的检测头占据了较大的计算量, 如图 5(a) 所示。针对这种情况, 本文采用 EMSConv 模块构建一个高效的具有多尺度信息共享参数的检测头, 其结构如图 5(b) 所示。

EMSConv 基于分组卷积思想, 对每组的通道数进行卷积特征提取操作。EMSConv 将通道分为两组, 一半通道生成廉价特征图, 另一半再分为两组: 其中一组采用 3×3 卷积提取特征, 另一组采用 5×5 卷积提取特征。这些特征在单一特征通道上完成, 各通道间信息独立。最后通过 1×1 逐点卷积交换通道信息, 对特征通道进行降维融合, 实现多尺度特征提取。

1.3 添加注意力机制

目前, 应用于 YOLOv8 图像检测的注意力机制主要有通道注意力机制、空间注意力机制与混合注意力机制。这些机制在低对比度、高噪声情形下的特征提取能力较弱, 且会增加模型复杂度。为了解决这些问题, 本文引入三重注意力机制, 通过 3 个分支结构捕捉跨维度信息, 能够自适应地突出关键特征, 其网络结构如图 5 所示。

输入张量被传递至三重注意力模块的各分支。第 1 个分支将输入张量 χ 绕着 H 轴逆时针旋转 90° , 得到张量 $\hat{\chi}_1$, 其形状为 $(W \times H \times C)$, 经 Z 池化降维后得到形状为 $(2 \times H \times C)$ 的张量 $\hat{\chi}_1^*$, 后通过 $k \times k$ 的卷积层和批归一化处理, 将其降维至 $(1 \times H \times C)$, 并经 Sigmoid 激活层 (σ) 生成注意力权重。最终绕着 H 轴顺时针旋转恢复原始输入形状。第 2 个分支绕着 W 轴逆时针旋转输入张量



(a) Detect 网络结构图



(b) Detect EMSConv 网络结构图

图 5 检测头网络结构图

Fig. 5 Network structure of the detection head

χ 得到形状为 $(H \times C \times W)$ 的张量 $\hat{\chi}_2$, 再经过 Z 池化降维至 $(2 \times C \times W)$ 的 $\hat{\chi}_2^*$, 后经卷积和批归一化处理, 输出形状为 $(1 \times C \times W)$ 的张量。随后经激活层生成注意力权重, 再旋转回原形状 χ 。第 3 个分支中, 张量 χ 通过 Z 池化压缩到 $(2 \times H \times W)$, 后将压缩张量传递给一个由卷积核大小 k 定义的标准卷积层和一个批归一化层, 最后经激活层生成形状为 $(1 \times H \times W)$ 的注意力权重。三重注意力机制的网络结构如图 6 所示。

对每个分支产生的形状为 $(C \times H \times W)$ 的精炼张量进行简单平均聚合。其输出应用张量 y 的公式为

$$y = \frac{1}{3} (\hat{\chi}_1 \sigma(\psi_1(\hat{\chi}_1^*)) + \hat{\chi}_2 \sigma(\psi_2(\hat{\chi}_2^*)) + \chi \sigma(\psi_3(\hat{\chi}_3))) \quad (1)$$

式中: σ 为 Sigmoid 激活函数; ψ_1, ψ_2 和 ψ_3 为由 3 个分支中卷积核大小为 k 的标准二维卷积层定义的注意力三元组。

上述过程中, Z 池化层 Z_{pool} 通过将平均池化特征 $\text{Avg}_{\text{Pool0d}}(\chi)$ 和最大池化特征 $\text{Max}_{\text{Pool0d}}(\chi)$ 在该维度上进行拼接, 将张量的第 0 维度缩减为 2。这使得该层在保留实际张量丰富表示的同时, 缩小其深度, 从而使后续的计算更为轻量化。其公式为

$$Z_{\text{pool}}(\chi) = [\text{Max}_{\text{Pool0d}}(\chi), \text{Avg}_{\text{Pool0d}}(\chi)] \quad (2)$$

