

DOI:10.19651/j.cnki.emt.2519093

# 基于 YOLO11-MDA 的水下垃圾多尺度目标检测方法<sup>\*</sup>

赵雪峰 任艺 仲兆满 仲晓敏

(江苏海洋大学计算机工程学院 连云港 222005)

**摘要:** 水下垃圾检测是维持水下生态系统平衡的关键技术。针对水下垃圾检测面临着目标尺度差异大的问题,提出一种基于 YOLO11 的水下垃圾检测算法 YOLO11-MDA。首先,提出了一种多域特征提取模块 MFEM,通过提取空域和频域的目标特征,能够从输入特征图中提取不同尺度特征,增强全局特征与局部信息的表达能力。其次,引入轻量级动态上采样 DySample 模块,融合上下文信息,提升上采样的质量和效率。最后,引入自适应阈值焦点分类损失 ATFL,降低多尺度样本分布不均衡对检测结果的影响,提高多尺度目标的检测精度。实验结果表明,相比基线模型, YOLO11-MDA 在 TrashCan 数据集和 Trash\_ICRA19 数据集的 mAP 分别达到了 91.4% 和 97%,提升 3.1% 和 10.7%,FPS 达到了 354.3 fps 的检测速度,充分说明改进的模型整体性能优于其他算法,为水下环境的自动化监测提供一种有效的解决方案。

**关键词:** 水下垃圾;多尺度目标检测;YOLO11;多分支卷积

**中图分类号:** TP391;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 510.40

## YOLO11-MDA based multi-scale target detection method for underwater trash

Zhao Xuefeng Ren Yi Zhong Zhaoman Zhong Xiaomin

(School of Computer Engineering, Jiangsu Ocean University, Lianyungang 222005, China)

**Abstract:** Underwater litter detection is a crucial technology for maintaining the balance of underwater ecosystems. To address the challenge of significant variations in target scales encountered in underwater litter detection, we propose the YOLO11-MDA based on YOLO11 is proposed. Firstly, a multidomain feature extraction module MFEM is proposed, which is capable of extracting different scales of features from the input feature map by extracting the target features in both spatial and frequency domains, and enhances the ability of expression of the global features and local information. Second, the lightweight dynamic up-sampling DySample module is introduced to integrate contextual information and improve the quality and efficiency of up-sampling. Finally, the adaptive threshold focused classification loss ATFL is introduced to reduce the impact of the uneven distribution of multi-scale samples on the detection results and improve the detection accuracy of multi-scale targets. The experimental results show that compared with the baseline model, the mAP of YOLO11-MDA in TrashCan dataset and Trash\_ICRA19 dataset reaches 91.4% and 97% respectively, which is an enhancement of 3.1% and 10.7%, and the FPS reaches the detection speed of 354.3 fps, which fully demonstrates that the overall performance of the improved model outperforms that of other algorithms, and it can provide an effective method for the automated monitoring of underwater environments.

**Keywords:** underwater trash; multi-scale target detection; YOLO11; multi-branch convolution

## 0 引言

水下垃圾,尤其是塑料垃圾,不仅破坏全球生物的多样性,还危及全球环境稳定<sup>[1]</sup>。大量研究揭示了水下垃圾对

超 700 个不同物种因为缠绕或误食水下垃圾而死亡<sup>[2-3]</sup>。得益于目标检测技术及自主水下航行器(autonomous underwater vehicles, AUVs)的发展,水下垃圾的自动检测取得重大突破。AUVs 通过水下摄像机获取目标的光学图

收稿日期:2025-06-13

\* 基金项目:国家自然科学基金(72174079)、江苏省“青蓝工程”优秀教学团队项目(2022-29)、江苏省自然科学基金青年基金(SBK2024041254)项目资助



高网络的检测精度,使其更适合部署在 AUVs 上。其次,在特征融合的上采样过程中,引入 DySample 模块,在不增加参数量的情况下,提升对小目标的检测能力。最后,引入 ATFL 作为新的损失函数,以降低类别不平衡对检测结果的影响,使模型达到最佳检测效果。改进 YOLO11 网络的整体设计如图 2 所示。

