

# 基于改进 YOLOv10s 的海洋牧场水下海参检测方法

刘侦龙<sup>1,3</sup>, 王骥<sup>2,3\*</sup>, 麦仁贵<sup>1,3</sup>

(1. 广东海洋大学数学与计算机学院, 湛江 524088; 2. 广东海洋大学电子与信息工程学院, 湛江 524088; 3. 广东省智慧海洋传感网及其装备工程技术研究中心, 湛江 524088)

**摘要:** 针对海洋牧场水下复杂环境中海参目标小且体表与背景区分难度大, 光线高度弱化, 图像存在大量噪声以及海参堆叠遮挡导致检测精度低的问题, 该研究提出了一种基于改进 YOLOv10s 模型的水下海参检测方法 YOLOv10-MECAS。该方法设计了中值增强的通道和空间注意力模块 MECAS (median-enhanced channel and spatial), 保留目标特征的同时减少图像噪声并通过多尺度深度卷积提升海参图像特征捕捉能力; 引入可切换空洞卷积 SAConv (switchable atrous convolution) 模块替换 SCDDown (spatial-channel decoupling downsampling) 模块中的 3×3 普通卷积模块, 在无需增大卷积核尺寸的前提下扩大了感受野, 增强模型捕获遮挡目标的特征能力; 采用基于暗通道先验的水下图像增强算法 UDCP (underwater dark channel prior) 对数据集图像进行增强, 优化对比度, 提高图像质量; 使用 MPDIoU (minimum points distance intersection over union) 损失函数, 减少由于样本差异性大引起的检测框失真, 提高模型鲁棒性。试验利用水下真实场景下采样的海参数据集对模型的性能进行了评价。结果显示, 在常规数据集上, 该模型的精确率、召回率、mAP<sub>0.5</sub> 分别达到 85.7%、81.5%、89.7%, 相比基线模型分别提高了 6.4、4.4、5.0 个百分点; 在增强数据集上, 该模型的精确率、召回率、mAP<sub>0.5</sub> 分别达到 86.4%、82.6%、90.4%, 相比基线模型分别提高了 3.3、2.1、4.8 个百分点。研究结果表明, 该研究提出的模型在复杂海洋牧场水下环境中, 能有效提高小目标海参的检测精度, 可为海参自动化捕捞提供理论支持。

**关键词:** 海参; YOLOv10s; 水下目标检测; 注意力模块; 损失函数; 水下图像增强

doi: 10.11975/j.issn.1002-6819.202412109

中图分类号: TP391.41; S97

文献标志码: A

文章编号: 1002-6819(2025)-10-0186-09

刘侦龙, 王骥, 麦仁贵. 基于改进 YOLOv10s 的海洋牧场水下海参检测方法[J]. 农业工程学报, 2025, 41(10): 186-194.

doi: 10.11975/j.issn.1002-6819.202412109 <http://www.tcsae.org>

LIU Zhenlong, WANG Ji, MAI Rengui. Detecting underwater sea cucumber in marine ranching using improved YOLOv10s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2025, 41(10): 186-194. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202412109 <http://www.tcsae.org>

## 0 引言

海洋牧场作为一种新型的海洋资源利用方式, 近年来迅速发展, 通过智能化装备实现了海洋生物资源的可持续开发和利用。海参作为高经济价值海产品, 是海洋牧场不可或缺的养殖物种<sup>[1]</sup>。海参的捕捞是海参养殖的重要环节, 由于水下作业的特殊性, 成熟海参捕捞难度很大, 目前主要依靠人工潜水进行捕捞, 这种方法效率低且捕捞人员容易患上减压病, 因此机器捕捞海参是未来发展的趋势<sup>[2]</sup>。机器捕捞海参利用摄像头拍摄水下海参图像, 通过目标检测技术精确识别与定位海参进行捕捞<sup>[3]</sup>。值得注意的是, 目标检测算法的识别精度直接影响捕捞的效率。水下环境光线弱, 尤其在深海或水质较浑浊的情况下, 光的传播受到严重限制, 导致摄像头捕捉到的图像存在大量噪声, 细节模糊<sup>[4]</sup>。另一方面, 可食用海参大多为深色, 弱光条件下常常与水底的泥沙、岩石等背景融为一体, 进一步增加了目标识别难度。另外, 在海洋牧场海参养殖区域, 由于海参养殖密度大, 呈现相互堆叠, 相互遮挡, 导致个体识别难度大, 影响

其定位与识别精度。

近年来应用于水下海参等小目标检测算法主要分为传统目标检测算法和基于深度学习的目标检测算法两种类型。传统目标检测方法主要依赖于从图像中提取目标的显著特征, 并通过分类器对其进行识别, 设计合理的特征对于目标检测的准确性至关重要。通常需要研究人员深入分析目标的属性, 如颜色、纹理和形状等, 从而更好地描述目标的特征信息。这些特征能够反映目标物体的固有属性, 帮助算法提高检测效果。在实际应用中, 算法通常通过手动设计的特征描述方法, 结合特征选择和分类器进行目标检测。算法会对提取的特征进行计算与处理, 最终通过输入图像获取对应的检测结果<sup>[5]</sup>。在海参检测方面, 崔尚等<sup>[6]</sup>使用改进的 Sobel 算子和形态学方法得到了海参的二值图像, 经过多次膨胀、腐蚀处理结合小目标移除算法, 锁定海参目标。李娟等<sup>[7]</sup>利用计算机视觉方法开展研究, 采用边缘检测技术识别水下海参目标, 通过分析海参刺的外形特征提取其轮廓, 进一步计算中心点坐标并应用椭圆拟合策略实现了水下海参目标的检测。以上为最有代表性的传统目标检测算法, 严重依赖于研究人员手工设计的特征提取算子, 在复杂场景应用时, 往往无法有效应对光照变化、遮挡和背景复杂等问题, 会导致模型泛化性差, 鲁棒性低。

随着人工智能技术的不断进步, 基于深度学习的目标检测技术逐渐得到业界重视, 其原理是通过卷积神经网络 (convolutional neural networks, CNN) 自动提取水

收稿日期: 2024-12-16 修订日期: 2025-03-24

基金项目: 广东省普通高校重点领域新一代信息技术专项 (2020ZDZX3008)