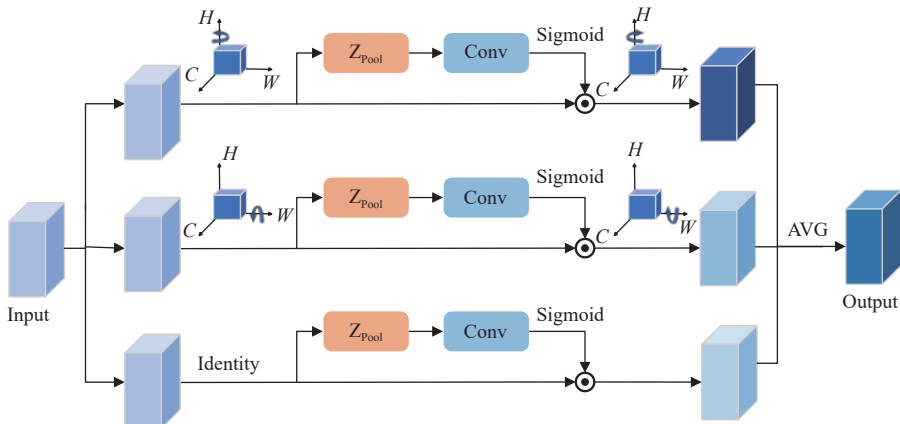


图 6 三重注意力机制网络结构图

Fig. 6 Network structure of the triple attention mechanism

式中,下标 $0d$ 为执行最大池化和平均池化操作的第0维度。例如,形状为 $(C \times H \times W)$ 的张量经过Z池化操作后,会变为形状为 $(2 \times H \times W)$ 的张量。

1.4 改进损失函数

在目标检测中,定位误差是指预测的边界框与真实边界框之间的位置偏差。边界框回归损失用于衡量和减少定位误差。相较于传统的L1或L2损失,交并比(IoU)成为目前评价模型预测边界框与真实边界框重叠程度的主流标准。交并比损失函数 L_{IoU} 计算公式为

$$L_{IoU} = \frac{|B^{pred} \cap B^{gt}|}{|B^{pred} \cup B^{gt}|} \quad (3)$$

式中: B^{pred} 为预测框(预测区域); B^{gt} 为真实框(真实区域)。

YOLOv8针对IoU的改进有许多,如距离交并比(DIoU)、完全交并比(CIoU)和广义交并比(GIoU)等。通过优化边界框损失,模型能够更精准地确定目标位置和大小。YOLOv8n模型中采用了 $L_{CIoU}^{[16]}$,其定义为

$$L_{CIoU} = L_{IoU} - \left(\frac{\rho^2(B^{pred}, B^{gt})}{c^2} + \alpha v \right) \quad (4)$$

式中: c 为包围预测框和真实框的最小外接框的对角线长度; α 为平衡参数; v 为衡量长宽比差异的度量; $\rho^2(B^{pred}, B^{gt})$ 为预测框与真实框中心点之间的欧氏距离。

然而,这些基于IoU改进的损失函数由于存在恒定的惩罚因子,导致先验框在回归过程中扩展,显著降低了收敛速度,并且无法充分反映锚框与目标框之间的差异,对尺度差异较大的目标、遮挡目标检测效果较差。强交并比(PIoU)根据目标大小自适应的惩罚因子和基于先验框质量的梯度调整函数,引导先验框沿着更有效路径进行回归。在海上红外图像目标检测中,PIoU相较于其他IoU变体直接最小化锚框和目标框四边之间的距离,提高了检测效率。PIoUv2引入了一个非单调的注意力机制和一个自适应惩罚因子,这些改进有助于更有效地进行复杂场景中的目标检测,如遮挡和目标密集的环境。因此,本文采用改进的PIoUv2。

PIoU定义为

$$L_{PIoU} = L_{IoU} - (1 - e^{-P^2}) \quad (5)$$

式中, P 为与目标大小相适应的惩罚因子,其定义为

$$P = \left(\frac{dw_1}{w_{gt}} + \frac{dw_2}{w_{gt}} + \frac{dh_1}{h_{gt}} + \frac{dh_2}{h_{gt}} \right) / 4 \quad (6)$$

式中: dw_1, dw_2, dh_1 和 dh_2 为预测框与目标框对应边缘之间距离的绝对值; w_{gt} 和 h_{gt} 分别为目标框的宽度和高度。

由于惩罚因子 P 的分母只取决于目标框的大小,与锚框和目标框中最小外部框的大小无关,所以使用 P 作为损失函数中的惩罚因子不会引起锚框膨胀问题。同时 P 对目标的大小具有适应性,除非锚框完全与目标框重叠,否则 P 不会退化为零。

通过向PIoU损失添加注意力层得到的PIoUv2损失增强了对中高置信度锚框的关注能力,提高了目标检测器的性能。注意力函数和PIoUv2损失定义为

$$u(x) = 3x \cdot e^{-x^2} \quad (7)$$

$$q = e^{-P}, q \in (0, 1] \quad (8)$$

$$L_{PIoUv2} = 1 - u(\lambda q) \cdot L_{PIoU} = 1 - 3 \cdot (\lambda q) \cdot e^{-(\lambda q)^2} \cdot (1 - L_{PIoU}) \quad (9)$$

