

DOI: 10.13382/j.jemi.B2407841

# 改进 YOLOv8n 的多尺度轻量化水下目标检测\*

苗力恒 田 莹

(辽宁科技大学计算机与软件工程学院 鞍山 114000)

**摘 要:**由于水下检测仪器的存储和计算资源有限,使用高参数量的模型会占用更多的存储空间和计算能力。为了解决这一问题,提出了一种基于 YOLOv8n 改进的目标检测模型—YOLOv8n-MAL,该模型在保持检测精度的同时,实现了整体的轻量化。首先,提出了一种轻量型多尺度注意力机制(MSCBAM),通过不同大小的卷积核和池化操作,能够从输入特征图中提取不同尺度的特征,可以提升模型在复杂场景中的鲁棒性,确保模型在面对不同类型的输入时都能维持较高的检测精度。然后,设计新的颈部模型 MSFFN,对比原颈部模型增强了模型多尺度融合的能力,加强了不同层级特征的交互,使得高层次和低层次特征能够更加充分地结合,这种跨层级融合能够更有效地利用网络的特征表示能力,避免信息在传递过程中的丢失或弱化,进而提升模型的检测效果。其次,提出了轻量多尺度卷积模块(LMSCM),并将该模块和部分卷积融入到 C2F 模块中构成 PC2F-LMS,通过引入更高效的卷积结构和轻量化设计,增强了特征提取和表达的能力。最后,使用 WIoU 优化原网络损失函数。实验结果表明,改进后的算法在 URPC 数据集上的平均精度均值(mAP) mAP@ 0.5 提高了 1.4%,与 YOLOv8n 算法相比参数量下降了 38.6%,为水下目标检测提供了有效的参考价值。

**关键词:**水下目标检测;YOLOv8n;轻量化;多尺度特征融合;计算机视觉

**中图分类号:** TN911.73;TP391      **文献标识码:** A      **国家标准学科分类代码:** 520.2040

## Improved YOLOv8n multi-scale and lightweight underwater target detection

Miao Liheng Tian Ying

(School of Computer and Software Engineering, University of Science and Technology Liaoning, Anshan 114000, China)

**Abstract:** Due to the limited storage and computing resources of underwater detection instruments, using high-parameter models will occupy more storage space and computing power. To solve this problem, a target detection model based on YOLOv8n, called YOLOv8n-MAL, was proposed, which achieves lightweight overall while maintaining detection accuracy. Firstly, a lightweight multi-scale convolutional block attention module (MSCBAM) is proposed, which can extract features of different scales from input feature graphs through convolution kernel and pooling operations of different sizes. It can improve the robustness of the model in complex scenarios and ensure that the model can maintain high detection accuracy in the face of different types of inputs. Then, a new neck model multi-scale feature fusion network (MSFFN) is designed. Compared with the original neck model, the multi-scale fusion capability of the model is enhanced, and the interaction of features at different levels is strengthened, so that the high-level and low-level features can be more fully combined. This cross-level fusion can make use of the feature representation ability of the network more effectively, avoid the loss or weakening of information in the process of transmission, and improve the detection effect of the model. Secondly, a lightweight multi-scale convolution module lightweight multi-scale convolution module (LMSCM) is proposed, and the module and some convolution modules are integrated into the C2F module to form PC2F-LMS. By introducing a more efficient convolution structure. Finally, the original network loss function is optimized using WIoU. The experimental results show that the average accuracy mAP@ 0.5 of the improved algorithm on the URPC dataset is increased by 1.4%, and the number of parameters is reduced by 38.6% compared with the YOLOv8n algorithm, which provides an effective reference value for underwater target detection.

**Keywords:** underwater target detection; YOLOv8n; lightweight; multi-scale feature fusion; computer vision

## 0 引言

地球超过 70% 的表面被水覆盖,海洋在地球生态系统中扮演着至关重要的角色。相比于陆地,海洋中蕴藏着大量尚未开发的资源,对海洋的研究具有深远意义。同时随着计算机视觉技术的进步,越来越多的研究人员开始利用该技术探索海洋,这已成为一种新的研究途径<sup>[1-2]</sup>。水下图像通常具有较高的分辨率并包含丰富的信息,这使得它们在近距离水下目标检测任务中尤为有利<sup>[3]</sup>。在计算机视觉的推动下,目标检测技术已逐渐成为海洋勘探领域的热点之一,并取得了显著进展,海洋检测仪器是国家整体海洋技术的体现,它的影响辐射到了海洋的经济发展,也为海洋资源开发做出了重要贡献<sup>[4-5]</sup>。

在水下目标检测领域仍然面临诸多挑战。首先,水下环境的复杂性,如不均匀的光照、低对比度、模糊和亮点等,这为水下目标检测带来了巨大的挑战<sup>[6]</sup>。其次,在水下生物通常体积较小且分布密集,导致目标的重叠和遮挡问题,尤其在小目标检测上,难度更大。最后,水下嵌入式仪器通常具有有限的计算和存储能力,这使得在这些仪器上部署大型模型变得困难<sup>[7]</sup>。寻找解决方案允许在复杂的水下环境中准确和快速地检测目标是一个迫切需要解决的问题。

基于深度学习的目标检测算法根据算法的流程主要分为双阶段算法以及单阶段算法两类。双阶段目标检测的代表性算法有 R-CNN<sup>[8-9]</sup>等。双阶段模型的算法流程主要是先选择出若干候选区域,然后对候选区域进行进一步的分类。单阶段目标检测算法的代表性算法有 YOLO 系列<sup>[10-14]</sup>和 RetinaNet<sup>[15]</sup>等。单阶段模型的算法流程主要是在图像上直接进行目标的分类以及调整边界框,大大减少了计算量,能够在保证一定的准确率的前提下大幅度的提高检测速度,更适合实际的应用。

许多学者对水下海洋目标检测进行大量研究。周新等<sup>[16]</sup>引入 GhostNet 网络和构建 C3CA 模块提高模型的特征提取和融合能力。引入 GhostNet 网络既有效地降低了参数量,但降低了模型的精确度,而 C3CA 模块的引入则弥补了 GhostNet 网络损失的精度。李卓润等<sup>[17]</sup>采用 MobileNetV3 作为新的特征提取网络进行轻量化处理,通过逐点卷积操作,在保证特征图通道和空间信息都完整的情况下,有效地降低参数量。杨婷等<sup>[18]</sup>基于 Faster-RCNN 提出的双向特征融合(PAFPN),通过在 FPN 的基础上添加了一个自底向上的传输路径,以增加底层信息的利用率,从而提高了准确率。涂万等<sup>[19]</sup>通过颈部网络