作者简介: 刘侦龙, 研究方向为人工智能。Email: 1945700701@qq.com

\*通信作者: 王骥, 教授, 硕士生导师, 研究方向为无线传感器网络、海洋物联网与人工智能。Email: 13902576499@163.com

下目标的特征信息，并通过训练数据的拟合使得网络能够自主学习目标的特征，省去了人工设计滤波器的环节，展现出更强的泛化能力和鲁棒性。根据工作流程差异，基于深度学习的目标检测算法通常分为两大类：两阶段检测算法（two-stage）和单阶段检测算法（one-stage）。前者以 R-CNN<sup>[8]</sup>、Fast R-CNN<sup>[9]</sup>、Faster R-CNN<sup>[10]</sup> 算法为代表，将检测任务分为候选区域提取和分类 2 个步骤；后者以 YOLO<sup>[11]</sup> 和 SSD<sup>[12]</sup> 等为代表，直接获得目标的类别概率和位置坐标而无需区域建议。近些年来，研究人员利用深度学习对水下生物的目标识别做出了大量研究。LIU 等<sup>[13]</sup> 提出了一种使用 Swin Transformer 作为骨干网络的 Faster R-CNN 模型，并加入了路径聚合网络，在 URPC（underwater robot perception challenge）数据集上平均精度达到 80.5%。SHI 等<sup>[14]</sup> 提出了一种使用 ResNet 作为骨干网络的 Faster R-CNN 模型，结合 Bi-FPN 结构增强特征提取能力和多尺度特征融合能力并使用 EIou 替换原有的 IoU，在 URPC 数据集上平均精度达到 88.9%。虽然两阶段目标检测算法在检测精度上具有一定优势，但其复杂的候选区域提取过程需要更多的计算资源和检测时间，不适用于水下实时检测任务。相比之下，单阶段目标检测算法需要的计算资源较少、检测速度更快，更适合水下实时检测任务。LI 等<sup>[15]</sup> 在 YOLOv3-ST 模型中嵌入 SEnet 注意力机制，对水草的检测精度比原模型提高了 22.3%。YANG 等<sup>[16]</sup> 以 YOLOv3 模型为基础，在特征提取网络的并行结构中融合全局语义信息，改善了水下目标的感知能力。SUN 等<sup>[17]</sup> 提出了一种用 MobileViT 作为骨干网络的 YOLOX 水下目标检测模型，并设计 DCA（double coordinate attention）模块，增强全局特征的提取的同时有效减少了模型的参数量。LI 等<sup>[18]</sup> 提出了一种改进的 YOLOv5s 水下扇贝检测模型，使用群卷积和逆残差模块代替骨干网，检测精度达到 88.2%。LIU 等<sup>[19]</sup> 提出了一种改进的 YOLOv7 水下目标检测模型，通过结合全局注意力机制与残差网络结构，有效提高了特征提取的效率并加快检测速度。FENG 等<sup>[20]</sup> 在 YOLOv8 模型上结合了高阶可变形注意模块和增强的空间金字塔池化模块，在探测水下物体数据集上平均精度达到 88.4%。ZHOU 等<sup>[21]</sup> 在 YOLOv8 模型上集成了跨阶段多分支模块和大核空间金字塔模块，在 URPC 数据集上平均精度达到 84.0%。

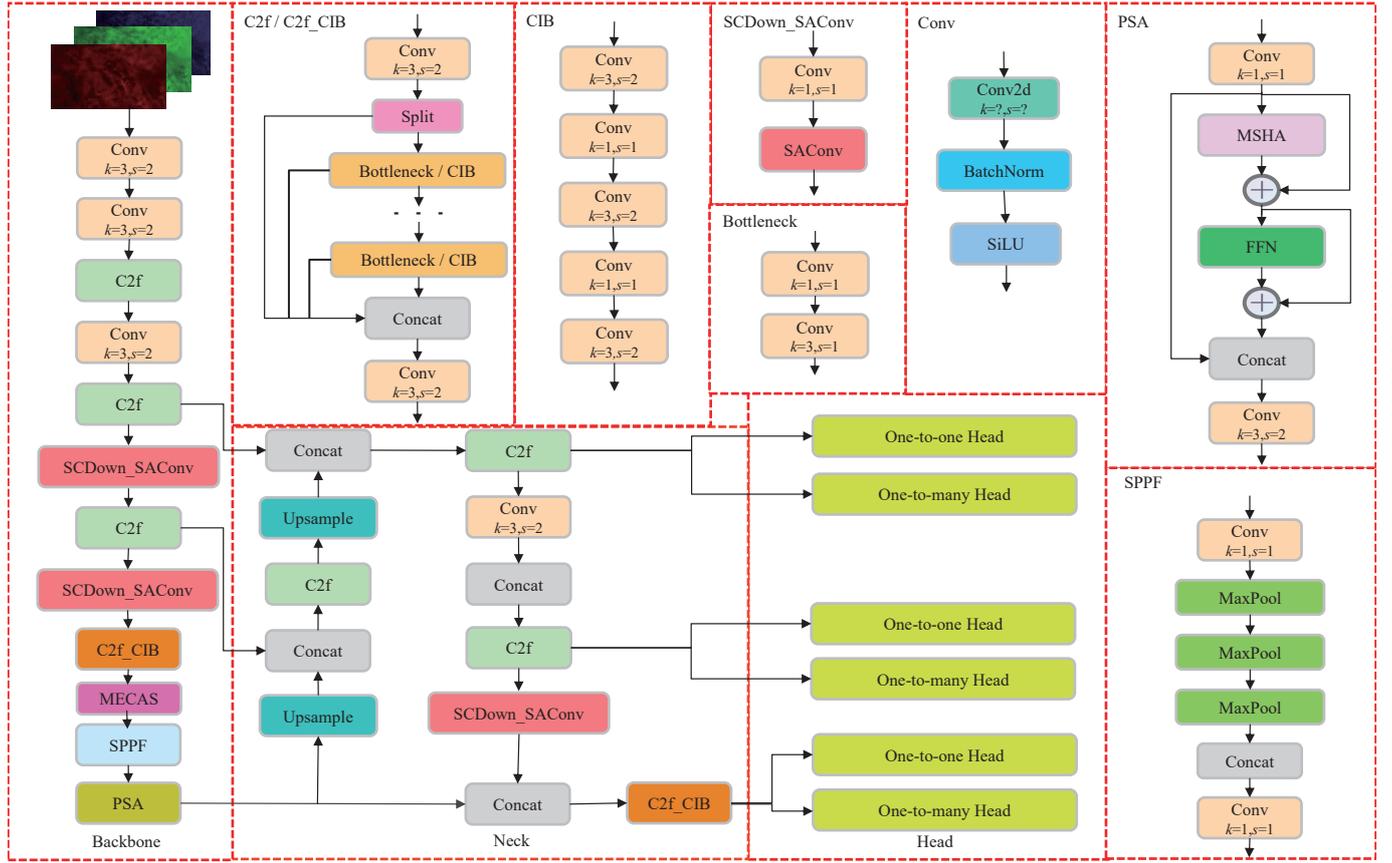
尽管上述文献的深度学习算法在水下目标检测中表现出良好潜力，但仍未有效解决海参图像噪声、海参体表颜色与背景融为一体的“保护色”问题，特别是因养殖密度大导致的海参相互遮挡等问题。针对以上问题，本研究提出了以 YOLOv10s 为基线模型改进的 YOLOv10-MECAS 模型，通过设计中值增强的通道和空间注意力模块 MECAS（median-enhanced channel and spatial），并引入可切换空洞卷积模块 SAConv（switchable atrous convolution），结合利用 MPDIoU（minimum points distance intersection over union）损失函数，优化模型水下恶劣环境的检测能力，进一步结合基于暗通道先验的水下图像增强算法 UDCP（underwater dark channel prior）进行水下海参图像质量提升，构建优质水下海参目标数据集对模型进行评估。

## 1 YOLOv10s 模型

在目标检测领域，YOLO 系列模型以良好的检测速度和准确率著称，但仍存在计算冗余和低参数利用率等问题。YOLOv10s<sup>[22]</sup> 的提出可以一定程度上改善这些问题。相较于以往优秀的 YOLO 系列模型（如 YOLOv8、YOLOv5），YOLOv10s 在检测精度和速度方面取得了更好的平衡。其网络结构主要由主干网络、颈部网络和头部网络构成。主干网络使用改进的 CSPDarknet53 网络<sup>[23]</sup>，引入了 SCDowN（spatial-channel decoupling downsampling）模块，CIB（compact inverted block）模块和 PSA（partial self-attention）模块。SCDowN 模块通过点卷积改变通道维度，再用深度卷积进行空间下采样，有效降低参数量并最大化信息保留。CIB 模块采用成本效益高的深度卷积和逐点卷积进行通道和空间混合，作为基本构建块嵌入到 C2f（cross stage partial network fusion）模块之中。PSA 模块低成本地融合全局表示学习能力，增强了模型性能，多头自注意力模块（multi-head self-attention module，MSHA）和前馈网络 FFN（feed-forward network）为其核心组成部分；颈部网络采用了路径聚合网络（path aggregation network，PAN）<sup>[24]</sup>。通过自上而下、自下而上的网络结构和横向连接融合高层与低层特征图，从而增强了特征的多样性和完整性，提高了小尺寸目标的检测效果。YOLOv10 通过在检测高层特征图前引入 SCDowN 模块和 C2f\_CIB 模块，在保持性能的同时提升了效率；头部网络设计了双重标签匹配检测头，即同时采用一对一匹配检测头和一对多匹配检测头。一对一匹配头避免了非极大值抑制（non-maximum suppression，NMS）后处理但会导致次优的精度<sup>[25]</sup>，一对多匹配头则能确保最优精度但增加推理时间和收敛速度<sup>[26]</sup>。YOLOv10s 结合两者优点，在训练过程中，充分利用一对多头部提供的丰富的监督信息，在推理过程中，只采用一对一匹配检测头进行预测，实现了在不损失精度的情况下加快了推理速度。

