王德兴,何　勇,袁红春. 基于 YOLOv8-BAN 模型的水下生物目标检测方法[J]. 江苏农业学报,2025,41(1):101-111.
doi:10.3969/j.issn.1000-4440.2025.01.012

基于 YOLOv8-BAN 模型的水下生物目标检测方法

王德兴,　何　勇,　袁红春
(上海海洋大学信息学院,上海　201306)

摘要:　水下目标检测技术对于自动化水下捕捞至关重要,可有效推动渔业的智能化发展。针对水下图像质量较差和小目标水下生物聚集导致漏检、误检等问题,本研究提出了一种基于改进 YOLOv8m 模型的水下生物目标检测模型——YOLOv8-BAN。该模型首先在骨干网络中嵌入双向路由自注意力机制,以增强网络的特征提取能力;其次在颈部结合自适应特征融合模块,优化特征融合效果,增强了模型对多尺度目标的检测能力;最后设计了一种小目标损失函数,通过精确标签分配进一步提升了水下生物小目标的检测精度。在 URPC2018 和 Brackish 数据集上的测试结果显示,YOLOv8-BAN 模型的平均检测精度分别达到 86.9%和 98.6%,较 YOLOv8m 分别提高了 3.5 个百分点和 3.3 个百分点;与其他 6 种模型相比,YOLOv8-BAN 模型具有更高的检测精度和较快的检测速度。本研究结果可为水下机器人进行水产捕捞作业提供了技术支持。

关键词:　水下生物;YOLOv8m;深度学习;小目标检测
中图分类号:　TP391　　**文献标识码:**　A　　**文章编号:**　1000-4440(2025)01-0101-11

Underwater biological target detection method based on YOLOv8-BAN model

WANG Dexing,　HE Yong,　YUAN Hongchun
(*College of Information Technology*, *Shanghai Ocean University*, *Shanghai* 201306, *China*)

Abstract:　Underwater target detection technology is crucial for the automation of underwater fishing, which can effectively promote the intelligent development of the fishing industry. Aiming at the problems of poor underwater image quality and missed and false detections caused by the aggregation of small target underwater organisms, this study proposed an underwater biological target detection method based on improved YOLOv8m model, namely YOLOv8-BAN. The model first embedded a bidirectional routing self-attention mechanism in the backbone network to enhance the network's feature extraction capability. Secondly, the adaptive feature fusion module was combined in the neck to optimize feature fusion effects, enhancing the model's detection capability for multi-scale targets. Finally, a small target loss function was designed to further improve the detection accuracy of small targets through precise label assignment. Experimental results on the URPC2018 and Brackish datasets showed that the average detection accuracy of YOLOv8-BAN model reached 86.9% and 98.6% respectively, which was 3.5 percentage points and 3.3 percentage points higher than that of YOLOv8m model. Compared with the other six models, the YOLOv8-BAN model had higher detection accuracy and faster detection speed. The results of this study can provide technical support for underwater robots to carry out aquaculture fishing operations.

Key words:　underwater organisms;　YOLOv8m;　deep learning;　small target detection

收稿日期:2024-04-03
基金项目:国家自然科学基金项目(41776142)
作者简介:王德兴(1971-),男,河北保定人,博士,副教授,研究方向为人工智能、模式识别和数据挖掘等。(E-mail)dxwang@shou.edu.cn
通讯作者:何　勇,(E-mail)2850035542@qq.com

水下生物目标检测技术是水产养殖智能化战略的核心组成部分,对于实现水下机器人自动化捕捞具有重要意义[1]。水下机器人在自动化捕捞中的

应用依赖于高效的水下生物目标检测技术[2]。然而,受水下环境和光照条件的影响,光学图像存在纹理特征信息不足、对比度低等问题[3],同时小目标生物的聚集也增加了检测的难度。因此,亟需设计一种高精度的、鲁棒性强的水下生物目标检测模型。

近年来,随着计算机视觉技术的发展,基于深度学习的检测方法被广泛应用于水下生物目标检测。深度学习目标检测方法分为两大类,即双阶段方法和单阶段方法。双阶段方法需要先生成候选区域,然后经过分类和回归得到检测结果。当前,在水下生物目标检测领域,许多研究者选择基于双阶段检测算法进行研究,尤其是针对经典的 Faster R-CNN 算法[4]进行改进和优化。袁红春等[5]提出了一种基于 Faster R-CNN 二次迁移学习和带色彩恢复的多尺度视网膜增强算法,该方法在水下小规模鱼类数据集上表现出较高的准确率。Liu 等[6]对 Faster R-CNN 进行了改进,将骨干网络替换为 Transformer 结构,并添加了聚合路径网络以增强特征提取能力,但该方法检测速度较慢。Lin 等[7]提出了一种基于 Faster R-CNN 的数据增强方法 RoIMix,将多张图片中感兴趣的区域进行融合,模拟水下生物的重叠和遮挡。Shi 等[8]将 Faster R-CNN 骨干网络进行改进,使用 ResNet 并引入 BI-FPN 特征金字塔结构以加强模型的特征提取能力。相比双阶段方法,单阶段方法速度优势明显,而且随着单阶段算法的不断迭代,其精度也能达到很高。目前研究人员主要基于 YOLO 系列模型[9-12]开展单阶段算法的研究。Guo 等[13]针对水下图像模糊、对比度低的问题,提出了一种改进自适应算法的 MSRP 图像增强算法,并和 YOLOv3 模型结合进行检测,但该模型骨干网络的特征提取能力较弱。Chen 等[14]在 YOLOv4 模型的基础上进行改进,通过增加残差块与通道注意力机制结合,增强骨干网络特征提取能力。Lei 等[15]基于 YOLOv5 模型进行改进,将 Swim Transformer 作为基本骨干网络并改进路径聚合网络 PANet,让网络更适用于模糊的水下图像。翟先一等[16]使用带色彩恢复的多尺度视网膜增强算法对图像进行预处理,并使用卷积注意力机制对海参进行检测,但该方法使用的数据集类别较少,仅对海参有较好的检测效果。Sun 等[17]使用 MobileT 作为 YOLOX 模型的骨干网络,提高算法的全局特征提取能力,减少了参数量,但该方法对小目标检测效果不佳。Yi 等[18]针对小目标检测漏检率高的问题,提出了一种基于 YOLOv7 模型改进的检测算法,该方法通过整合 SENet 注意力机制,增强 FPN 金字塔结构,合并 EIOU 损失函数,集中了小目标的更多关键特征信息,提高了小目标的检测精度。