引入 ECAM 注意力机制,能够加强对模糊的水下生物的认识能力,同时添加检测头用于提高精度,但整体参数量也随之增长。

上述研究在提升精度或降低参数量方面取得了一定进展,但往往难以兼顾两者。基于此提出了一种多尺度轻量化水下目标检测模型 YOLOv8n-MAL。首先,在 YOLOv8n 的主干网络中引入了卷积块注意力模块(CBAM)与多尺度模块相结合的注意力机制,能够有效提升不同尺度目标的检测性能。其次,通过重构颈部网络,增强了模型对浅层信息的捕捉能力,从而提高了对细小和遮挡目标的检测精度。最后,优化了 C2F 模块内部结构,可以进一步减少参数量,并增强模型的表达能力。最后,采用了 WIoU 损失函数,解决了重叠目标检测困难的问题。实验结果表明,提出的模型在显著降低参数量的同时,依然保持了较高的精确度,实现了精度与轻量化的良好平衡。

## 1 YOLOv8 网络模型

YOLOv8 模型由 3 个主要部分组成:主干、颈部和头部。较大的模型通过混合上采样和数据增强技术能够丰富数据集,提升模型的泛化能力和鲁棒性。骨干网络负责从图像中提取信息,并将其提供给颈部和头部。包括多个 Conv、C2F 模块和最后的 SPPF 模块。颈部部分主要是利用主干网络提取的特征来促进特征融合,并采用了 FPN<sup>[20]</sup>+PAN<sup>[21]</sup>结构,增强了跨多个尺度的语义表达和定位能力。

为了满足不同的需求,YOLOv8 根据网络的深度和宽度被划分为不同的版本,即 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x。与以往 YOLO 版本相比,YOLOv8 在 COCO 数据集上显示出了优越的性能。本文考虑到 YOLOv8n 是 YOLO 系列最轻量的版本,能够在较低的计算成本下保持较好的检测精度,特别适用于水下目标检测任务,故选择 YOLOv8n 作为研究对象。

## 2 改进 YOLOv8n 检测模型

为了提高水下目标检测仪器的检测效率,复杂的模型通常需要消耗更多的计算资源,导致仪器能耗的增加。提出了一种改进 YOLOv8n 的目标检测模型—YOLOv8n-MAL,结构如图 1 所示。首先,提出一种基于 CBAM 改进的轻量化多尺度的注意力机制(multi-scale convolutional block attention module, MSCBAM)。将 MSCBAM 融入到主干网络后,不仅提高检测精度和鲁棒性,还进一步增强

模型的特征提取和表达能力。其次,对颈部网络进行重构,形成多尺度特征融合网络(multi-scale feature fusion network, MSFFN),通过多尺度特征的高效融合,增强特征信息的提取。然后,设计了轻量化多尺度卷积模块(lightweight multi-scale convolution module, LMSCM)替换 C2F 中 Bottleneck 的第 2 个卷积,并引入部分卷积以提高

精度,最终构成 PC2F-LMS(lightweight multi-scale fusion of PConv-based C2F)。最后,采用 WIoU 替代原有的 CIoU 损失函数,该改进使模型在重叠目标检测任务中表现更为出色,优化了训练过程中的收敛速度。整体来看,改进后的模型在确保检测精度,显著减少了参数量,为水下检测仪器的稳定运行提供了更加可靠的支持。

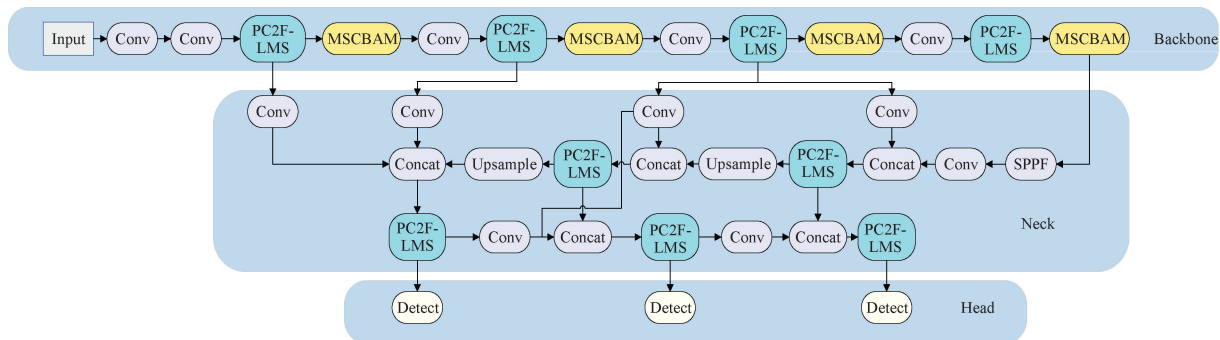


图 1 YOLOv8n-MAL 模型

Fig. 1 YOLOv8n-MAL model

## 2.1 轻量型多尺度注意力机制

在水下目标检测任务中,复杂环境以及生物间的重叠增加了任务难度。尽管高精度的检测模型能够提升检测效果,但往往伴随着较大的参数量。这会导致计算速度减慢、处理时间延长,进而增加检测延迟,为水下检测仪器带来额外负担。吴兴辉等<sup>[22]</sup>通过在主干网络添加 CBAM 注意力机制,捕捉特征信息,增强卷积神经网络的表达能力,使模型可以有效地降低参数量并关注到重要信息。

CBAM<sup>[23]</sup>注意力机制由两个顺序的子模块组成:通道注意力模块和空间注意力模块。首先输入特征图经过通道注意力模块得到  $X'$ ,随后经过空间注意力模块得到  $X''$ ,组成了 CBAM 的基本框架如式(1)~(4)所示。

