

融合多尺度注意力机制和改进特征融合的轻量化水面小目标检测模型

仲尚¹, 马丽^{1,2*}, 刘文哲¹, 李雨豪¹

(1.河北地质大学信息工程学院, 河北 石家庄 052161; 2.河北地质大学智能传感物联网技术河北省工程研究中心, 河北 石家庄 052161)

摘要:在复杂水面场景下,针对小目标检测精度较低、漏检率高和计算资源有限的问题,提出一种融合多尺度注意力机制和改进特征融合的轻量化水面小目标检测模型。根据中心度理论,设计新的主干网络,利用多尺度注意力机制增强模型的特征提取能力,利用部分卷积从减少特征图冗余的角度改进颈部网络,有效降低模型计算量。使用大型可分离核注意力模块改进快速空间金字塔池化模块,提升模型的特征融合能力。实验结果表明,与其他模型相比,本文模型的检测精度高、漏检率低、参数量少。

关键词:小目标检测;特征融合;多尺度注意力机制;特征图冗余

中图分类号:TP391 **文献标志码:**A

引用格式:仲尚,马丽,刘文哲,等.融合多尺度注意力机制和改进特征融合的轻量化水面小目标检测模型[J].山东大学学报(理学版),2026,61(1):15-25.

Lightweight water surface small object detection model with multi-scale attention mechanism and improved feature fusion

ZHONG Shang¹, MA Li^{1,2*}, LIU Wenzhe¹, LI Yuhao¹

(1. College of Information Engineering, Hebei GEO University, Shijiazhuang 052161, Hebei, China; 2. Intelligent Sensor Network Engineering Research Center of Hebei Province, Hebei GEO University, Shijiazhuang 052161, Hebei, China)

Abstract: In complex water surface scenarios, addressing the issues of low detection accuracy, high missed detection rates, and limited computational resources for small target detection, this paper proposes a lightweight water surface small object detection model with multi-scale attention mechanism and improved feature fusion. Based on centerness theory, a new backbone network is designed, leveraging the multi-scale attention mechanism to enhance the model's feature extraction capabilities. Partial convolution is used to improve the neck network by reducing feature map redundancy, effectively lowering the model's computational load. A large separable kernel attention module is employed to improve the spatial pyramid pooling module, enhancing the model's feature fusion ability. Experimental results demonstrate that, compared to other models, the proposed model achieves higher detection accuracy, lower missed detection rates, and fewer parameters.

Key words: small target detection; feature fusion; multi-scale attention mechanism; feature map redundancy

0 引言

基于深度学习的水面目标检测技术的发展^[1-3],使得小型无人艇应用于应急响应、检测水面污染和蜉蝣生物分布等方面。由于水面环境复杂且变化多端,因此要求水面目标检测技术具有更高的信息处理能力。

基于深度学习的诸多模型在目标检测领域有很大突破^[4-7],主要分为2类:以基于区域的卷积神经网络 RCNN(regions with convolutional neural network)为代表的2阶段模型,以及 YOLO(you only look once)为

收稿日期:2024-10-29;网络出版时间:2025-05-07

基金项目:河北省高等学校科学技术研究重点资助项目(ZD2018043);河北省教育科学规划课题一般资助课题项目(2303121);河北地质大学博士基金资助项目(BQ2017045)

第一作者:仲尚(1999—),男,硕士研究生,研究方向为机器学习、目标检测。E-mail:2371019647@qq.com

*通信作者:马丽(1977—),女,副教授,硕士生导师,博士,研究方向为知识发现与粒计算、机器学习。E-mail:mali_new@163.com

代表的 1 阶段模型。戚玲珑等^[8]通过改进多路卷积模块降低有效特征被忽视的情况,提升模型在水面小目标检测任务中的性能。面对水面漂浮物小目标检测任务, Lin 等^[9]在 YOLOv5 的主干网络中使用马赛克数据增强策略,有效提升模型的鲁棒性。王林^[10]使用深度可分离卷积模块对 YOLOv4 网络进行改进,在海上船舶目标检测任务中实现高精度的实时检测。以上工作未考虑模型计算量大、实施困难的问题。本文从模型检测精度和轻量化考虑,提出一种融合多尺度注意力机制和改进特征融合的轻量化水面小目标检测模型。选用 YOLO 模型的第 8 个版本 YOLOv8s 模型作为基础目标检测框架,先提升 YOLOv8s 模型检测精度,再对 YOLOv8s 模型进行轻量化处理的设计理念,提升 YOLOv8s 模型的特征提取能力,且减少了特征冗余计算。使用大尺度可分离卷积核模块,避免在主干网络使用过深层次网络提取特征,增大快速空间金字塔池化模块感受野,提升 YOLOv8s 模型特征融合能力。将改进后的模型命名为 YOLO-多尺度快速可分离模型(YOLO-multi-scale fast separable model, MFS_YOLO)。

1 YOLOv8 模型简介

YOLOv8 模型是 1 阶段目标检测模型,根据大小可划分为 n 、 s 、 m 、 l 和 x 共 5 种规格。与 YOLOv5 模型相比, YOLOv8 模型的主干网络第 1 次的卷积核大小由 6×6 缩减到 3×3 ,瓶颈模块的预处理卷积次数也由 3 次变成 2 次,同时 YOLOv8 模型还借鉴了 YOLOv7 模型的多堆叠结构。为了加快计算速度,在颈部网络部分, YOLOv8 模型也不再对主干部分获得的特征图进行卷积操作。在最后的检测头部分, YOLOv8 模型使用解耦头结构,将待检测目标的分类和回归进行分离。

2 MFS_YOLO 模型

2.1 网络重构

为提升模型检测的泛化能力, YOLOv8s 模型使用无锚框机制替换有锚框机制,且采用基于目标中心位置的距离回归,即预测目标框四边距离目标中心的距离。反映预测框像素点距离目标框中心远近程度的中心度为

$$C = \sqrt{\frac{\min(l, r) \min(t, b)}{\max(l, r) \max(t, b)}}, \quad (1)$$

式中: l 、 r 、 t 和 b 分别为预测框像素点到目标框左边界、右边界、上边界和下边界的距离, \min 表示取小操作, \max 表示取大操作。 C 的计算示例如图 1 所示,图中浅蓝色矩形框为目标真实框, 2 条黑色双向箭头的交点为预测框像素点。

