doi:10.6041/j.issn.1000-1298.2025.01.037

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

张明华　黄基萍　宋　巍　肖启华　赵丹枫
(上海海洋大学信息学院, 上海 201306)

摘要: 针对水下目标检测面临着生物尺度变化大以及样本不均衡的问题,本文提出一种水下生物多尺度目标检测方法(Underwater multi-scale - YOLO v7,UMS - YOLO v7)。首先,设计一种由可切换空洞卷积组成的特征提取模块,该模块可在不同大小的感受野上捕获多尺度目标特征,使得提取的特征信息更加全面;其次,使用轻量级的上采样算子融合上下文信息,提高模型对目标的特征学习能力;最后,通过结合 Wise - IoU 和归一化 Wasserstein 距离两种相似性度量,提高了不同尺度目标的定位精度,同时降低了多尺度样本分布不均衡对模型的影响。实验结果表明,该模型相较于当前其他模型在检测精度方面表现出明显的提升,在 RUOD 和 DUO 数据集上平均精度均值分别达到 64.5% 和 68.9%。与 YOLO v7 模型相比,UMS - YOLO v7 提高了多种尺度目标检测精度,在 DUO 数据集上,针对大、中、小 3 种尺度目标平均精度均值分别提升 8.3、4.8、12.5 个百分点,其中小目标提升效果最为显著。与现有的其他模型相比,改进的模型具有更高的检测精度,更适用于水下生物多尺度目标检测任务,并且针对不同数据分布的样本具有泛化性和鲁棒性。

关键词: 水下生物; 多尺度目标检测; YOLO v7; 空洞卷积; 上采样算子; 相似性度量
中图分类号: TP391.41　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)01-0388-09　　**OSID**:

Multi-scale Object Detection Method for Underwater Organisms under Unbalanced Samples Based on UMS - YOLO v7

ZHANG Minghua　HUANG Jiping　SONG Wei　XIAO Qihua　ZHAO Danfeng
(*College of Information Technology*, *Shanghai Ocean University*, *Shanghai* 201306, *China*)

Abstract: In response to the challenges posed by significant variations in biological scales and the issue of sample imbalance in underwater object detection, a multi-scale object detection method for underwater organisms (UMS - YOLO v7) was proposed. Firstly, a feature extraction module was designed, comprising switchable atrous convolutions. This module captured multi-scale target features across various receptive field sizes, ensuring a more comprehensive extraction of feature information. Secondly, a lightweight universal upsampling operator was employed to fuse contextual information, enhancing the model's ability to learn features for objects. Finally, by combining two similarity metrics, Wise - IoU and normalized Wasserstein distance, the localization accuracy of targets at different scales was improved, simultaneously mitigated the impact of uneven distribution of multi-scale samples on the model. The experimental results demonstrated that the proposed model significantly enhanced detection accuracy compared with other current models, with average accuracies of 64.5% and 68.9% on the RUOD and DUO datasets, respectively. Compared with the YOLO v7 model, UMS - YOLO v7 improved multi-scale object detection accuracy, and precise detection of underwater organisms can also be achieved in complex underwater environments. On the DUO dataset, the average accuracy for large, medium, and small-scale objects was respectively increased by 8.3 percentage points, 4.8 percentage points, and 12.5 percentage points, respectively, with the most notable improvement observed for small objects. In comparison with other existing models, the improved model exhibited higher detection accuracy, and it was better suited

收稿日期: 2024 - 01 - 25　修回日期: 2024 - 04 - 01
基金项目: 国家自然科学基金项目(61972240、42106190)
作者简介: 张明华(1977—),女,副教授,博士,主要从事目标检测与跟踪研究,E-mail: mhzhang@shou.edu.cn
通信作者: 赵丹枫(1982—),女,副教授,博士,主要从事图计算理论和海洋大数据分析研究,E-mail: dfzhao@shou.edu.cn

for underwater biological multi-scale object detection tasks. Additionally, it exhibited generalization, robustness, and adaptability for samples with different data distributions.