$$X' = M_c(X) \cdot X \quad (1)$$

$$X'' = M_s(X') \cdot X' \quad (2)$$

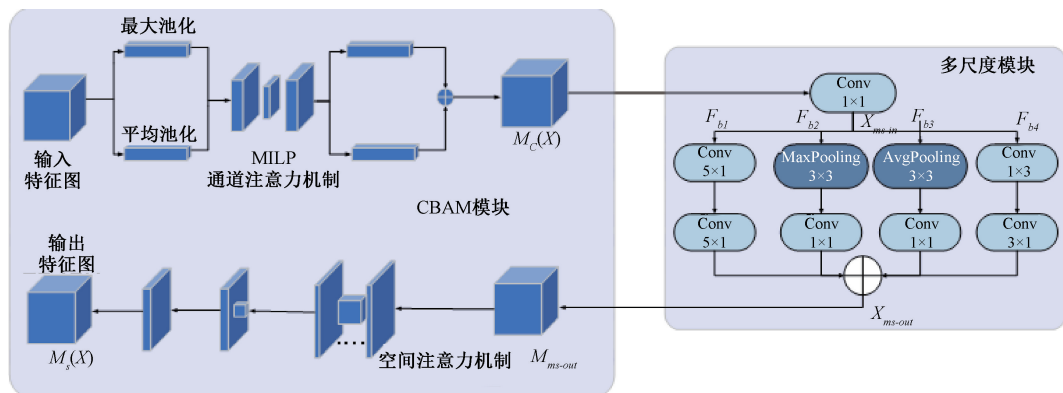


图 2 轻量型多尺度注意力机制

Fig. 2 MSCBAM

$$M_c(X) = \sigma(MLP(F_{avg}) + MLP(F_{max})) \quad (3)$$

$$M_s(X') = \sigma(f^7(F_{cat})) \quad (4)$$

式中:  $M_c(X)$  是生成的通道注意力权重;  $M_s(X')$  是生成的空间注意力权重。

CBAM 虽然能够有效地降低参数量,但采用的是单一尺度的卷积操作来捕捉空间注意力。这限制了其仅能在固定的感受野下进行特征提取,导致在处理需要捕捉多尺度信息的复杂场景表现不足,从而导致水下目标检测的效果不佳。提出了一种基于 CBAM 改进的轻量化 MSCBAM。有效地捕捉多尺度信息,增强了模型在不同目标大小和复杂场景中的表现,同时提升了其对多样化输入的鲁棒性,减少了通过增加网络深度来提升性能的需求,从而进一步减少了参数量。MSCBAM 注意力机制如图 2 所示。

$M_c(X)$  之后经过一个  $1 \times 1$  的卷积, 相当于在空间维度上扩展全连接层, 允许跨通道信息的交互和融合。然后将通道数分成 4 个分支得到  $P_{1,2,3,4} \in R^{w \times h \times \frac{c}{4}}$ , 其中两个分支利用两个维度为  $1 \times 3$  和  $3 \times 1$  以及  $1 \times 5$  和  $5 \times 1$  的顺序卷积操作, 横向和纵向的卷积有助于捕捉到细粒度的局部特征和更大范围的特征, 更大的感受野的卷积核能够更好地处理全局信息。 $P_2 \in R^{w \times h \times \frac{c}{4}}$  使用  $3 \times 3$  最大池化来捕捉局部的最强特征响应, 有助于突出重要的边缘和角点信息。最大池化后,  $1 \times 1$  卷积重新调整通道数, 并整合强响应特征, 为最终的特征融合提供更多的判别信息。 $P_3 \in R^{w \times h \times \frac{c}{4}}$  使用平均池化, 平均池化能够聚合小尺度上的局部信息, 减少噪声影响, 从而得到更加平滑的特征。平均池化后, 通过  $1 \times 1$  卷积进一步调整和组合通道信息, 提取重要的局部特征信息, 并压缩输出的通道数。

MSCBAM 中的多尺度模块通过捕捉和融合不同尺度的特征信息, 增强了模型对多尺度特征的表达能力, 提高了模型对复杂场景和多样化目标的适应能力。这使得 MSCBAM 在处理具有不同尺度和复杂背景的图像时表现更加出色。多尺度模块公式如式 (5) ~ (11) 所示。

$$X_{ms-in} = Conv_{1 \times 1}(X_c) \quad (5)$$

$$F_{b1} = Conv_{5 \times 1}(Conv_{1 \times 5}(X_{ms-in}/4)) \quad (6)$$

$$F_{b2} = Conv_{1 \times 1}(MaxPool_{3 \times 3}(X_{ms-in}/4)) \quad (7)$$

$$F_{b3} = Conv_{1 \times 1}(AvgPool_{3 \times 3}(X_{ms-in}/4)) \quad (8)$$

$$F_{b4} = Conv_{3 \times 1}(Conv_{1 \times 3}(X_{ms-in}/4)) \quad (9)$$

$$X_{ms-out} = (Concat[F_{b1}, F_{b2}, F_{b3}, F_{b4}]) \quad (10)$$

$$X'' = M_s(X_{ms-out}) \cdot X_{ms-out} \quad (11)$$

式中:  $X_{ms-in}$  是  $X_c$  经过  $1 \times 1$  卷积的输出;  $F_{b1}, F_{b2}, F_{b3}, F_{b4}$  分别是 4 个分支的输出;  $X_{ms-out}$  是由 4 个分支通过 Concat 操作所得到的;  $X''$  是  $X_{ms-out}$  经过空间注意力的输出。

## 2.2 重构颈部结构

YOLOv8n 颈部网络模型如图 3 所示, 采用 FPN+PAN 结构进行特征融合。FPN 通过自上而下的路径和横向连接, 将不同尺度的特征融合, 使得高层次的信息能够传递到低分辨率的特征图中。而 PAN 采用自下而上的路径, 将浅层特征逐级传递到高层, 并通过横向连接实现特征融合, 从而有效提高小目标的检测能力。图 3 中输出 Q3 然后上采样与 P4 融合生成 Q2, 同理 Q2 与 P3 融合生成 Q1, 通过进一步操作得到 R1, R1 在自下而上的路径与 Q2 融合, 得到 R2, 同理得到 R3。

为了提高模型的性能, 对原有的颈部模型进行了重构, 提出了一种新的多尺度特征融合网络结构 MSFFN, 如图 4 所示。与原有颈部模型相比, MSFFN 通

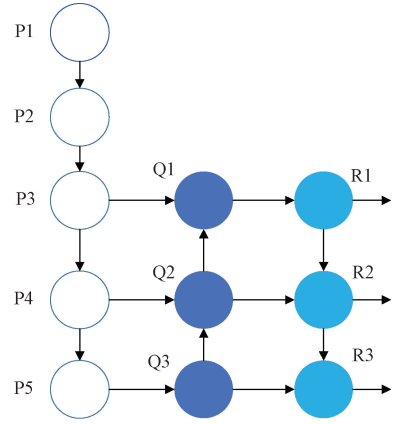


图 3 原颈部模型

Fig. 3 Original neck model

过引入浅层特征, 增强了模型对小目标的捕捉能力。此外, MSFFN 增加了卷积层的数量, 并在各特征层之间插入了额外的卷积操作, 以强化特征提取。具体而言, 特征层 C3 和 C4 在同一层上相连, P4 和 P5 通过融合 C4 和 C5 的特征生成了新的特征层 Q3, 随后 Q3 与 C3 进行融合, 生成 Q2。Q1 则通过融合 P2、P3 与 C1、C2 的特征生成。接下来, R1 沿自上而下的路径与 Q2 及 C3 融合, 生成 R2, 随后 R2 与 Q3 融合得到 R3。最终, 颈部网络输出 3 个尺度的特征图, R1 ( $80 \times 80$ )、R2 ( $40 \times 40$ )、R3 ( $20 \times 20$ )。