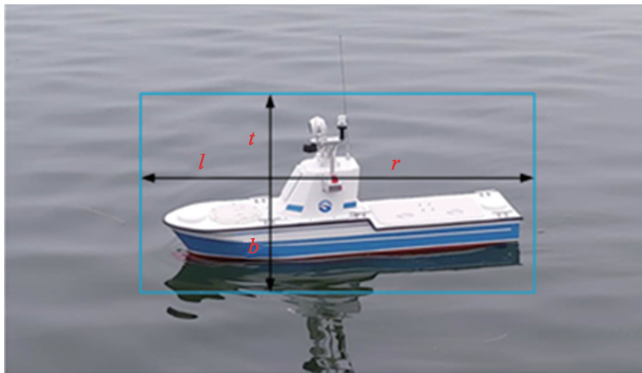


图 1 中心度的计算示例
Fig.1 Calculation example of centerness

对于图像中的任意一个目标 A ,假设它的真实框尺寸大小为 $x \times y$,其中 x 表示真实框的宽, y 表示真实框的高。假设预测目标框像素点与目标真实框在横向和纵向的偏差分别为 x_0 、 y_0 ,其中 x_0 、 y_0 分别表示预测框像素点距离目标真实框中心位置的水平距离、垂直距离。由于预测框像素点的偏向对中心度无影响,假设 $l < r$, $t < b$,

即预测框像素点偏向左上方,此时, A 的中心度 $C_A = \sqrt{\frac{lt}{rb}}$,其中 $l = \frac{1}{2}x - x_0$, $r = \frac{1}{2}x + x_0$, $t = \frac{1}{2}y - y_0$, $b = \frac{1}{2}y + y_0$ 。

假设存在 1 个与目标 A 真实框中心重合的目标 B , B 的中心度为 C_B , 真实框尺寸为 $(x-2x') \times (y-2y')$, 其中, $2x'$ 与 $2y'$ 分别为在水平和垂直方向上 B 相较于 A 在真实框尺寸上的差值, 则 B 的中心度为 $C_B = \sqrt{\frac{(l-x')(t-y')}{(r-x')(b-y')}}$, $C_B - C_A = \sqrt{\frac{(l-x')(t-y')}{(r-x')(b-y')}} - \sqrt{\frac{lt}{rb}} < 0$ 。

对于小目标而言, 在相同的偏差 (x_0, y_0) 下, 中心度会更小, 容易产生质量较低检测结果, 说明对于小目标检测, 原有的 3 种尺度的检测头不能满足要求。为此, 须要增加专门用于小目标的检测头, 本文将此种模型命名为 YOLOv8s-P2 模型。然而, YOLOv8s-P2 模型在提升模型检测精度的同时, 会大大增加模型的计算量。小目标包含的特征信息较少, 过度的深层卷积会消除小目标的特征信息, 不利于特征融合, 影响检测效果。本文重构 YOLOv8s 模型, 移除 YOLOv8s 模型主干网络最后一组的卷积模块和瓶颈模块, 并将 YOLOv8s 模型颈部网络的特征金字塔结构和路径聚合网络进行整体上移, 对 YOLOv8s 模型重构的模型命名为 YOLOv8s-Better 模型。

2.2 多尺度注意力机制

为了增强模型的特征提取能力, 本文在 YOLOv8s-Better 模型中融入了高效的多尺度注意力 (efficient multi-scale attention, EMA) 模块^[11]。与常规的注意力机制不同, EMA 模块保留了每个通道的信息, 通过跨空间学习方法融合了 2 个平行子网的输出特征图, 结构如图 2 所示。图 2 中 H 、 W 和 C 分别表示特征图的高度、宽度和通道数, Softmax 表示使用 Softmax 激活函数, Sigmoid 表示使用 Sigmoid 激活函数, \oplus 表示特征图的相加操作。依据 EMA 模块网络架构设计, 特征计算过程分为特征分组、3 条并行子网和跨空间学习。

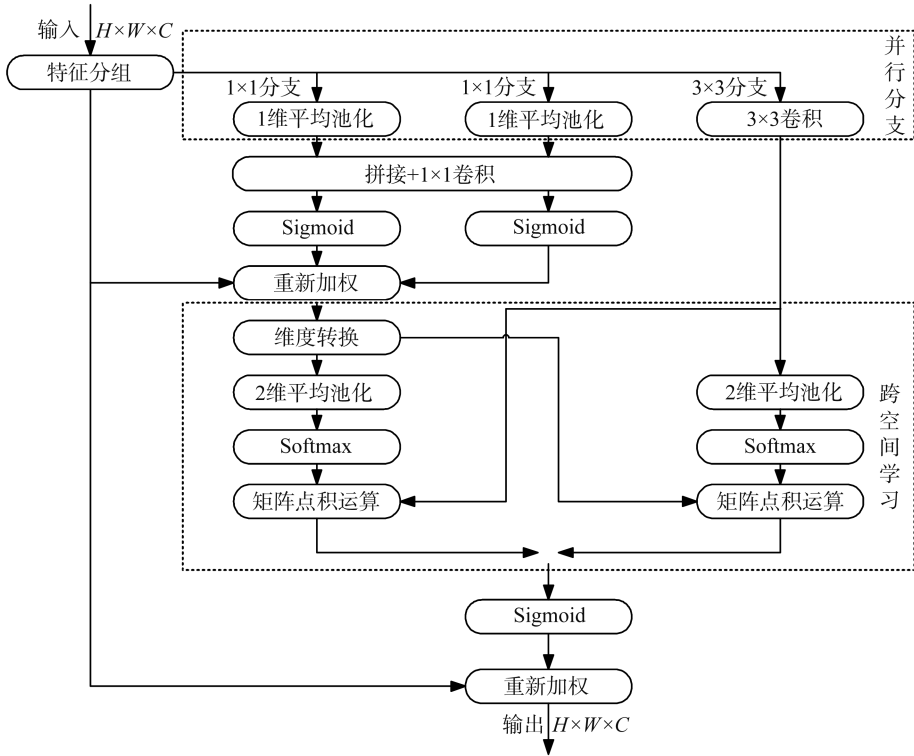


图 2 EMA 模块结构图

Fig.2 The structure diagram of EMA module

对于给定的输入特征图 $X \in \mathbf{R}^{H \times W \times C}$, EMA 模块对 X 在通道维度方向上进行特征分组, 生成 G 个子特征, 即 $\{X_i | X = X_1 \cup X_2 \cup \dots \cup X_i \cup \dots \cup X_G, G \leq C\}$, 其中 $\{X_i | X_i \in \mathbf{R}^{H \times W \times C // G}, i = 1, 2, \dots, G\}$ ($//$ 表示除运算), 且对 $\forall i, j \in [1, G]$, 当 $i \neq j$ 时, $X_i \neq X_j$ 。