Key words: underwater organisms; multi-scale object detection; YOLO v7; atrous convolution; upsampling operator; similarity metrics

0　引言

海洋生物是宝贵的海洋资源,其生物多样性直接反映了海洋生态的健康状况。水下生物的显著变化,如数量和种类减少等,通常与自然环境变化、水质污染以及过度的渔业捕捞等相关。通过分析不同海洋生物的聚集和分布模式,海洋学者能够深入研究生物之间的相互作用,以便对海洋生态系统有更加深刻的了解。水下目标检测作为获取海洋信息的重要途径,广泛应用于海底生物监测、水下机器人等多种场景中。

在早期目标检测的研究过程中,工作主要集中在包含单个鱼类目标的清晰图像上[1-2],但研究方法因过于依赖手工制作的特征而具有极大的局限性。卷积神经网络的发展为光学图像的水下目标检测提供了方向[3],但现有研究方法[4-13]未讨论针对多尺度目标的样本不均衡问题,忽略了在实际应用场景中目标尺度多变且分布差异大的情况,泛化能力较弱[14]。从目前的研究情况可以看出,由于水下生物的移动性,导致目标的大小和形状不断变化,多尺度目标成为了水下目标检测任务中的难点之一。此外,由于水下环境的复杂性和采集设备的局限性,水下数据的获取存在一定困难,特别是水下不同尺度目标的数据分布不均衡,导致模型在训练过程倾向于学习数量占比多的尺度目标。

针对上述问题,本文提出一种水下生物多尺度目标检测模型(Underwater multi-scale－YOLO v7, UMS－YOLO v7)。使用特征提取模块 SACBS(Switchable atrous convolution and batch normalization and silu)捕获不同尺度的特征;采用轻量级上采样算子 CARAFE(Content-aware reassembly of features)[15],根据输入特征的内容动态生成自适应上采样核,以充分利用上下文信息,使得模型更有效地理解目标与周围区域的比例关系。提出综合 WIoU(Wise－IoU)[16]和归一化 Wasserstein 距离(Normalized Wasserstein distance, NWD)[17]两种度量方式的损失函数,进一步优化边界框的回归过程,根据不同尺度目标的占比,调整损失函数对目标的真实框和预测框位置偏移的惩罚,提高网络对不同尺度目标的学习能力,以适应样本数据的不均衡性,满足水下场景检测的需求。

1　水下生物多尺度目标检测方法

1.1　整体网络结构

为解决水下目标尺度多样和数据分布不均的问题,提出基于 YOLO v7 的水下多尺度目标检测模型(UMS－YOLO v7),网络整体结构如图 1 所示。在输入端将所有原始图像的尺寸变换为 640 像素×640 像素作为模型的输入,输入图像通过改进后的骨干网络进行纹理、形状以及空间特征的提取,在 S－ELAN 网络结构中插入 SACBS 特征提取模块以获取更加全面的特征信息,并将多尺度信息输入 Neck 层中。在 Neck 层提取并融合包含浅层的位置信息和深层的语义信息的特征图,并使用 CARAFE 进行上采样操作,增强网络对特征图的学习能力。在损失函数部分,结合 WIoU 与 NWD 两种度量,提高网络对多种尺度目标特征以及不均衡数据的处理能力。最后,将 3 种不同尺度的特征图输入检测头中进行目标预测。

1.2　特征提取模块(SACBS)

在水下目标检测过程中,为了精确提取不同尺度目标的有用信息,提出 SACBS 模块,并将其插入到骨干网络的 S－ELAN 模块中。SACBS 模块主要由卷积层、批归一化层(Batch normalization, BN)和激活函数层组成。输入特征首先通过卷积层,即具有两种扩张率的可切换空洞卷积(Switchable atrous convolution, SAC)[18],生成输出特征图,再经过 BN 层对每个特征通道进行归一化操作,使得特征分布保持平稳,加速训练收敛,最后使用激活函数加强对非线性特征的学习,提高特征的表达能力。