式中, $u(\lambda q)$ 为注意力函数,其中用测量锚框质量的质量因子 q 替换了惩罚因子 P , q 的范围为 $(0, 1]$ 。当 $q = 1$ 时,则 $P = 0$,意味着锚框与目标框完全对齐。随着 P 增加, q 逐渐减小,表示低质量的锚框。非单调注意力函数 $u(\lambda q)$ 增强了PIoU关注中等质量锚框的能力。

2 实验结果与分析

2.1 实验环境和训练参数

实验环境配置如下:硬件设备CPU为AMD Ryzen 75700X8-Core Processor@3.40 GHz, GPU为NVIDIA GeForce RTX 3070;软件包括PyTorch 2.2.1深度学习框架、CUDA11.8图形加速器和Python 3.8.18。在使用MITD-YOLO模型检测海上船舶红外图像时,其训练的超参数设置为默认值,具体如表1所示。

2.2 数据集

本文采用的是烟台艾睿光电科技有限公司公开发布的红外海上船舶数据集。该数据集包括8402张红外图像,涵盖邮轮、散货船、军舰、帆船、小艇、集装箱船、渔船共7类典型目标。数据集按照7:2:1的比例严格划分为训练集、验证集与测试集,各子集图片数量分别为5881,1680和841张。

表1 训练参数设置

Table 1 Training parameters setting

参数	数值
输入分辨率	640×640
学习率浮点数(lr)	0.01
动量	0.937
权重衰减	0.0005
批大小	8
工作线程数	4
训练轮数	100
优化器	SGD
锚框交并比	0.7

为验证算法的有效性,测试模型的泛化能力,本文补充采集了海上红外图像,扩充海上船舶红外数据集。补充图像具有以下特点,一是红外目标尺度小,仅占几个至几十个像素;二是目标背景复杂,包含港口、狭水道等交通繁忙水域的场景;三是单张图像中包含多个目标,且目标尺度大小不一。共补充海上红外图像 387 张,其中包含小目标的图像 308 张,背景复杂的图像 79 张(图 7)。补充数据集仅用于模型性能验证,不参与模型训练。



(a) 小目标



(b) 复杂背景目标



(c) 多目标

图7 补充数据集图片示例

Fig. 7 Example images of the supplemental dataset

2.3 评价指标

实验中模型评价指标为精确率 (precision)、召回率 (recall)、平均精度均值 (mean of average precision)。

精确率 P 是指模型预测为正类的样本中,实际为正类的比例,计算公式为

$$P = \frac{T_P}{T_P + F_P} \quad (10)$$

式中: T_P 为图片中正确预测部件的数量; F_P 为图片中错误预测部件的数量。

召回率 R 是指所有正类样本中,被模型正确预测为正类的比例,计算公式为

$$R = \frac{T_P}{T_P + F_N} \quad (11)$$

式中, F_N 为图片中漏检的数量。

平均精度均值 mAP 是所有类别平均精确率 AP 的平均值,其反映了模型在多个类别和多个阈值上的精确度和召回率的平均效果,计算公式为

$$AP = \int_0^1 P(R) \cdot dR \quad (12)$$

$$mAP = \frac{1}{n} \left(\sum_{i=1}^n AP_i \right) \quad (13)$$

式中: n 为总类别数; AP_i 为第 i 个分类的平均精确率。

选取帧率 (frames per second, FPS) 作为衡量算法运行速率的指标,即单位时间内模型完成任务的图像个数。

2.4 对比实验

为验证改进算法对海上红外船舶数据集目标检测的优越性,本文开展了对比试验。采用上述评价指标,在同一平台进行训练,并采用相同的训练集、验证集和测试集。

2.4.1 原模型与改进模型的对比

与原 YOLOv8n 模型对比实验结果见表 2,可见 MITD-YOLO 模型的检测效果有所改善,其精确率 P 提升了 2.3%,召回率 R 提升了 1.7%,平均精度 $mAP@0.5$ 提升 2.2%,FPS 达到 125.5,相较于 YOLOv8n 下降了 7.5%。

表2 对比实验结果

Table 2 Results of the comparative experiments

模型	$P/\%$	$R/\%$	$mAP@0.5/\%$	FPS
YOLOv8n	89.4	80.7	86.7	135.7
MITD-YOLO	91.7	82.4	88.9	125.5