效仿 Inception 的多分支结构和多尺度卷积操作, 特征分组后, EMA 模块生成 3 条并行子网, 其中 2 条为 1×1 分支, 1 条为 3×3 分支。2 条 1×1 分支子网分别在水平和垂直方向上对通道进行编码, 分别得到特征图 $Z_C^H(H)$ 和特征图 $Z_C^W(W)$, 用以捕捉水平方向和垂直方向上的长距离依赖关系, 即

$$Z_C^H(H) = \frac{1}{W} \sum_{i=0}^W X_C(H, i), \tag{2}$$

$$Z_c^w(W) = \frac{1}{H} \sum_{j=0}^H X_c(j, W), \quad (3)$$

式中: X_c 代表第 C 个通道的输入特征, i 为特征在水平方向的位置, j 为特征在垂直方向的位置。

对特征编码后的特征图进行拼接, 利用共享的 1×1 卷积对拼接后的特征图进行不降维的提取特征, 得到特征图 $D \in \mathbf{R}^{1 \times (H+W) \times C // G}$, 即

$$D = \text{Conv}_{1 \times 1}(\text{Concat}[Z_c^H(H), Z_c^W(W)]), \quad (4)$$

式中: Concat 表示拼接操作, $\text{Conv}_{1 \times 1}$ 表示卷积核大小为 1×1 的卷积操作。对特征图 D 进行分解后, 使用 Sigmoid 函数进行非线性变换。 3×3 分支通过 3×3 大小的卷积进行跨通道交互, 扩大特征空间。

在 2 条 1×1 分支子网完成重新加权和维度转换后, 利用二维平均池化操作编码全局空间信息, 得到特征图 Z_c , 即

$$Z_c = \frac{\sum_m^H \sum_n^W X_c(m, n)}{HW}, \quad (5)$$

式中 (m, n) 表示特征所在的位置。利用 Softmax 函数对 Z_c 进行拟合, 再与 3×3 并行分支的输出进行矩阵点积运算, 得到 1 张空间注意力图。类似地, 在 3×3 分支中, 通过二维全局平均池化操作和 Softmax 函数, 得到另外 1 张保留精确空间位置信息的空间注意力图。最后, 根据各个组内的输出特征图, 得到 2 个空间注意力图权重值的集合, 并通过 Sigmoid 函数进行特征输出。

2.3 快速瓶颈模块

在融入多尺度注意力模块后, 提高了 YOLOv8s-Better 模型的检测精度, 但是 YOLOv8s-Better 模型的计算量也相应增加。而在 YOLOv8s-Better 模型中, 瓶颈模块中存在着大量的残差结构和堆叠操作, 导致瓶颈模块结构较为复杂, 计算量较大。本文从减少计算量的角度对瓶颈模块进行改进, 构造轻量化的瓶颈模块。本文引入 FasterNet^[12] 中的 FasterNet Block 模块替换瓶颈模块中的 Bottleneck 模块, 减少特征计算过程中的特征图冗余, 在保持 YOLOv8s-Better 模型检测精度的同时, 减少计算量。将改进后的瓶颈模块命名为快速瓶颈模块, 其结构如图 3 所示。图 3 中 C_{in} 表示输入特征图的通道数, C_{out} 表示输出特征图的通道数, n 表示 FasterNet Block 模块的个数, ReLU 表示线性修正单元 (rectified linear unit, ReLU), \oplus 表示特征图的相加操作。此外, PConv 表示部分卷积 (partial convolution, PConv), C_p 表示参与卷积运算的通道数, k 为 PConv 的卷积核大小, $*$ 表示输入特征图使用 PConv 进行卷积运算。