## 2 YOLOv10s 模型改进

在特定应用场景中，YOLOv10s 仍存在一些不足之处。例如，在水下弱光环境中，海参的体积较小且与背景区分困难，同时海参之间可能会相互堆叠。此外，水下环境中拍摄的海参图像还存在噪声问题，这使得 YOLOv10s 对水下真实环境中海参目标的适应性显得不足。本文以 YOLOv10s 为基线模型提出了 YOLOv10-MECAS 水下目标检测模型，其网络结构如图 1 所示。首先设计中值增强的通道和空间注意力模块 MECAS，将其嵌入到主干网络中，MECAS 模块通过中值池化保留水下图像重要特征信息的同时去除噪声，并通过多尺度深度卷积提升特征捕捉能力，提高模型在复杂水下环境中的检测能力；其次引入可切换空洞卷积 SAConv<sup>[27]</sup> 来替换 SCDowN 模块的 3×3 普通卷积，在无需增大卷积核尺寸的前提下扩大感受野，以捕获更广泛的遮挡目标特征，提升模型对遮挡目标的检测精度；最后使用 MPDIoU 损失函数<sup>[28]</sup> 替换原损失函数（complete intersection over union，CIoU），减少由于样本差异性大引起的检测框失真，提高模型鲁棒性。



注: Split 为拆分操作; Concat 为拼接操作; Conv2 d 为卷积操作;  $k$  为卷积核大小;  $s$  为步幅; BatchNorm 为批归一化操作; SiLU 为激活函数; MSHA 为多头自注意力模块; FFN 为前馈网络; MaxPool 为最大池化操作; Upsample 为上采样操作; One-to-one Head 为一对一匹配检测头; One-to-many Head 为一对多匹配检测头; SAConv 为可切换空洞卷积; MECAS 为中度增强空间和通道注意力模块。

Note: Split is the splitting operation; Concat is the concatenation operation; Conv2 d is the convolution operation; where  $k$  is the kernel size;  $s$  is the stride; BatchNorm is the batch normalization operation; SiLU is the activation function; MSHA is the multi-head self-attention module; FFN is the feed-forward network; MaxPool is the max pooling operation; Upsample is the upsampling operation; One-to-one Head is the one-to-one matching detection head; One-to-many Head is the one-to-many matching detection head; SAConv is the switchable atrous convolution; MECAS is the Median-enhanced channel and spatial module.

图 1 YOLOv10-MECAS 网络结构图

Fig.1 YOLOv10-MECAS network structure

## 2.1 MECAS 模块

注意力机制是深度学习中的核心技术之一, 能够帮助模型自动识别并聚焦于输入数据中最重要的部分, 从而提高性能和计算效率。本文提出的 MECAS 模块通过中度池化在保留图像重要特征信息的同时去除噪声。此外, 模块采用了多尺度卷积以捕捉和融合不同尺度的特征, 从而增强模型在存在噪声和目标遮挡情况下的目标识别能力。MECAS 模块由通道注意力和空间注意力两部分组成, 通过残差连接结合以上两种注意力机制, 能够提供更丰富的特征表示, 从而提高水下目标检测的精度和鲁棒性。其结构如图 2 所示。

### 2.1.1 通道注意力

传统的通道注意力机制通常依赖全局平均池化和最大池化来提取特征信息, 但这些方法在噪声较多的图像中表现较差。MECAS 模块在通道注意力通过引入中度池化, 结合平均池化和最大池化, 能够有效去除噪声并保留重要特征信息。通道注意力整体流程如下:

1) 输入特征图  $F \in R^{C \times H \times W}$  经过平均池化 (AvgPool)、中度池化 (MedianPool) 和最大池化 (MaxPool), 得到 3 个池化结果  $F_{avg}^c$ 、 $F_{max}^c$ 、 $F_{median}^c$ 。每个结果的尺寸均为  $R^{c \times 1 \times 1}$ 。

$$F_{avg}^c = \text{AvgPool}(F) \quad (1)$$

$$F_{max}^c = \text{MaxPool}(F) \quad (2)$$

$$F_{median}^c = \text{MedianPool}(F) \quad (3)$$

2) 每个池化结果通过共享的多层感知机 MLP (multilayer perceptron), MLP 包含 2 个  $1 \times 1$  卷积层和一个 ReLU 激活函数。相加最终经过 Sigmoid 激活函数生成通道特征图  $F^c$ 。

$$F^c = \sigma(\text{MLP}(F_{avg}^c) \oplus \text{MLP}(F_{max}^c) \oplus \text{MLP}(F_{median}^c)) \quad (4)$$

式中  $\sigma$  为 Sigmoid 激活函数。

3) 将通道特征图  $F^c$  与输入特征图  $F$  进行元素级相乘, 得到加权特征图  $F'$ 。

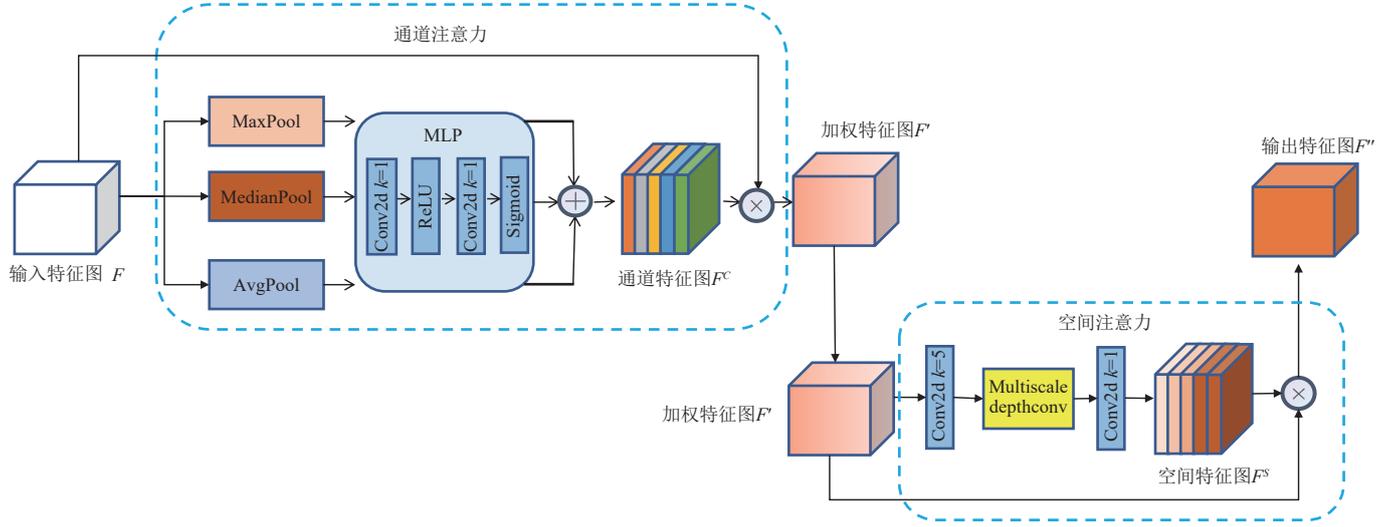
$$F' = F^c \otimes F \quad (5)$$

### 2.1.2 空间注意力

在复杂的水下检测场景中, 目标的特征信息可能会由于其他目标或背景的遮挡而无法完全获取。传统的通道注意力使用的单一尺度的卷积核对被遮挡的目标表现出较差的检测效果, 容易出现漏检或误检的情况。为此, MECAS 模块在空间注意力采用多尺度深度卷积来捕捉不同尺度和方向的特征信息, 从而提升模型对遮挡目标的检测能力以及精度。通过多尺度卷积提取的特征图生