尽管基于深度学习方法在水下生物目标检测任务上已经获得了显著的精度和速度提升,但仍然存在一些问题。首先,现有方法使用的数据集数量较少或者种类单一,导致模型泛化性不足。其次,现有方法对于水下目标尤其是小目标检测精度不足,这主要是因为水下图像质量不佳,导致目标特征难以提取。同时,大部分网络在多尺度融合过程中主要使用简单的元素相加,容易携带不同特征层的矛盾信息。此外,基于交并比(Intersection over Union,IOU)改进的损失函数对于小目标位置偏差较为敏感,难以对小目标进行精准定位。针对以上问题,本研究提出了一种基于改进 YOLOv8m 模型的水下生物目标检测模型 YOLOv8-BAN 模型。该模型首先在骨干网络中嵌入双向路由自注意力机制(Bi-Level Routing Attention, BRA),用于提高网络特征提取能力。其次在颈部结合自适应特征融合网络(Adaptive Feature Fusion,AFF),提高不同尺度目标尤其是小目标的检测精度。最后设计了 NWD-CIOU损失函数,替换原始的完全交并比(Complete-IOU,CIOU)损失函数,对小目标进行更准确的标签分配,进一步提高小目标的定位精度。为了让模型具有较强的泛化性,本研究在两个公共数据集上进行消融试验和对比试验,以验证改进模型的有效性。

1 材料与方法

1.1 YOLOv8-BAN 模型的网络结构

为了保证检测的实时性,本研究使用 YOLOv8m 模型作为基础模型,并根据所提出的改进方法,将其命名为 YOLOv8-BAN 模型。YOLOv8-BAN 模型网络架构由 3 个主要部分组成,分别是骨干网络(Backbone)、颈部网络(Neck)以及检测头(Head),网络结构如图 1 所示。Backbone 采用了一系列卷积和反卷积来提取特征,同时使用残差连接和瓶颈结构来缩减网络大小并提高性能。Backbone 部分采用了 C2F 模块作为基本的构成单元,与 YOLOv5s 模型的 C3 模块相比,C2F 模块具有更少的参数和更优秀的特征提取能力。同时为了增强在水下环境的特征提取能力,嵌入了 BRA 双向路由自注意力机制。Neck 部分增加一个 4 倍下采样的浅层特征层,

使用 4 个特征层结合 AFF 网络进行自适应特征空间融合,将融合后的 4 个特征作为检测头进行检测。Head 负责最终的目标检测和分类任务,包括一个检测头和一个分类头,检测头包括一系列的反卷积层和池化层,用于生成检测结果;分类头采用全局池化对每个特征层进行分类。YOLOv8 模型使用 CIOU 作为边界框定位损失函数,由于该函数不利于小目标的检测,因此本研究使用一种边界框距离度量标准 NWD,与 CIOU 结合设计了 NWD-CIOU 损失函数,用来提高小目标在底层标签分配中的准确性。