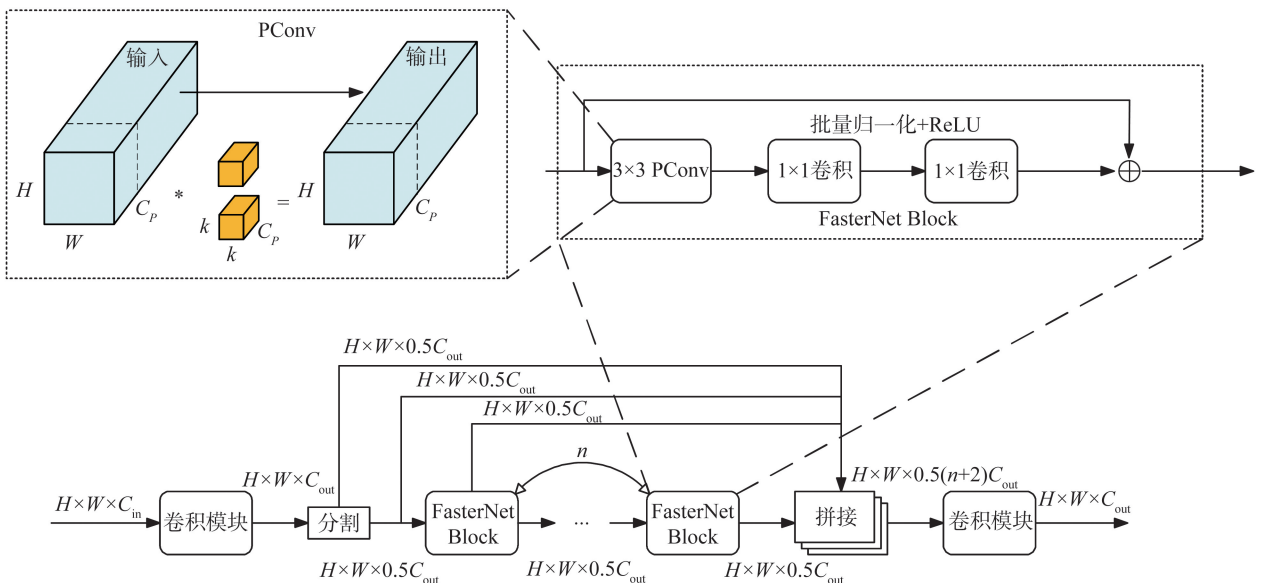


图 3 快速瓶颈模块
Fig.3 Fast bottleneck module

与常规卷积不同的是, PConv 仅在部分通道上进行卷积操作并提取特征, 在其余通道上操作保持不变。设常规卷积操作的计算量为 G , 即

$$G = C^2HWk^2. \quad (6)$$

而在 PConv 中,以 $C_p = 1/2C$ 为例,即一半通道进行卷积,另一半保持不变,此时 PConv 的计算量大大减少,仅为常规卷积的 1/4。在保证特征多样性的同时,加快模型计算速度、降低延迟,仅在 PConv 后的第一个 1×1 卷积层后使用批归一化操作和 ReLU 激活函数。

2.4 融入大型可分离卷积核注意力模块

鉴于小目标由于尺寸较小且形状不规则,导致小目标易受背景干扰,本文快速空间金字塔池化 (spatial pyramid pooling fast, SPPF) 模块中融入大型可分离卷积核注意力 (large separable kernel attention, LSKA)^[13] 模块,能有效增大模型的感受野,提升特征融合效果。将改进后的 SPPF 模块命名为 SPPF_LSKA 模块,其网络结构如如图 4 所示,图中 \otimes 表示哈达玛积操作。

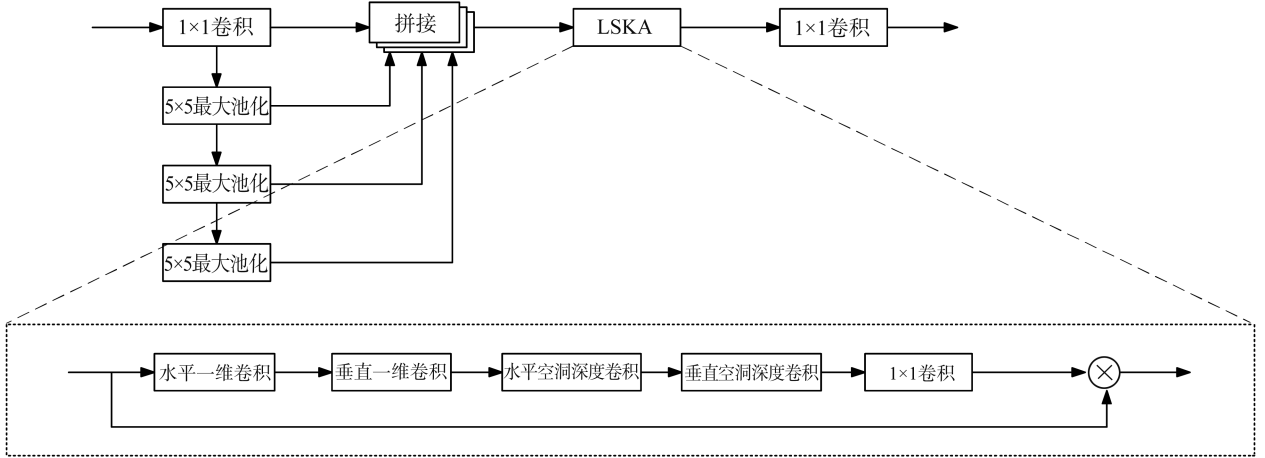


图 4 SPPF_LSKA 结构图
Fig.4 The structure diagram of SPPF_LSKA

在 LSKA 模块中,深度卷积的二维卷积核分解为级联的水平与垂直的一维卷积核,能减轻大卷积核带来的参数量二次增长问题,一维卷积特征计算过程为

$$Z_{C1} = \sum_{H,W} D_{(2d-1) \times 1}^C * \left(\sum_{H,W} D_{1 \times (2d-1)}^C * C_{in} \right), \quad (7)$$

$$Z_{C2} = \sum_{H,W} D_{F(k/d) \times 1}^C * \left(\sum_{H,W} D_{1 \times F(k/d)}^C * Z_{C1} \right), \quad (8)$$

式中, d 表示空洞率, $*$ 表示卷积操作, F 表示向下取整操作, C_{in} 为特征输入, D 为卷积核大小。在接收特征输入后,由 $(2d-1) \times 1$ 与 $1 \times (2d-1)$ 一维卷积核特征对特征输入进行深度可分离卷积,捕获特征输入的局部空间信息,得到特征图 Z_{C1} 。再由 $F(k/d) \times 1$ 和 $1 \times F(k/d)$ 空洞深度卷积捕获特征图 Z_{C1} 的全局空间信息,得到特征图 Z_{C2} 。

在 2 种深度卷积完成特征提取后,对得到的特征图 Z_{C2} 进行 1×1 卷积操作,得到特征图 Z_{C3} 。最后,由特征图 Z_{C3} 与特征输入 C_{in} 进行哈达玛积操作得到特征输出 C_{out} ,特征计算过程为

$$Z_{C3} = D_{1 \times 1} * Z_{C2}, \quad (9)$$

$$C_{out} = Z_{C3} \otimes C_{in}. \quad (10)$$

改进后的 SPPF 模块具有更多尺度的感受野,促进了局部特征和全局特征的融合,加强了网络对小目标的语义信息提取能力,有效降低了 YOLOv8s-Better 模型对于小目标的漏检率。

2.5 MFS_YOLO 模型结构

本文以 YOLOv8s 模型作为基础目标检测框架,经过改进得到 MFS_YOLO 模型,网络结构如图 5 所示。图 5 中,卷积模块由卷积层、批归一化层和激活函数层组成,瓶颈模块则由 2 个卷积模块和若干个 Bottleneck 模块组成,颈部网络的上采样表示将特征图尺度放大 2 倍,而拼接的含义是将特征图进行通道维度的叠加。输入端首先会对输入的图片进行预处理操作,包括增强数据和调整尺寸等策略,保证输入数据的一致性。由主干网络对输入端处理过后的图像信息进行特征提取,为颈部网络提供具有结构层次的、高级的特征表示,颈部网络通过特征金字塔结构和路径聚合网络将这些特征图转化为适用于目标检测的多尺度特征图,最后由多尺度检测头综合处理来自不同尺度的特征图,确保 MFS_YOLO 模型能够捕捉不同层次的目标信息。