成空间注意力图，最终与加权后的特征图相乘得到输出特征图。空间注意力处理流程如下：

1) 输入加权特征图  $F'$  通过一个  $5 \times 5$  的深度卷积层提取基础特征，通过不同尺寸的深度卷积层进一步提取多尺度特征并相加，随后通过  $1 \times 1$  卷积层生成空间特征图  $F^s$ 。



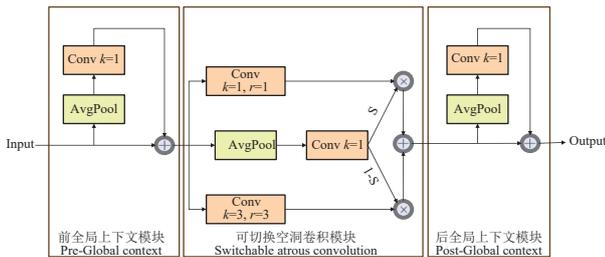
注：MaxPool 为最大池化操作；MedianPool 为中线池化操作；AvgPool 为平均池化操作；ReLU, Sigmoid 为激活函数；Multiscale depthconv 为多尺度深度卷积模块；⊕ 为特征图相加操作；⊗ 为特征图相乘操作。  
 Note: MaxPool is the max pooling operation; MedianPool is the median pooling operation; AvgPool is the average pooling operation; ReLU and Sigmoid are activation functions; Multiscale depthconv is the multiscale depthwise convolution module; ⊕ is the feature map addition operation; ⊗ is the feature map multiplication operation.

图 2 MECAS 模块结构图  
 Fig.2 MECAS module structure

### 2.2 SAConv 模块

传统卷积的感受野由固定的卷积核大小决定，容易忽略局部空间的特征信息。而 SAConv 模块通过引入动态调节机制，在相同的输入特征上应用不同的空洞率进行卷积，并通过开关函数来融合不同卷积的结果，从而扩大卷积操作的感受野，捕获更大范围的目标特征，提升模型的特征表达能力。基于此，本文将 SAConv 模块替换 SCDown 模块中的  $3 \times 3$  传统卷积模块，解决了传统卷积在复杂水下场景中提取特征不充分的问题。SAConv 模块的架构由 3 个主要部分组成：可切换空洞卷积模块以及附加在其前后的 2 个全局上下文模块，整体结构如图 3 所示。SAConv 可表示为

$$y = S(x) \cdot \text{Conv}(x, \omega, r) + (1 - S(x)) \cdot \text{Conv}(x, \omega + \Delta\omega, r) \quad (8)$$



注：r 为空洞率；S 为开关函数。  
 Note: r is the dilation rate; S is the switch function.

图 3 SAConv 整体结构图  
 Fig.3 Overall structure of SAConv

$$F^s = D_1 \left( \sum_{i=1}^n D_i (D_s(F')) \right) \quad (6)$$
 式中  $D_i$  为不同大小卷积核的卷积操作， $i$  为卷积核大小， $n$  为卷积核大小最大值。

2) 将空间特征图  $F^s$  与加权特征图  $F'$  相乘，得到输出特征图  $F''$ 。

$$F'' = F^s \otimes F' \quad (7)$$

式中  $x$  和  $y$  表示输入和输出特征图， $\text{Conv}(x, \omega, r)$  表示权重为  $\omega$ ，空洞率为  $r$  的卷积运算， $\Delta\omega$  为可训练权值， $S(\cdot)$  是融合不同空洞率卷积结果的开关函数。

### 2.3 边界框回归损失函数 MPDIoU

在目标检测任务中，损失函数的设计对检测结果的精度具有重要影响。其主要功能是通过优化预测框与真实框之间的位置差异，使得模型能够生成更接近真实框位置的预测框，从而提升检测的准确性。YOLOv10s 采用的是 CIoU 损失函数，由于 CIoU 在处理纵横比时其描述相对不够清晰，可能导致优化结果不够合理。此外，在实际检测场景中，样本之间的差异性对其影响较大，使得模型难以针对多样化的检测目标进行自适应调整，进而影响到模型的收敛速度。为解决这些问题，本文将 CIoU 替换为 MPDIoU，有效改善了因样本差异性大而引起的检测框失真，提高模型的鲁棒性。MPDIoU 引入最小点距离的度量方式，对损失函数进行了重新设计，其核心思想是直接预测边界框与真实框的左上角和右下角之间的最小点距离。这种方法减少了损失函数的自由度，使得优化过程更高效，并能够更好地适应水下海参目标多样化的场景。MPDIoU 的具体计算如下：

$$d_1^2 = (x_1^B - x_1^A)^2 + (y_1^B - y_1^A)^2 \quad (9)$$

$$d_2^2 = (x_2^B - x_2^A)^2 + (y_2^B - y_2^A)^2 \quad (10)$$

$$\text{MPDIoU} = \frac{A \cap B}{A \cup B} + \frac{d_1^2}{w^2 + h^2} + \frac{d_2^2}{w^2 + h^2} \quad (11)$$

式中 $A$ 表示真实框； $B$ 表示预测框； $(x_1^A, y_1^A)$ 、 $(x_2^A, y_2^A)$ 分别为真实框左上角和右下角； $(x_1^B, y_1^B)$ 、 $(x_2^B, y_2^B)$ 分别为预测框左上角与右下角； $d_1$ 为两框左上角之间距离； $d_2$ 为两框右下角之间距离； $w$ 为图片宽度； $h$ 为图片高度； $\cap$ 表示交集运算； $\cup$ 表示并集运算。

## 2.4 水下图像增强算法 UDCP

由于水中光线的吸收与散射作用，水下图像通常会呈现出对比度下降、色彩失真、画面模糊等问题。这些问题显著增加了水下目标检测的难度，导致检测模型在提取目标特征时面临更大的挑战。因此，本文使用基于暗通道优先算法（dark channel prior, DCP）<sup>[29]</sup>改进的水下暗通道优先算法 UDCP<sup>[30]</sup>增强水下图像。DCP 算法利用暗通道先验方法估计图像透射率，从而实现地面雾化图像的恢复。但当 DCP 算法用于水下图像时，会产生过度增强的问题。UDCP 算法考虑到在水下环境中的光传播，红光的能量衰减非常迅速，相较于绿光和蓝光有明显差异。因此，UDCP 算法通过校正红光影响，优化了图像的复原效果。UDCP 算法中，暗通道仅考虑蓝色和绿色通道，UDCP 算法暗通道的计算如下：

$$J^{\text{UDCP}}(x) = \min_{y \in \Omega(x)} \left( \min_{C \in G, B} J^C(y) \right) \quad (12)$$

式中 $J^{\text{UDCP}}(x)$ 表示位置 $x$ 处的暗通道值， $J^C(y)$ 表示位置 $y$ 处的颜色通道值（G、B分别表示绿色和蓝色通道）， $\Omega(x)$ 是定义以 $x$ 为中心的局部区域。UDCP 算法借鉴了 DCP 算法的透射率估计算法，并针对水下图像特点进行了优化。其透射率的计算式如下：

$$\tilde{t}(x) = 1 - \min_{y \in \Omega(x)} \left( \min_{C \in G, B} \frac{I^C(y)}{A^C} \right) \quad (13)$$

式中 $A$ 是大气光值， $I^C$ 是每个颜色通道的像素值。原图和 UDCP 算法处理之后的图片对比如图 4 所示，可以看出 UDCP 算法使水下图像的对比度显著提升，使得水下图像中的细节更加清晰，物体边缘轮廓更加明显，物体与背景更加容易区分，更加有利于进行物体目标检测。