YOLOv8n 模型和 MITD-YOLO 模型的训练过程如图 8 所示,可见改进模型的整体性能更优,拟合效果更好。

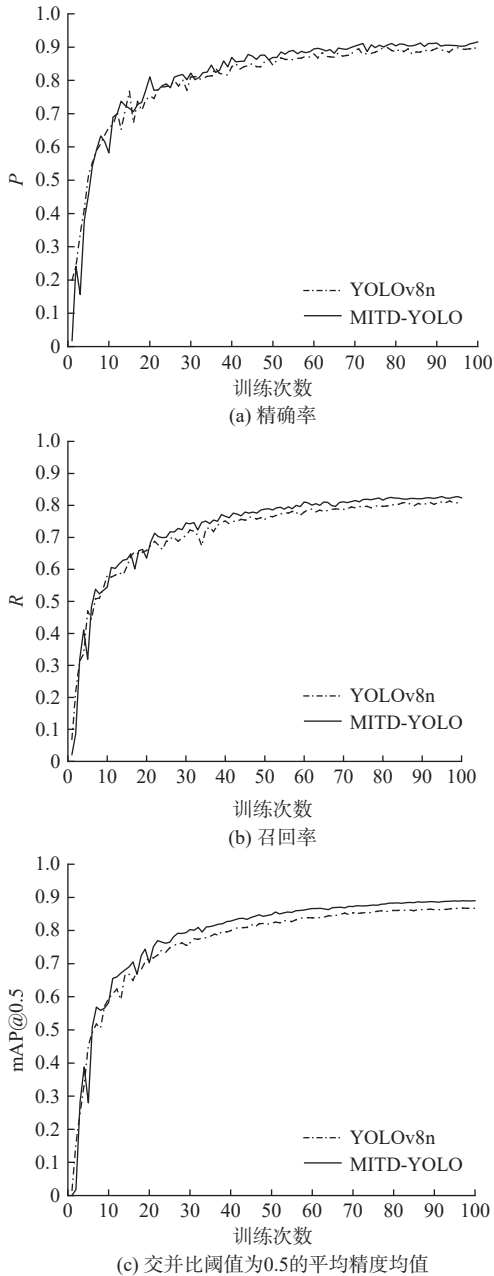


图 8 改进效果对比

Fig. 8 Comparison of improved effects

2.4.2 不同损失函数 IoU 对比

将 PIoUv2 损失函数和常用的损失函数 CIoU, DIoU^[17], GIoU^[18], SIoU^[19] 和 Inner-CIoU^[20] 进行实验对比,结果如表 3 所示。由表可见,相比使用 CIoU 损失函数的原 YOLOv8n 模型,使用 PIoUv2 损失函数的 YOLOv8n 模型检测准确率更高,其精确率 P 、召回率 R 和平均精度 $mAP@0.5$ 分别提升 1.3%、1.3% 和 1.6%,FPS 基本保持不变。这说明使用 PIoUv2 损失函数能够使模型对边界框的回归更加稳定,预测精度更高。

表 3 改进损失函数后模型性能对比

Table 3 Model performance comparison after improving the loss function

模型	$P/\%$	$R/\%$	$mAP@0.5/\%$	FPS
CIoU	89.4	80.7	86.7	135.7
DIoU	89.9	80.9	86.5	139.7
GIoU	89.9	80.7	86.2	134.0
SIoU	89.9	81.6	86.9	136.5
MPDIoU	89.8	80.3	86.0	136.4
Inner-CIoU	90.4	80.5	86.0	132.6
PIoUv2	90.7	82.0	88.3	135.5

2.4.3 与其他模型的对比

为进一步验证所提模型的优越性,与当前主流目标检测模型进行了比较,包括双阶段目标检测算法中的 Faster-RCNN^[21] 和 Mask-RCNN^[22] 算法,以及单阶段目标检测 YOLO 系列算法中的 YOLOv5, YOLOv7, YOLOv8n, YOLOv10 和最新 YOLOv11 模型。具体检测指标如表 4 所示。

表 4 不同模型的检测性能对比

Table 4 Detection performance comparison of different models

模型	$P/\%$	$R/\%$	$mAP@0.5/\%$	FPS
Faster-RCNN	87.7	77.0	83.6	94.35
Mask-RCNN	88.0	79.5	85.8	82.70
YOLOv5	88.9	81.9	87.5	129.80
YOLOv7	91.3	88.7	89.0	86.75
YOLOv8n	89.4	80.7	86.7	135.70
YOLOv10	82.6	75.2	82.5	112.65
YOLOv11	89.3	79.3	85.8	96.90
MITD-YOLO	91.7	82.4	88.9	125.5