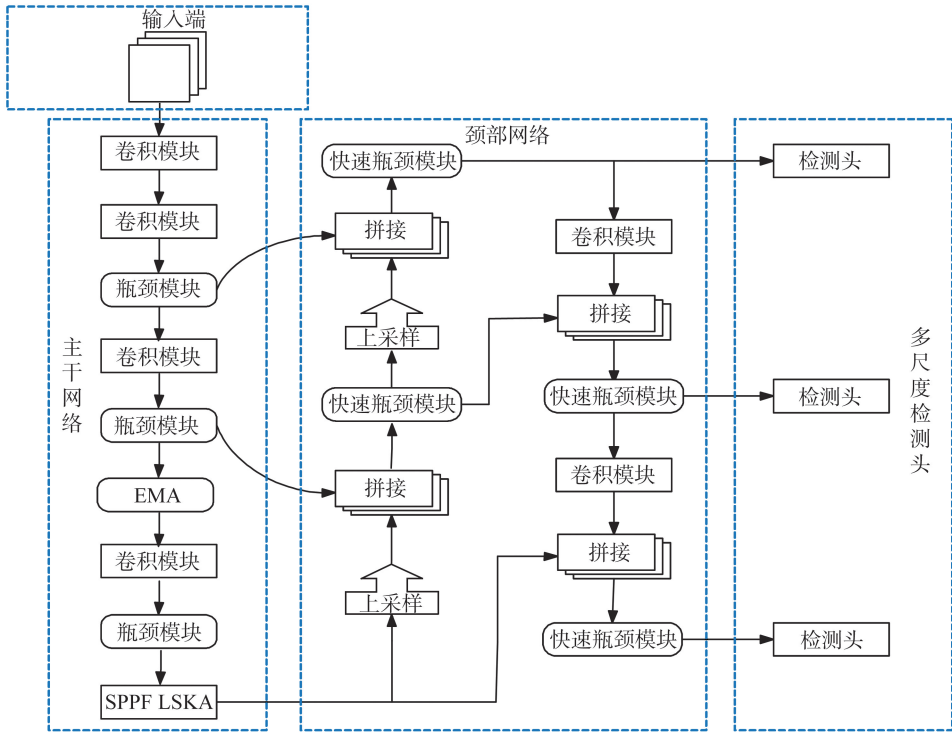


图 5 MFS_YOLO 模型网络结构图

Fig.5 The network structure diagram of MFS_YOLO model

3 实验结果与分析

3.1 实验数据集与实验环境

实验数据集选用水面目标检测数据集 (water surface object detection dataset, WSODD)^[14], 该数据集由 7 467 张不同水面环境、气候条件和拍摄时间的水面图像组成, 与实际检测环境较为相符。水面环境包括海洋、湖泊和河流 3 种, 图像比例分别为 24%、55% 和 21%, 具体取景地点为渤海、黄海、玄武湖、南海子湖和长江。此外, 拍摄天气也有多种, 包括晴天、多云和雾天。拍摄时间也包括白天和傍晚。WSODD 包含了 14 种常见的目标类别, 共有 21 911 个实例, 实例分布如图 6 所示。WSODD 将面积比例小于 10% 的实例定义为小目标, 小目标实例超过一半, 为本文针对水面小目标检测的改进研究提供了良好的支持。为满足实验需求, 本文按照 8:1:1 的比例对数据集进行了划分。

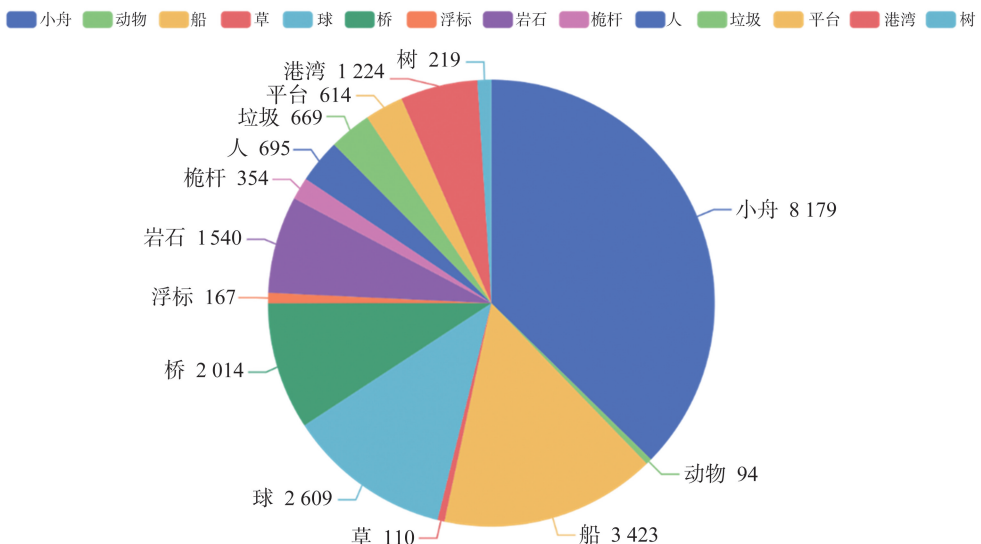


图 6 数据集实例分布

Fig.6 Instance distribution of dataset

本实验采用 Linux 操作系统,Pytorch 框架版本为 1.12.1,Python 版本为 3.8。使用随机梯度下降算法训练,训练时设置批量大小为 8,轮数为 200。