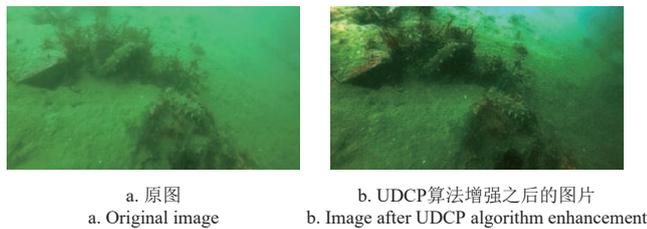


图 4 水下原图与增强算法处理后图像对比

Fig.4 Comparison between the original underwater image and the enhanced image after algorithm processing

## 3 试验与结果分析

### 3.1 试验设置

硬件环境方面，使用了 24 GB 显存的 RTX4090GPU，搭配 AMD EPYC 9 654 处理器。软件环境方面，操作系统为 Ubuntu 20.04，深度学习框架为 PyTorch，配置为 PyTorch 2.0、Python 3.8 和 CUDA 11.8。训练过程中，总批次为 200，batch size 设置为 16，采用 SGD 优化器，初始学习率为 0.01，并在 OneCycleLR 策略下调整为

0.000 1，权重衰减系数为 0.000 5，动量为 0.937。

### 3.2 水下海参图像数据集

本研究自建的原始数据集中图像数据来源于湛江市坡头区南三深海水产养殖场在水深 5 m 左右场景拍摄的海参图片或从现场监测视频中抽取的帧图。本次数据采集设备为大疆 Osmo Action 4。在数据采集过程中，特别关注了图像的稳定性，避免因水流或镜头抖动造成图像模糊。

数据集中海参图像呈现出多样性，具体体现在以下几种情况：海参相互遮挡情况、图像包括不同尺寸的海参、海参的体表颜色与背景色相近、水下光线不足造成图像包含噪声。共收集 2 730 张具有代表性的水下海参图像，图像部分展示如图 5 所示。采用 LabelImg 软件对海参图像中的目标进行标注，并将标注类别命名为“sea\_cucumber”，生成对应的标签文件。数据集按照 8 : 1 : 1 比例随机划分训练集、验证集以及测试集。应用 UDCP 图像增强算法对原始数据集进行处理，构建增强数据集。

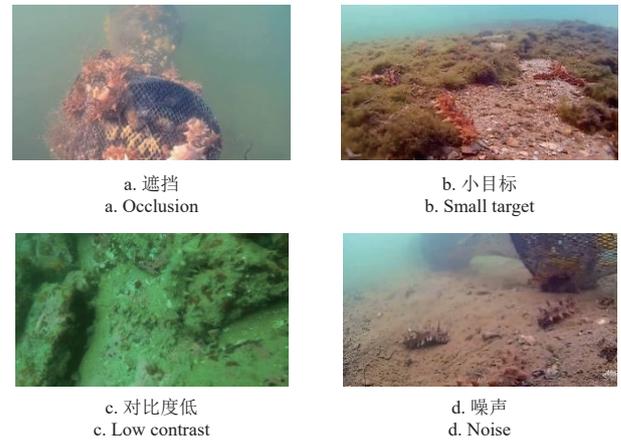


图 5 水下海参图像数据集示例

Fig.5 Example of underwater sea cucumber image dataset

### 3.3 评价指标

本文评价指标采用精确率（ $P$ ），召回率（ $R$ ），平均精度均值（mean average precision, mAP）、所需浮点运算数（giga floating-point operations per second, GFLOPs）、参数数量和帧率（frames per seconds, FPS），计算式如下：

$$P = \frac{T_P}{T_P + F_P} \quad (14)$$

$$R = \frac{T_P}{T_P + F_N} \quad (15)$$

$$\text{mAP} = \frac{\sum_{i=1}^N \int_0^1 P(R) dR}{N} \quad (16)$$

式中 $T_P$ 表示模型成功检测为海参且实际确为海参的样本数量； $F_P$ 指模型检测为海参目标但实际并非海参的样本数量； $F_N$ 则表示模型未能检测为海参目标但实际为海参的样本数量； $N$ 为检测类别数量， $N=1$ 。

### 3.4 注意力模块对比试验

为验证本文提出的 MECAS 模块的优越性，进行注意力模块的对比试验。本文选择 LSKA（large selective kernel attention）、CA（coordinate attention）、ECA

(efficient channel attention) 作为对比模块。以 YOLOv10s 为基线模型, MECAS 模块和对比模块嵌入到 YOLOv10s 的同一位置进行验证, 试验结果见表 1 所示, 结果表明, 嵌入 MECAS 模块后, 模型的  $P$ 、 $R$ 、 $mAP_{0.5}$  分别达到了 84.3%、82.6%、88.3%, 在模型参数量和计算量与 YOLOv10s 保持持平的情况下, 这些指标都超过了对比注意力模块。图 6 为 YOLOv10s 模型添加不同注意力模块热力图对比, 展示了不同注意力模块在聚焦于水下海参图像内的目标区域上的效果。由于水下复杂环境干扰, 原模型对海参目标区图像域缺乏关注度。LSKA 模块虽然对海参目标区域增加了关注度, 但未能完全的与背景进行区分。ECA 和 CA 模块在一定程度上区分了目标区域和背景, 但与 MECAS 模块相比, 对目标区域分配的注意力仍然不足。虽然 CA 和 ECA 模块都从不同的角度优化了注意力, 使更多的注意力集中到目标区域, 但仍然依赖于单一的卷积核进行特征提取, 缺乏多尺度特征信息的融合, 导致检测性能不佳。LSKA 模块动态调整空间感受野, 可以更好地模拟各种目标的测距上下文, 但受水下环境噪声干扰影响大。与上述注意模块相比, MECAS 模块有效地利用了中值池化降低水下环境中的噪声影响。不仅如此, MECAS 模块集成了水下海参的多尺度特征信息, 建立了通道域和空间域的长距离依赖关系。因此, MECAS 模块更加关注水下海参的目标区域, 减轻了背景干扰, 提高了检测精度。

表 1 注意力模块对比  
Table 1 Comparison of attention modules

模型 Model	$P$ /%	$R$ /%	$mAP_{0.5}$ /%	参数量 Params/M	计算量 GFLOPs
YOLOv10s	79.3	77.1	84.7	8.06	24.8
YOLOv10s+ECA	83.9	80.5	85.9	8.16	25.4
YOLOv10s+CA	83.2	80.4	85.3	8.10	24.9
YOLOv10s+LSKA	83.1	80.6	85.2	8.06	24.8
YOLOv10s+MECAS	84.3	82.6	88.3	8.06	24.8

注:  $P$  为精确率;  $R$  为召回率;  $mAP_{0.5}$  为 IoU 阈值是 0.5 的平均精度均值; IoU 为交并比。

Note:  $P$  is the precision;  $R$  is the recall;  $mAP_{0.5}$  is the mean average precision when IoU threshold is 0.5; IoU is intersection over union.

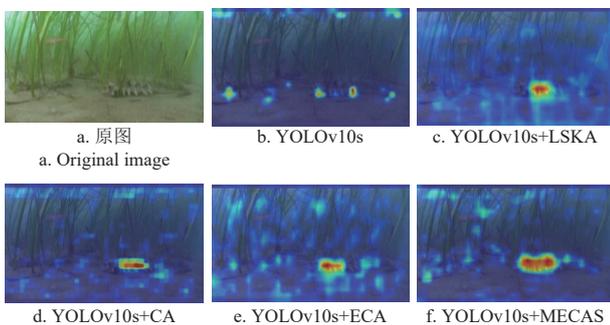


图 6 YOLOv10s 添加不同注意力模块热力图对比

Fig.6 YOLOv10s model adding different attention modules heatmap comparison

### 3.5 损失函数对比试验

为了验证本文替换损失函数操作的有效性, 设计了损失函数对比试验, 对 YOLOv10s 原始的 CIoU 以及现有的损失函数 EIou、GIou、DIou、SIou 和 MPDIou 进行对比试验, 结果见表 2。

与常用的回归损失函数 CIoU、EIou、GIou、DIou 和 SIou 相比, 引入 MPDIou 损失函数后的模型 (YOLOv10s+MPDIou) 在  $mAP_{0.5}$  精度上达到了最高值 85.6%。与原始模型 (YOLOv10s+CIoU) 相比,  $mAP_{0.5}$  和  $mAP_{0.5-0.95}$  分别提升了 0.9、0.3 个百分点。

表 2 不同 IoU 损失函数模型性能对比

Table 2 Performance comparison of different IoU loss function models

模型 Model	$mAP_{0.5}$ /%	$mAP_{0.5-0.95}$ /%
YOLOv10s+CIoU	84.7	56.3
YOLOv10s+EIou	84.7	56.7
YOLOv10s+GIou	84.9	56.8
YOLOv10s+DIou	83.8	56.2
YOLOv10s+SIou	85.0	56.6
YOLOv10s+MPDIou	85.6	56.9

注:  $mAP_{0.5-0.95}$  为 IoU 阈值是 0.5-0.95 的平均精度均值。

Note:  $mAP_{0.5-0.95}$  is the mean average precision when IoU threshold is 0.5-0.95.

