

DOI:10.19651/j.cnki.emt.2416011

# MEAS-YOLO: 改进 YOLOv5 的 水下目标智能检测算法 \*

赵永胜 严志远 毛瑞霞 吴 彭 朱宏娜

(西南交通大学物理科学与技术学院 成都 610031)

**摘要:** 针对水下光学图像目标检测中图像背景复杂、各尺度目标共存且分布广泛的问题,本文提出一种水下目标检测模型 MEAS-YOLO。首先,采用 Mosaic 和 Mixup 算法实现训练样本数据增强,提高模型的泛化能力;其次,将高效多尺度注意力机制 EMA 与 YOLOv5 网络中的骨干层融合,提高模型的特征提取能力;同时,引入自适应空间特征融合 ASFF 结构,使模型充分融合不同尺度的特征。最后采用 SIoU 损失函数,进一步提高检测精度。实验结果表明,本模型在全国水下机器人大赛 URPC2020 数据集上 mAP 达到 86.4%,较改进前提升 2.1%。本模型具有较高检测精度和小模型参数,可为水下目标高效检测提供技术支撑。

**关键词:** 水下目标检测;YOLOv5;注意力机制;自适应空间特征融合;SIoU

中图分类号: TP391; TN919.8 文献标识码: A 国家标准学科分类代码: 510.1050

## MEAS-YOLO: Improved underwater intelligent target detection algorithm of YOLOv5

Zhao Yongsheng Yan Zhiyuan Mao Ruixia Wu Zhang Zhu Hongna

(School of Physical Science and Technology, Southwest Jiaotong University, Chengdu 610031, China)

**Abstract:** Due to complex image backgrounds, multiscale coexistence and wide distribution of targets in underwater optical image target detection, an underwater target detection algorithm named MEAS-YOLO is proposed here. Firstly, this algorithm augments training samples to achieve data enhancement by utilizing the Mosaic and Mixup algorithms. Secondly, the efficient multi-scale attention module is integrated with the YOLOv5 backbone section to enhance the model's feature extraction capabilities. Simultaneously, the adaptively spatial feature fusion structure is introduced to enable the model to fully integrate features of different scales. Finally, the SIoU is used in the network model to improve detection accuracy. Experimental results demonstrate that our model has a mAP of 86.4% on the URPC 2020 dataset, improving the mAP by 2.1% than that of original model. This model exhibits high detection accuracy and lower model params, which provides a new support for precise underwater target detection.

**Keywords:** underwater target detection; YOLOv5; attention mechanism; adaptively spatial feature fusion; SIoU

## 0 引言

二十大报告提出:“发展海洋经济,保护海洋生态环境,加快建设海洋强国”<sup>[1]</sup>。海洋蕴藏着丰富的矿产、海洋生物等资源,需应用水下目标检测技术进行海底探矿、石油平台监测、海洋目标探察和海洋环境监测等工作<sup>[2-3]</sup>。因此,开展水下目标高效检测技术研究具有重要意义。

近年来, AI 技术快速发展,基于深度学习的目标检测方法相对成熟且检测效果较好<sup>[4-5]</sup>。然而,水下目标检测

模型易受到水下复杂环境影响,可能导致其特征信息少,检测精度低<sup>[6]</sup>。此外,水下目标尺度变化大且目标间存在遮挡,易造成漏检和错检。针对水下小目标难以高效检测的问题,开展了诸多研究,史朋飞等<sup>[7]</sup>在 YOLOv4 的路径聚合网络中添加同层跳跃连接和跨尺度连接结构,以增强网络在多尺度下的特征融合能力,提高模型的检测精度,但对于小目标的检测效果较差;罗飞等<sup>[8]</sup>将 FReLU 激活函数和 CBAM 注意力机制进行融合,并将其添加至 YOLOv5 骨干网络中,以增强对小目标的检测精度,但对

收稿日期:2024-05-10

\* 基金项目:中央高校基本科研业务费专项资金(202310613083)资助

于遮挡目标的检测能力较差; Lei 等<sup>[2]</sup>引入 Swin Transformer 作为 YOLOv5 的基本骨干网络, 增强了目标的特征提取能力, 同时改进多分辨率特征融合方法, 能够更有效地融合不同分辨率的图像, 但其在不同水下环境中的泛化能力较差; Chen 等<sup>[9]</sup>基于 YOLOv7 模型提出了一种 Underwater-YCC 检测模型, 采用 Conv2Former 作为水下模糊图像网络的颈部组件, 以提高对低质量图像的检测效果, 但检测时间较长, 实时性较差; Yi 等<sup>[10]</sup>提出一种基于 YOLOv8n 的改进模型 USSTD-YOLOv8n, 其采用 CARAFE 作为一种新的上采样方法, 在低水下能见度下实现更准确的特征重建, 一定程度上解决了由于水下环境能见度低导致的小目标检测精度低的问题, 但其实际应用效果需得到进一步的验证。实际中, 水下光学图像除了小目标占比多的特点外, 还存在目标跨尺度范围大且小目标分布密集的情况, 因此进一步提升复杂条件下的目标检测能力至关重要。