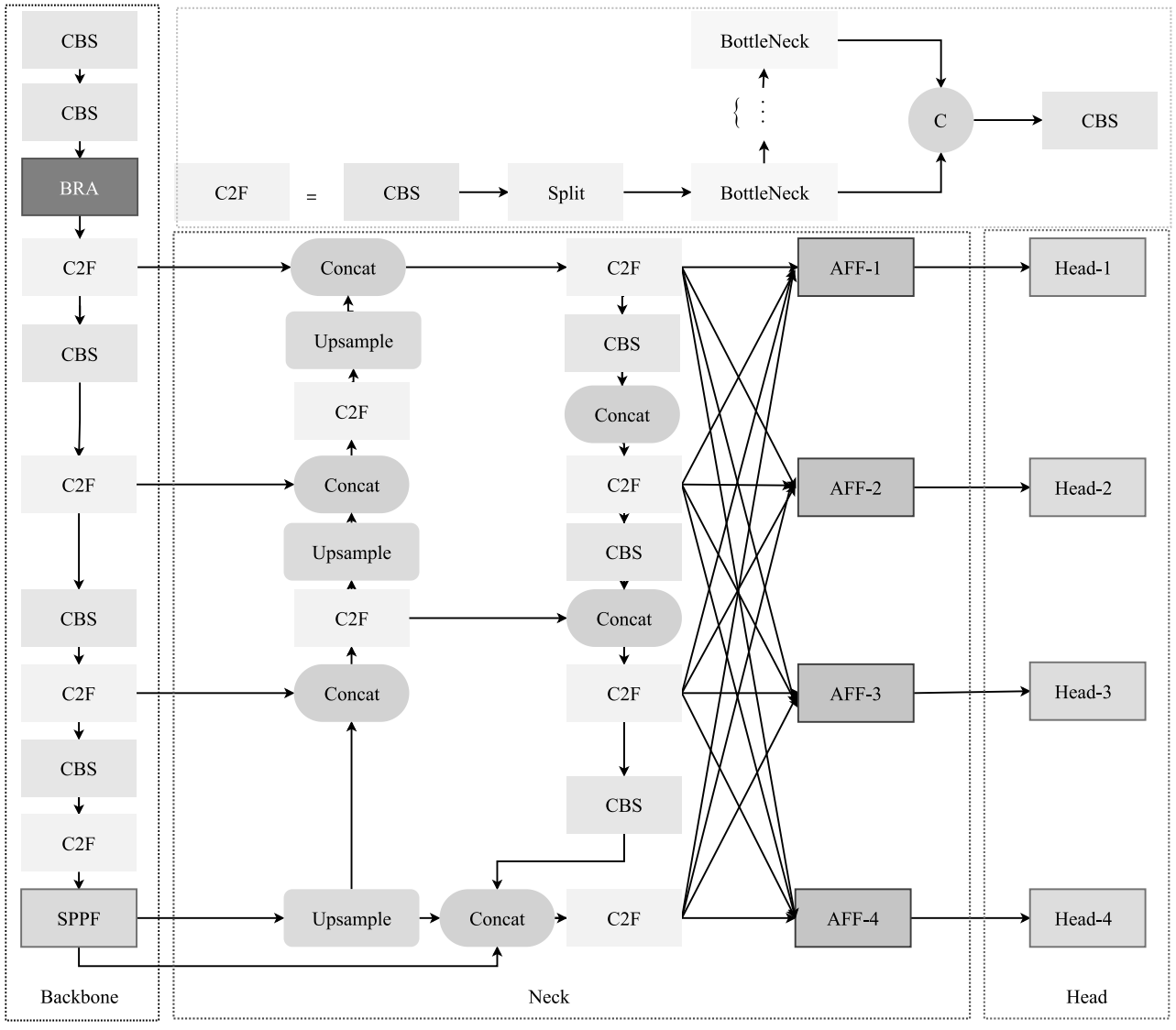


图 1 YOLOv8-BAN 模型的网络结构

Fig.1 The network structure of YOLOv8-BAN model

1.1.1 双向路由自注意力机制 针对水下环境中图像对比度低、模糊和失真等问题,本研究在骨干网络中嵌入了 BRA[19] 自注意力机制,以增强骨干网络的特征提取能力。这种机制使得模型能够更有效地捕捉并利用有限的目标特征,从而在复杂的水下环境中提高检测效果。BRA 本质是一种自注意力机制的变体(图 2),它将多头自注意力的计算分为两个阶段,第一阶段进行粗粒度的注意力计算,该模块基于稀疏采样而非下采样,可以保留细粒度的细节。第二阶段基于第一阶段的稀疏注意力结果进行细粒度的注意力计算。在第一阶段中将给定的 $H\times W$ 维图像划分为 $S\times S$ 个非重叠区域,然后对每个非重叠区域进行自注意力计算,得到查询 Q、键 K 和值 V。然后构建有向图来找到关注关系,即每个给定区域

应该关注的区域。具体来说,通过对 Q 和 K 应用每个区域的平均值来得到区域查询 Q^r 和键 K^r,然后通过 Q^r 和转置 K^r 之间的矩阵乘法,得到区域到区域亲和图像的邻接矩阵 A^r, 其中邻接矩阵中的每个数值表示两个区域之间的语义关联程度,其计算公式为:

$$A^r = Q^r (K^r)^{\mathrm{T}} \tag{1}$$

公式(1)中 r 表示计算的区域(region),T 表示转置符号。

为了更加高效地定位有价值的键值对进行全局参与,在粗粒度的区域级别中过滤掉不相关的键值对,只需要保留对每个区域关联程度最大的键值对,从而得到了一个路由索引矩阵 $\boldsymbol{I}^r$,其公式为:

$$\boldsymbol{I}^r = \mathrm{topkIndex}(A^r) \tag{2}$$

在第二阶段中,根据第一阶段得到的邻接矩阵 $\boldsymbol{I}^r$ 进行细粒度的自注意力计算。对于第 i 个区域中的每个查询,让它仅仅关注 $\boldsymbol{I}^r$ 中第 i 行的前 k 个区域的并集中所有的键值对,为了方便操作首先需要收集所有的 K 和 V,公式为:

$$K^g = \mathrm{gather}(K, I^r) \tag{3}$$

$$V^g = \mathrm{gather}(V, I^r) \tag{4}$$

公式(3)中 K^g 和公式(4)中 V^g 是 $\boldsymbol{I}^r$ 中所有区域收集到的键值张量。

最后将注意力集中在收集的键值对上,其计算公式为:

$$O = \mathrm{Attention}(Q, K^g, V^g) + LEC(V) \tag{5}$$

公式(5)中引入了一个局部上下文增强术语 $LCE(V)$[20], 可以增强 V 中相邻像素之间的联系。其中,函数 $LCE(\cdot)$ 使用深度可分离卷积进行参数化,卷积核大小设置为 5。

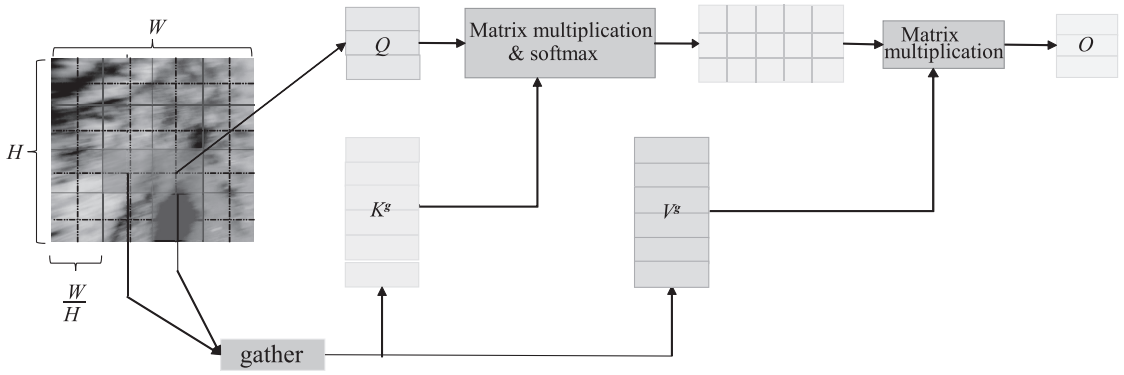


图 2 自注意力机制 BRA 的结构

Fig.2 The structure of self-attention mechanism BRA

1.1.2 自适应特征融合网络 YOLOv8 颈部网络使用 3 种尺度不同的特征层进行融合,分别是20×20、40×40 和80×80,然而在水下场景中,图像比较模糊且存在不同尺度的密集目标,这些目标的语义信息和位置信息更多集中在更浅的特征层,仅使用 3 个较深的特征层容易出现漏检或者误检。为此本研究在颈部特征融合过程中增加了一个160×160 的浅层特征层,以获得更多的特征信息,然后设计了 AFF 网络,将 4 个不同尺度大小的特征层进行自适应特征融合。该方法是训练过程中学习不同层次特征的最佳融合方法,融合过程中可以过滤掉携带矛盾的其他层的特征信息,从而缓解学习目标不一致的问题。

AFF 结构如图 3 所示,其核心思想是自适应学习每个尺度上特征图的融合空间权重,主要分为两个步骤,即特征缩放和自适应融合。先将特征图进行缩放,其中第 l 层将其他特征层通过上采样或者下采样的方式调整到和该层特征图大小。对于上采样使用1×1 的卷积层将特征图像通道压缩到和第 l 层相同,然后使用插值法提高分辨率;对于下采样则使用步长为 2 的3×3 卷积层修改通道数量和分辨率,最后进行特征融合,以第 l 层输出特征图像为例,其融合公式如下:

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1\to l} + \beta_{ij}^l \cdot x_{ij}^{2\to l} + \gamma_{ij}^l \cdot x_{ij}^{3\to l} + \eta_{ij}^l \cdot x_{ij}^{4\to l} \tag{6}$$