### 3.6 消融试验

为了验证 YOLOv10-MECAS 模型每一处改进的有效性, 在相同的试验环境下开展消融试验, 结果如表 3 所示。试验 1 为基线模型, 基线模型对水下真实环境中海参目标的检测性能较差。试验 2 在主干网络中嵌入了 MECAS 模块。相比于试验 1, 试验 2 的  $P$ 、 $R$ 、 $mAP_{0.5}$  分别提升了 5.0、5.6、3.6 个百分点。MECAS 通过残差连接结合中值增强通道注意力模块和空间注意力模块, 去除图像噪声的同时提供更丰富的特征表示, 明显提高水下目标检测模型的检测精度和鲁棒性。试验 3 引入了 SAConv 模块替换 SCDOWN 模块的  $3 \times 3$  普通卷积模块。相比于试验 1, 试验 3 的  $P$ 、 $R$ 、 $mAP_{0.5}$  分别提升了 3.3、3.0、0.7 个百分点。SAConv 模块扩大卷积操作的感受野, 捕获更大范围的目标特征。试验 4 采用 MPDIou 作为 YOLOv10s 的损失函数。相比于试验 1, 试验 4 的  $P$ 、 $R$ 、 $mAP_{0.5}$  分别提升了 2.6、3.0、0.9 个百分点, MPDIou 可以有效改善因样本差异性大而引起的检测框失真, 提高模型的鲁棒性。试验 8 除将 MECAS 模块嵌入主干网络外引入 SAConv 模块和 MPDIou 损失函数。相比于试验 1, 试验 8 的  $P$ 、 $R$ 、 $mAP_{0.5}$  分别提升了 6.4、4.4、5.0 个百分点。试验结果表明, YOLOv10-MECAS 目标检测模型有效应对了水下场景中复杂背景和目標遮挡等挑战。

表 3 消融试验对比结果

Table 3 Comparative results of ablation experiment

序号 No.	MECAS	SAConv	MPDIou	$P$ /%	$R$ /%	$mAP_{0.5}$ /%
1	×	×	×	79.3	77.1	84.7
2	√	×	×	84.3	82.6	88.3
3	×	√	×	82.6	80.1	85.4
4	×	×	√	81.9	79.5	85.6
5	√	√	×	85.0	80.9	89.2
6	√	×	√	84.5	82.4	88.8
7	×	√	√	82.0	80.5	86.2
8	√	√	√	85.7	81.5	89.7

注: √表示使用该改进; ×表示不使用该改进; 基线模型为 YOLOv10s。

Note: √ represents using this improvement; × represents not using this improvement; the baseline model is YOLOv10s.

### 3.7 原始数据集上对比试验

为验证 YOLOv10-MECAS 模型的优势, 本文在相同的试验环境和参数配置下, 将其与 Faster-RCNN、SSD、

YOLOv5s、YOLOv7、YOLOv8s、YOLOv9s 以及YOLOv10s、YOLOv11s 在原始数据集上进行对比试验。试验结果如表4所示。

由表4可以发现，YOLOv10-MECAS 相较于未改进的YOLOv10s模型，虽然计算量稍有增加，但在参数量和帧率大致保持持平条件下  $mAP_{0.5}$  提高了5.0个百分点。与其他对比模型相比，YOLOv10-MECAS 帧率大幅提高。与Faster-RCNN、SSD、YOLOv7相比，YOLOv10-MECAS 在减少计算量与参数量的同时， $mAP_{0.5}$  分别提升了16.5、15.4、6.1个百分点。YOLOv10-MECAS 与YOLOv5s、YOLOv8s、YOLOv9s、YOLOv11s相比，参数量和计算量都相差不大， $mAP_{0.5}$  分别提升了5.5、5.3、5.8、4.6个百分点。由此证明，YOLOv10-MECAS 整体性能优于其他对比模型。原始数据集下， $P-R$  曲线对比情况如图7所示，也证明了YOLOv10-MECAS 在精度方面高于对比模型。综上，通过对检测精度与计算效率的综合评估，YOLOv10-MECAS 更契合水下目标检测的实际需求。

表4 原始数据集下不同模型性能对比结果

Table 4 Performance comparison results of different models under original dataset

模型 Model	P/%	R/%	$mAP_{0.5}/\%$	参数量 Params/M	计算量 GFLOPs	帧率 FPS/(帧·s <sup>-1</sup> )
Faster-RCNN	75.2	68.9	73.2	41.30	251.4	13.5
SSD	79.3	67.2	74.3	26.30	62.8	29.3
YOLOv5s	80.9	79.3	84.2	7.20	16.5	150.8
YOLOv7	79.7	78.7	83.6	36.49	103.5	101.2
YOLOv8s	81.1	79.7	84.4	11.20	28.6	120.3
YOLOv9s	80.1	78.5	83.9	7.25	26.7	170.6
YOLOv10s	79.3	77.1	84.7	8.06	21.6	255.7
YOLOv11s	81.8	78.9	85.1	9.4	21.5	181.5
YOLOv10-MECAS	85.7	81.5	89.7	8.07	24.9	235.6

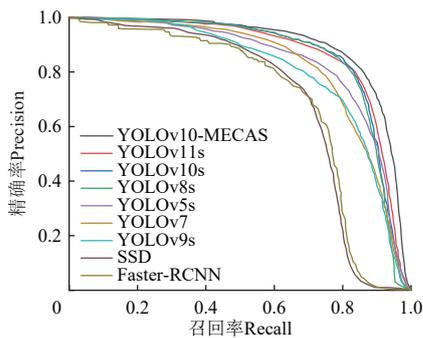


图7 原始数据集下不同模型的  $P-R$  曲线对比

Fig.7 Comparison of  $P-R$  curves of different models in the original dataset

### 3.8 增强数据集上对比试验

水下光学图像存在低对比度和色彩失真的问题，这会导致图像清晰度下降，增加了目标检测的难度。为改善这一问题，本文采用UDCP图像增强算法对原始数据集进行处理形成增强数据集。在增强数据集上，使用与表4相同的模型进行对比试验，试验结果如表5所示。

根据表5的试验结果可以看出，YOLOv10-MECAS 在增强数据集上明显优于其他检测模型。YOLOv10-MECAS 与Faster-RCNN、SSD、YOLOv5s、YOLOv7、

YOLOv8s、YOLOv9s、YOLOv10s、YOLOv11s相比，在参数量和计算量都有优势的前提下， $mAP_{0.5}$  均有提升，分别提升了16.1、15.4、5.5、5.9、5.2、6.1、4.8、4.3个百分点。增强数据集下， $P-R$  曲线对比情况如图8所示，证明了YOLOv10-MECAS 在精度方面高于对比模型。

表5 增强数据集下不同模型性能对比结果

Table 5 Performance comparison results of different models under enhanced dataset