图 2 为 SAC 示意图,由两个全局上下文模块和可切换扩张率的空洞卷积组成。在全局上下文模块中,输入特征首先在全局平均池化层中压缩,之后经过一个卷积层进行特征提取,输出的特征再添加入主干中,充分学习全局上下文信息。在 SAC 模块中包含扩张率为 1 和 3 的两种 3×3 空洞卷积,实现对不同尺度目标的特征捕捉。

普通卷积转换为 SAC 的过程表达式为

$$\mathrm{Conv}(x,w,1)\rightarrow S(x)\times\mathrm{Conv}(x,w,1)+(1-S(x))\times\mathrm{Conv}(x,w+\Delta w,r) \tag{1}$$

式中　$S(x)$——切换函数,由平均池化层和卷积层组成

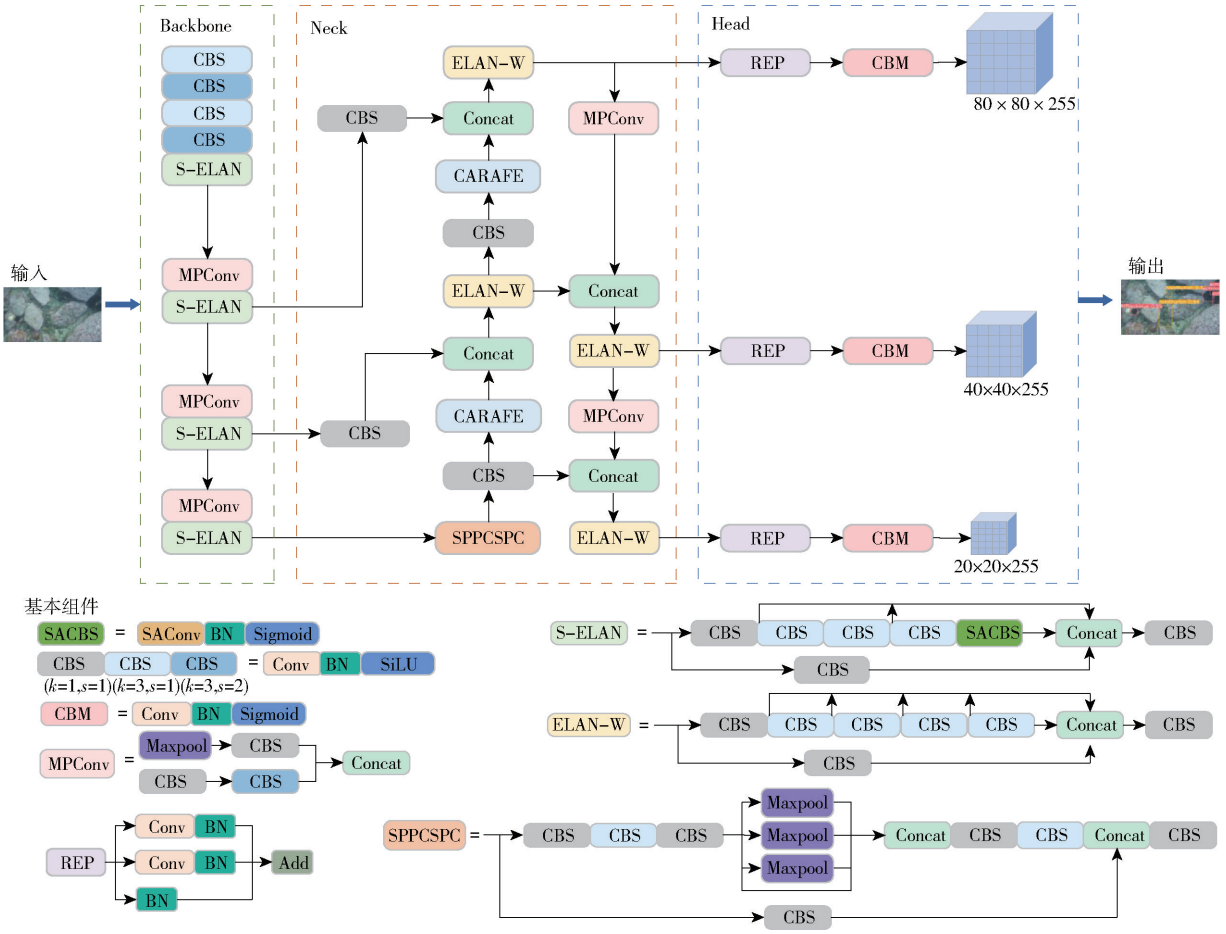


图 1 UMS－YOLO v7 网络结构图

Fig. 1 Architecture for UMS－YOLO v7 network

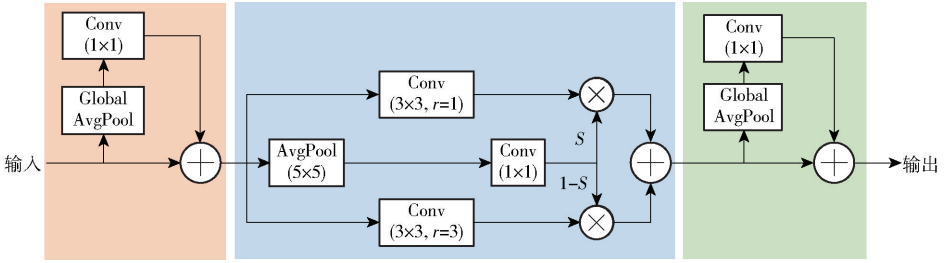


图 2 可切换空洞卷积示意图

Fig. 2 Switchable atrous convolution