3.2 评价指标

本实验选用平均精度均值 A_{mp} 评估模型的检测精度,选用计算量、参数量和模型权重评估模型的大小,平均对数漏检率评估模型的漏检率,帧率评估模型实际检测中的实时性。查准比例 P 衡量模型判断正确的能力,查全比例 R 衡量模型找出正样本的能力,即

$$P = \frac{T_p}{T_p + F_p}, \quad (11)$$

$$R = \frac{T_p}{T_p + F_N}, \quad (12)$$

式中: T_p 表示将正样本预测为正样本的个体数量, F_p 表示将负样本预测为正样本的个体数量, F_N 表示将正样本预测为负样本的个体数量。平均精度 A_p 是 P - R 曲线与坐标轴围成的面积大小,即

$$A_p = \int_0^1 P(R) d(R). \quad (13)$$

A_p 表示 1 个目标类别的平均精度, A_{mp} 表示 m 个目标类别的平均精度均值,即

$$A_{mp} = \frac{1}{m} \sum_{i=1}^m A_{p,i}, \quad (14)$$

式中: $A_{p,i}$ 为第 i 个类别的 A_p 值,将交并比为 0.5 时的 A_{mp} 记为 $A_{mp,1}$,交并比为 0.5~0.95 的 A_{mp} 均值记为 $A_{mp,2}$ 。

3.3 三种模型性能对比

YOLOv8s、YOLOv8s-P2、YOLOv8s-Better 模型对比实验结果如表 1 所示。检测头尺度 A、B、C、D 分别为 160×160 、 80×80 、 40×40 、 20×20 ,分别用于检测微小目标、小目标、中目标、大目标。由表 1 可知,与 YOLOv8s 模型相比,YOLOv8s-Better 模型的检测精度分别提高 3.72% 和 0.71%,计算量增加 1.428×10^9 ;与 YOLOv8s-P2 模型相比,YOLOv8s-Better 模型的检测精度分别降低 0.11% 和 0.21%,计算量减少 6.901×10^9 。

表 1 三种模型对比

Table 1 Comparison of three models

模型	$A_{mp,1}/\%$	$A_{mp,2}/\%$	计算量/ 10^9	检测头尺度
YOLOv8s	78.98	45.89	28.675	(B,C,D)
YOLOv8s-P2	82.81	46.81	37.004	(A,B,C,D)
YOLOv8s-Better	82.70	46.60	30.103	(A,B,C)

3.4 不同位置 EMA 模块对比实验分析

以 YOLOv8s-Better 模型为基准模型,分别在骨干网络前 3 个瓶颈模块后添加 EMA 模块进行对比实验,将 3 个位置的 EMA 模块分别记为 EMA1、EMA2 和 EMA3 模块,在 3 个位置添加 EMA 模块后的 YOLOv8s-Better 模型分别记为 YOLOv8s-Better+EMA1、YOLOv8s-Better+EMA2 和 YOLOv8s-Better+EMA3 模型,对比实验结果如表 2 所示。由表 2 可知,与 YOLOv8s-Better+EMA3 模型相比,YOLOv8s-Better+EMA2 模型检测精度高,与 YOLOv8s-Better+EMA1 模型相比,YOLOv8s-Better+EMA2 模型计算量小、参数量少。

表 2 EMA 添加位置对比实验结果

Table 2 The experimental results of the comparison of EMA addition positions

模型	$A_{mp,1}/\%$	$A_{mp,2}/\%$	计算量/ 10^9	参数量/ 10^6
YOLOv8s-Better	82.70	46.60	30.103	3.310
YOLOv8s-Better+EMA1	83.64	46.92	30.352	3.320
YOLOv8s-Better+EMA2	83.58	46.99	30.346	3.312
YOLOv8s-Better+EMA3	83.35	46.75	30.343	3.311

3.5 不同注意力机制对比实验分析

为了验证 EMA 模块的优异性,采用多种注意力机制进行实验对比,包括全局注意力机制(global attention mechanism, GAM)、挤压与激发(squeeze-and-excitation, SE)^[15]、卷积注意力模块(convolutional block

attention module, CBAM)^[16]。以 YOLOv8s-Better 模型为基准模型,将在 YOLOv8s-Better 模型主干网络的第 2 个瓶颈模块后添加不同注意力机制,分别记为 YOLOv8s-Better+EMA 模型、YOLOv8s-Better+GAM 模型、YOLOv8s-Better+SE 模型和 YOLOv8s-Better+CBAM 模型,实验结果如表 3 所示。由表 3 可知,相较于 YOLOv8s-Better+GAM 模型,YOLOv8s-Better+EMA 模型的计算量小、参数量少。相较于 YOLOv8s-Better+SE 模型和 YOLOv8s-Better+CBAM 模型,YOLOv8s-Better+EMA 模型的检测精度高。

表 3 不同注意力机制对比实验结果

Table 3 The experimental results of the comparison of different attention mechanisms

模型	$A_{mP,1}/\%$	$A_{mP,2}/\%$	计算量/ 10^9	参数量/ 10^6
YOLOv8s-Better+EMA	83.58	46.99	30.346	3.312
YOLOv8s-Better+GAM	83.62	47.06	35.354	3.720
YOLOv8s-Better+SE	83.05	46.71	30.105	3.312
YOLOv8s-Better+CBAM	83.11	46.82	30.105	3.314

3.6 不同位置瓶颈模块改进对比实验分析

以 YOLOv8s-Better 为基准模型,将 YOLOv8s-Better 模型主干网络的瓶颈模块替换为快速瓶颈模块,改进后模型记为 YOLOv8s-Better+C2f1 模型;将 YOLOv8s-Better 模型颈部网络的瓶颈模块替换为快速瓶颈模块,改进后模型记为 YOLOv8s-Better+C2f2 模型;将 YOLOv8s-Better 模型整个网络的瓶颈模块替换为快速瓶颈模块,改进后模型记为 YOLOv8s-Better+C2f3 模型,对比实验结果如表 4 所示。YOLOv8s-Better+C2f2 模型的检测精度高于 YOLOv8s-Better+C2f1 模型和 YOLOv8s-Better+C2f3 模型;相较于 YOLOv8s-Better 模型,YOLOv8s-Better+C2f2 模型计算量小、参数量少。

表 4 改进瓶颈模块后对比实验结果

Table 4 The experimental results after improving the bottleneck module

模型	$A_{mP,1}/\%$	$A_{mP,2}/\%$	计算量/ 10^9	参数量/ 10^6
YOLOv8s-Better	82.70	46.60	30.103	3.310
YOLOv8s-Better+C2f1	82.96	46.33	26.580	2.745
YOLOv8s-Better+C2f2	83.08	46.41	27.285	2.966
YOLOv8s-Better+C2f3	81.49	45.73	23.762	2.401