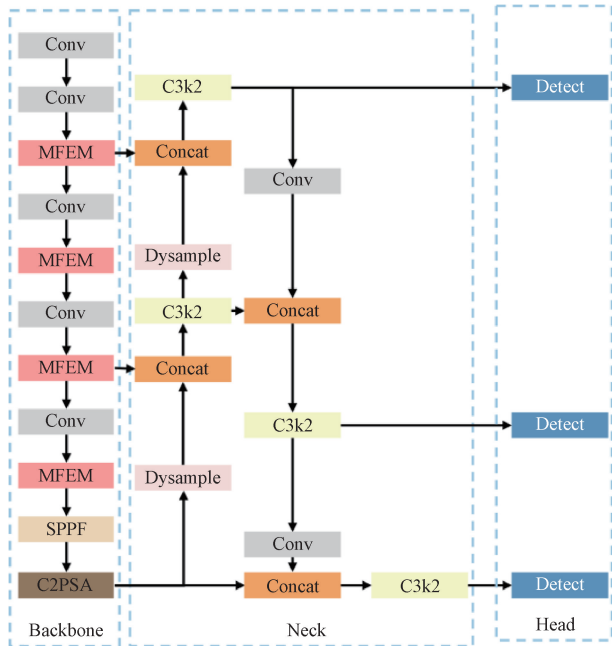


图 2 YOLO11-MDA 网络结构  
Fig. 2 YOLO11-MDA network structure

### 2.1 多域特征提取模块 MFEM

由于水下的特殊环境,导致水下图像往往对比度低,色彩失真,但频域特征能够有效增强目标特征,同时降低水下图像质量差所带来的影响。小波变换(wavelet transform, WT)<sup>[16]</sup>将图像分解为低频子带与高频子带。低频自带保留图像的整体结构与光照信息,降低水下光照不足导致的局部特征缺失。逆小波变换(inverse wavelet transform, IWT)在恢复特征时,通过低频自带重建图像的低分辨率表示,从而抑制背景干扰,降低水下色偏和噪声对检测效果的影响。此外小波变换能在保留全局信息的同时通过不同尺度的分解丰富在频域中的特征信息,深层次的提取频域中的局部和全局信息。YOLO11 的 C3k2 模块在进行多尺度局部特征融合过程中,因为浅层特征利用率不足而丢失高频细节信息,尤其是在处理形态各异和尺度变换较大的特征输入时,C3k2 模块无法有效整合不同尺度的信息,这限制了它在多尺度目标检测的表现。此外,C3k2 采用普通卷积对图像进行特征提取,在增加模型计算量的同时存在大量冗余信息。

因此提出 MFEM 来代替 C3k2 模块。该模块参考 IDC<sup>[17]</sup>的思想,MFEM 模块由特征处理分支与跳跃连接组成。输入特征经过卷积进行处理,然后馈送到 MFEM 的

4 个分支中学习局部与全局特征,保持大感受野的同时,保留丰富的多尺度特征信息。MFEM 模块如图 3 所示。

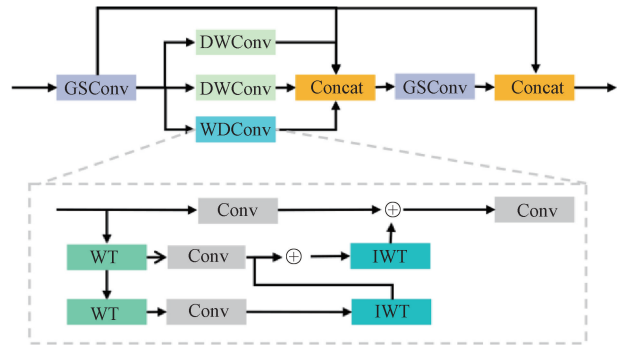


图 3 MFEM 结构  
Fig. 3 MFEM structure

输入特征  $X$  首先经过  $GSConv$ <sup>[18]</sup> 进行处理。 $GSConv$  卷积,是一种有效衡量模型准确性和速度的模块。 $GSConv$  首先进行一个卷积,然后使用  $DWConv$ ,在减少参数量的同时提高网络检测速度。然后将两个卷积结果进行拼接,使用均匀混合策略,完成通道间的信息融合。实现检测速度和精度的平衡。其结构如图 4 所示。

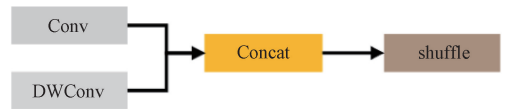


图 4 GSConv 模块  
Fig. 4 GSConv module

沿着通道维度分解为双域特征处理分支与跳跃连接分支,通过分组操作可以减少模块的计算量,从而提高模型效率。

$$X_{WT}, X_W, X_H, X_{short} = Split(X) \quad (1)$$

双域特征提取分支中的频域特征提取分支采用卷积核大小为  $3 \times 3$  的  $WDCConv$ ,通过多级 WT 将输入特征划分为低频分量与高频分量,在不同的频率分量中分别进行卷积操作后利用逆小波变换重新组合结果,增强表达低频形状信息和捕捉高频纹理细节的能力。对于输入特征,首先使用核函数进行深度卷积运算,随后采用因子为 2 的标准下采样操作,具体通过式(2)的滤波器完成步长为 2 的深度卷积。

$$f_{LL} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad f_{LH} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad (2)$$

$$f_{HL} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \quad f_{HH} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

其中,  $f_{LL}$  表示低通滤波器,  $f_{LH}$ 、 $f_{HL}$  和  $f_{HH}$  构成高通滤波器。针对每个输入通道,经卷积运算后得到的输出结果如下:

$$[X_{LL}, X_{LH}, X_{HL}, X_{HH}] = Conv([f_{LL}, f_{LH}, f_{HL}, f_{HH}], X) \quad (3)$$