Faster-RCNN 使用区域提议网络(RPN)高效生成目标提议,由于其计算复杂性,FPS 相对较低。Mask-RCNN 在 Faster-RCNN 的基础上增加一个分支用于生成目标的掩码,适用于实例分割任务,召回率和平均精度均值表现更优,但处理速度(FPS)略低。YOLOv5 是 YOLO 系列中的快速检测模型,其采用多尺度训练、更好的锚定机制和更高效的网络结构,在保持较高检测精度的同时大幅提升了处理速度。YOLOv7 使用更深更复杂的网络结构和更细致的特征提取策略,在召回率和平均精度均值上表现最好,但牺牲了一定的速度。YOLOv8n 是一种超级轻量模型,在保持极高推理速度的同时,仍能提供相对不错的检测精度。YOLOv10 由于模型在特定数据集或配置下的适应性不如其他版本,虽然其 FPS 表现良好,但精确度和召回率相对较低,可能需要进一步优化模型架构或训练过程来提升准确度。YOLOv11 在精确度、召回率和平均精度均值上

均表现稳定,但其设计更注重在多种性能指标上取得平衡,而不是单一指标的极端优化,适合需要较均衡性能的应用场景。

尽管 YOLOv7 的准确率较高,但其模型参数较多,检测速度较慢。改进后的模型在检测速度 FPS 方面优于 YOLOv7 模型,召回率优于 YOLOv8n

模型。

2.5 消融实验

为验证改进模型中每个改进模块对海上船舶红外图像目标检测性能的影响,对海上红外船舶数据集进行消融实验,结果如表 5 所示。

表 5 消融实验结果

Table 5 Results of the ablation experiments

改进模块				检测结果			
C2f_DBB	EMSCConv	Triple Attention	PloUv2	$P/\%$	$R/\%$	mAP@0.5/%	FPS
×	×	×	×	89.4	80.7	86.7	135.7
√	×	×	×	90.6	79.7	86.2	135.1
×	√	×	×	89.5	80.6	87.0	136.9
×	×	√	×	90.7	81.2	87.1	125.1
×	×	×	√	90.7	82.0	88.3	135.5
√	√	×	×	90.2	80.4	85.9	130.9
√	√	√	×	91.0	81.8	88.1	124.6
√	√	√	√	91.7	82.4	88.9	125.5

1) 添加 C2f-DBB 后,相对于 YOLOv8n 模型, YOLOv8n 模型+C2f-DBB 模块的目标检测精度有所提升,其精确率 P 、召回率 R 、平均精度 mAP@0.5 分别达到 90.6%、79.7% 和 86.2%, FPS 为 135.1。这说明 C2f-DBB 模块的多分支结构及多尺度特征提取能够提供更丰富的特征表示和更强的泛化能力,通过多样性增强和特征融合机制,能更好地捕捉目标特征,提升海上船舶红外图像中复杂场景和多样化目标的检测精度。

2) 使用 EMSCConv 替换检测头卷积后,虽然精确率 P 有所下降,但是召回率 R 和平均精度 mAP@0.5 有所提升。

3) 引入 Triple Attention 注意力机制后,模型通过使用三分支结构在不同维度信息交互计算注意力权重,能够提高海上船舶红外图像目标检测的稳定性,使模型效果有明显提升。

4) 改进 PloUv2 损失函数后,模型对交并比度量更合理、对目标边界框的回归更加准确,且能提供平滑的梯度信息,在海上船舶红外数据集测试集上的精确率 P 、召回率 R 和平均精度 mAP@0.5 分别达到 90.7%、82% 和 88.3%, FPS 达到 135.5。这证明了 PloUv2 损失函数能够提高复杂海上船舶红外图像检测的精度和稳定性。

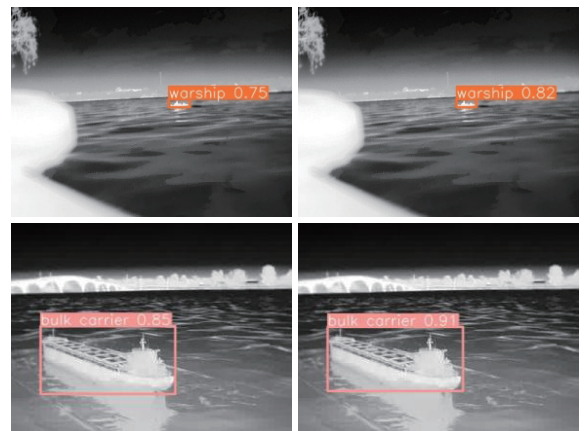
综上所述,同时使用 C2f-DBB 模块、EMSCConv 卷积、三重注意力机制、PloUv2 损失函数,经过 100 轮迭代训练,训练得到的 MITD-YOLO 模型的精确率 P 、召回率 R 和平均精度 mAP@0.5 比