x——输入　　w——初始权重

Δw——可训练的权重

r——扩张率

输入的特征以两种不同的扩张率进行卷积操作，从而得到不同感受野的特征信息。改进后的主干网络可以适应不同尺度的特征，提高水下多尺度目标的检测精度。

1.3 轻量级上采样算子(CARAFE)

在网络模型中，常用的上采样方法为最邻近插值和双线性插值，然而这些方法仅考虑了特征图上的像素值，而忽略了上下文信息。本文引入轻量级上采样算子(CARAFE)，该算子充分学习了特征图中空间信息和语义信息，以便在较大感受野中聚合上下文信息，并且没有引入过多的参数量和计算量。该采样方法可以更加精准地获取水下小目标的特征信息，提升网络对水下目标的识别性能。

如图 3 所示，CARAFE 分为上采样核预测和内容感知重组 2 个模块，首先根据每个目标位置的内容预测其重组上采样核 k_{up}，再使用预测的上采样核进行特征重组。在上采样核预测模块，首先使用 1×1 的卷积层将通道从 C 压缩到 C_m，以便减小后续步骤的计算量，再基于输入特征生成重组核，实现内容编码操作，充分利用更大感受野中的上下文信息。最后在重组核进行特征映射前，使用 softmax 函数在空间上进行归一化处理。在内容感知重组模块中，对于每个输出像素，首先确定其在输入特征图中