该输出包含4个通道,且每个通道在各空间维度上的分辨率均为原始输入特征 $X$ 的一半。其中, $X_{LL}$ 表示输入特征 $X$ 的低频成分, $X_{LH}, X_{HL}, X_{HH}$ 分别对应水平、垂直和对角方向的高频成分。

4种滤波器形成正交基,使用转置卷积实现逆小波变换(IWT),将分解后的频率带重新组合为原始空间域的信号,其过程为WT操作的逆操作,可以表示为:

$$X = \text{Conv-transposed}([f_{LL}, f_{LH}, f_{HL}, f_{HH}], [X_{LL}, X_{LH}, X_{HL}, X_{HH}]) \quad (4)$$

其中, *transposed* 为转置操作。

WDCConv 通过在不同的频率分量上进行卷积,处理不同频率带的信息,实现对图像多尺度信息的提取,这不仅增强了模型的感受野有效捕捉不同尺度的特征,还避免了过度参数化。此外,通过对频率信息的分解与重构,采用WDCConv后模型对目标形状更加敏感,能够通过过滤复杂水下环境中的高频噪声信息,增强模型对水下干扰的鲁棒性。该分支输出表示为:

$$X'_{WT} = \text{Conv}(IWT(\text{Conv}(W, WT(X_{WT})))) \quad (5)$$

其中,  $W, WT, IWT$  分别表示卷积核权值,小波变换与逆小波变换。

特征提取分支的空域分支分别采用大小为 $1 \times 11$ 和 $11 \times 1$ 的DWConv,在宽度和高度上进行卷积操作,高效捕获宽度与高度方向上的全局信息。通过两个正交的带状卷积,使网络学习更全面的特征表示,提升对不同尺度、形状的检测能力。

$$X'_W = \text{DWConv}_{1 \times k_b}^{g \times k_b} g(X_W) \quad (6)$$

$$X'_H = \text{DWConv}_{k_b \times 1}^{g \times k_b} g(X_H) \quad (7)$$

$$X'_e = X_e \quad (8)$$

其中,  $g$  是卷积分支的通道数,  $k_b$  表示大小为11的卷积核。最后将4个分支的输出 $X'_{WT}, X'_W, X'_H, X'_e$ 按通道维度进行拼接,即:

$$X' = \text{Concat}(X'_{WT}, X'_W, X'_H, X'_e) \quad (9)$$

MFEM模块能够高效识别出水下垃圾,并增强对输入图像多尺度信息的理解,提升模型对局部和全局信息的感知能力,提高模型在水下环境中检测多尺度目标的表现。MFEM不仅能够多角度捕捉特征,获取更大范围的全局特征,提高在复杂环境中对特征的提取能力,并降低模型复杂度,提高推理速度。

## 2.2 DySample 上采样

在水下垃圾检测任务中,上采样模块对恢复特征图的边界细节和目标形状至关重要。然而,YOLO11中的上采样层使用最近邻插值算法进行上采样操作,但该方法仅依赖特征图的像素位置,忽略了上下文信息,难以有效表达不同尺度目标的特征。为解决这些问题,引入基于动态点采样的轻量化上采样模块DySample<sup>[19]</sup>,处理特征的细节和语义信息,更精准地获取水下多尺度目标的特征信息,提高模型对水下垃圾的识别能力,其结构如图5所示。

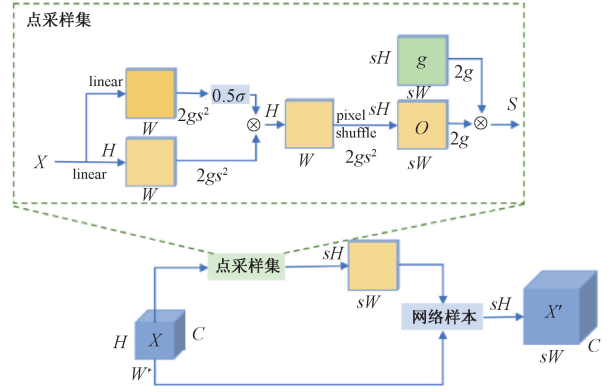


图5 DySample 结构  
Fig. 5 DySample structure

动态点采样能够根据特征信息,灵活确定采样点的位置,更有效地获取特征信息,减少冗余采样,提高采样效率。DySample从动态点采样的角度设计上采样策略,通过动态范围因子和偏移量生产机制重新采样原始特征图,并使用像素重建(pixel shuffle)模块进行特征图重塑。该方法显著降低参数数量和计算复杂度的同时,增强了特征细节的表达能力。将输入的特征图,其大小为 $C \times H \times W$ ,通过采样点生成器转换成大小为 $2 \times sH \times sW$ 的采样集。接着,使用grid sample函数对输入的特征图进行重新采样,最终获得尺寸为 $C \times sH \times sW$ 的上采样特征图。网络采样定义为:

$$X' = \text{grid sample}(X, S) \quad (10)$$

其中,  $S$  为给定的上采样比例因子,  $X$  为输入特征,  $X'$  为输出特征图。

采样点生成器是实现高效上采样的核心组件,它通过线性层、动态采样因子以及pixel shuffle子像素卷积,生成用于调整特征图采样位置的偏移量,其尺寸为 $2s^2 \times H \times W$ 。这些偏移量与原始网络相加,形成最终的采样集 $S$ 。动态采样因子增强了偏移量的灵活性,并有效缓解了偏移量之间的重叠问题。其实现为:

$$O = \text{linear}(X) \quad (11)$$

$$S = G + O \quad (12)$$

其中,  $O$  表示生成偏移量,  $G$  表示原始网络。

## 2.3 自适应阈值焦点损失 ATFL

YOLO11采用BCE Loss作为分类损失,尽管简单有效,但无法解决样本不平衡问题。BCE损失函数公式如式(13)所示。

$$L_{BCE} = -((1-y)\log(1-p) + y\log(p)) \quad (13)$$

式中:  $p$  表示模型的预测概率;  $y$  表示真实标签。

针对样本不平衡的问题,焦点损失(focal loss, FL)<sup>[20]</sup>通过引入调制因子 $(1-p_i)^\gamma$ 来减少简单样本的分类损失。FL函数定义如式(14)所示。

$$FL = -(1-p_i)^\gamma \log(p_i) \quad (14)$$

其中,  $\gamma$  是一个可调参数,用来调节调制因子的强度。

FL 损失函数通过调制因子减少简单样本损失的同时也减少了困难样本的损失值,这不利于困难样本的学习。为解决此问题,ATFL<sup>[19]</sup>在 FL 基础上做了进一步优化,有效减少简单样本的损失权重,增加分配给困难样本的损失权重。

ATFL<sup>[21]</sup>根据每个样本特征自适应地调整损失权重,根据预测概率值动态调整损失权重,来适应不同的样本。通过设置阈值将背景与目标解耦,增强与目标相关的损失并降低背景相关的损失,迫使模型更加关注目标特征,从而缓解样本不平衡。此外,ATFL 还采用了自适应设计来减少因调整超参数而导致的时间消耗。

$$\hat{p}_t = 0.05 \times \frac{1}{t-1} \sum_{i=0}^{t-1} \hat{p}_i + 0.95 \times p_t \quad (15)$$

其中,  $\hat{p}_t$  表示模型训练进度的预测值,  $p_t$  表示当前平均预测值,  $t$  为训练轮数,  $\hat{p}_i$  表示每轮训练的平均预测概率值,  $\lambda$  表示超参数。

本文采用的水下垃圾目标检测数据集包含不同光照条件下的复杂场景,挑选水下强光与弱光、偏绿、偏蓝环境下的 4 张对比图,图 6 为不同光照条件下的复杂场景对比图,图 6(a)为强光环境,图 6(b)为水下昏暗环境,图 6(c)为正常光照的水下场景,图 6(d)为偏绿环境。这些光照条件导致背景环境多变,增加了水下垃圾检测的难度。在这种情况下,ATFL 通过解耦目标与背景,利用自适应机制调整损失权重,确保模型关注目标特征而非背景特征,从而提高模型在复杂背景下的鲁棒性,确保模型在不同光照条件下均能保持良好表现。

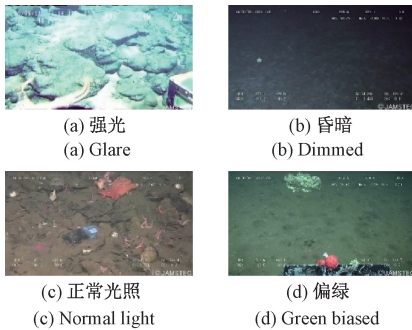


图 6 不同水下场景对比图

Fig. 6 Comparison of different water scenes

### 3 实验设计与结果分析

#### 3.1 数据集

因为水下生物可能会影响水下垃圾检测的效果,因此采用包含水下生物与水下垃圾的两个公开的水下垃圾数据集 TrashCan<sup>[22]</sup>和 Trash\_ICRA19<sup>[23]</sup>作为实验数据。将两个数据集分别按照 8:1:1 的比例随机划分为训练集、验证集和测试集。

TrashCan 数据集是专为水下垃圾实例分割设计的,它

包含了 7 212 张详细标注的图像,包含塑料、橡胶、遥控潜水器、各种海底生物等 16 个类别。图 7 展示了 TrashCan 数据中 bio、plastic、fabric 和 rov 4 个主要类别的部分图片。