原 YOLOv8n 模型分别提高 2.3%、1.7% 和 2.2%,检测速度 FPS 达到 125.5。

2.6 可视化分析

为验证改进 MITD-YOLO 模型的性能,图 9~图 12 分别为 YOLOv8n 模型和改进的 MITD-YOLO 模型在同一实验条件和参数下训练后,针对不同场景海上红外目标的检测效果。

原 YOLOv8n 模型对军舰和散货船的检测置信度分别为 75% 和 85%。如图 9 所示,MITD-YOLO 模型对于军舰的检测置信度达到 82%,对于散货船的检测置信度达到 91%,这证明 MITD-YOLO 模型不仅使整体检测精度有显著提升,对于不同种类船舶的检测效果也有所改善。



(a) YOLOv8n 模型

(b) MITD-YOLOv8 模型

图 9 场景 1 对比实验

Fig. 9 Comparative experiment of scenario 1

图 10 所展示的 YOLOv8n 模型对于海上红外船舶图像中的小目标存在大量漏检问题,而 MITD-YOLO 模型识别出了这些微小船舶目标,显著提升了对小目标的检测效果,能够更精准地定位海上船舶红外图像中不同尺度的目标。

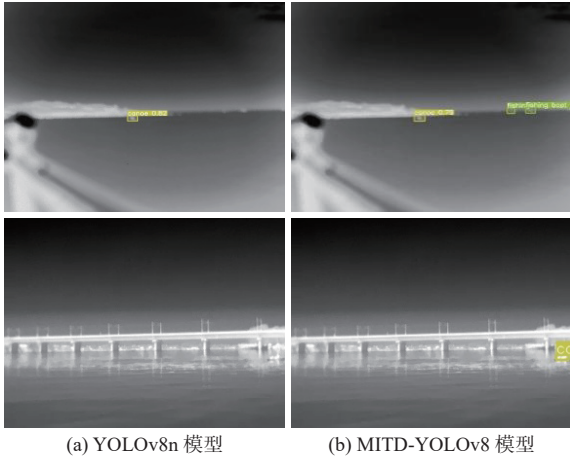


图 10 场景 2 对比实验

Fig. 10 Comparative experiment of scenario 2

图 11 展示了 YOLOv8n 模型对于海上红外船舶数据集存在的误检情况,其损失函数的局限性导致锚框回归出现误差,使同一个目标被多次识别。改进后的 MITD-YOLO 模型优化了损失函数,有效降低了船舶目标被误检的概率。

图 11 和图 12 中背景复杂,包含了在海上复杂环境中容易混淆的背景目标和在港口中的建筑物等。由图可见, YOLOv8n 模型存在混淆背景信息,误将背景中的干扰识别成船舶目标,并错误标注船舶种类;而 MITD-YOLO 海上船舶红外图像目标检测模型能有效区分背景信息和船舶目标。在复杂背景条件下,改进后的模型能够更准确地检测和识别船舶目标,在实际应用中对于多场景需求具有重要意义。

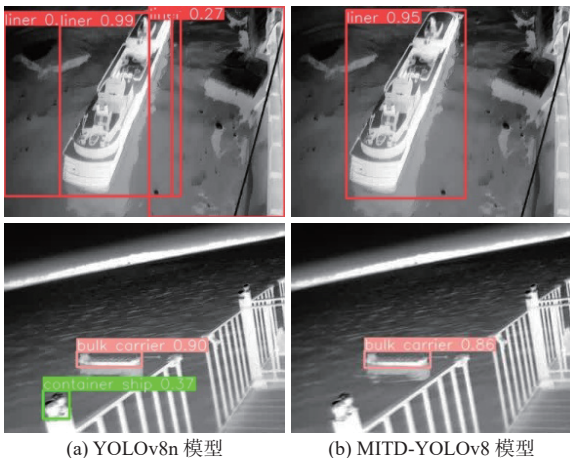


图 11 场景 3 对比实验

Fig. 11 Comparative experiment of scenario 3

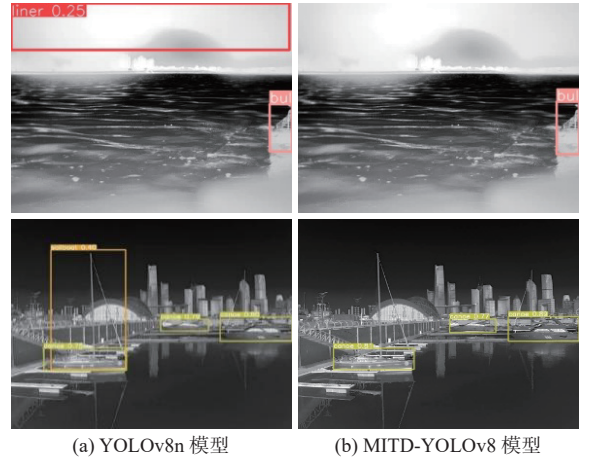


图 12 场景 4 对比实验

Fig. 12 Comparative experiment of scenario 4

3 结论

通过引入 C2f_DBB 以及 EMSConv 卷积,添加三重注意力机制,替换 PIoUv2 损失函数等改进方法,本文提出了 MITD-YOLO 模型。实验结果表明,该算法显著提高了在复杂背景下船舶目标的检测和识别准确率,降低了对船舶红外图像中小目标的误检和漏检率。