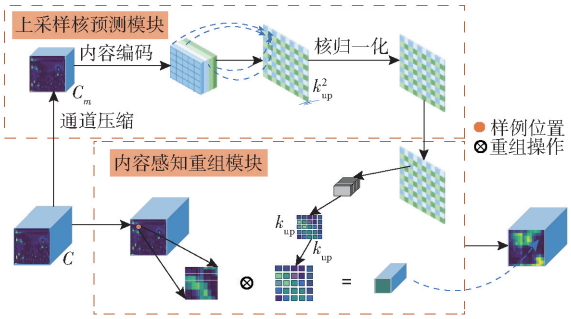


图 3　CARAFE 总体框架图

Fig. 3　Overall framework of CARAFE

的对应位置，然后选择以该输入位置为中心的 $k_{up} \times k_{up}$ 区域，与相应的像素值进行重组，以获得最终的输出像素值。

1.4　结合 WIoU 和 NWD 度量的回归损失函数

在目标检测任务中，损失函数的设计直接影响检测性能，回归的准确性则直接影响检测精度。其中，常用于度量检测精度的一个指标是交并比（Intersection over union，IoU），该指标衡量真实框与预测框之间的重叠程度。如图 4 所示（最小封闭框为包含预测框和真实框最小面积的边框），真实框 $\boldsymbol{B}_{gt} = (x_{gt}, y_{gt}, w_{gt}, h_{gt})$ 和预测框 $\boldsymbol{B}_p = (x_p, y_p, w_p, h_p)$，其中 (x_{gt}, y_{gt})、(x_p, y_p) 对应框的中心坐标，(w_{gt}, h_{gt})、(w_p, h_p) 为框的宽度和高度。真实框与预测框之间的并集面积 S_u 为

$$S_u = w_p h_p + w_{gt} h_{gt} - W_i H_i \tag{2}$$

式中　W_i、H_i——边界框重叠区域的宽、高

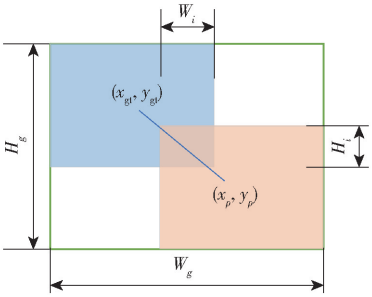


图 4　最小封闭框

Fig. 4　The smallest enclosing box

基于 IoU 的边界框回归损失 L_{IoU} 为

$$L_{IoU} = 1 - \frac{W_i H_i}{S_u} \tag{3}$$

但是，IoU 对不同尺度对象的敏感性存在显著差异，对于多尺度目标检测而言，可能导致大量小目标漏检的情况，并且水下不同尺度目标的数据分布不均衡。针对上述问题，本文方法在提高检测精度的基础上，也能保证不大幅增加计算量，提出 WIoU 和 NWD 结合策略来处理含有多种尺度目标的数据集，更加全面地考虑和训练数据集中不同尺度目标。引入两种度量方式的回归损失函数 L_{reg}，公式为

$$L_{reg} = \gamma L_{NWD} + (1 - \gamma) L_{WIoU} \tag{4}$$

其中　$$L_{NWD} = 1 - \exp\left(- \frac{\sqrt{W_2^2(N_{gt}, N_p)}}{m} \right) \tag{5}$$

$$L_{WIoU} = F \exp \frac{(x_p - x_{gt})^2 + (y_p - y_{gt})^2}{(W_g^2 + H_g^2)^2} L_{IoU} \tag{6}$$

式中　γ——度量因子

L_{NWD}——NWD 度量损失

L_{WIoU}——WIoU 度量损失

N_{gt}——真实框二维高斯分布

N_p——预测框二维高斯分布

F——非单调聚焦系数

m——数据集中目标尺度平均绝对值

W_g——最小封闭框的宽

H_g——最小封闭框的高

度量因子根据不同数据集中的小目标占比进行设值，以平衡不同的数据分布情况，当数据集中的小目标较多时，适当地增大参数值，以便损失函数更倾向于 NWD 度量，反之则减小。$W_2^2(N_{gt}, N_p)$ 表示 N_{gt}、N_p 之间的二阶 Wasserstein 距离，计算公式为

$$W_2^2(N_{gt}, N_p) = \left\| \left(\left[x_{gt} \quad y_{gt} \quad \frac{w_{gt}}{2} \quad \frac{h_{gt}}{2} \right]^{\mathrm{T}}, \left[x_p \quad y_p \quad \frac{w_p}{2} \quad \frac{h_p}{2} \right]^{\mathrm{T}} \right) \right\|_2^2 \tag{7}$$