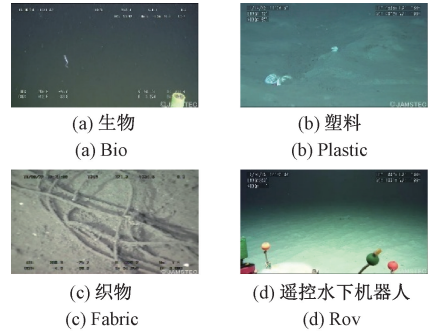


图 7 TrashCan 数据集样图

Fig. 7 Sample images of TrashCan dataset

Trash\_ICRA19 数据集主要包含深海废弃物的图像数据。图像在质量、深度、场景等方面存在显著差异。此外,该数据集中的图像清晰度和光线条件也各不相同。数据集将海洋废弃物主要分为 3 个类别:塑料(plastic)、生物(bio)、遥控水下机器人(rov)。图 8 展示了 Trash\_ICRA19 数据集中 bio、plastic 和 rov 类别的部分图片。



图 8 Trash\_ICRA19 数据类别示例图

Fig. 8 Sample images of the Trash\_ICRA19 dataset

为了提升模型的泛化能力,本文采用了 Mosaic 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、裁剪和拼接等操作后,将相应部分交叉拼接,将拼接完成后的图像缩放为指定输入特征尺寸大小,传输到 YOLO11 网络模型进行训练。这种数据增强方法增加数据的多样性、丰富图像的背景、提高模型检测多尺度目标的能力。

#### 3.2 实验环境

软件环境采用 Win11 操作系统。开发环境 Python3. 10. 16, CUDA12. 1, Pytorch2. 5. 0 深度学习框架开发环境。硬件环境采用 NVIDIA GeForce RTX4060 显卡, 8 G 显存, 处理器为 Intel(R) Core(TM) i7-14650HX 2. 20 GHz。训练过程中使用 YOLO11 作为基线模型, 调整图像分辨率至 640×640, 迭代批次为 32, 采用 SGD 优化器对模型进行优化, 初始学习率设为 0. 01, 共训练 200 个周期。

#### 3.3 评价指标

为全面评估模型性能, 本文采用以下指标: 精度 (precision, P)、召回率 (recall, R)、平均精度 (mean average

precision, mAP)、浮点运算量(giga floating point operations per second, GFLOPs)和检测速度(frame per second, FPS)。其中, P 衡量模型在所有预测为阳性的实例中正确的比例,用以反映模型的预测可靠性。R 衡量模型在所有实际为阳性的实例中正确预测的比例,用以反映模型的覆盖能力。mAP 是衡量检测精度的核心指标,用于评估模型在多类别目标检测中综合表现。GFLOPs 和 FPS 用于衡量模型的运行效率和复杂度。上述指标的计算公式如式(16)~(20)所示。

$$P = \frac{TP}{TP + FP} \quad (16)$$

$$R = \frac{TP}{TP + FN} \quad (17)$$

$$AP = \int_0^1 P(R) dR \quad (18)$$

$$mAP = \frac{\sum_1^n AP_i}{n} \quad (19)$$

$$FPS = \frac{1\ 000}{Pre + I + N} \quad (20)$$

式中: TP 为正确预测的阳性实例数, FP 为错误预测的阳性实例数, FN 为未能正确预测的阳性实例数, n 为检测类别总数, AP<sub>i</sub> 为第 i 个类别的平均精度, Pre 为图像前处理时间, I 为模型推理时间, N 为后续处理时间。

### 3.4 消融实验

通过消融实验进一步验证 YOLO11-MDA 算法的有效性,在 TrashCan 数据集上选择性地添加模块,以直观地观察不同的改进对模型性能的影响,实验结果如表 1 所示,其中 ✓ 表示使用了对应的改进策略。

表 1 消融实验

Table 1 Ablation experiment

| BaseLine | MFEM | DySample | ATFL | P            | R            | mAP          | GFLOPs     |
|----------|------|----------|------|--------------|--------------|--------------|------------|
| ✓        | ×    | ×        | ×    | 0.864        | 0.829        | 0.883        | 6.6        |
| ✓        | ✓    | ×        | ×    | 0.905        | 0.828        | 0.893        | 6.0        |
| ✓        | ×    | ✓        | ×    | 0.901        | 0.853        | 0.898        | 6.6        |
| ✓        | ×    | ×        | ✓    | 0.890        | 0.838        | 0.888        | 6.9        |
| ✓        | ✓    | ✓        | ×    | 0.887        | 0.858        | 0.904        | 6.0        |
| ✓        | ✓    | ✓        | ✓    | <b>0.913</b> | <b>0.854</b> | <b>0.914</b> | <b>6.0</b> |

表 1 的消融实验表明可知,不同的改进策略对模型的多尺度目标检测性能由不同的提升效果。MFEM 能高效的从多角度提取目标的特征信息,增强特征的表示,准确率提升了 4.1%,mAP 提升 1%,且 GFLOPs 有所降低,表明 MFEM 模块在增强了模型的特征提取能力的同时,降低了计算复杂度。当引入 DySample 时 mAP 提高了 1.5%,GFLOPs 有所降低,说明 DySample 能在提高检测精度的同时降低计算量,通过改变上采样模块降低特征信息的损失。引入 ATFL 损失函数时模型的 mAP 提升 0.5%,说明降低类别不均衡能有效提高模型的检测精度。同时添加 MFEM 模块、DySample 算子 R 提升至 85.8%,mAP 提升至 90.4%,说明这两个模块在多尺度处理上的互补性。采用 MFEM 模块、DySample 和优化后的损失函数能够显著提升 mAP,提升了 3.1%。这说明 MFEM 模块、DySample 算子和 ATFL 形成叠加增益效应,在平衡计算量的同时,大幅提升了模型对水下多尺度目标的检测效果,验证了模型的合理性。