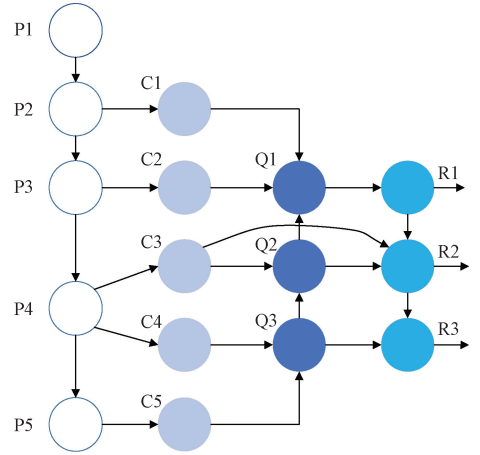


图 4 MSFFN 模型

Fig. 4 MSFFN model

改进后的颈部网络结构极大提升了特征提取与融合的灵活性, 特别是在处理多尺度和复杂背景的小目标时, 显著增强了模型的检测能力, 同时有效减少了小目标检测丢失的问题。此外, MSFFN 通过多层次特征融合, 强化了跨尺度的信息交互, 使不同尺度的目标特征得以充分表达, 增强了复杂场景下的检测精度和鲁棒性。

MSFFN 的设计不仅减少了参数量,还提高了模型的特征利用效率,尤其适用于计算资源受限的水下检测环境<sup>[24]</sup>。

### 2.3 PC2F-LMS 模块

C2F 模块在 YOLOv8n 模型中的作用是通过改进特征融合,增强模型在复杂背景或细节丰富目标下的检测能力。然而,由于水下仪器的存储和计算能力有限,检测模型需要尽可能地轻量化,以便在有限的硬件资源下依然保持高效的目标检测性能。所以提出了 LMSCM,如图 5 所示。

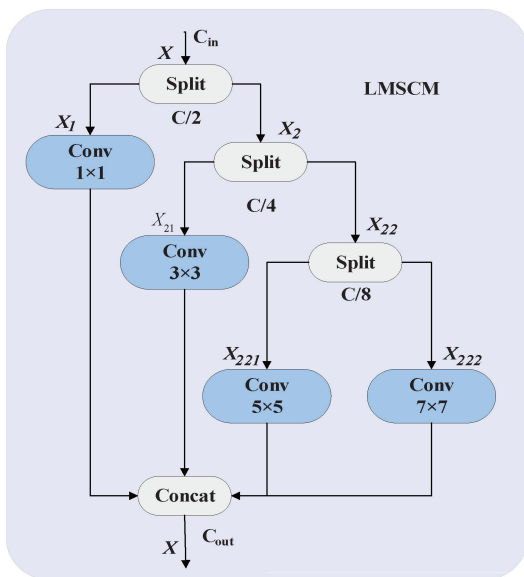


图 5 LMSCM 模块

Fig. 5 LMSCM model

LMSCM 的整体结构类似 SPPELAN<sup>[25]</sup> 的金字塔结构。假设  $C$  是输入的通道数;  $X$  为输入特征图; 在这个模块中  $C$  被分成了 3 部分  $C_1 \in X^{w \times h \times \frac{C}{2}}$ 、 $C_2 \in X^{w \times h \times \frac{C}{4}}$ 、 $C_3 \in X^{w \times h \times \frac{C}{8}}$ , 同时特征图也被分割 3 次, 第 1 次分割成  $X_1$ 、 $X_2$ 、 $X_3$  输入到  $1 \times 1$  卷积中提取局部特征, 而  $X_2$  继续被分割成  $X_{21}$  和  $X_{22}$ , 其中  $X_{21}$  输入到  $3 \times 3$  卷积中提取中等尺度的特征, 它捕捉了较大的上下文信息, 同时保持了良好的局部细节,  $X_{22}$  继续分割成  $X_{221}$  和  $X_{222}$  分别输入  $5 \times 5$  和  $7 \times 7$  卷积, 通过上述卷积可以提取更大尺度的特征, 这些特征适合捕捉大型或模糊的水下目标。这种金字塔结构通过多尺度卷积操作, 从输入特征中同时提取不同尺度的信息, 显著增强对多样化水下目标的识别能力。即便涉及多个卷积操作, 得益于并行计算, 该结构有效减少了计算冗余, 提高整体计算效率, 更加适应水下复杂环境的多样性挑战。

因此对 C2F 模块进行了两项改进以提升水下目标

检测性能。首先, 将 Bottleneck 结构的第 2 个标准卷积层替换为 LMSCM, 构建 LMSCM-Bottleneck 结构。该设计通过多尺度卷积核捕获丰富的空间上下文信息, 在保持参数效率的同时增强了模型对目标尺度变化的适应性, 显著提升了特征提取能力。

其次由于水下环境的复杂性, 被检测目标常常受到遮挡的影响。为此在 C2F 模块中引入了部分卷积 (PConv)<sup>[26]</sup>, PConv 能够有效处理稀疏数据, 通过忽略噪声较大的区域, 避免在无意义的零值上进行无效计算, 从而优化特征提取过程。不仅增强了模型在处理遮挡目标时的鲁棒性, 还显著提高了水下目标检测的整体精度。

改进后的 PC2F-LMS 模块如图 6 所示。结合了 LMSCM-Bottleneck 和 PConv 的优势, 能够更高效地捕捉和融合多尺度特征, 确保每个特征块的输出得到充分整合, 进一步增强了模型的表达能力与特征学习能力。此外, 该模块对水下环境的遮挡问题进行了优化, 在提升检测精度的同时降低了参数量, 使其更适用于存储资源受限的检测设备。

### 2.4 替换损失函数

YOLOv8n 的损失函数是一个多任务损失函数, 主要包括类别损失、置信度损失和边界框损失函数 3 个部分。类别损失通常使用多元交叉熵损失, 置信度损失通常使用二元交叉熵损失, YOLOv8n 通常使用 CIoU。CIoU 公式如式 (12) 所示。

$$CIoU = 1 - IoU + \frac{p^2(b, b^g)}{c^2} + a \cdot v \quad (12)$$

式中:  $IoU$  是预测框与真实框的交并比;  $p^2(b, b^g)$  是预测框的中心点  $b$  与真实框的中心点  $b^g$  之间的欧氏距离;  $c$  是包围预测框和真实框的最小闭包区域的对角线长度;  $a$  是一个衡量参数, 用来调整宽高比对损失的影响;  $v$  是宽高比的一致性度量, 用来衡量预测框和真实框宽高比的差异。

但是 CIoU 对不同目标的差异没有灵活的调节机制, 在一些复杂的环境中, CIoU 不能很好地适应这些变化。所以选择使用 WIoU<sup>[27]</sup> 替代 CIoU。

WIoU 在 IoU 计算时引入了权重机制, 可以针对在水下不同大小的目标进行更加细致的优化, 更好地平衡模型的训练, 确保所有目标都能被良好检测, 而不是因为某些被遮挡的目标而导致模型性能下降。WIoU 公式如式 (13) 所示。

$$WIoU = IoU - \lambda_1 \cdot \frac{p^2(b, b^g)}{c^2} - \lambda_2 \cdot v \quad (13)$$