为确保高质量标注锚框和低质量标注锚框均被分配相对较小的梯度增益，有助于模型集中关注普通质量的锚框，并防止低质量标注导致梯度损害，设置非单调聚焦系数 F，公式为

$$F = \frac{\beta}{\delta \alpha^{\beta - \delta}} \tag{8}$$

其中　$$\beta = \frac{L_{IoU}}{\overline{L}_{IoU}} \tag{9}$$

式中　α、δ——超参数，调控系数 F，分别取 1.9、3

β——离群度　　$\overline{L}_{IoU}$——L_{IoU} 动态平均值

当离群度较低时为高质量标注锚框，反之则为低质量标注锚框。$\overline{L}_{IoU}$ 使得在判断锚框质量时，其标准也能够动态变化，可以在不同情况下获得合适的梯度分配策略。

2　实验与结果分析

2.1　数据集和实验环境

使用 RUOD[19]、DUO[20] 两个水下光学图像数据集进行训练和测试。RUOD 数据集全部来自水下场景中，涵盖大规模的数据，包含 14 000 幅水下图像，共有 10 个分类：海参（holothurian）、海胆（echinus）、扇贝（scallop）、海星（starfish）、鱼（fish）、珊瑚

(corals)、潜水员(diver)、墨鱼(cuttlefish)、海龟(turtle)和水母(jellyfish),包含不同类别、尺度和密度。DUO 数据集是真实水下环境中拍摄的图像,共有 7 782 幅水下图像,包含海参、海胆、扇贝和海星 4 种类型的水下生物,平均每幅图像含有 9.57 个检测对象,并包括多种尺度目标。两个数据集中不同类别目标数量如图 5 所示。

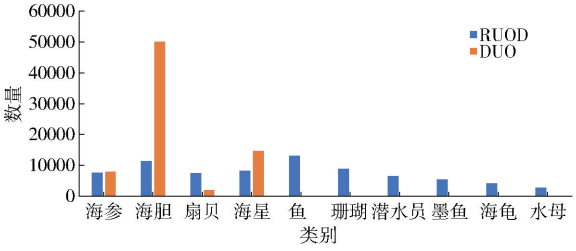


图 5 不同类别目标数量

Fig. 5 Quantities of different object categories

根据 MS COCO 数据集[21]中对目标尺度划分的定义,图 6 为两个数据集中不同尺度目标的分布情况,其中 RUOD 中大(Large)、中(Medium)、小(Small)尺度目标比例约为 61%、33%、6%,DUO 中大、中、小尺度目标比例约为 37%、59%、4%,两个数据集之间存在着一定的分布差异,并且不同尺度目标占比差异大,出现了明显的失衡。

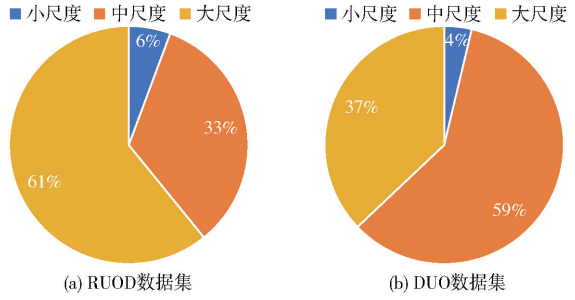


图 6 不同尺度目标占比

Fig. 6 Proportion of different scale objects

本实验在 Ubuntu 16.04 系统下进行,显卡为 NVIDIA RTX 2080ti,内存 11 GB,CUDA 版本为 11.4,pytorch 版本 1.10,编译环境为 python 3.6。在训练过程中,使用 Adam 优化器,初始学习率为 0.001。