在公式(6)中,l 表示融合的层数,$x_{ij}^{k\to l}$ 表示第 k 个输入特征层($k=1,2,3,4$)在第 l 层融合后在(i,j)位置上输出的特征结果,α_{ij}^l、β_{ij}^l、γ_{ij}^l 和 η_{ij}^l 分别代表对于不同层的权重系数,并且对于权重参数满足 $\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l + \eta_{ij}^l = 1$,$\alpha_{ij}^l$、$\beta_{ij}^l$、$\gamma_{ij}^l$ 和 $\eta_{ij}^l \in [0,1]$,其中权重参

数 α_{ij}^{l} 定义为:

$$\alpha_{ij}^{l}=\frac{e^{\lambda^{l}_{\alpha_{ij}}}}{e^{\lambda^{l}_{\alpha_{ij}}}+e^{\lambda^{l}_{\beta_{ij}}}+e^{\lambda^{l}_{\gamma_{ij}}}+e^{\lambda^{l}_{\eta_{ij}}}} \tag{7}$$

公式(7)中 $e^{\lambda^{l}_{\alpha_{ij}}}$、$e^{\lambda^{l}_{\beta_{ij}}}$、$e^{\lambda^{l}_{\gamma_{ij}}}$、$e^{\lambda^{l}_{\eta_{ij}}}$ 都是控制参数,通过 1×1 的卷积核与 $x^{1\to l}$、$x^{2\to l}$、$x^{3\to l}$、$x^{4\to l}$ 分别学习得到,其他权重参数以此类推。通过该方式进行自适应特征融合后得到 4 个输出特征层,后续使这 4 个特征层作为检测头进行检测。

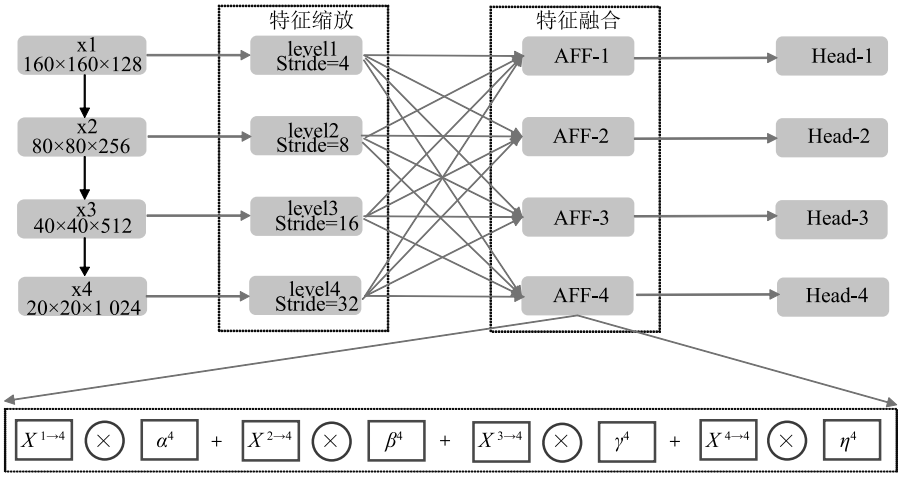


图 3 自适应特征融合网络(AFF)结构

Fig.3 The structure of the adaptive feature fusion network (AFF)

1.1.3 小目标损失函数(NWD-CIOU) YOLOv8 中的损失函数包括 3 个部分,即边界框定位损失、置信度损失和分类损失。其中边界框定位损失默认使用完全交并比 CIOU[21] 为度量标准,CIOU 是对 IOU 的改进,然而这两种度量标准对于小目标的位置偏差都较为敏感,导致小目标在标签分配中可能无法匹配到正样本或者正样本数量太少,降低模型的性能。考虑到 CIOU 在衡量小目标边界框距离时可能不是最佳选择,本研究提出了一种改进的 CIOU 损失函数,命名为 NWD-CIOU。这种新的损失函数旨在更准确地评估并优化小目标的边界框预测,提升小目标检测的精度。

应用 NWD-CIOU 首先需要对边界框进行高斯分布建模。对于较小物体的边界框,由于物体不是严格意义的矩形,所以存在一些前景像素和背景像素,各自分布在边界框的中间和边界[22]。为了描述边界框中不同像素的权重,对边界框进行二维高斯分布建模,其中最中间的像素有最高权重,权值大小从中心到边界逐渐降低。对于边界框 $R(cx,cy,\omega,h)$,其中 (cx,cy)、ω 和 h 分别表示为边界框的中心坐标、宽度和高度。其内接圆的方程式为:

$$\frac{(x-\mu_x)^2}{\sigma_x^{\ 2}}+\frac{(y-\mu_y)^2}{\sigma_y^{\ 2}}=1 \tag{8}$$

公式(8)中 μ_x 和 μ_y 是椭圆的中心坐标,σ_x 和 σ_y 表示 x 轴和 y 轴的半轴长度。因此 $\mu_x=cx$,$\mu_y=cy$,$\sigma_x=\omega/2$,$\sigma_y=h/2$。二维高斯分布的概率密度函数为:

$$f(X|\mu,\Sigma)=\frac{\exp\left[-\frac{1}{2}(X-\mu)^{T}\Sigma^{-1}(X-\mu)\right]}{2\pi|\Sigma|^{\frac{1}{2}}} \tag{9}$$

公式(9)中 exp 表示以 e 为底的指数函数,X、μ 和 Σ 分别表示高斯分布的坐标、平均向量和协方差矩阵。

进行二维高斯分布建模后,使用最优运输理论中的 Wasserstein 距离[23]来衡量两个边界框的距离。对于 $\mu_1=N(m_1,\Sigma_1)$ 和 $\mu_2=N(m_2,\Sigma_2)$ 两个二维高斯,两者之间的二阶 Wasserstein 距离定义为:

$$W_2^2(\mu_1,\mu_2)=\|m_1-m_2\|_2^2+\|\Sigma_1^{1/2}-\Sigma_2^{1/2}\| \tag{10}$$

公式(10)中 m_1 和 m_2 表示高斯分布的均值向量,$\|\cdot\|_F$ 表示 Frobenius 范数。

对于两个边界框,距离度量又可以表示为:

$$W_2^2(N_a,N_b)=\left\|\left[\left(cx_a,cy_a,\frac{\omega_a}{2},\frac{h_a}{2}\right)^{T},\left(cx_b,cy_b,\frac{\omega_b}{2},\frac{h_b}{2}\right)^{T}\right]\right\|_2^2 \tag{11}$$