式中:  $\lambda_1$  和  $\lambda_2$  为权重系数, 用于调节中心点距离和宽高比向的影响。

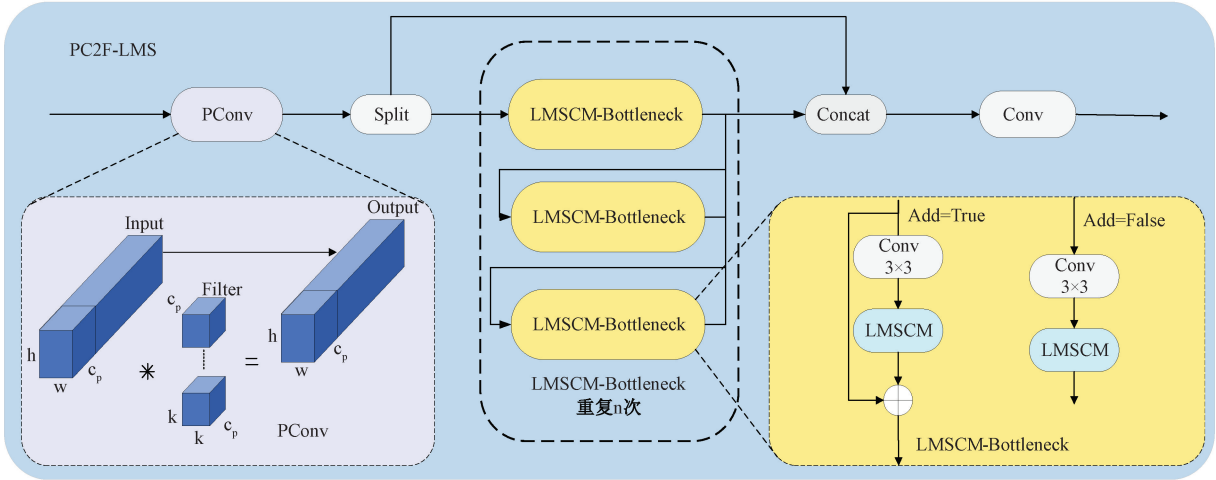


图 6 PC2F-LMS 模块

Fig. 6 PC2F-LMS model

3 实验结果与分析

3.1 实验环境及配置

实验环境如表 1 所示。模型开发使用 Python 语言。在训练过程中。基于 YOLOv8n 作为基线模型进行训练。输入图像尺寸设为 640×640, 优化器采用 SGB 函数。模型训练了 300 个 epoch, Batchsize 为 16, 学习率为 0.01, 并使用余弦退火算法进行调整。

表 1 实验环境

Table 1 Experimental environment

| 主要环境    | 配置                        |
|---------|---------------------------|
| 操作系统    | Win10                     |
| 显卡      | NVIDIA GeForce RTX 4070Ti |
| Python  | 3. 8. 16                  |
| Pytorch | 1. 11. 0                  |

3.2 数据集介绍

数据集选取水下机器人目标抓取大赛 (URPC) 中的数据集, 该数据集是在大连獐子岛海洋牧场实地拍摄, 还原了最真实的水下环境。其中选取了 6 670 张图片作为数据集使用, 图片中包含不同数量和大小的大海参 (holothurian)、海星 (starfish)、海胆 (echinus)、扇贝 (scallop)。为了确保模型的评估的公平性和科学性, 把 6 670 张图片按照 7 : 1 : 2 的比例随机划分, 分成训练集、验证集和测试集, 其中训练集包含 4 469 张图片, 验证集包括 667 张图片, 测试集包括 1 334 张图片。

3.3 评价指标

分析实验结果时采用目标检测常用评估指标, 分别

为精确度 (precision)、召回率 (recall) 和平均精度均值 (mean average precision, mAP)、参数量 (parameters)、浮点计算量, 其计算公式如式 (14) ~ (17) 所示。

$$Precision = \frac{TP}{TP + FP}$$
 (14)

$$Recall = \frac{TP}{TP + FN}$$
 (15)

$$AP = \int_0^1 P(R) dR$$
 (16)

$$mAP = \frac{AP_1 + AP_2 + \cdots + AP_n}{n}$$
 (17)

式中: TP 为真正例, 模型正确的识别目标; FP 为假正例, 模型没有识别出正确目标, 反而识别成其他目标, 但能识别出目标位置; FN 为假反例, 目标存在, 但并未识别出来; n 代表类别数; mAP 就是指多个类别 AP<sub>n</sub> 的平均值。此外, 在此基础上加入参数量和计算量作为评价指标。参数量越小, 说明模型需要的存储的权重数据更少, 占用内存空间更小。计算量低意味着模型在推理时所需的计算量较少。

3.4 实验结果

为验证所提模型的性能, 采用划分后的 URPC 数据集进行实验, 并通过上一节的评估指标进行衡量。将提出的 YOLOv8n-MAL 模型与其他模型进行对比, 实验结果如表 2 所示。

Faster-RCNN 是一种双阶段的目标检测算法, 模型复杂导致训练和推理时间较长, 同时较高的参数量和计算量也不适用于水下检测仪器。

表2 与其他模型对比实验结果

Table 2 Results compared with other models

| 模型          | mAP@ 0.5/%  | mAP@ 0.5 :  |                      | Parameters/ | 浮点数/ |
|-------------|-------------|-------------|----------------------|-------------|------|
|             |             | 0.95/%      | ( ×10 <sup>6</sup> ) | GFLOPs      |      |
| Faster-RCNN | 67.8        | 49.4        | 136.8                | 368.9       |      |
| SSD         | 75.4        | 38.8        | 25.1                 | 40.5        |      |
| YOLOv3-tiny | 78.9        | 55.8        | 12.1                 | 18.9        |      |
| YOLOv5n     | 81.9        | 61.3        | 2.5                  | 7.2         |      |
| YOLOv6n     | 80.1        | 60          | 4.2                  | 11.8        |      |
| YOLOv8n     | 82.1        | 61.8        | 3.0                  | 8.1         |      |
| YOLOv10n    | 82.6        | 62.1        | 2.7                  | 8.2         |      |
| YOLOv11     | 82.2        | 62          | 2.6                  | <b>6.4</b>  |      |
| 文献[ 28 ]    | 83.4        | 62.4        | 2.6                  | 7.3         |      |
| YOLOv8n-MAL | <b>83.5</b> | <b>62.6</b> | <b>1.8</b>           | 7.2         |      |