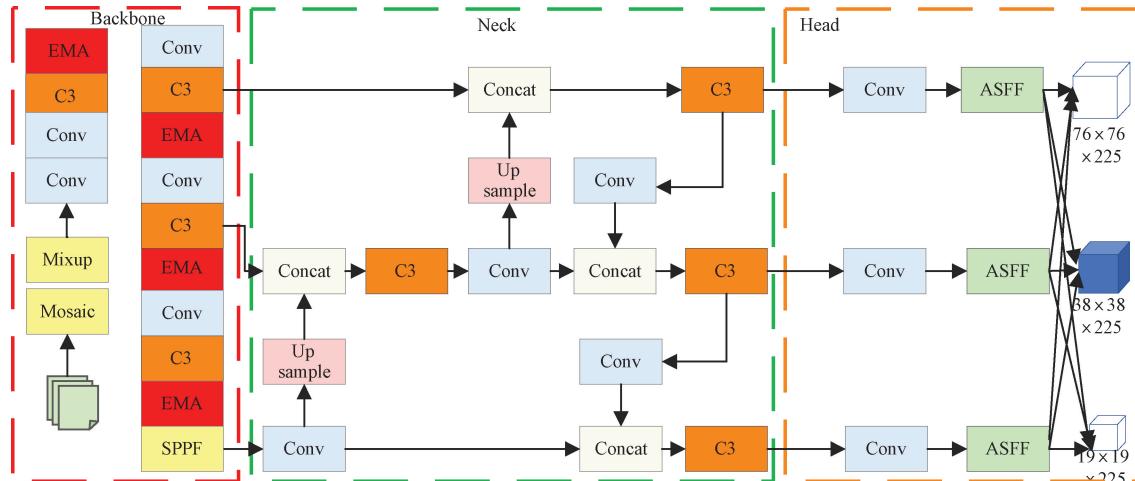


图 1 改进后的 YOLOv5 网络结构

Fig. 1 The structure of improved YOLOv5 network

针对以往大部分方法只使用单一的数据增强实现模型改进的情况, 本文采用 Mosaic 和 Mixup 混合算法实现训练样本数据增强。该方法结合了两种数据增强技术的优点, 通过多步骤的图像变换和融合, 提高数据集的多样性和丰富度, 增强了模型的泛化能力, 特别是在处理水下复杂背景和多尺度目标方面效果显著。

常见注意力机制如 CA<sup>[16]</sup>、SE<sup>[17]</sup> 和 ECA<sup>[18]</sup> 主要通过压缩和扩展通道维度来提取特征, 在一定程度上可能导致空间信息的丢失, 而本文在 Backbone 层加入 EMA 注意力机制, 其能够在保留每个通道信息的同时减少降维导致的信息丢失, 从而更有效地突出关键信息, 抑制背景干扰。

原有的 FPN 结构在处理不同特征层时缺乏灵活性, 并且在特征图融合时可能会引入一些冗余信息, 影响检测的精度和效率<sup>[19]</sup>, 而本文在 Head 层采用 ASFF 自适应空间特征融合, 通过自适应策略调整特征层之间的融合比

YOLOv5 因其优良的检测能力、模型大小和检测速度而被广泛应用于各类目标检测领域<sup>[11]</sup>, 本文基于 YOLOv5 模型, 提出一种基于注意力机制和自适应空间特征融合的水下目标检测模型。首先通过在 YOLOv5 输入端增加 Mixup-Mosaic 混合数据增强<sup>[12]</sup>, 提高模型的泛化能力。其次, 在主干提取网络增加高效多尺度注意力机制<sup>[13]</sup> (efficient multi-scale attention, EMA), 减少背景干扰, 提高模型的特征提取能力。在头部网络增加自适应空间特征融合 (adaptively spatial feature fusion, ASFF)<sup>[14]</sup>, 提高模型的特征融合能力, 实现了不同尺度特征的充分融合。最后采用 SIoU<sup>[15]</sup> 代替原网络中的损失函数, 进一步提高水下目标检测精度。

## 1 MEAS-YOLO 模型

### 1.1 网络整体结构

MEAS-YOLO 模型网络结构如图 1 所示。

例, 提升了不同尺度特征的融合能力, 使不同层的特征能够更有效地融合。

YOLOv5 原模型中采用 CIoU 作为损失函数<sup>[20]</sup>, 其综合考虑了位置、形状和方向等多个因素, 但其不能有效处理预测框和真实框的长和宽呈线性关系的情况; EIoU 是<sup>[21]</sup>在 CIoU 的惩罚项基础上将预测框和真实框的纵横比的影响因子分离, 分别计算预测框和真实框的长和宽, 从而解决 CIoU 存在的问题; 而本文采用的 SIoU 损失函数相比于以上的损失函数对目标的几何特征具有更全面的考虑, 对细长和不规则目标有较高的检测精度。

### 1.2 Mixup-Mosaic 混合数据增强

为提升模型的泛化能力, 本文设计了结合 Mixup 与 Mosaic 的混合数据增强方法。首先从数据集中随机挑选 4 张图像依次进行随机缩放、裁剪和拼接, 组合成 1 张新图像, 之后从重新拼接后的数据集中随机抽取两张图像  $x_A$