3.7 消融实验结果分析

以 YOLOv8s-Better 为基准模型进行消融实验,实验结果如表 5 所示。表 5 中,E 代表 YOLOv8s-Better 模型,F 代表 YOLOv8s-Better+EMA2 模型,G 代表 YOLOv8s-Better+C2f2 模型,H 代表 YOLOv8s-Better+SPPF_LSKA 模型,I 代表 YOLOv8s-Better+EMA2+C2f2 模型,J 代表 YOLOv8s-Better+EMA2+C2f2+SPPF_LSKA 模型。

表 5 网络重构后的消融实验结果

Table 5 Ablation experimental results after network reconstruction

模型	$A_{mP,1}/\%$	$A_{mP,2}/\%$	计算量/ 10^9	参数量/ 10^6	权重大小/ 10^6
YOLOv8s	78.98	45.89	28.675	11.141	42.6
E	82.70	46.60	30.103	3.310	12.7
F	83.58	46.99	30.346	3.312	12.8
G	83.08	46.41	27.285	2.966	11.4
H	83.79	47.02	30.968	3.582	13.8
I	84.77	47.13	27.528	2.968	11.5
J	85.46	47.35	28.393	3.241	12.5

由表 5 可知,相较于 YOLOv8s 模型,网络重构得到的 YOLOv8s-Better 模型在检测精度指标上分别提高 3.72% 和 0.71%,参数量减少 70%,计算量增加 1.428×10^9 ,说明本文的网络重构整体上非常有效。相较于 YOLOv8s-Better 模型,YOLOv8s-Better+EMA2 模型的检测精度分别提高 0.88% 和 0.39%,计算量和参数量分别增加 0.243×10^9 和 0.002×10^6 ,说明多尺度注意力机制可以有效替代深层网络。相较于 YOLOv8s-Better 模型,YOLOv8s-Better+C2f2 模型的计算量小,参数量少,检测精度高,说明改进后的快速瓶颈模块能够减少特征图冗余。相较于 YOLOv8s-Better 模型,YOLOv8s-Better+LSKA 模型的检测精度高。

为了更好地验证 MFS_YOLO 模型的有效性,与 YOLOv8s 模型对比研究,实验结果如图 7 所示。由图 7 可知,在港湾和树两类目标中 MFS_YOLO 模型的检测精度下降,其余 12 类目标中 MFS_YOLO 模型的检测精度提高,其中检测目标为小目标岩石时,检测精度提高最大,为 19.60%。此外,MFS_YOLO 模型的平均对数漏检率比 YOLOv8s 模型低。