公式(11)中 a 和 b 代表两个边界框。然而,这

个距离度量并不能直接用于相似度的计算,需要对其进行归一化,获得归一化的 Wasserstein 距离(Normalized Wasserstein distance,NWD),将其作为边界框的度量标准,其公式如下:

$$NWD(N_a,N_b)=\exp\left[-\frac{\sqrt{W_2^2(N_a,N_b)}}{C}\right] \quad (12)$$

公式(12)中 a 和 b 代表两个边界框,C 是一个和数据集相关的常数(数据集的平均大小)。如果仅以 NWD 度量方式作为模型的定位损失函数,对于包含不同尺度大小的数据集可能达不到更好的检测效果,为此本研究将 NWD 和 CIOU 两种度量标准进行结合,引入一个比例因子,设计了新的 NWD-CIOU 损失函数,即:

$$LOSS_{\text{NWD-CIOU}}=(1-\mu)LOSS_{\text{NWD}}+\mu LOSS_{\text{CIOU}} \quad (13)$$

公式(13)中 μ 值为超参数,经过多次试验后该值取 0.2 达到最佳。和 CIOU 相比,NWD-CIOU 考虑到了小目标由于位置偏差过于敏感导致标签分配失败的问题,提升了模型对小目标的检测效果。和仅使用 NWD 度量标准相比,在包含不同尺度大小的数据集上,NWD-CIOU 能获得更高的精度,并且可以加快模型训练的收敛速度。

1.2 试验数据集

本试验所用到的数据集来自公开的 URPC2018 数据集和 Brackish 数据集。其中 URPC2018 数据集有3 701张图片,包含海星、海参等 4 种海洋生物,部分数据集图片如图 4 所示。Brackish 数据集总共有 14 518张图片,包含鱼类、螃蟹等 6 种海洋生物,部分数据集图片如图 5 所示。本研究将两个数据集均按照8∶1∶1 的比例划分为训练集、验证集和测试集进行后续试验。

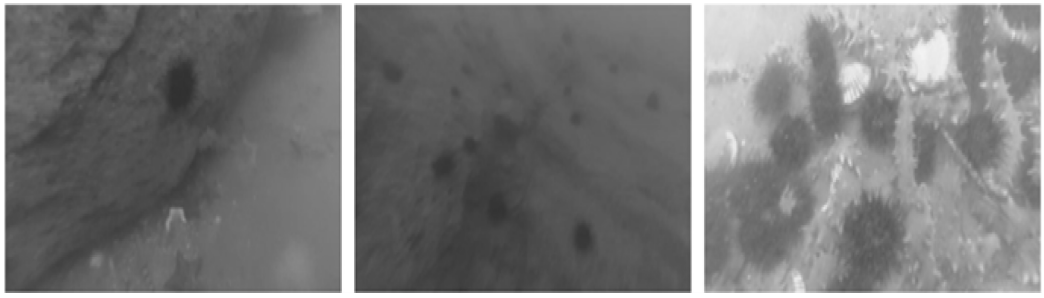


图 4 部分 URPC2018 数据集图片

Fig.4 Images from the URPC2018 dataset

图 5 部分 Brackish 数据集图片

Fig.5 Images from the Brackish dataset

1.3 试验设置

本研究的模型构建在 PyTorch 深度学习框架之上,并在 Ubuntu 20.04 操作系统环境下进行试验。具体而言,PyTorch 版本为 1.8,搭配的 Python 版本是 3.8。模型训练任务在配备 NVIDIA GeForce RTX 3090(24 GB 显存)的 GPU 上执行。为了确保模型训练的一致性和可重复性,输入图像被标准化为 640×640 像素。在训练过程中,Batch-Size 设置为 16,总共进行 300 个 epoch 训练。初始学习率设定为 0.01,并采用余弦退火策略对学习率进行调整。优化过程中,动量参数设置为 0.937,选用 SGD(随机梯度下降)作为优化器。

1.4 评价指标

为了评价模型的检测精度,本研究采用平均精确率(*mAP*)、精准率(*P*)和召回率(*R*)作为评价指标。其中 *AP* 表示单类标签的平均精确率,*mAP* 表示所有类别标签的平均精确率,*IOU* 取值为 0.5。精准率表示在预测的所有正样本中实际也是正样本的概率。召回率表示实际为正样本被预测为正样本的概率。为了评价模型的检测速度,选取每 1 s 检测帧数(*FPS*)作为检测速度的评价标准。上述指标的计算公式分别为:

$$P=\frac{TP}{TP+FP} \tag{14}$$

$$R=\frac{TP}{TP+FN} \tag{15}$$

$$AP=\int_0^1 P\mathrm{d}R \tag{16}$$

$$mAP=\frac{\sum_n^N AP_n}{N} \tag{17}$$

$$FPS=\frac{F_T}{T_C} \tag{18}$$

公式(14)和(15)中 *TP* 表示被预测为正样本的数量,*FP* 表示预测为负样本的数量,*FN* 表示实际为正样本被预测为负样本的数量;公式(17)中 *N* 表示类别数量;公式(18)中 F_T 表示总帧数;T_C 表示检测时间。

2 结果与分析

2.1 消融试验结果

为了验证改进模块的有效性,对本研究方法使用的 BRA、AFF 和 NWD-CIOU 3 个模块进行消融试验,在 URPC2018 和 Brackish 两个数据集上的试验结果如表 1 和表 2 所示。可以看到,当网络中添加 BRA 自注意力模块后,与 YOLOv8m 模型相比 *mAP* 分别提升 1.4 个百分点和 1.5 个百分点,这说明 BRA 使用了细粒度的自注意力机制,建立远程的上下文特征依赖,捕获最显著特征,从而增强了网络的特征提取效果。当加入 AFF 之后,与 YOLOv8m 模型相比 *mAP* 均提升 1.2 个百分点,这说明 AFF 通过增加浅层特征层进行自适应特征融合,更加充分地利用不同尺度特征层的位置信息和语义信息,提高不同尺度目标的检测效果。将 NWD 与 CIOU 结合作为边界框损失函数后,与 YOLOv8m 模型相比 *mAP* 分别提升了 1.8 个百分点和 1.5 个百分点,这一改进在两个不同的数据集上都得到了验证。图 6 中的边界框损失曲线图直观展示了改进措施的效果。在训练过程中,使用 NWD 的模型与使用 CIOU 的模型相比,显示出更明显的训练损失下降。这一结果表明,NWD 在处理小目标时性能更佳,能够实现更精确的标签分配,从而提高对小目标的检测精度。当 NWD 与 CIOU 结合使用时,模型的损失下降最为显著。由此可见,将 NWD 与 CIOU 结合能够充分利用两种度量标准的优势,平衡对不同尺度目标的检测性能,提升模型的整体检测效果。从表 1 和表 2 还可以发现,将全部模块添加之后,模型在两个数据集上的 *mAP* 达到最高,分别为 86.9% 和 98.6%,与 YOLOv8m 模型相比分别提高了 3.5 个百分点和 3.3 个百分点。