和  $x_B$  进行叠加, 叠加过程中引入权衡两个图像比例的比例因子  $\lambda$  进行加权处理, 其中  $\lambda \in [0,1]$ 。最后得到加权以后的图像, 表达式为:

$$\tilde{x} = \lambda x_A + (1 - \lambda) x_B \quad (1)$$

Mixup-Mosaic 混合数据增强的流程示意图如图 2 所示。

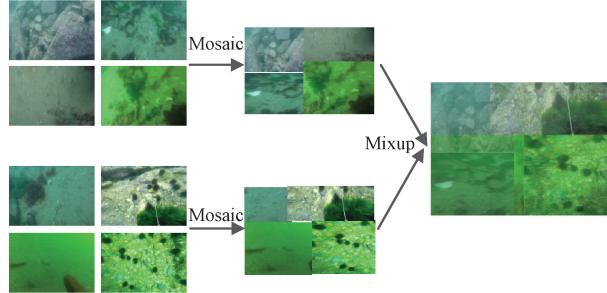


图 2 Mixup-Mosaic 混合数据增强

Fig. 2 Mixup-Mosaic hybrid data enhancement

### 1.3 EMA 注意力机制

EMA 注意力机制将部分通道维度重塑为批处理维度, 避免因卷积导致的降维。同时通过跨空间学习方法融合了两个并行子网的输出特征映射, 在保留通道信息的同时使网络更高效, 其结构如图 3 所示。

首先, 对于输入特征图  $X \in R^{C \times H \times W}$ , 其中  $C, H, W$  分别为输入特征映射的通道数量、高度和宽度。为学习不同语义, 按照跨通道维度方向将  $X$  划分为  $G$  个子特征, 然后其通过两条在  $1 \times 1$  分支上和一条在  $3 \times 3$  分支上的平行路径来提取不同特征图的注意力权重。在  $1 \times 1$  分支中, 采用两个一维全局平均池化操作, 分别沿垂直和平方向对通道进行编码, 再将两个编码特征连接, 经过  $1 \times 1$  卷积分解为两个向量后, 利用非线性 Sigmoid 函数进行线性拟合, 最后将两条分支的注意力特征进行聚合。在跨空间学习部分, 首先将  $1 \times 1$  分支输出转换为对应的维度形状  $R_3^{1 \times C//G} \times R_1^{C//G \times H \times W}$ , 之后采用二维全局平均池化对  $1 \times 1$  支

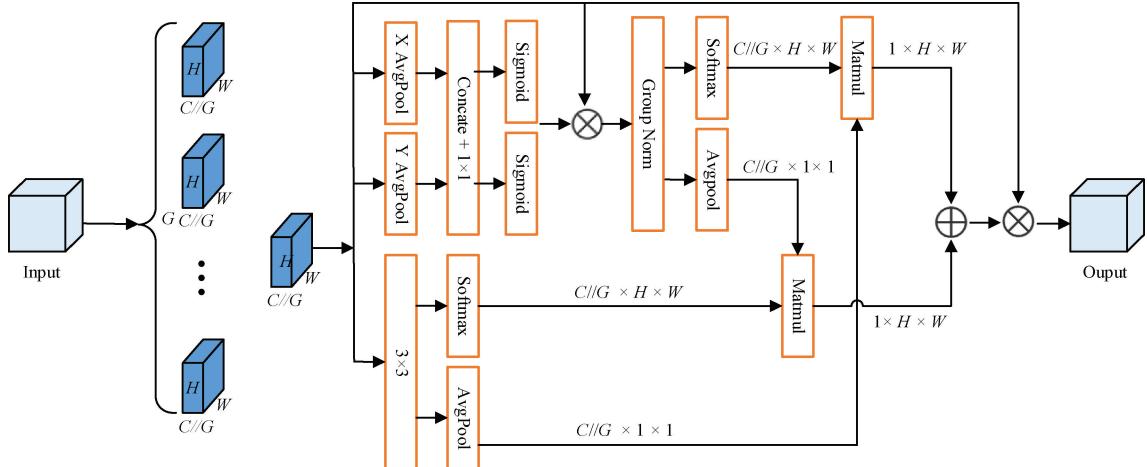


图 3 EMA 结构图

Fig. 3 The structure of EMA

路输出进行全局空间信息编码, 其计算公式如下:

$$Z_c = \frac{1}{H \times W} \sum_j^H \sum_i^W x_c(i, j) \quad (2)$$

在二维全局平均池化输出部分采用非线性 Softmax 函数拟合, 通过将上述并行处理输出与矩阵进行点积运算, 得到第 1 个空间注意力图。在  $3 \times 3$  分支中, 通过  $3 \times 3$  卷积捕获局部跨通道信息, 将其输出同样利用二维全局平均池化编码全局空间信息, 在通道特征联合激活机制前直接转换为相应的维度形状  $R_3^{1 \times C//G} \times R_1^{C//G \times H \times W}$ , 由此输出保留整个精确空间位置信息的第 2 个空间注意图。最后, 将每组内输出特征映射计算为生成的两个空间注意力权重值的集合, 再经过 Sigmoid 函数和乘法操作后输出结果。

注意力机制的主要功能是通过对特征图进行加权重构, 突出其中的关键信息, 同时抑制次要信息。而在 YOLOv5 网络中, 特征提取主要在 Backbone 层进行。因

此, 本文将 EMA 注意力机制添加到 Backbone 层 C3 模块之后。

### 1.4 ASFF 自适应空间特征融合

原 YOLOv5 网络采用 FPN 结构, 该融合方式存在降低多尺度特征融合能力的问题, 易导致网络检测精度低。因此, 本文引入 ASFF 自适应空间特征融合结构, ASFF 通过自适应方法调整不同特征层之间的融合比例, 将高层特征和底层特征进行融合, 实现不同尺度特征的充分利用。