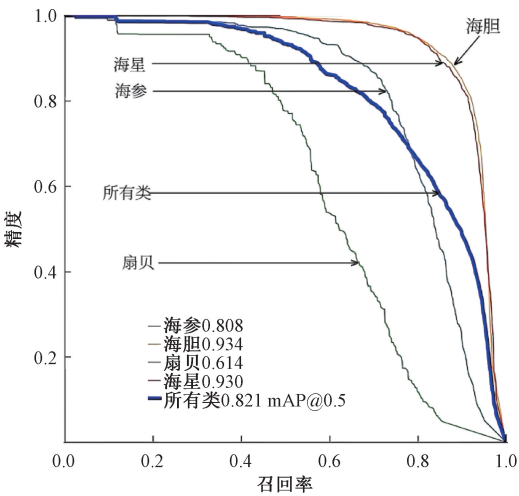
SSD 和 YOLO 系列都属于单阶段的目标检测算法。相较于双阶段检测算法,单阶段目标检测算法的主要优势在于速度快、结构简单、训练和部署更加方便,尤其适合实时应用和资源受限的仪器。但 SSD 对比 YOLO 系列性能还是没有优势。YOLOv8n-MAL 与 YOLOv3-tiny、

YOLOv5n、YOLOv6n 以及最新提出的 YOLOv10n 和 YOLOv11 进行比较,除了计算量低于 YOLOv11,其他性能均超过各模型。

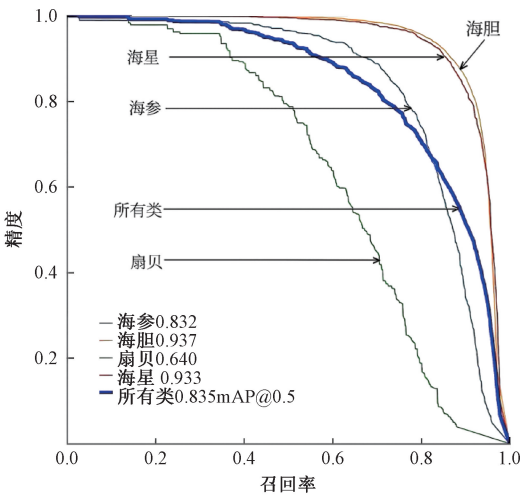
为了更充分地验证改进的有效性,与其他参考文献进行了对比。文献[28]基于 YOLOv7 改进,结果表明改进后的模型在精度和计算量方面具有一定优势,且参数量相比该文献减少了 26%,为存储资源受限的水下检测设备提供了更好的支持。

通过实验对比能够显示出,与目前主流模型相比,YOLOv8n-MAL 精度有一定程度的提高,计算量稍有降幅,参数量大幅下降,更适合水下目标检测仪器的部署和应用。

图 7(a)为 YOLOv8n 原模型的 PR 曲线,图 7(b)为改进后模型的 PR 曲线。由图 7 可见,改进后的 YOLOv8n-MAL 模型每条曲线所围成的面积都比原模型有着或多或少的增加,这说明改进的模型更好的检测能力。



(a) YOLOv8n的PR曲线  
(a) The PR curve of YOLOv8n



(b) YOLOv8n-MAL的PR曲线  
(b) The PR curve of YOLOv8n-MAL

图7 改进前后的PR曲线

Fig. 7 PR curves before and after improvement

3.5 消融实验

为了更直观地展示各个改进模块对模型的影响,对相关创新点进行消融实验,实验结果如表3所示。在原模型主干网络添加了4个MSCBAM注意力机制,多尺度的特征融合增强了捕捉不同大小信息的能力,虽然计算量没有明显变化,但 mAP 小幅度提高,参数量较原模型下降了 23.3%。将原模型的颈部网络模型替换成 MSFFN 后,整体性能均有所提升,MSFFN 通过融合不同层次的信息,增强主干网络的表达能力,并通过主干网络之后的卷积层调整通道数,降低模型参数量。PC2F-LMS

中的轻量化多尺度模块能够有效地融合了不同尺度的特征,能让模型更加适应不同复杂的环境。WiIoU 在计算时引入了权重,在不影响其他性能的情况下,有效地提高了 mAP 值。通过对提出的模块逐个添加并进行实验,表明每个改进的模块对模型的性能都有一定的提升,最后 mAP 提升了 1.4%,参数量下降了 38.6%,说明改进的模型在水下检测上有更大的优势。

3.6 不同颈部结构的对比实验

不同的颈部网络结构,会带来不同的效果。为了验证所设计 MSFFN 的有效性,现通过改变特征融合的方式

表 3 消融实验  
Table 3 Ablation experiment

| MSCBAM | MSFFN | PC2F-LMS | WIoU | mAP@ 0. 5/% | Parameters/( ×10 <sup>6</sup> ) | 浮点数/GFLOPs |
|--------|-------|----------|------|-------------|---------------------------------|------------|
| ✓      | ✓     | ✓        | ✓    | 82. 1       | 3. 0                            | 8. 1       |
|        |       |          |      | 82. 9       | 2. 3                            | 8. 0       |
|        |       |          |      | 82. 9       | 2. 1                            | 7. 0       |
|        |       |          |      | 82. 5       | 2. 8                            | 7. 9       |
| ✓      | ✓     | ✓        | ✓    | 82. 8       | 3. 0                            | 8. 1       |
|        |       |          |      | 83. 0       | 1. 9                            | 7. 3       |
|        |       |          |      | 83. 2       | 1. 8                            | 7. 2       |
|        |       |          |      | 83. 5       | 1. 8                            | 7. 2       |

来进行验证。实验结果如表 4 所示。

表 4 不同颈部网络对比  
Table 4 Comparison of different neck networks

| 模型        | mAP@ 0. 5/% | Parameters/( ×10 <sup>6</sup> ) | 浮点数/GFLOPs |
|-----------|-------------|---------------------------------|------------|
| Version A | 82. 6       | 2. 1                            | 7. 0       |
| Version B | 82. 5       | 2. 2                            | 7. 5       |
| Version C | 82. 7       | 2. 0                            | 7. 1       |
| MSFFN     | 82. 9       | 2. 1                            | 7. 0       |

表 4 中, Version A 为 R3 是由 C4、Q3 和 R2 融合得到; Version B 为删除 C4, P5 通过 C5 将特征输入到 Q3; Version C 为在 P4 和 P5 之间再加个特征层 P#, P# 通过 C4 与 C5 进行特征融合得到 Q3。

多余的特征层和跳跃融合可能导致信息冗余, 从而导致 mAP 值下降。相比之下, MSFFN 在特征融合与模型复杂度之间能够达到更好的平衡。

3. 7 不同注意力机制对比实验

为验证 MSCBAM 注意力机制的有效性, 本文将 YOLOv8n 与多种注意力机制结合进行对比分析。实验结果如表 5 所示。尽管 MSCBAM 的计算量略高, 但其精度提升显著, 且模型复杂度较低, 适合对精度要求较高且需要控制模型复杂度的任务。综上所述, MSCBAM 在精度、参数量和计算量之间取得了更好的平衡, 是一种更高效的注意力机制。