1) 采用多种尺度的卷积以及多分支并行结构,对尺度差别较大、对比度低、噪声明显的海上红外图像目标识别效果有显著改善。

2) 在损失函数中提高先验框的回归效率,能够提高复杂海上船舶红外图像检测框的精度和稳定性。

3) 与 YOLOv8n 算法相比,采用本文构建的 MITD-YOLO 模型后,目标检测精确率和召回率分别提升 2.3% 和 1.7%, mAP@0.5 提升 2.2%,并保证了检测速度,具有较强实用性。

本文提出的 MITD-YOLO 模型为海上智能感知系统提供了技术支持,有助于提升海上船舶航行安全。未来还可以研究多模态数据融合以进一步提高检测性能,并探索模型压缩技术以适应资源受限的海上设备。

参考文献:

- [1] ZHU D Y, TANG J W, FU X X, et al. Detection of infrared small target based on background subtraction local contrast measure and Gaussian structural similarity [J]. Heliyon, 2023, 9(6): e16998.
- [2] LU Z L, LIU S X, YILAHUN H, et al. Infrared small target detection based on background estimation and scale fusion[J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 1-5.
- [3] ZHANG Z Z, DING C, GAO Z S, et al. ANLPT: self-adaptive and non-local patch-tensor model for infrared

- small target detection[J]. *Remote Sensing*, 2023, 15(4): 1021.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH: IEEE, 2014: 580–587. DOI:10.1109/CVPR.2014.81.
- [5] GUPTA V, MANDLOI A, PAWAR S, et al. Deep learning based object detection using mask RCNN[M]//AWASTHI S, SINGH NARUKA M, PRAKASH YADAV S, et al. *AI and IoT-based Intelligent Health Care & Sanitation*. Singapore: Bentham Science Publishers, 2023: 207–221.
- [6] OU J H, WANG J G, XUE J, et al. Infrared image target detection of substation electrical equipment using an improved faster R-CNN[J]. *IEEE Transactions on Power Delivery*, 2023, 38(1): 387–396.
- [7] ZHAO X F, XIA Y T, XU M Y, et al. An infrared small vehicle target detection method based on deep learning[C]//Third International Seminar on Artificial Intelligence, Networking, and Information Technology (AINIT 2022). Shanghai: SPIE, 2023. DOI:10.1117/12.2667313.
- [8] 陈德海, 邵恒, 张军令. 改进 YOLOv3 的船舶检测算法研究[J]. *现代电子技术*, 2023, 46(2): 101–106.
CHEN D H, SHAO H, ZHANG J L. Research on improved YOLOv3 ship detection algorithm[J]. *Modern Electronics Technique*, 2023, 46(2): 101–106 (in Chinese).
- [9] WANG D, DU H Q, MA Z F. Object detection in infrared images using modified YOLOV4 models and an image enhancement module[C]//Fourteenth International Conference on Graphics and Image Processing (ICGIP 2022). Nanjing: SPIE, 2023. DOI:10.1117/12.2680173.
- [10] 张炳焱, 张闯, 石振男, 等. 基于 YOLO-FNC 模型的轻量化船舶检测方法[J]. *中国舰船研究*, 2024, 19(5): 180–187.
ZHANG B Y, ZHANG C, SHI Z N, et al. Lightweight ship detection method based on YOLO-FNC model[J]. *Chinese Journal of Ship Research*, 2024, 19(5): 180–187 (in both Chinese and English).
- [11] HU S M, ZHAO F, LU H Z, et al. Improving YOLOv7-tiny for infrared and visible light image object detection on drones[J]. *Remote Sensing*, 2023, 15(13): 3214.
- [12] 张瑶, 陈姚节. 改进 YOLOv8 的水面小目标检测算法[J]. *计算机系统应用*, 2024, 33(4): 152–161.
ZHANG Y, CHEN Y J. Improved YOLOv8 algorithm for small surface object detection on water surface[J]. *Computer Systems & Applications*, 2024, 33(4): 152–161 (in Chinese).
- [13] DING X H, ZHANG X Y, HAN J G, et al. Diverse branch block: building a convolution as an inception-like unit[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN: IEEE, 2021: 10881–10890. DOI:10.1109/CVPR46437.2021.01074.
- [14] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: convolutional triplet attention module [C]//2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI: IEEE, 2021: 3138–3147. DOI:10.1109/WACV48630.2021.00318.
- [15] 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.
- [16] ZHENG Z 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*, 2022, 52(8): 5874–5886.
- [17] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 12993–13000. DOI:10.1609/aaai.v34i07.6999.
- [18] REZATOFIHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA: IEEE, 2019: 658–666. DOI:10.1109/CVPR.2019.00075.
- [19] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[Z/OL]. (2022-05-25)[2024-07-26]. <https://arxiv.org/abs/2205.12740>.
- [20] ZHANG H, XU C, ZHANG S J. Inner-IoU: more effective intersection over union loss with auxiliary bounding box[Z/OL]. (2023-11-14)[2024-07-25]. <https://arxiv.org/abs/2311.02877>.
- [21] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1131–1149.
- [22] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[Z/OL]. (2018-01-24)[2025-02-26]. <https://arxiv.org/abs/1703.06870>.