表 1 在 URPC2018 数据集上的消融试验结果

Table 1 Results of the ablation experiments on the URPC2018 dataset

模型	精准率(%)	召回率(%)	平均精确率(%)	检测速度(帧,1 s)
YOLOv8m	84.3	77.1	83.4	118.4
YOLOv8m+BRA	85.3	78.4	84.8	110.2
YOLOv8m+AFF	84.9	78.7	84.6	114.6
YOLOv8m+NWD-CIOU	84.5	79.1	85.2	118.4
YOLOv8m+BRA+AFF	86.4	79.6	85.5	105.7
YOLOv8m+BRA+NWD-CIOU	86.8	80.7	86.1	110.2
YOLOv8m+AFF+NWD-CIOU	86.3	81.0	86.2	110.2
YOLOv8m+BRA+AFF+NWD-CIOU	87.1	81.5	86.9	105.7

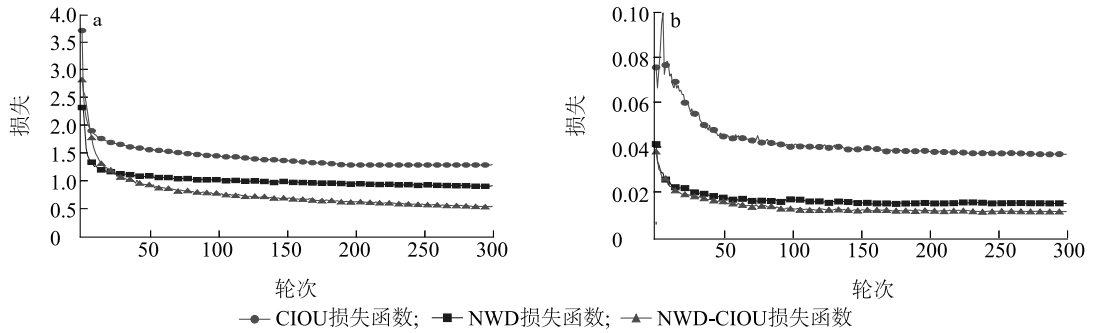
为了更直观地展示模型各模块对检测效果的影响,通过逐一添加 BRA 自注意力模块、AFF 模块和 NWD-CIOU 损失函数来进行试验。在 URPC2018 和 Brackish 两个数据集中分别随机抽取一张图片,并生

成了相关的热力图,结果如图 7 和图 8 所示。可以看出,使用 YOLOv8m 基础模型时,热力图上检测到的目标数量较少,且覆盖的区域较小,表明有些目标没有被模型准确识别。这一结果说明 YOLOv8m 骨干网络对于水下复杂场景下的特征提取能力不足。加入 BRA 模块后,热力图中检测到的目标区域扩大且更集中在目标周围,同时检测到的目标数量增加,这表明 BRA 通过其细粒度的自注意力机制,建立了远程的上下文连接,让网络更容易关注到目标的最显著特征。在增加 BRA 模块的基础上增加 AFF 模块后热力图中检测到的目标数量增多,且检测到的区域更集中于实际目标上,但是仍然存在较小目标没有被关注。这说明 AFF 模块通过自适应特征融合增强了对不同尺度目标的识别能力,且增加的浅层特征层包含了更多的位置信息和语义信息,扩大了检测范围,从而对不同尺度目标识别更为精准。在增加 BRA 模块和 AFF 模块基础上再使用 NWD-CIOU 损失函数后,热力图中检测到的目标数量继续增加,且检测到的区域更加精确地集中在目标的中心位置。这表明 NWD-CIOU 损失函数在小目标的定位上更为精确,提高了小目标标签分配的准确性,从而提高了模型对小目标的检测精度。

表 2 在 Brackish 数据集上的消融试验结果

Table 2 Results of the ablation experiments on the Brackish dataset

模型	精准率(%)	召回率(%)	平均精确率(%)	检测速度(帧,1 s)
YOLOv8m	94.2	93.5	95.3	114.5
YOLOv8m+BRA	95.3	94.9	96.8	104.4
YOLOv8m+AFF	95.4	94.4	96.5	110.2
YOLOv8m+NWD-CIOU	95.9	94.7	96.8	114.5
YOLOv8m+BRA+AFF	96.6	95.9	97.6	100.4
YOLOv8m+BRA+NWD-CIOU	96.5	95.7	97.4	104.4
YOLOv8m+AFF+NWD-CIOU	96.9	96.3	98.1	110.2
YOLOv8m+BRA+AFF+NWD-CIOU	97.2	97.4	98.6	100.4



a:URPC2018 数据集;b:Brackish 数据集。

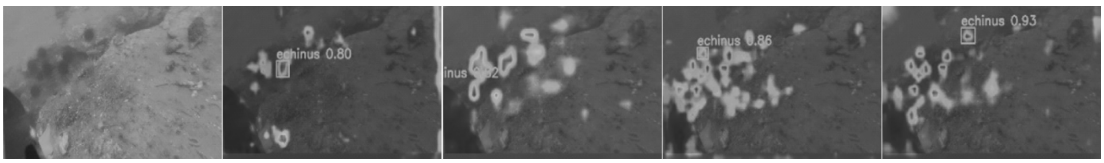
图 6 不同损失函数的训练损失曲线

Fig.6 Training loss curves of different loss functions

2.2 对比试验结果

为了客观评估本研究方法的性能,应用 YOLOv8-BAN 模型和一些经典的目标检测模型在 URPC2018 和 Brackish 两个数据集上进行定量和定性对比试验,对比的经典目标检测模型有 Faster R-CNN[4]、YOLO V5s[24]、YOLOX[25]、ViTDet[26] 和 YOLOv7[27] 等模型。

定量对比试验结果如表 3、表 4 所示。Faster R-CNN 模型由于采用了两阶段检测设计,导致其检测速度较慢;同时,由于未能利用多尺度特征层进行特征融合,其检测平均精度也最低,分别只有 73.5% 和 84.4%。YOLOv5s 模型虽然因其轻量化设计,在两个数据集上的检测速度最高,分别为 1 s 124.6 帧和 118.6 帧,但网络深度和复杂度的不足限制了其特征提取能力,因此其平均检测精度与 YOLOv8-BAN 模型相比分别低 5.8 个百分点和 5.2 个百分点。YOLOX 模型的平均检测精度与 YOLOv8-BAN 模型相