表 5 不同注意力机制对比  
Table 5 Comparison of different attention mechanisms

| Attention | mAP@ 0. 5/% | Parameters/( ×10 <sup>6</sup> ) | 浮点数/GFLOPs |
|-----------|-------------|---------------------------------|------------|
| SE        | 82. 5       | 2. 5                            | 7. 8       |
| ECA       | 82. 4       | 2. 3                            | 7. 6       |
| MSCBAM    | 82. 9       | 2. 3                            | 8. 0       |

3. 8 PC2F-LMS 模块的多尺度特征提取能力验证

特征图可视化对比如图 8 所示。清晰地展示了 PC2F-LMS 模块在多尺度特征提取中的显著贡献。通过

对比图 8(a) 和(b) 可以观察到, PC2F-LMS 模块能够更精确地高亮目标区域, 尤其是在处理不完整或复杂目标时表现出更强的鲁棒性。这表明 PC2F-LMS 模块通过其多尺度特征提取机制, 能够更有效地捕捉局部细节和全局上下文信息, 从而提升特征表达的准确性和丰富性。PC2F-LMS 模块结合 PConv 和 LMSCM-Bottleneck, 实现了多层次特征的融合与优化, 显著增强了模型在目标检测和分类任务中的性能。

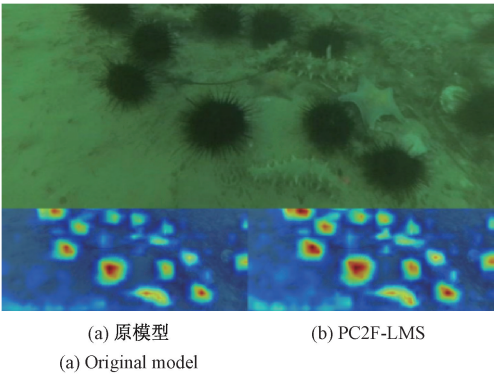


图 8 特征图可视化对比  
Fig. 8 Visualization comparison chart of feature maps

3. 9 实验结果

图 9 和 10 所示为 YOLOv8n 与 YOLOv8n-MAL 模型的检测结果。在图 9 中, 场景 A 为海胆(echinus)后方的被遮挡海参(holothurian)未被检测出。而在场景 B, 由于生物与环境高度融合, 左侧的海参和右侧的海星(starfish)也未被检测出。图 10 所示为改进后的模型模型在相同场景下的表现。该模型在检测精度上有明显提升, 成功检测到了场景 A 中被遮挡的海参, 并在场景 B 中识别出与背景高度融合的海参和海星。

实验证实, 在复杂和昏暗的水下环境中, YOLOv8n-MAL 模型相比于原模型, 表现出了更强的鲁棒性和精确度, 能够更准确地检测被遮挡或漏检的生物, 显然更适合水下检测仪器, 能够更有效地应对实际水下环境中的挑战。



图 9 YOLOv8n 检测结果

Fig. 9 YOLOv8n test results



图 10 YOLOv8n-MAL 检测结果

Fig. 10 YOLOv8n-MAL test results

3. 10 资源受限硬件的实验分析

为验证模型在资源受限环境下的性能,在两种不同硬件平台上进行了对比实验。相较于原始 YOLOv8n 模型,YOLOv8n-MAL 在 RTX 4070Ti 和 GTX 1080Ti 上均实现了更低的参数量,显著降低了模型的存储成本。同时,YOLOv8n-MAL 在两种硬件平台上的检测精度均优于原始模型,分别提升至 83.5%和 82.4%,展现了更优的检测性能。尽管帧率略有下降,但在资源受限的场景下,低参数量的设计能够有效减少存储需求,使 YOLOv8n-MAL 具备更高的性价比。该模型在降低计算成本的同时保持了较高的检测精度,适用于高效推理任务。实验结果如表 6 所示。

表 6 不同硬件下的数据对比

| Table 6 Comparison of data across different hardware |            |                                  |                |            |
|------------------------------------------------------|------------|----------------------------------|----------------|------------|
| Model+<br>Hardware                                   | mAP@ 0.5/% | Parameters/<br>( $\times 10^6$ ) | 浮点数/<br>GFLOPs | 帧率/<br>fps |
| YOLOv8n-MAL<br>RTX 4070Ti                            | 83.5       | 1.8                              | 7.2            | 67.1       |
| YOLOv8n-MAL<br>GTX 1080Ti                            | 82.5       | 1.8                              | 7.2            | 51.2       |
| YOLOv8n<br>RTX 4070Ti                                | 82.1       | 3.0                              | 8.0            | 80.3       |
| YOLOv8n<br>GTX 1080Ti                                | 81.5       | 3.0                              | 8.0            | 65.8       |

4 结 论

为解决现有水下检测模型参数量过大、对硬件存储资源要求高的问题,本文提出了一种基于 YOLOv8n 多尺度特征融合的轻量化水下目标检测模型。首先,基于 CBAM 改进的轻量化多尺度的注意力机制提出的 MSCBAM,以更好地捕捉不同尺度的信息。其次,重构了原模型的颈部网络,提出 MSFFN 模型,通过在主干与颈部网络之间添加卷积层,实现多层次特征融合,提升了特征利用效率,显著减少了模型参数量。随后,提出 PC2F-LMS 模块,利用金字塔结构更好地捕捉并融合多尺度特征。最后,替换损失函数为 WIoU,进一步提高了模型精度。实验结果表明,在相同实验条件下,YOLOv8n-MAL 模型的 mAP@ 0.5 值较原模型提升了 1.4%,参数量减少了 38.6%。该模型的轻量化设计不仅降低了对硬件资源的依赖,还提升了在复杂水下环境中执行目标检测任务时的效率和可靠性。

尽管所提出的 YOLOv8n-MAL 模型在参数量压缩和检测精度提升方面取得了显著进展,但仍存在一定局限性。首先,轻量化设计可能导致高层特征表达能力下降,影响超小目标、远距离目标及遮挡目标的检测精度。其次,尽管模型在精度和参数量上优于原模型,但由于多尺度特征融合引入了更复杂的计算路径,增加了数据流动和特征交互的开销,推理速度有所下降,这在高实时性要

求的场景中可能成为瓶颈。因此,未来改进的重点在于平衡检测精度与计算效率。一方面,需进一步增强模型对超小目标和复杂背景的适应能力;另一方面,需探索更高效的特征融合策略,减少计算冗余以提升推理速度,同时保持低参数数量的优势。