MITD-YOLO: an improved YOLOv8n-based method for maritime infrared target detection

YANG Xuefeng^{*1,2}, LIU Jiayao¹, ZHOU Changhua¹

1 School of Shipping and Naval Architecture, Chongqing Jiaotong University, Chongqing 400074, China

2 National Engineering Laboratory of Transport Safety and Emergency Informatics, Beijing 100011, China

Abstract: [**Objective**] Complex backgrounds, significant target size variations, and severe sea clutter in maritime infrared imagery often result in missed or false detections. To address this challenge, an improved method based on YOLOv8n, termed maritime infrared target detection-YOLO (MITD-YOLO), is proposed to enhance target detection accuracy in maritime infrared images. [**Method**] MITD-YOLO incorporates a diverse branch module (DBB) and enhanced multi-scale convolution (EMSCConv) to leverage multi-scale convolutions, enabling the model to more effectively capture complex features. A triple attention mechanism is employed to facilitate spatial and channel-wise feature interaction, thereby improving key feature extraction. Additionally, the powerful-IoUv2 (PIoUv2) loss function is introduced to address the anchor box expansion problem, leading to improved detection accuracy and enhanced model robustness. [**Results**] Experimental results show that the improved model significantly enhances the efficiency of maritime infrared target detection, with a 2.3% increase in precision and a 1.7% increase in recall. The model achieves an average precision of 88.9%, and 132.8 FPS, outperforming the original model. [**Conclusion**] MITD-YOLO enhances maritime infrared target detection performance and provides a more reliable target detection technology for applications such as maritime surveillance and ship navigation, contributing to the advancement of intelligent maritime systems.

Key words: target tracking; infrared target detection; multi-scale convolution; triple attention mechanism

相关论文

- [1] 白翎, 肖进丽, 甘浪雄, 等. 一种基于改进 YOLOv12 的轻量化船舶目标检测算法: SEE-YOLO[J]. 中国舰船研究, 2025, 20(6): 1–9. <https://www.ship-research.com/article/doi/10.19693/j.issn.1673-3185.04659>
- [2] 石廷超. 基于改进 YOLOv3 的复杂海空场景舰船目标智能识别算法 [J]. 中国舰船研究, 2025, 20(5): 1–12. <https://www.ship-research.com/article/doi/10.19693/j.issn.1673-3185.04538>
- [3] 严荣慧, 郭前, 雷鸣, 等. 基于特征融合及混合注意力的小目标船舶识别 [J]. 中国舰船研究, 2024, 19(6): 284–292. <https://www.ship-research.com/article/doi/10.19693/j.issn.1673-3185.03489>
- [4] 李刚, 张磊, 王鑫, 等. 多特征融合的无人艇视觉小目标鲁棒跟踪 [J]. 中国舰船研究, 2024, 19(5): 65–78. <https://www.ship-research.com/article/doi/10.19693/j.issn.1673-3185.03394>
- [5] 张炳焱, 张闯, 石振男, 等. 基于 YOLO-FNC 模型的轻量化船舶检测方法 [J]. 中国舰船研究, 2024, 19(5): 180–187. <https://www.ship-research.com/article/doi/10.19693/j.issn.1673-3185.03487>
- [6] 马枫, 石子慧, 孙杰, 等. 自注意力机制驱动的轻量化高鲁棒船舶目标检测方法 [J]. 中国舰船研究, 2024, 19(5): 188–199. <https://www.ship-research.com/article/doi/10.19693/j.issn.1673-3185.03389>