模型 Model	P/%	R/%	$mAP_{0.5}/\%$	参数量 Params/M	计算量 GFLOPs	帧率 FPS/(帧·s <sup>-1</sup> )
Faster-RCNN	75.9	69.4	74.3	41.30	251.4	11.9
SSD	79.5	67.6	75.0	26.30	62.8	28.1
YOLOv5s	81.3	79.5	84.9	7.20	16.5	147.4
YOLOv7	80.7	78.9	84.5	36.49	103.5	100.1
YOLOv8s	82.9	80.1	85.2	11.20	28.6	117.5
YOLOv9s	81.9	78.6	84.3	7.25	26.7	173.4
YOLOv10s	83.1	80.5	85.6	8.06	21.6	251.3
YOLOv11s	83.9	81.2	86.1	9.4	21.5	176.3
YOLOv10-MECAS	86.4	82.6	90.4	8.07	24.9	231.9

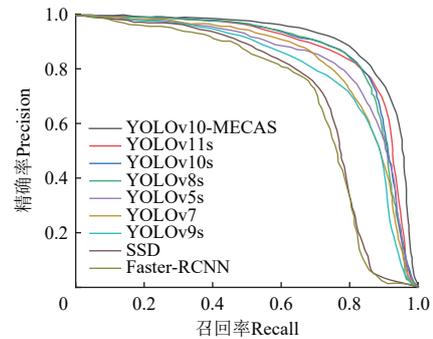


图8 增强数据集下不同模型的  $P-R$  曲线对比

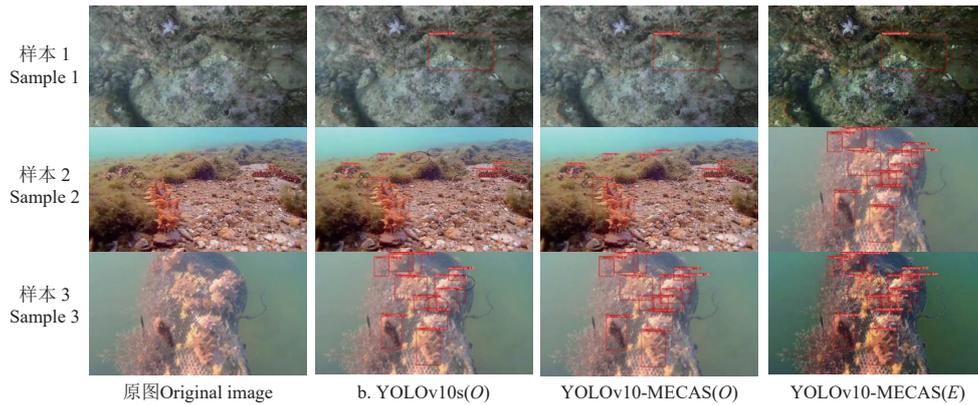
Fig.8 Comparison of  $P-R$  curves of different models in the enhanced dataset

结合表4和表5的结果分析，YOLOv10-MECAS 与对比模型在增强数据集上的  $mAP_{0.5}$  相比原始数据集均有提升。具体而言，Faster-RCNN、SSD、YOLOv5s、YOLOv7、YOLOv8s、YOLOv9s、YOLOv10s、YOLOv11s、YOLOv10-MECAS 的  $mAP_{0.5}$  分别提升了1.1、0.7、0.7、0.9、0.8、0.4、0.9、1.0、0.7个百分点。这一结果表明，图像增强算法在水下目标检测任务中具有明显提升精度作用。

综上，本文提出的目标检测模型在增强数据集上表现出优于传统目标检测模型的性能优势。

### 3.9 改进模型检测效果试验

为了验证YOLOv10-MECAS 在实际检测中的效果，选择验证集图像进行检测验证，检测效果如图9所示。可以看出，YOLOv10s 在水下海参图像数据集上检测性能低下，尤其在小目标检测和目标堆叠情况下漏检严重，如图中黑色椭圆所突出显示的。相比之下，YOLOv10-MECAS 在原始数据集和增强数据集中均有良好的表现，在处理小目标检测、密集目标分布以及目标重叠等复杂场景时表现出更优异的效果。试验结果充分证明了改进模型的有效性，展现出更高的检测准确率和更强的模型稳定性。



注：O 为原始数据集，E 为增强数据集。  
Note: O is the original dataset; E is the enhanced dataset.

图 9 YOLOv10s 与 YOLO-MECAS 检测效果

Fig.9 Detection results with YOLOv10s and YOLOv10-MECAS

## 4 结论

本研究针对海洋牧场水下复杂环境中海参目标检测任务，以 YOLOv10s 作为基线模型进行优化，提出了 YOLOv10-MECAS 水下海参检测模型。YOLOv10-MECAS 有效解决了水下复杂环境中海参目标小与背景区分难度大，光线高度弱化，图像存在大量噪声以及海参堆叠遮挡导致检测精度低等问题。试验结果表明，在原始数据集上 YOLOv10-MECAS 与 Faster-RCNN、SSD、YOLOv5s、YOLOv7、YOLOv8s、YOLOv9s、YOLOv10s、YOLOv11s 模型相比  $mAP_{0.5}$  分别提升了 16.5、15.4、5.5、6.1、5.3、5.8、5.0、4.6 个百分点；在增强数据集上，YOLOv10-MECAS 与 Faster-RCNN、SSD、YOLOv5s、YOLOv7、YOLOv8s、YOLOv9s、YOLOv10s、YOLOv11s 模型相比  $mAP_{0.5}$  分别提升了 16.1、15.4、5.5、5.9、5.2、6.1、4.8、4.3 个百分点。此外也证明了图像增强算法在水下目标检测领域是有必要的。综上，YOLOv10-MECAS 在海洋牧场水下复杂环境中海参目标检测有优异的表现，为海参自动化捕捞提供视觉检测基础。

尽管本研究在海洋牧场水下目标检测中取得了良好的效果，但仍存在一些局限性。首先，本研究未涉及面向水下场景中快速移动目标的检测应用，亦未考虑优化模型的效率及其运行速度。因此，未来的研究将集中于提升模型效率，以降低计算资源的消耗，并加快模型推理速度。此外，还将探讨针对水下快速移动目标引起的图像模糊现象的增强方法，以增强模型在水下恶劣环境下对快速移动目标的识别能力。

### 【参考文献】

- [1] 镇帅, 林远山, 盛亦凡, 等. 基于双目视觉的海参体积测量方法[J]. 农业工程学报, 2024, 40(21): 165-174.  
ZHEN Shuai, LIN Yuanshan, SHENG Yifan, et al. Volume measurement method of sea cucumber based on binocular vision[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(21): 165-174. (in Chinese with English abstract)
- [2] GUAN Z, HOU C, ZHOU S, et al. Research on underwater target recognition technology based on neural network[J]. Wireless Communications and Mobile Computing, 2022, 2022(1): 4197178.
- [3] 周丽芹, 郭太远, 姜迁里, 等. 基于改进 YOLOv3-Tiny 的端到端海参自吸捕方法研究[J]. 中国海洋大学学报(自然