## 参考文献

- [ 1 ] VORKAPIC A, POBAR M, KOS I M. A computer vision approach to estimate the localized sea state[J]. *Ocean Engineering*, 2024, 309(P1): 118318.
- [ 2 ] 裴凯洋, 张胜茂, 樊伟, 等. 基于计算机视觉的鱼类视频跟踪技术应用研究进展[J]. *海洋渔业*, 2022, 44 ( 5 ): 640-647.  
PEI K Y, ZHANG SH M, FAN W, et al. Fish video tracking based on computer vision technology application research progress[J]. *Journal of Marine fishery*, 2022, 44 ( 5 ): 640-647.
- [ 3 ] JIAN M, LIU X, LUO H, et al. Underwater image processing and analysis: A review[J]. *Signal Processing Image Communication*, 2021, 91(5):116088.
- [ 4 ] YANG Q, ZHOU J, MENG H, et al. Real-time marine target recognition method of underwater inspection robot based on computer vision [J]. *International Journal of Advanced Robotic Systems*. 2023; 20 ( 5 ), DOI: 10.1177/17298806231199845.
- [ 5 ] 彭伟, 厉运周, 高艳波, 等. 我国海洋观测仪器设备体系化发展研究[J]. *仪器仪表学报*, 2023, 44(12): 88-100.  
PENG W, LI Y ZH, GAO Y B, et al. Research on the systematic development of marine observation equipment in China [J]. *Chinese Journal of Scientific Instrument*, 2023, 44(12): 88-100.
- [ 6 ] WANG Y, SONG W, FORTINO G, et al. An experimental-based review of image enhancement and image restoration methods for underwater imaging [J]. *arXiv*, 2019, DOI: 10.1109/ACCESS.2019.2932130.
- [ 7 ] KUBACKA M, GAJEWSKI L, BURCHACZ M, et al. Many faces of survey equipment failures during marine research at sea—Risk analysis [J]. *Plos One*, 2022, 17(8): e0272960.
- [ 8 ] BHARATI P, PRAMANIK A. Deep learning techniques R-CNN to mask R-CNN: A survey [J]. *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019, 2020: 657-668*.
- [ 9 ] HE K, GKIOXARI G, PIOTR DOLLÁR, et al. Mask R-CNN [C]. *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2961-2969.
- [ 10 ] JIANG P, ERGU D, LIU F, et al. A review of YOLO algorithm developments [J]. 2021, DOI: 10.1016/j.procs.2022.01.135.
- [ 11 ] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. *ArXiv preprint arXiv: 1804.02767*, 2018.
- [ 12 ] VARGHESE R, SAMBATH M. YOLOv8: A novel object detection algorithm with enhanced performance and robustness [C]. 2024 *International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*.
- [ 13 ] WANGAO, CHEN HUI, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[J]. *ArXiv preprint arXiv: 2405.14458*, 2024.
- [ 14 ] GAŠPAROVIĆ, BORIS, MAUSA G, RUKAVINA J, et al. Evaluating YOLOv5, YOLOv6, YOLOv7, and YOLOv8 in underwater environment: Is there real improvement [C]. 2023 *8th International Conference on Smart and Sustainable Technologies (SpliTech)*. IEEE, 2023:1-4.
- [ 15 ] LI Y X, REN F B. Light-weight retinanet for object detection [J]. *ArXiv preprint arXiv: 1905.10011*, 2019.
- [ 16 ] 周新, 张春堂, 樊春玲. 基于 YOLOv5\_PGS 的轻量级水下生物识别目标检测[J]. *电子测量技术*, 2023, 46(21): 168-175.  
ZHOU X, ZHANG CH T, FAN CH L. Lightweight underwater biometric target detection based on YOLOv5\_PGS [J]. *Electronic Measurement Technology*, 2023, 46(21): 168-175.
- [ 17 ] 李卓润, 李波, 邱鹏程, 等. 轻量化 YOLOv7-tiny 的水下压印字符识别 [J]. *国外电子测量技术*, 2024, 43(4): 162-169.  
LI ZH R, LI B, QIU P CH, et al. Underwater imprinting character recognition of lightweight YOLOv7-tiny [J]. *Foreign Electronic Measurement Technology*, 2024, 43(4): 162-169.
- [ 18 ] 杨婷, 高武奇, 王鹏, 等. 自动色阶与双向特征融合的水下目标检测算法[J]. *激光与光电子学进展*, 2023, 60(6): 122-133.  
YANG T, GAO W Q, WANG P, et al. Underwater target detection algorithm based on automatic color level and bidirectional feature fusion [J]. *Laser & Optoelectronics Progress*, 2023, 60(6): 122-133.
- [ 19 ] 涂万, 于红, 张鹏, 等. 基于通道非降维与空间协调注意力的改进 YOLOv8 养殖鱼群检测[J]. *大连海洋大学学报*, 2023, 38(4): 717-725.  
TU W, YU H, ZHANG P, et al. Improved detection of cultured fish in YOLOv8 based on channel non-dimensionality reduction and spatial coordination attention [J]. *Journal of Dalian Ocean University*, 2023, 38(4): 717-725.

[20]

LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[21]

LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[22]

吴兴辉, 何赞泽, 周辉, 等. 改进 YOLOv7 算法下的监控水域环境人员识别研究[J]. 电子测量与仪器学报, 2023, 37(5): 20-27.  
WU X H, HE Y Z, ZHOU H, et al. Research on environmental personnel identification of monitored waters based on improved YOLOv7 algorithm [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(5): 20-27.

[23]

WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.

[24]

LIANG C, YU Y Y, ZHE N W, et al. Underwater target detection lightweight algorithm based on multi-scale feature fusion [J]. Journal of Marine Science and Engineering, 2023, 11(2): 320-320.

[25]

WANG C Y, YEH I H, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information [J]. ArXiv preprint arXiv: 2402.13616, 2024.

[26]

PARK S, YEO Y J, SHIN Y G. PConv: Simple yet effective convolutional layer for generative adversarial network[J]. Neural Computing and Applications, 2022, 34(9): 7113-7124.

[27]

TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism [J]. ArXiv preprint arXiv: 2301.10051, 2023.

[28]

唐鲁婷, 黄洪琼. 基于 YOLOv7 的轻量化水下目标检测算法[J]. 电光与控制, 2024, 31(9): 92-97.  
TANG L T, HUANG H Q. Lightweight underwater target detection algorithm based on YOLOv7[J]. Electro-optics & Control, 2024, 31(9): 92-97.

作者简介



苗力恒, 2023 年于辽宁科技大学获得学士学位, 现为辽宁科技大学硕士研究生, 主要研究方向为计算机视觉、深度学习。  
E-mail: 1113596430@qq.com

Miao Liheng received his B. Sc. degree from University of Science and Technology Liaoning in 2023. Now he is a M. Sc. candidate at University of Science and Technology Liaoning. His main research interests include computer vision and deep learning.



田莹(通信作者), 2008 年于沈阳工业大学获得博士学位, 现为辽宁科技大学教授, 主要研究方向为计算机视觉、深度学习、模式识别等。  
E-mail: t\_tianying@126.com

Tian Ying (Corresponding author) received her Ph. D. degree from Shenyang University of Technology in 2008. Now she is a professor at University of Science and Technology Liaoning. Her main research interests include computer vision and deep learning, and pattern recognition, etc.