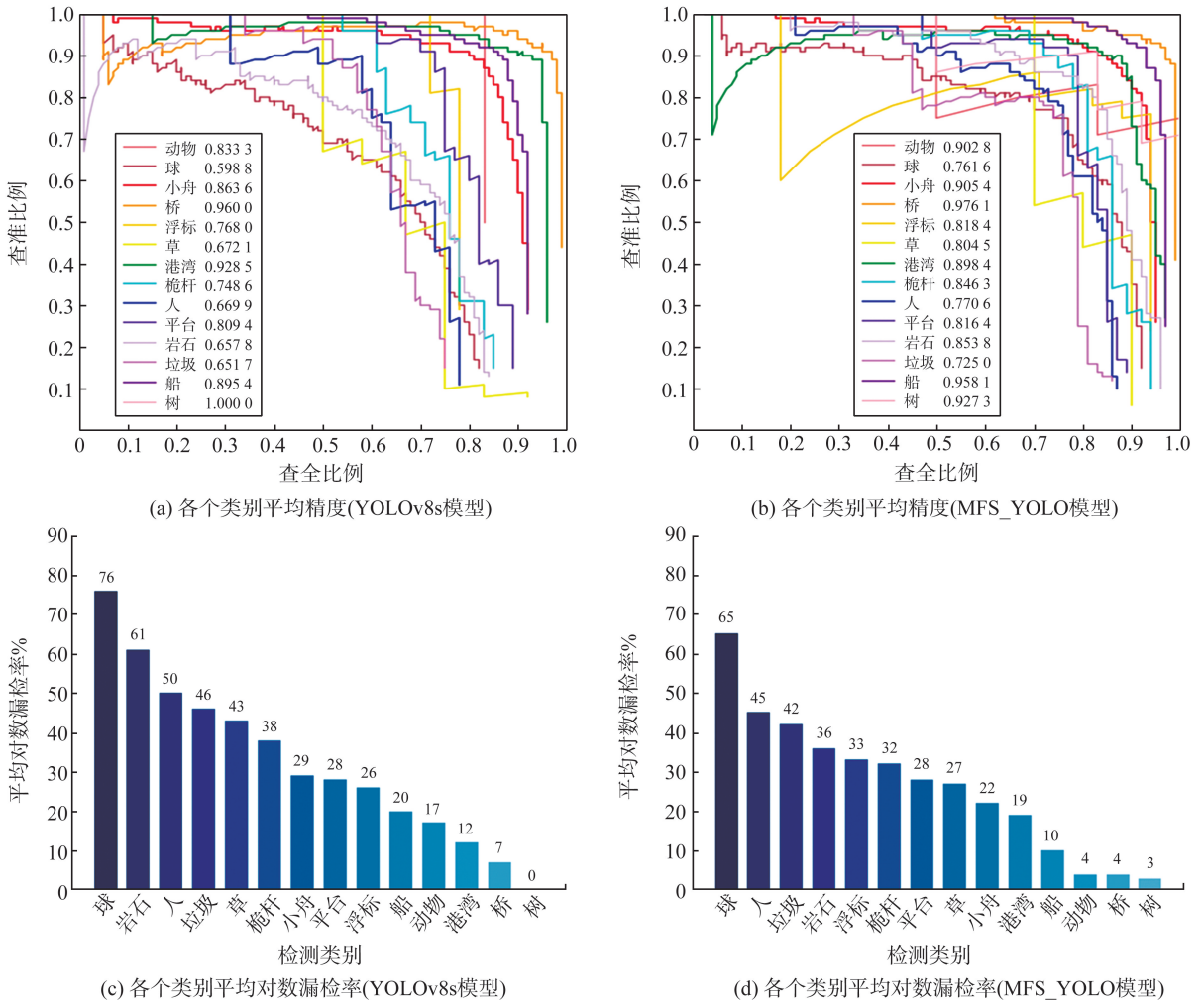


图 7 平均精度和平均对数漏检率对比

Fig.7 Comparison of average precision and log-average miss rate

3.8 与其它模型对比

为进一步说明本文模型的优越性,再与常见的目标检测模型进行了实验对比,包括单阶段多目标检测器(single shot multibox detector, SSD)模型、基于区域的更快的卷积神经网络(regions with convolutional neural network, Faster-RCNN)模型,以及其它 4 个不同版本的 YOLO 模型,对比结果如表 6 所示。由表 6 可知,MFS_YOLO 模型的检测精度最高、参数量最少,帧率也大于 30,满足实时检测需求。

表 6 多种目标检测模型对比

Table 6 Comparison of multiple target detection models

模型	$A_{mP,1}/\%$	$A_{mP,2}/\%$	计算量/ 10^9	参数量/ 10^6	帧率
YOLOv3	55.68	21.52	65.692	61.594	42
YOLOv5s	49.74	24.92	16.588	7.099	60
YOLOv7	59.88	31.20	105.337	37.265	30
SSD	49.32	23.24	62.050	25.349	73
Faster-RCNN	47.68	21.94	941.083	28.409	13
YOLOv8s	78.98	45.89	28.675	11.141	80
MFS_YOLO	85.46	47.35	28.393	3.241	62

3.9 检测结果可视化

为了更直观地体现出改进后的模型在检测性能上的优异,在不同环境下进行测试,测试结果如图 8 所示。从图 8 可以看出,MFS_YOLO 模型可以检测出更多的目标。例如在晴天环境下,MFS_YOLO 模型不仅检测出了正面的 4 个小舟和 1 个桥,还检测出被遮挡住的 1 个小舟。在多云环境下,MFS_YOLO 模型检测出了更小的目标岩石。在雾天和傍晚环境下,光线较暗,且水面还有反射光线的干扰,MFS_YOLO 模型依然能检测出更多的远处的小目标。综合以上对比,可以看出 MFS_YOLO 模型有效地降低了模型的漏检率。



(a) 晴天(YOLOv8s模型)



(b) 晴天(MFS_YOLO模型)



(c) 多云(YOLOv8s模型)



(d) 多云(MFS_YOLO模型)



(e) 雾天(YOLOv8s模型)



(f) 雾天(MFS_YOLO模型)



(g) 傍晚(YOLOv8s模型)



(h) 傍晚(MFS_YOLO模型)

图 8 不同环境下检测效果对比

Fig.8 Comparison of detection effects in different environments

4 结论

本文对 YOLOv8s 模型改进,在水面小目标检测任务中,改进后的 MFS_YOLO 模型的性能提高。根据中心度理论,考虑深层网络对模型可能带来的负面影响,重构 YOLOv8s 模型主干网络,调整颈部网络特征金字塔结构和路径聚合网络,得到 YOLOv8s-Better 模型。在主干网络引入多尺度注意力机制,增强 YOLOv8s-Better 模型的特征提取能力。改进颈部网络的瓶颈模块,降低了 YOLOv8s-Better 模型的计算量和参数量。在 SPPF 模块中引入大尺度可分离核,能增大 YOLOv8s-Better 模型的感受视野,提高 YOLOv8s-

Better 模型的特征融合能力。实验表明,改进后的 MFS_YOLO 模型提高了检测精度,计算量小、参数量少、漏检率低,能够检测出更多水面小目标。

参考文献:

- [1] GIRSHICK R. Fast R-CNN[C] // Proceedings of the IEEE International Conference on Computer Vision (ICCV). Boston: IEEE, 2015:1440-1448.
- [2] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016:779-788.
- [3] LIU W, ANGUELOW D, ERHAN D, et al. SSD: single shot multibox detector[C] // Proceedings of the 14th European Conference on Computer Vision (ECCV). Berlin: Springer, 2016:21-37.
- [4] 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):1137-1149.
- [5] TAO Kong, YAO Anbang, CHEN Yurong, et al. HyperNet: Towards accurate region proposal generation and joint object detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016:845-853.
- [6] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C] // Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017:936-944.
- [7] CHEN C Y, LIU M Y, TUZEL O, et al. R-CNN for small object detection[C] // Proceedings of the 13th Asian Conference on Computer Vision. Berlin: Springer, 2016:214-230.
- [8] 戚玲珑, 高建瓴. 基于改进 YOLOv7 的小目标检测[J]. 计算机工程, 2023, 49(1):41-48.
QI Linglong, GAO Jianling. Small target detection based on improved YOLOv7[J]. Computer Engineering, 2023, 49(1):41-48.
- [9] LIN Feng, HOU Tian, JIN Qiannan, et al. Improved YOLO based detection on algorithm for floating debris in waterway[J]. Entropy, 2021, 23(9):1111.
- [10] 王林, 汪钰婷. 基于加强特征融合的轻量化船舶目标检测[J]. 计算机系统应用, 2023, 32(2):288-294.
WANG Lin, WANG Yuting. Lightweight ship target detection based on enhanced feature fusion[J]. Computer System Applications, 2023, 32(2):288-294.
- [11] OUYANG Daliang, HE Su, ZHAN Jian, et al. Efficient multi-scale attention module with cross-spatial learning[C] // Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP). Rhodes Island: IEEE, 2023:1-5.
- [12] CHEN J R, KAO S, HE H, et al. Run, don't walk: chasing higher flops for faster neural networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023:12021-12031.
- [13] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: rethinking the large kernel attention design in CNN [J]. Expert Systems with Applications, 2024, 236:121352.
- [14] ZHOU Zhiguo, SUN Jiaen, YU Jiabao, et al. An image-based benchmark dataset and a novel object detector for water surface object detection[J]. Frontiers in Neurorobotics, 2021, 15:723336.
- [15] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2018:7132-7141.
- [16] WOO S H, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C] // Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018:3-19.

(编辑:陈丽萍)