- 科学版), 2023, 53(10): 115-120.
- ZHOU Liqin, GUO Taiyuan, JIANG Qianli, et al. Research on the end-to-end self-capture method of sea cucumber based on the improved YOLOv3-Tiny[J]. Journal of Ocean University of China (Natural Science Edition), 2023, 53(10): 115-120. (in Chinese with English abstract)
- [4] XU S, ZHANG M, SONG W, et al. A systematic review and analysis of deep learning-based underwater object detection[J]. Neurocomputing, 2023, 527: 204-232.
- [5] ZOU Z, CHEN K, SHI Z, et al. Object detection in 20 years: A survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [6] 崔尚, 段志威, 李国平, 等. 基于 Sobel 改进算子的海参图像识别研究[J]. 电脑知识与技术, 2018, 14(8): 145-146.  
CUI Shang, DUAN Zhiwei, LI Guoping, et al. Research on sea cucumber image recognition based on Sobel improved operator[J]. Computer Knowledge and Technology, 2018, 14(8): 145-146. (in Chinese with English abstract)
- [7] 李娟, 朱学岩, 葛凤丽, 等. 基于计算机视觉的水下海参检测研究方法[J]. 中国农机化学报, 2020, 41(7): 171-177.  
LI Juan, ZHU Xueyan, GE Fengli, et al. Research on underwater sea cucumber detection method based on computer vision[J]. Journal of Agricultural Mechanization Research, 2020, 41(7): 171-177. (in Chinese with English abstract)
- [8] GIRSHICK R, DONAHUE J, DARELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus: IEEE, 2014: 580-587.
- [9] GIRSHICK R. Fast R-CNN[C]//IEEE International Conference on Computer Vision (ICCV), Santiago: IEEE, 2015: 1440-1448.
- [10] REN S, HE K, 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, 2016, 39(6): 1137-1149.
- [11] 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, Las Vegas: IEEE, 2016: 779-788.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam: Springer, 2016: 21-37.
- [13] LIU J, LIU S, XU S, et al. Two-stage underwater object detection network using Swin Transformer[J]. IEEE Access, 2022, 10: 117235-117247.
- [14] SHI P, XU X, NI J, et al. Underwater biological detection algorithm based on improved Faster-RCNN[J]. Water, 2021, 13(17): 2420.
- [15] LI J, ZHU Y, CHEN M, et al. Research on underwater small target detection algorithm based on improved YOLOv3[C]//2022 16th IEEE International Conference on Signal Processing (ICSP), Xi'an: IEEE, 2022, 1: 76-80.
- [16] YANG Y, CHEN L, ZHANG J, et al. UGC-YOLO:

- Underwater environment object detection based on YOLO with a global context block[J]. *Journal of Ocean University of China*, 2023, 22(3): 665-674.
- [17] SUN Y, ZHENG W, 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] LI S, LI C, YANG Y, et al. Underwater scallop recognition algorithm using improved YOLOv5[J]. *Aquacultural Engineering*, 2022, 98: 102273.
- [19] LIU K, SUN Q, SUN D, et al. Underwater target detection based on improved YOLOv7[J]. *Journal of Marine Science and Engineering*, 2023, 11(3): 677.
- [20] FENG J, JIN T. CEH-YOLO: A composite enhanced YOLO-based model for underwater object detection[J]. *Ecological Informatics*, 2024, 82: 102758.
- [21] ZHOU H, KONG M, YUAN H, et al. Real-time underwater object detection technology for complex underwater environments based on deep learning[J]. *Ecological Informatics*, 2024, 82: 102680.
- [22] WANG A, CHEN H, LIU L, et al. YOLOv10: Real-time end-to-end object detection[EB/OL]. (2024-10-30) [2025-03-18] <https://arxiv.org/abs/2405.14458>.
- [23] BOCHKOVSKIY A, WANG CY, LIAO HYM. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-21) [2025-03-18] <https://arxiv.org/abs/2004.10934>.
- [24] 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, Salt Lake City: IEEE, 2018: 8759-8768.
- [25] ZONG Z, SONG G, LIU Y. DETRs with collaborative hybrid assignments training[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris: IEEE, 2023: 6748-6758.
- [26] CHEN Y, CHEN Q, HU Q, et al. DATE: Dual assignment for end-to-end fully convolutional object detection[EB/OL]. (2022-11-28) [2025-03-18] <https://arxiv.org/abs/2211.13859>.
- [27] HANG Q, JIANG Z, LU Q, et al. Split to be slim: An overlooked redundancy in vanilla convolution[EB/OL]. (2020-06-22) [2025-03-18] <https://arxiv.org/abs/2006.12085>.
- [28] MA S, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression[EB/OL]. (2023-07-17) [2025-03-18] <https://arxiv.org/abs/2307.07662>.
- [29] HE K, SUN J, TANG X. Single image haze removal using dark channel prior[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 33(12): 2341-2353.
- [30] DREWS P, NASCIMENTO E, MORAES F, et al. Transmission estimation in underwater single images[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, Sydney: IEEE, 2013: 825-830.

## Detecting underwater sea cucumber in marine ranching using improved YOLOv10s

LIU Zhenlong<sup>1,3</sup>, WANG Ji<sup>2,3\*</sup>, MAI Rengui<sup>1,3</sup>

(1. School of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang 524088, China; 2. School of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang 524088, China; 3. Guangdong Engineering Technology Research Center for Smart Ocean Sensor Networks and Equipment, Zhanjiang 524088, China)

**Abstract:** Sea cucumber is required to accurately and rapidly detect under the underwater complex environments of ocean Sea cucumber is required to accurately and rapidly detect under the underwater complex environments of ocean ranching. However, the small size of the sea cucumbers target is difficult to distinguish from the background. Particularly, the challenge of detection can also be found under the weak lighting, serious noise, and occlusion due to the overlapping sea cucumbers. Therefore, this study aims to propose the YOLOv10-MECAS model improved by the YOLOv10s baseline to enhance the performance of detection. A median pooling enhanced channel attention and a spatial attention were used to design the MECAS (median-enhanced channel and spatial attention) module. MECAS effectively retained the target features and reduced the image noise. Sea cucumber features were then captured using multiscale depth wise convolution. Additionally, the SACConv (switchable atrous convolution) module was introduced to replace the standard 3×3 convolutional module in the SCDown (spatial-channel decoupling downsampling) module. The receptive field was expanded without increasing the convolution kernel size. Thereby the model was improved to capture the features of occluded targets. An enhancement algorithm was employed on the underwater image using UDCP (the underwater dark channel prior). The dataset images were enhanced to significantly optimize the contrast for the high image quality. Furthermore, the MPDIoU (minimum points distance intersection over union) regression loss function was adopted to reduce the distortion of detection boxes caused by large sample variability. Thereby the robustness of the model was enhanced. An experiment was carried out to evaluate the performance of the improved model. A dataset of sea cucumbers was sampled from real underwater scenarios. The experimental results show that the better performance of the improved model was achieved on the original dataset, with a prediction precision of 85.7%, recall of 81.5%, and the mean average precision at IoU (intersection over union) 0.5 of 89.7%, indicating the improvement by 6.4%, 4.4%, and 5.0% over the baseline model. Compared with the comparison models Faster-RCNN, SSD, YOLOv5s, YOLOv7, YOLOv8s, YOLOv9s and YOLOv11s, while maintaining advantages in terms of the number of parameters, GFLOPs (giga floating-point operations per second), and FPS (frames per seconds), the mean average precision at IoU 0.5 had been improved by 16.5, 15.4, 5.5, 6.1, 5.3, 5.8, and 4.6 percentage points. On the enhanced dataset by UDCP algorithm, the improved model was achieved in a prediction precision of 86.4%, recall of 82.6%, and the mean average precision at IoU 0.5 of 90.4%, indicating the improvement of 3.3%, 2.1%, and 4.8% over the baseline model. Compared with the comparison models Faster-RCNN, SSD, YOLOv5s, YOLOv7, YOLOv8s, YOLOv9s and YOLOv11s, the mean average precision at IoU 0.5 had been improved by 16.1, 15.4, 5.5, 5.9, 5.2, 6.1, and 4.3 percentage points. The MECAS module into the YOLOv10s also outperformed the combination of current mainstream attention modules, such as LSKA (large selective kernel attention), CA (coordinate attention) and ECA (efficient channel attention) in underwater sea cucumber detection. Finally, the experiment verified that the YOLOv10s with the MPDIoU also performed better than that with CIoU (complete intersection over union), EIoU (enhanced intersection over union), GIoU (generalized intersection over union), DIoU (distance intersection over union), and SIoU (scaled intersection over union). Consequently, the detection accuracy of small target sea cucumbers was effectively improved in complex underwater environments. The finding can provide a theoretical basis to detect the sea cucumber during harvesting.

**Keywords:** sea cucumber; YOLOv10s; underwater target detection; attention module; loss function; underwater image enhancement