如图 4 所示, Neck 层首先输出 3 个不同尺度的特征层, 之后经过恒等缩放和自适应融合得到 3 个 ASFF 层, 分别对应 ASFF-1、ASFF-2、ASFF-3。恒等缩放包括  $1/2$  下采样、 $1/4$  下采样和上采样。自适应融合为不同层级输出与可学习权重系数  $\alpha^3, \beta^3, \gamma^3$  乘积的和。以 ASFF-3 为例, 融合的 ASFF-3 计算表达式如下:

$$y_{ij}^l = \alpha_{ij}^l \times X_{ij}^{1 \rightarrow l} + \beta_{ij}^l \times X_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \times X_{ij}^{3 \rightarrow l} \quad (3)$$

其中,  $y_{ij}^l$  为特征图  $y^l$  再  $(i, j)$  处的输出,  $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l$  为 3 个层级特征图的可学习权重系数, 满足以下关系式:

$$\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1 \quad (4)$$

$$\alpha_{ij}^l = \frac{e^{\lambda_{\alpha_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (5)$$

$$\beta_{ij}^l = \frac{e^{\lambda_{\beta_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (6)$$

$$\gamma_{ij}^l = \frac{e^{\lambda_{\gamma_{ij}}^l}}{e^{\lambda_{\alpha_{ij}}^l} + e^{\lambda_{\beta_{ij}}^l} + e^{\lambda_{\gamma_{ij}}^l}} \quad (7)$$

其中,  $X_{ij}^{1 \rightarrow l}, X_{ij}^{2 \rightarrow l}, X_{ij}^{3 \rightarrow l}$  为某个位置特征图的输出, 权重系数  $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l$  是通过 resize 函数统一尺寸后, 对应层的特征图经过  $1 \times 1$  卷积确定, 再将权重系数进行拼接通过 Softmax 函数归一化, 将其输出调整到  $[0, 1]$  区间。

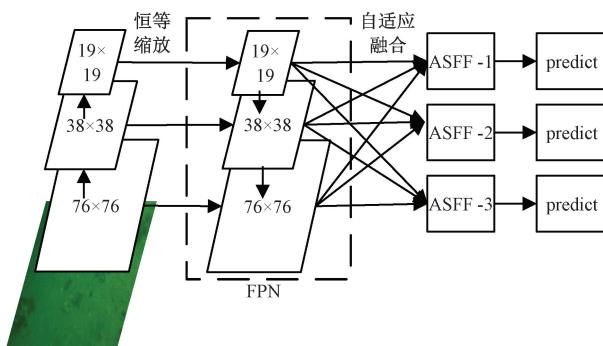
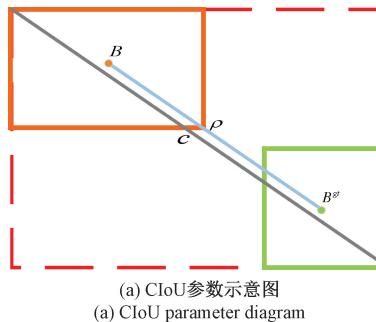
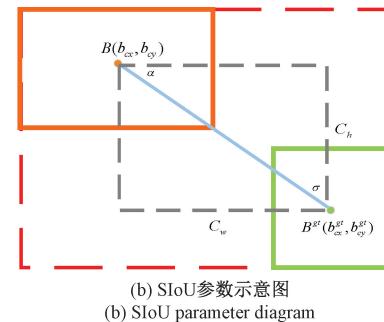


图 4 ASFF 结构

Fig. 4 The structure of ASFF



(a) CloU 参数示意图  
(a) CloU parameter diagram



(b) SIoU 参数示意图  
(b) SIoU parameter diagram

图 5 参数示意图

Fig. 5 Parameter diagram

## 2 实验对比与结果分析

### 2.1 实验数据集

本文实验采用 URPC2020 数据集<sup>[22]</sup>, 包括各深度下不同亮度、背景、对比度、模糊度和色彩偏差的图像, 能够充分模拟实际水下生物目标检测。数据集中包含海参

### 1.5 SIoU 损失函数

在 YOLOv5 模型中, 采用 CIou Loss 作为损失函数。CIou Loss 计算公式如下所示, 其参数示意图如图 5(a)所示。

$$L_{\text{CIoU}} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (8)$$

式中:  $IoU$  为真实框和预测框交并比;  $\alpha$  为权重系数;  $\rho^2(b, b^{gt})$  为预测框和真实框中心点之间的欧式距离;  $c$  表示真实框的最小包围框与预测框最短对角线长度;  $v$  表示长宽比的特性。 $\alpha$  和  $v$  的计算公式如下:

$$v = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h})^2 \quad (9)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (10)$$

CIoU 对预测框和真实框的重叠面积、中心点距离以及宽高比进行了充分考虑, 但其计算复杂性高、模型收敛慢, 在预测框与真实框不重叠时无法正确反映其紧密程度。因此, 本文采用 SIoU 损失函数替换原网络中的 CIoU 损失函数。SIoU 损失函数由交并比损失、形状损失、角度损失、距离损失构成, 其参数示意图如图 5(b)所示。

定义角度损失  $\Delta$  为:

$$\Delta = 1 - 2 \sin^2(\arcsin x - \frac{\pi}{4}) \quad (11)$$

定义形状损失  $\Omega$  为:

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^\theta \quad (12)$$

定义距离损失  $\Delta$  为:

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma p_t}) \quad (13)$$

则最终 SIoU 损失函数定义为:

$$L_{\text{SIOU}} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (14)$$

(holothurian)、扇贝 (scallop)、海星 (starfish) 和海胆 (echinus) 四类水下生物目标, 共有 5 543 张图像, 数据集按 8 : 2 划分, 随机抽取 4 434 张作为训练集, 1 109 张作为测试集。

本文所有实验均在操作系统 Windows10 下进行。实验硬件设备: CPU 为 AMD EPYC 7742, GPU 为 NVIDIA

GeForce RTX 3090; 实验环境: 深度学习框架为 Pytorch 1.11.0,CUDA 的版本为 11.3; 实验参数: 学习率(learning rate)设置为 0.01, 批次大小(batch size)设置为 64, 训练轮次(epoch)设置为 150。

## 2.2 评价指标

本实验采用平均精度均值(mean average precision, mAP)对模型进行评估, 平均精度均值与准确率和召回率有关, 计算公式如下:

$$AP = \int_0^1 P(R) dR \quad (15)$$

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (16)$$

其中,  $P$  为准确率,  $R$  为召回率, 计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (17)$$

$$R = \frac{TP}{TP + FN} \quad (18)$$

其中,  $TP$  表示分类器判定是正类的正样本数量,  $FP$  表示分类器判定是正类的负样本数量。  $FN$  表示分类器判定是负类的正样本数量。

## 2.3 实验结果及分析

### 1) MEAS-YOLO 模型实验结果

MEAS-YOLO 在 URPC2020 上进行训练, 得到的混淆矩阵如图 6 所示。在图中行表示真实标签, 列表示预测类别, 对角线数据表示正确的检测率。改进后的模型对海参、海胆、扇贝和海星四类目标的检测精度分别为 78%、91%、83% 和 87%, 其中分类错误主要是目标和背景混淆造成的。如图 6 所示, 最后一列表示将真实背景预测为其他目标的比例, 错检为背景的比例高于错检为其他目标;

最后一行表示将真实目标预测为背景的比例, 四类目标的漏检率均超过 5%。

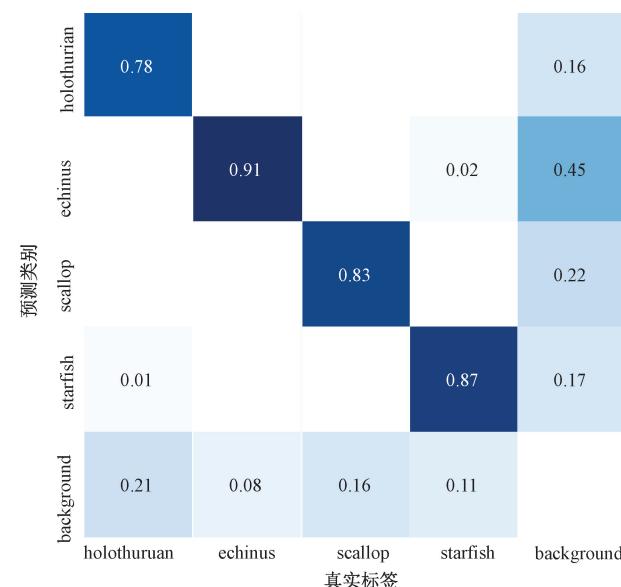


图 6 MEAS-YOLO 模型检测结果混淆矩阵

Fig. 6 Confusion matrix with MEAS-YOLO algorithm

为验证 MEAS-YOLO 模型的优越性, 将改进前后的模型进行对比分析, 结果如表 1 所示。MEAS-YOLO 模型的 mAP 有明显的提升, 较原网络, 提升 2.1%。并且对于海参、扇贝、海星和海胆四类目标, mAP 均有明显的提升, 分别提升 2.8%、1.1%、3.2% 和 1.5%。虽然引入 EMA 注意力机制和 ASFF 自适应空间特征融合带来计算量和模型大小增加, 但本模型仍属于轻量化模型, 满足在水下目标检测中模型部署的需求。

表 1 YOLOv5s 模型改进前后的结果对比

Table 1 Comparison of results before and after the improvement of YOLOv5s

模型	海参/%	扇贝/%	海星/%	海胆/%	mAP/%	模型计算量/G	模型大小/M
YOLOv5s	75.8	85.2	85.1	90.9	84.3	15.8	13.8
MEAS-YOLO	<b>78.6</b>	<b>86.3</b>	<b>88.3</b>	<b>92.4</b>	<b>86.4</b>	25.3	24.3

采用梯度加权类激活映射(gradient-weighted class activation mapping, Grad-CAM)技术<sup>[23]</sup>对原 YOLOv5s 和 MEAS-YOLO 模型进行可视化分析, 结果如图 7 所示。该激活图能够识别并突出显示图像中预测目标的位置, 有助于理解进行目标检测时模型所关注的区域。

根据图 7 可明显看出, MEAS-YOLO 模型能够抑制背景的干扰, 降低在复杂水下环境下的漏检率。

### 2) 消融实验

消融实验旨在验证各个改进模块的作用, 实验结果如表 2 所示, 其中√代表采用该模块。从表 2 实验结果可见, 引入 Mixup-Mosaic 混合数据增强之后, 显著提升了网络