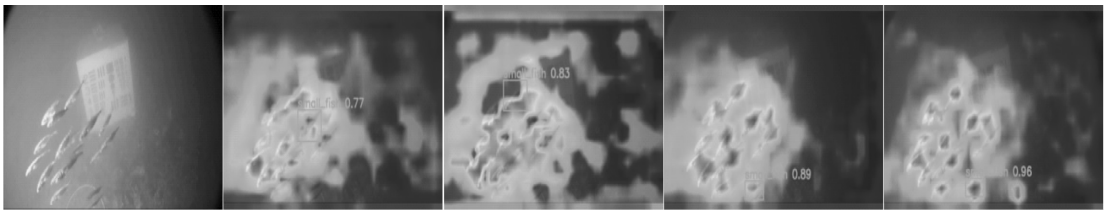


a b c d e

a:原始图片;b:使用 YOLOv8m 模型处理的图片;c:YOLOv8m 模型添加 BRA 模块后处理的图片;d:添加 BRA 模块后的 YOLOv8m 模型进一步加入 AFF 模块后处理的图片;e:YOLOv8m 模型添加 BRA 模块、AFF 模块、NWD-CIOU 损失函数后处理的图片。

图 7 URPC 2018 测试集中部分图片热力图

Fig.7 The heat map of some images in URPC 2018 test set



a b c d e

a:原始图片;b:使用 YOLOv8m 模型处理的图片;c:YOLOv8m 模型添加 BRA 模块后处理的图片;d:添加 BRA 模块后的 YOLOv8m 模型进一步加入 AFF 模块后处理的图片;e:YOLOv8m 模型添加 BRA 模块、AFF 模块、NWD-CIOU 损失函数后处理的图片。

图 8 Brackish 测试集中部分图片热力图

Fig.8 The heat map of some images in Brackish test set

比也有 3.7 个百分点和 3.0 个百分点的差距;检测速度虽然高于 Faster R-CNN 模型和 ViTDet 模型,但实时性表现一般。ViTDet 模型基于 ViT 模型进行了改进,增强了网络特征提取能力,但在小目标的标签分配上仍有误差,平均检测精度比 YOLOv8-BAN 模型低 1.7 个百分点和 1.0 个百分点;同时因为模型参数量和计算量较大,导致其检测速度较慢,难以满足实时性要求。YOLOv7 模型在检测精度和实时性方面表现尚可,平均检测精度分别为 82.6%和 94.6%,但 YOLOv7 模型使用 CIOU 边界框定位损失函数,导致其对小目标检测效果一般。YOLOv8-BAN 模型与另外 6 个模型相比平均检测精度最高,分别达到 86.9%和 98.6%,相比 YOLOv8m 模型分别提升 3.5 个百分点和 3.3 个百分点。这一显著提升归功于本研究提出的 3 个改进模块,其中 BRA 自注意力机制增强了其网络特征提取能力,让网络更加关注目标的最显著区域;AFF 模块使用自适应特征融合的方式,减少了不同特征层融合产生的矛盾信息,提高了融合效果,让网络对不同尺度目标检测精度提升;NWD-CIOU 损失函数提高了小目标在底层标签分配过程中的准确性,让小目标瞄框可以分配到更多的正样本目标,从而提高对小目标的检测精度。

表 3 不同模型在 URPC2018 数据集上的测试结果

Table 3 Experimental results of different models on the URPC2018 test set

模型	检测精准率(%)	召回率(%)	平均检测精度(%)	检测速度(帧,1 s)
Faster R-CNN	75.3	71.4	73.5	25.9
YOLOv5s	83.9	78.7	81.1	124.6
YOLOX	83.7	79.5	83.2	70.1
ViTDet	86.4	81.4	85.2	34.5
YOLOv7	84.8	79.4	82.6	94.6
YOLOv8m	84.3	77.1	83.4	118.4
YOLOv8-BAN	87.1	81.5	86.9	105.7

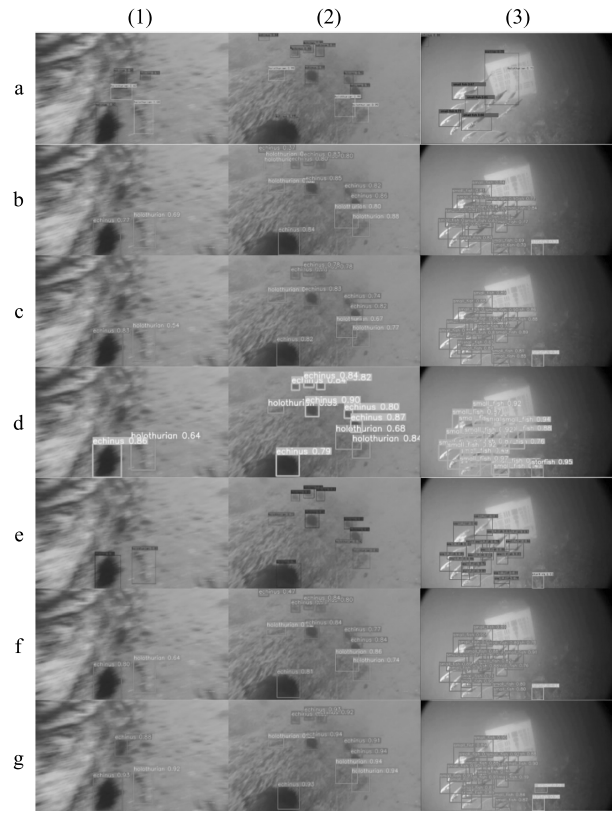
表 4 不同模型在 Brackish 数据集上的测试结果

Table 4 Experimental results of different models on the Brackish test set

模型	检测精准率(%)	召回率(%)	平均检测精度(%)	检测速度(帧,1 s)
Faster R-CNN	82.3	83.1	84.4	24.4
YOLOv5s	92.1	92.7	93.4	118.6
YOLOX	94.5	94.7	95.6	67.6
ViTDet	96.8	96.9	97.6	26.7
YOLOv7	94.4	93.3	94.6	90.2
YOLOv8m	94.2	93.5	95.3	114.5
YOLOv8-BAN	97.2	97.4	98.6	100.4