### 3.5 对比实验

为了验证所提网络模型的有效性,采用多阶段目标检测网络 Faster R-CNN、单阶段目标检测网络 YOLOv8、YOLOv10、YOLO11、RT-DERT、以及一种针对水下垃圾检测的改进模型进行对比实验。各个模型在两个数据集

上的实验结果如表 2 所示,与其他模型相比,本文方法在检测准确性方面具有明显优势,在两个水下数据集上都获得了最高的 mAP,分别达到 91.4% 和 97%。YOLO11-MDA 模型与基线模型相比,检测精度提升了 4.9%,且降低了 0.5 GFLOPs,同时 FPS 处于前列。与同类型的 YOLOv5、YOLOv8、YOLOv10 等模型相比,检测精度有着显著优势。与最新改进的 YOLOv8-GST 相比,在实验环境相同的环境下,YOLO11-MDA 的 mAP 和 FPS 有着显著优势。综合各项指标,YOLO11-MDA 是水下垃圾分类与检测任务中表现最佳的模型。

为了直观评估 YOLO11-MDA 在水下垃圾检测任务中的性能提升,我们随机挑选了 4 张图片。通过比较各个模型在这些图片上的检测结果,深入理解本文提出方法在多尺度目标和多类别目标检测方面的优势。对比实验结果如表 3 所示。

表 3 中图片 A 显示了水下多尺度目标的检测结果,其中包括 6 个大尺度、2 个中尺度以及 6 个小尺度水下塑料目标;图片 B 显示了水下不同类别目标的检测结果,其中有 2 个塑料垃圾和 4 个其他类别垃圾;图片 C 显示了不同尺度和不同类别的检测结果,包含 1 个 rov 类别、1 个 plastic 类别、1 个 bio 类别,分别为大尺度目标、小尺度目标以及中尺度目标;图片 D 选取非数据集内的水下垃圾的检

表 2 对比实验  
Table 2 Comparative experiments

| 数据集          | 模型           | P            | R            | mAP          | GFLOPs     | FPS          |
|--------------|--------------|--------------|--------------|--------------|------------|--------------|
| TrashCan     | Faster R-CNN | 0.627        | 0.843        | 0.818        | 104.4      | 34.4         |
|              | YOLOv5       | 0.840        | 0.764        | 0.819        | 7.1        | 337.3        |
|              | YOLOv8       | 0.841        | 0.817        | 0.880        | 8.1        | 334.2        |
|              | YOLOv10      | 0.792        | 0.787        | 0.853        | 8.3        | 323.4        |
|              | YOLO11       | 0.864        | 0.767        | 0.883        | 6.3        | 320.6        |
|              | YOLOv8-GST   | 0.873        | 0.745        | 0.846        | <b>4.9</b> | 332.7        |
|              | RT-DETR      | 0.834        | 0.790        | 0.855        | 103.9      | 51.2         |
|              | 本文方法         | <b>0.913</b> | <b>0.854</b> | <b>0.914</b> | 5.8        | <b>354.3</b> |
| Trash-ICRA19 | Faster R-CNN | 0.767        | 0.874        | 0.839        | 104.4      | 49.8         |
|              | YOLOv5       | 0.916        | 0.933        | 0.959        | 7.1        | 362.5        |
|              | YOLOv8       | 0.811        | 0.904        | 0.917        | 8.1        | 345.8        |
|              | YOLOv10      | 0.880        | 0.905        | 0.943        | 8.3        | 358.5        |
|              | YOLO11       | 0.718        | 0.911        | 0.863        | 6.3        | 362.4        |
|              | YOLOv8-GST   | 0.898        | 0.841        | 0.911        | <b>4.9</b> | 364.7        |
|              | R-DETR       | 0.807        | 0.924        | 0.937        | 103.9      | 51.4         |
|              | 本文方法         | <b>0.923</b> | <b>0.951</b> | <b>0.970</b> | 5.8        | <b>372.7</b> |

表 3 不同模型目标检测效果对比  
Table 3 Comparison of different models for object detection


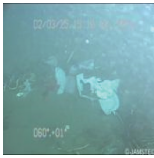



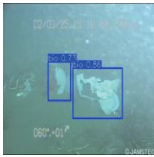

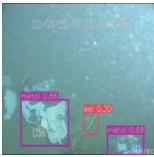
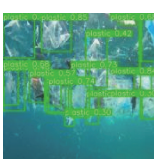