的性能, mAP 提高 1.3%; 将 FPN 结构替换为 ASFF 后, mAP 提高 0.2%, 改善了网络对不同尺度特征的融合能力; 将损失函数替换为 SIoU 之后, mAP 提高 0.2%, 海参等细长型目标的检测精度得到有效提升。将四个改进模块共同加入模型之中, mAP 相较于原网络提高 2.1%, 特别对于小目标、遮挡目标和密集目标的检测精度有较高提升。

为进一步验证注意力机制改进的有效性, 将 EMA 注意力机制与 CBAM<sup>[24]</sup>、NAM<sup>[25]</sup>、SE、ECA、CA 和 SimAM<sup>[26]</sup>注意力机制进行对比。采用相同的训练参数和训练设备在 URPC2020 数据集上进行训练, 结果如表 2 所示。

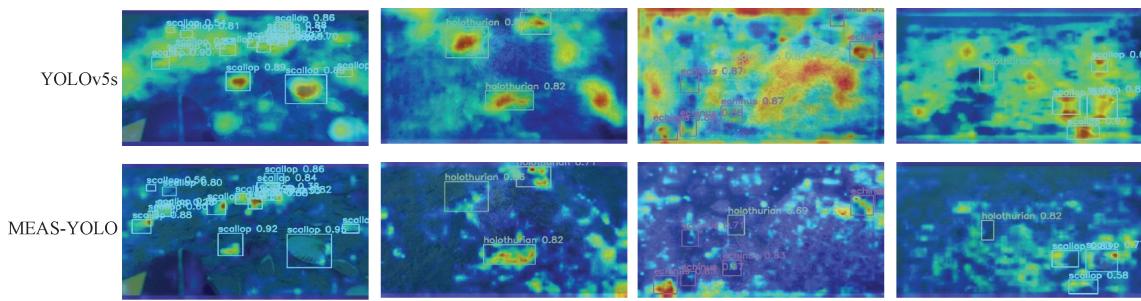


图 7 Grad-CAM 可视化结果  
Fig. 7 Grad-CAM visualization results

表 2 消融实验结果

Table 2 Results of ablation experiment

Mixup-Mosaic	ASFF	SIoU	EMA	mAP/%	海参/%	扇贝/%	海星/%	海胆/%
—	—	—	—	84.3	75.8	85.2	85.1	90.9
✓	—	—	—	85.6	76.5	86.1	87.5	92.4
✓	✓	—	—	85.8	77.4	86.4	87.9	91.4
✓	✓	✓	—	86.0	78.4	86.2	87.5	91.8
✓	✓	✓	✓	<b>86.4</b>	<b>78.6</b>	<b>86.3</b>	<b>88.3</b>	<b>92.4</b>

通过表 3 可知,与 YOLOv5s 相比,添加 CBAM 和 NAM 注意力机制后,mAP 分别提高 0.3% 和 0.2%,提升效果不显著,表明这两种注意力机制在该数据集表现较差。添加 SE、ECA 和 CA 注意力机制后,mAP 分别提高 0.4%、0.7% 和 0.5%,提升幅度较小,添加 CA 注意力机制后模型大小增加 0.6 MB。SE、ECA 和 CA 属于通道注意力机制,在提取通道注意力的过程中,空间维度信息被压缩,易损失空间位置信息。SimAM 是一种轻量级、无参数的注意力机制,嵌入网络后,mAP 提升 0.6%。EMA 注意力机制能够保留每个通道的信息,避免降维导致的问题,提高目标检测的性能。在 YOLOv5s 的 Backbone 融合 EMA 注意力机制之后,mAP 提高 0.9%,表明该注意力机制有效提升了模型的检测精度。

表 3 注意力机制对比实验

Table 3 Comparative experiment of attention mechanism

模型	mAP/%	模型大小/M
YOLOv5s	84.3	13.8
+CBAM	84.6	13.8
+NAM	84.5	13.8
+SE	84.7	13.8
+ECA	85.0	13.8
+CA	84.8	14.4
+SimAM	84.9	13.8
+EMA	<b>85.3</b>	<b>13.9</b>

检和误检情况,本实验分别采用 YOLOv5s、YOLOv8s 和 MEAS-YOLO 模型对 3 个代表性测试样本进行检测。结果如图 8 所示,红色圆圈代表漏检和错检。对第 1 行较小目标样本,YOLOv8s 存在漏检。对第 2 行细长型目标样本,MEAS-YOLO 能够检测到原 YOLOv5s 和 YOLOv8s 遗漏的目标海参。对第 3 行水下低可见度样本,MEAS-YOLO 模型相较 YOLOv5s 和 YOLOv8s 错检、漏检更少。

为进一步验证 MEAS-YOLO 模型水下目标检测性能,将 MEAS-YOLO 模型与 Faster R-CNN<sup>[27]</sup>、SSD<sup>[28]</sup>、YOLOv3<sup>[29]</sup>、YOLOv5s、YOLOv8s 等常规目标检测模型以及 3 种先进的水下目标检测模型进行对比分析。文献[30]提出了一种水下检测模型 YOLOv5-SRR,其引入 SPD-Block 和 RepBottleneck-ResNets 模块,提高小目标的特征学习能力和重叠目标的检测精度。文献[31]在 YOLOv7 在主干部分加入了 RFM,提升了模型的特征提取能力;在颈部部分采用了 FAFF 模块,实现了跨通道信息交互,有效提高了目标检测的精度,同时采用了一种训练泛式,通过两次训练达到全局最优解来解决网络参数增加后模型不易收敛到最优状态的问题。文献[32]在 YOLOv5 骨干网络中设计 C3-CA 模块,在颈部采用双向特征金字塔网络结构,提高模型的特征提取能力和不同层次特征融合能力,减少了目标信息的丢失。不同模型对比实验结果如表 4 所示。