为了定性评价和比较不同检测模型在实际水下环境中对水生生物的检测效果,本研究选择了 3 张具有代表性的水下图片进行测试。在 URPC2018 数据集中选取了 2 张图片,即图 9(1)和图 9(2);在 Brackish 数据集中选取了 1 张图片,即图 9(3)。其中图 9(1)受背景干扰,目标特征不明显;图 9(2)包含不同尺度的目标;图 9(3)则存在密集小目标。测试结果显示,YOLOv8-BAN 模型检测到了图 9(1)中所有目标,并且置信度超过另外 6 种模型,其他模型则产生了漏检或者误检。这一结果突出了 YOLOv8-BAN 模型在特征提取方面的超强能力,尤其是加入的 BRA 自注意力机制能够有效建立远程特征之间的联结,使网络更加关注目标的关键特征。YOLOv8-BAN 模型也能够检测到图 9(2)中所有目标,而且置信度比其他模型都高,这进一步验证了 AFF 模块的有效性,该模块能够在多尺度特征融合过程中增加目标的特征信息,并过滤掉携带矛盾信息的特征,从而解决学习目标不一致的问题,提高了不同尺度目标的检测效果。针对图 9(3),YOLOv8-BAN 模型检测到了所有的小目标,其他 6 种模型都产生了漏检的现象。这一结果再次验证了 NWD-CIOU 损失函数在提高小目标检测精度方面的作用,该损失函数提高了小目标标签分配的准确性,可以对小目标进行精准定位。

综上所述,经过定量和定性的对比试验,充分验证了本研究提出的模型对于水下生物目标检测任务的适用性,对于水下密集目标和小目标都具有良好的检测效果。



a:Faster R-CNN 模型检测结果;b:YOLOv5s 模型检测结果;c:ViTDet 模型检测结果;d:YOLOv7 模型检测结果;e:YOLOX 模型检测结果;f:YOLOv8m 模型检测结果;g:YOLOv8-BAN 模型检测结果。

图 9 不同模型对 3 张水下图片的检测结果

Fig.9 Detection results of three underwater images by different models

3 结 论

水下生物目标检测技术赋予水下机器人精确捕捞的能力,并辅助进行水生生物的统计监测,为水产养殖的智能化转型提供了坚实的技术支持。为了应对水下图像质量差和小目标生物聚集所带来的检测精度低的挑战,本研究通过改进 YOLOv8m 模型,获得 YOLOv8-BAN 模型。该模型首先在骨干网络中集成了 BRA 自注意力机制,以捕获目标间的长距离特征关联,使网络更加聚焦于目标的关键特征;其次,通过在颈部网络中结合 AFF 模块进行自适应特征融合,有效降低了特征融合过程中的矛盾信息,提升了对不同尺寸目标的检测效果;此外,本研究将 NWD 和 CIOU 两种边界框距离度量标准相结合,设计了 NWD-CIOU 损失函数,完成了对小目标更精准的标签分配。在 URPC2018 和 Brackish 两个水下公共数据集上的测试结果表明,YOLOv8-BAN 模型取得了良好的检测效果,能够为水下生物目标检测的自动化和智能化提供技术支撑。

参考文献:

[1] FAYAZ S, PARAH S A, QURESHI G J, et al. Underwater object detection:architectures and algorithms-a comprehensive review[J]. Multimedia Tools and Applications,2022,81(1):20871-20916.

[2] 许裕良,杜江辉,雷泽宇,等. 水下机器人在渔业中的应用现状与关键技术综述[J]. 机器人,2023,45(1):110-128.

[3] XU S B, ZHANG M H, SONG W, et al. A systematic review and analysis of deep learning-based underwater object detection[J]. Neurocomputing,2023,527:204-232.

[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] 袁红春,张 硕. 基于 Faster R-CNN 和图像增强的水下鱼类目标检测方法[J]. 大连海洋大学学报,2020,35(4):612-619.

[6] LIU J, LIU S, XU S J, et al. Two-stage underwater object detection network using swin transformer[J]. IEEE Access,2022,10:117235-117247.

[7] LIN W H, ZHONG J X, LIU S, et al. Roimix: proposal-fusion among multiple images for underwater object detection[C]. Barcelona:ICASSP,2020.

[8] SHI P, XU X, NI J, et al. Underwater biological detection algorithm based on improved faster-RCNN[J]. Water,2021,13(17):2420.

[9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once:unified, real-time object detection[C]. Las Vegas: IEEE, 2016.

[10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]. Honolulu:IEEE,2017.

[11] REDMON J, FARHADI A. Yolov3:an incremental improvement[C]. Salt Lake City:CVPR,2018.

[12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M, et al. Yolov4: optimal speed and accuracy of object detection[C]. Seattle: CVPR,2020.

[13] GUO T, WEI Y, SHAO H, et al. Research on underwater target detection method based on improved MSRCP and YOLOv3[C]. Nashville:IEEE,2021.

[14] CHEN L Y, ZHENG M C, DUAN S Q, et al. Underwater target recognition based on improved YOLOv4 neural network[J]. Electronics,2021,10(14):1634.

[15] LEI F, TANG F, LI S. Underwater target detection algorithm based on improved YOLOv5[J]. Journal of Marine Science and Engineering,2022,10(3):310.

[16] 翟先一,魏鸿磊,韩美奇,等. 基于改进 YOLO 卷积神经网络的水下海参检测[J]. 江苏农业学报,2023,39(7):1543-1553.

[17] SUN Y, ZHENG W X, DU X, et al. Underwater small target detection based on YOLOX combined with mobileViT and double coordinate attention[J]. Journal of Marine Science and Engineering, 2023,11(6):1178.

[18] YI W G, WANG B. Research on underwater small target detection algorithm based on improved YOLOv7[J]. IEEE Access, 2023, 11:66818-66827.

[19] ZHU L, WANG X, KE Z, et al. BiFormer:vision transformer with Bi-level routing attention[C]. Vancouver:IEEE,2023.

[20] REN S, ZHOU D, HE S, et al. Shunted self-attention via multi-scale token aggregation[C]. New Orleans:IEEE,2022.

[21] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss:faster and better learning for bounding box regression[C]. New York:AAAI, 2020.

[22] XU C, WANG J W, YANG W, et al. Detecting tiny objects in aerial images:a normalized Wasserstein distance and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022,190:79-93.

[23] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[C]. Sydney:ICML,2017.

[24] WANG D D, HE D J. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning[J]. Biosystems Engineering,2021,210:271-281.

[25] HE Q, XU A, YE Z, et al. Object detection based on lightweight YOLOX for autonomous driving[J]. Sensors,2023,23(17):7596.

[26] LI Y H, MAO H Z, GIRSHICK R, et al. Exploring plain vision transformer backbones for object detection[C]. Tel Aviv:ECCV, 2022.

[27] 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]. Vancouver:IEEE,2023.

(责任编辑:黄克玲)