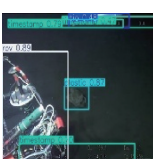
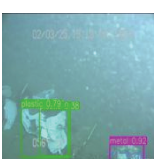




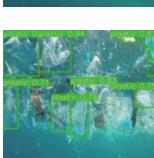

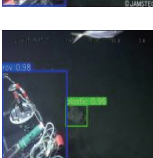


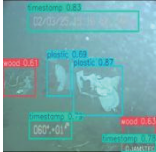

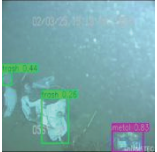
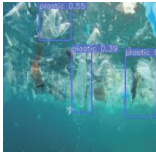


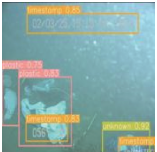
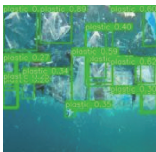







| 模型           | A   | B   | C   | D   |
|--------------|---|---|---|---|
| 原图           |  |  |  |  |
| Faster R-CNN |  |  |  |  |
| YOLOv5       |  |  |  |  |
| YOLOv8       |  |  |  |  |
| YOLOv10      |  |  |  |  |

表 3(续)  
Table 3 (continued)

| 模型         | A   | B   | C   | D   |
|------------|---|---|---|---|
| YOLO11     |  |  |  |  |
| YOLOv8-GST |  |  |  |  |
| RT-DERT    |  |  |  |  |
| 本文方法       |  |  |  |  |

测结果,包含 3 个类别目标。各个模型在图片 A、B、C 和 D 上的检测效果如表 3 所示。根据图片 A 的检测结果, Faster R-CNN 的检测效果最差,仅用一个边界框定位所有水下垃圾。YOLOv8-GST 对水下多尺度垃圾检测结果较差,仅识别出 3 个大尺度目标。YOLOv8、YOLOv10、YOLO11 表现较好,但也有漏检的情况。YOLOv5、RT-DETR 与 YOLO11-MDA 能够准确检测出不同尺度的水下,且 YOLO11-MDA 的置信度较高。因此, YOLO11-MDA 在多尺度目标的检测上最优。根据图片 B 的检测结果, YOLOv5、YOLOv8、YOLO11、YOLOv8-GST 在水下不同类别的检测上表现最差,只检测出 2 个塑料垃圾类别与 2 个其他垃圾类别。RT-DERT 表现较好,但将背景识别为其他类别垃圾,存在误检现象。YOLO11-MDA 准确识别出所有类别。根据图片 C 的检测结果, Faster R-CNN 对多尺度多类别的检测效果最差,将 rov 误识别为 plastic,并将 timestamp 误识别为 rubber,误检率高。YOLO11 和 YOLOv8-GST 出现漏检,未准确识别出生物类别。YOLOv5 与 YOLOv8 表现较好,但出现了重复检测的问题。RT-DERT 与本文方法表现效果最好。根据图片 D 的检测结果, YOLOv10、RT-DERT 表现较差,未能检测出所有类别, Faster R-CNN、YOLOv5、YOLOv8、YOLOv8-GST、YOLO11、均出现大量漏检现象,而本文方法检测效果最好,可见模型的泛化性能。综合不同情况下的检测结果, YOLO11-MDA 实现多尺度、多类别的水下垃圾检测效

果最好。

## 4 结 论

针对水下垃圾检测中多尺度目标导致的检测效果差的问题,提出 YOLO11-MDA 目标检测算法,通过一系列实验验证了其有效性。实验结果表明, YOLO11-MDA 在水下垃圾数据集 TrashCan 和 Trash\_ICRA19 上检测性能显著提升,在多尺度与多类别目标检测上精度表现优异。采用 MFEM 模块,提高对多尺度目标的特征提取能力。该模型还采用 DySample 上采样算子,增强水下垃圾的语义信息和特征表达能力,提高目标检测的鲁棒性。采用 ATFL 损失函数,降低水下垃圾分布不均衡,进一步提高检测准确度。在 TrashCan 数据集和 Trash\_ICRA19 数据集上进行了大量的实验,并于其他模型进行了对比。然而,当前模型存在计算复杂度较高的问题。未来的研究可以对模型进行剪枝,使其保持检测精度的同时实现轻量化,从而更适合水下垃圾检测的实际应用。

## 参考文献

- [1] PRAVED P H, NEETHU K V, NANDAN S B, et al. Multidimensional risk assessment of marine litter pollution in the ecologically fragile coral atolls of india [J]. Journal of Environmental Management, 2025, 376:124578.
- [2] FERREIRA N, PIRODDI C, SERPETTI N, et al.