由表 4 可知,MEAS-YOLO 模型与 Faster R-CNN、SSD 和 YOLOv3 相比属于轻量级的网络模型,且检测精度分别提升 23.7%、32.1% 和 3.2%,提升显著;改进后的 YOLOv5 模型较原 YOLOv5s 模型,mAP 提高 2.1%;相

### 3) 对比实验

为验证 MEAS-YOLO 模型能降低水下特殊场景的漏

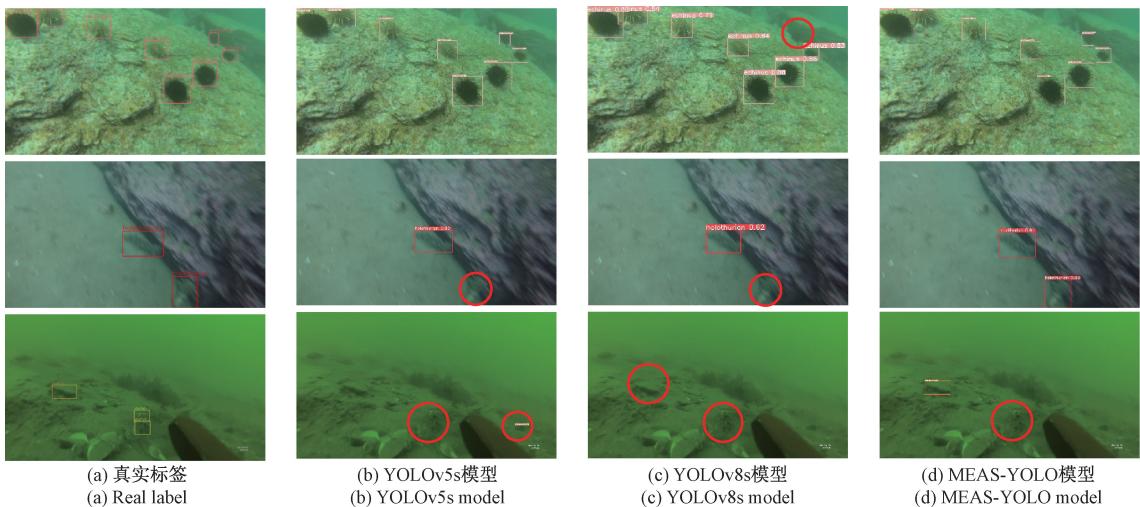


图 8 不同算法检测结果对比

Fig. 8 Comparison of detection results from different algorithms

表 4 不同模型对比实验结果

Table 4 Comparative experimental results of different algorithms

模型	mAP/%	模型大小/M
Faster R-CNN	62.7	315.0
SSD	54.3	103.0
YOLOv3	83.2	123.5
YOLOv5s	84.3	13.8
YOLOv5-SRR <sup>[30]</sup>	83.6	17.9
YOLOv8s	84.6	21.5
RF-YOLO <sup>[31]</sup>	86.3	—
Improved YOLOv5 <sup>[32]</sup>	83.5	—
MEAS-YOLO(ours)	<b>86.4</b>	24.3

比目前 YOLO 系列的较新模型 YOLOv8s, 该模型 mAP 提高 1.8%; 为进一步验证该模型的优势, 本文选择 3 种先进的水下目标检测模型进行对比。该模型相较于 YOLOv5-SRR 模型, 在保持模型较小的同时, 该模型 mAP 提高 2.8%, 实现了更高的检测精度; 相较于 RF-YOLO 模型, mAP 提高 0.1%; 相较于文献[32], mAP 提高 2.9%, 该模型具有更高的检测和小模型参数。

### 3 结 论

本文提出了一种水下目标检测模型 MEAS-YOLO。该模型通过在输入端采用 Mixup-Mosaic 混合数据增强, 提高模型的泛化能力; 在 YOLOv5 的主干网络增加 EMA 注意力机制, 提高模型的特征提取能力; 在颈部网络增加 ASFF 自适应空间特征融合, 提高模型的特征融合能力; 进一步采用 SIoU Loss 作为网络模型的定位损失函数, 提高对细长型目标的检测精度。实验表明, 改进的 YOLOv5 模

型检测效果明显优于原始 YOLOv5 模型, 在 URPC2020 数据集上 mAP 达到 86.4%, 相较于 YOLOv3、YOLOv8、Faster R-CNN 和 SSD 等检测方法, 本模型的 mAP 分别提高 3.2%、1.8%、23.7% 和 32.1%, 对复杂水下场景下小目标、密集目标、遮挡目标检测的精度有一定的提升。