- Potential risk of macro-plastic on the megafauna of two semi-enclosed European seas[J]. *Marine Pollution Bulletin*, 2025, 213:117683.
- [3] RUPASINGHE H P A, PERERA I J J U N, SANDARUWAN R D C, et al. Coastal beach ecosystems contaminated by marine litter: Impact on coastal biodiversity, tourism, and environmental sustainability [J]. *Environmental Pollution*, 2025, 372:126006.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580-587.
- [5] SHENG W SH, YU X F, LIN J Y, et al. Faster RCNN target detection algorithm integrating CBAM and FPN[J]. *Applied Sciences*, 2023, 13(12): 6913.
- [6] 罗偲, 李凯扬, 吴吉花, 等. 基于对抗注意力机制的水下遮挡目标检测算法[J]. *计算机工程*, 2024, 50(10): 313-321.
- LUO C, LI K Y, WU J H, et al. Underwater occlusion target detection algorithm based on adversarial attention mechanism[J]. *Computer Engineering*, 2024, 50(10): 313-321.
- [7] 王坤, 邵崇洲. Focus Meta R-CNN: 水下垃圾小样本目标检测算法[J]. *计算机工程与应用*, 2025, 61(18): 231-240.
- WANG K, SHAO CH ZH. Focus meta R-CNN: Few-shot object detection algorithm for underwater debris[J]. *Computer Engineering and Applications*, 2025, 61(18): 231-240.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. *Computer Vision-ECCV 2016*, 2016: 21-37.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779-788.
- [10] JIANG L Y, LIU F H, LYU J W, et al. GST-YOLO: A lightweight visual detection algorithm for underwater garbage detection [J]. *Journal of Real-Time Image Processing*, 2024, 21:114.
- [11] 付均尚, 田莹. 采用多信息残差融合和多尺度特征表达的水下目标检测[J]. *计算机工程与应用*, 2025, 61(11):272-283.
- FU J SH, TIAN Y. Underwater target detection using multi-information residual fusion and multi-scale feature expression [J]. *Computer Engineering and Applications*, 2025, 61(11):272-283.
- [12] HUANG J H, FANG CH, ZHENG X G, et al. YOLOv8-UC: An improved YOLOv8-based underwater object detection algorithm [J]. *IEEE Access*, 2024, 12:172186-172195.
- [13] 张明华, 黄基萍, 宋巍, 等. 基于 UMS-YOLO v7 的面向样本不均衡的水下生物多尺度目标检测方法[J]. *农业机械学报*, 2025, 56(1):388-396, 409.
- ZHANG M H, HUANG J P, SONG W, et al. Multi-scale object detection method for underwater organisms under unbalanced samples based on UMS-YOLO v7[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2025, 56(1):388-396, 409.
- [14] 王书朋, 李凡. 抗混叠与多尺度特征融合的水下目标检测算法[J]. *计算机工程与应用*, 2025, 61(18): 209-217.
- WANG SH P, LI F. Underwater object detection algorithm with anti-aliasing and multi-scale feature fusion[J]. *Computer Engineering and Applications*, 2025, 61(18):209-217.
- [15] 苗力恒, 田莹. 改进 YOLOv8n 的多尺度轻量化水下目标检测[J]. *电子测量与仪器学报*, 2025, 39(4): 141-151.
- MIAO L H, TIAN Y. Improved YOLOv8n multi-scale and lightweight underwater target detection[J]. *Journal of Electronic Measurement and Instrumentation*, 2025, 39(4):141-151.
- [16] FINDER S E, AMOYAL R, TREISTER E, et al. Wavelet convolutions for large receptive fields [C]. *Computer Vision-ECCV 2024*, 2024: 363-380.
- [17] YU W H, ZHOU P, YAN SH CH, et al. Inceptionnext: When inception meets convnext [C]. *2014 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024:5672-5683.
- [18] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: A lightweight-design for real-time detector architectures [J]. *Journal of Real-Time Image Processing* 2024, DOI:10.1007/s11554-024-01436-6.
- [19] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample[C]. *2023 IEEE/CVF*

- International Conference on Computer Vision, 2023: 6004-6014.
- [20] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. 2017 IEEE International Conference on Computer Vision, 2017: 2999-3007.
- [21] YANG B, ZHANG X Y, ZHANG J, et al. EFLNet: Enhancing feature learning network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 62:1-11.
- [22] HONG J, FULTON M, SATTER J. Trashcan: A semantically segmented dataset towards visual detection of marine debris[J]. ArXiv preprint arXiv: 2007. 08097, 2020.
- [23] SAJI S, MANIKANDAN M S, ZHOU J, et al. Underwater debris detection using visual images and YOLOv8n for marine pollution monitoring[C]. 2024 IEEE 19th Conference on Industrial Electronics and Applications, 2024: 10664718.

### 作者简介

**赵雪峰**(通信作者), 博士, 主要研究方向为数字图像处理 and 计算机视觉、多模态情感分析。

E-mail: zhaoxf@jou. edu. cn

**任艺**, 硕士研究生, 主要研究方向为目标检测。

E-mail: ry1838843470@163. com

**仲兆满**, 博士, 主要研究方向为互联网舆情监测、多模态情感分析。

E-mail: zhongzhaoman@163. com

**仲晓敏**, 硕士, 主要研究方向为软件缺陷预测。

E-mail: zhongxm@jou. edu. cn