### 参考文献

- [1] 孙健. 新时代建设海洋强国战略研究述评[J]. 青岛科技大学学报, 2020, 36(2): 99-104.
- [2] SUN J. Review of strategic research on building a maritime power in the new era[J]. Journal of Qingdao University of Science and Technology, 2020, 36(2): 99-104.
- [3] LEI F, TANG F F, LI SH H. Underwater target detection algorithm based on improved YOLOv5[J]. Journal of Marine Science and Engineering, 2022, 10(3): 310.
- [4] TENG B, ZHAO H J. Underwater target recognition methods based on the framework of deep learning: A survey[J]. International Journal of Advanced Robotic Systems, 2020, 17(6): 90-110.
- [5] 卫策, 吕进, 曲晨阳. 改进 YOLOv5s 的复杂交通场景下目标检测算法[J]. 电子测量技术, 2024, 47(2): 121-130.
- [6] WEI C, LYU J, QU CH Y. Improved object detection algorithm for complex traffic scenes in YOLOv5s[J]. Electronic Measurement Technology, 2024, 47(2): 121-130.
- [7] 张蕊, 高诗博, 赵霞, 等. 基于改进 YOLOv5s 的无人驾驶夜间车辆目标检测算法[J]. 电子测量技术, 2023, 46(17): 87-93.
- [8] ZHANG R, GAO SH B, ZHAO X, et al. Algorithm on nighttime target detection for unmanned vehicles based on an improved YOLOv5s [J]. Electronic Measurement Technology, 2023, 46(17): 87-93.
- [9] XU SH 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.
- [7] 史朋飞, 韩松, 倪建军, 等. 结合数据增强和改进 YOLOv4 的水下目标检测算法[J]. 电子测量与仪器学报, 2022, 36(3): 113-121.
- SHI P F, HAN S, NI J J, et al. Underwater object detection algorithm combining data enhancement and improved YOLOv4[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 113-121.
- [8] 罗飞, 王润峰. 基于 YOLOv5 水下目标检测算法研究与改进[J]. 通信与信息技术, 2024(1): 34-40.
- LUO F, WANG R F. Research and improvement of underwater target detection algorithm based on YOLOv5 [J]. Communication and Information Technology, 2024(1): 34-40.
- [9] CHEN X, YUAN M J H, YANG Q, et al. Underwater-YCC: underwater target detection optimization algorithm based on YOLOv7 [J]. Journal of Marine Science and Engineering, 2023, 11(5): 995.
- [10] YI W G, YANG J W, YAN L W. Research on underwater small target detection technology based on single-stage USSTD-YOLOv8n [J]. IEEE Access, 2024. DOI: 10.1109/ACCESS.2024.3400962.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [12] KAUR P, KHEHRA B S, MAVI E B S. Data augmentation for object detection: a review[C]. 2021 IEEE International Midwest Symposium on Circuits and Systems, 2021: 537-543.
- [13] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing, 2023: 1-5.
- [14] LIU S T, HUANG D, WANG Y H. Learning spatial fusion for single-shot object detection[C]. Computer Vision and Pattern Recognition, 2019.
- [15] DU SH J, ZHANG B F, ZHANG P, et al. An improved bounding box regression loss function based on CIoU loss for multi-scale object detection [C]. 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning, 2021: 92-98.
- [16] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [18] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11531-11539.
- [19] ZHANG D, ZHANG H W, TANG J H, et al. Feature pyramid transformer[C]. Computer Vision-ECCV 2020: 16th European Conference, 2020: 323-339.
- [20] ZHENG ZH H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [J]. AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [21] ZHANG Y F, REN W Q, ZHANG ZH, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [22] LIU CH W, LI H J, WANG SH CH, et al. A dataset and benchmark of underwater object detection for robot picking[C]. 2021 IEEE International Conference on Multimedia & Expo Workshops, 2021: 1-6.
- [23] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization [C]. IEEE International Conference on Computer Vision, 2017: 618-626.
- [24] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]. European Conference on Computer Vision, 2018: 3-19.
- [25] LIU Y CH, SHAO Z R, TENG Y Y, et al. NAM: Normalization-based attention module [J]. ArXiv preprint arXiv:2111.12419, 2021.
- [26] YANG L X, ZHANG R Y, LI L D, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]. International Conference on Machine Learning, 2021: 11863-11874.
- [27] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C]. 28th International Conference on Neural Information Processing Systems, 2015.
- [28] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, 2016: 21-37.
- [29] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv preprint arXiv: 1804.02767, 2018.
- [30] SHI J Y, WU W J. YOLOv5-SRR: Enhancing YOLOv5 for effective underwater target detection[C]. International Conference on Multimedia Modeling, 2024: 147-158.
- [31] XU H P, HE Z G, CHEN SH. Receptive field enhancement and attention feature fusion network for underwater object detection[J]. Journal of Electronic Imaging, 2024, 33(3): 033007.
- [32] YANG J T, SUN L H, WANG G. Improving the underwater object detection algorithm for YOLOv5 [C]. Third International Conference on Electronic Information Engineering and Data Processing, 2024, 13184: 1812-1818.

### 作者简介

- 赵永胜, 学士, 主要研究方向为计算机视觉、目标检测等。  
E-mail: zhao Yongsheng@my.swjtu.edu.cn
- 严志远, 学士, 主要研究方向为深度学习、光电检测。  
E-mail: yan zhiyuan@my.swjtu.edu.cn
- 毛瑞霞, 学士, 主要研究方向为信号处理。  
E-mail: ruixiam@163.com
- 吴彰, 学士, 主要研究方向计算机视觉、图像处理。  
E-mail: wz2021115126@my.swjtu.edu.cn
- 朱宏娜(通信作者), 教授, 主要研究方向为机器学习及智能信号处理、光纤通信。  
E-mail: hnzhu@home.swjtu.edu.cn