2.2 实验结果分析

2.2.1 评价指标

使用平均精度均值(Mean average precision,mAP)指标来评价模型的性能,用来衡量多个类别的平均精度。精确率(Precision,P)是指在所有检测出的目标中检测正确的概率,也称为查准率,召回率(Recall,R)是指所有正样本中正确识别的概率,也称为查全率,平均精度(Average precision,AP)是某个类别在不同召回率下的平均精确率。此外,参数量指模型中参数总和,用以评估模型大小。

2.2.2 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[22]、Cascade R-CNN[23]、Grid R-CNN[24]和 RepPoints[25],单阶段目标检测网络 RetinaNet[26]、FreeAnchor[27]、FSAF[28]、FCOS[29]、ATSS[30]和 SSD[31],以及针对小目标检测的 SSD-EMB[32]和其他 YOLO 系列算法进行对比实验。如表 1 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,在这两个水下数据集上都实现了更高的平均精度。本文方法在检测框交并比阈值为[0.5,0.95]时,mAP 分别为 64.5% 和 68.9%,当阈值设为 0.5 时,mAP50 分别达到 88.5% 和 88.0%。但在参数量方面,该模型仍存在一定的改进空间。

图 7 中对比不同算法在 DUO 数据上各类目标的 mAP,由于扇贝类数量过少,在所有方法中,该

表 1 不同方法在 RUOD 和 DUO 数据集上的检测结果

Tab. 1 Detection results of different methods on RUOD and DUO datasets

模型	主干	RUOD		DUO		参数量
		mAP/%	mAP50/%	mAP/%	mAP50/%	
Faster R-CNN	ResNetXt101	52.8	81.8	53.8	75.4	6.013×10^7
Cascade R-CNN	ResNetXt101	54.8	81.1	56.0	76.1	8.793×10^7
Grid R-CNN	ResNetXt101	53.7	81.1	55.6	75.6	8.324×10^7
RepPoints	ResNet101	55.4	83.7	55.4	79.0	5.560×10^7
RetinaNet	ResNetXt101	50.7	79.3	50.4	71.7	5.516×10^7
FreeAnchor	ResNetXt101	55.0	82.4	54.6	76.9	5.516×10^7
FSAF	ResNetXt101	48.7	78.5	54.6	78.7	5.501×10^7
FCOS	ResNetXt101	50.7	79.5	53.2	77.3	5.078×10^7
ATSS	ResNet101	52.9	80.3	57.6	79.4	5.089×10^7
SSD	VGG	43.4	73.4	49.0	70.1	2.732×10^7
SSD-EMB	VGG	51.2	77.8	56.5	79.8	3.156×10^7
YOLO v5	CSPDarkNet	58.9	84.4	63.6	83.0	7.030×10^6
YOLO v8[33]	CSPDarkNet	57.4	83.4	64.3	83.4	1.121×10^7
本文方法	CSPNet	64.5	88.5	68.9	88.0	3.749×10^7

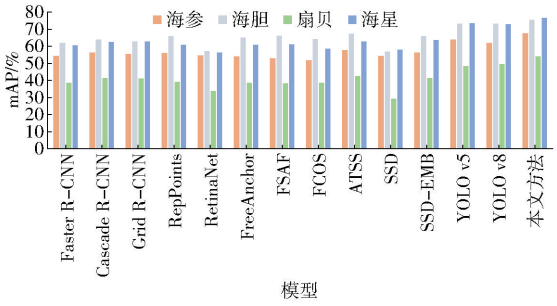


图 7　不同类别目标检测精度对比

Fig. 7　Comparison of detection accuracy of different categories of objects

类精度都低于其他类。相较于其他模型，本文方法在海参、海胆、扇贝和海星类上都取得了最高的平均检测精度。

表 2 比较了不同模型对不同尺度目标的检测结果。在 RUOD 数据集上，本文方法 3 种尺度目标的平均精度均值分别达到 20.1%、48.8% 和 70.9%，在 DUO 数据集上，平均精度均值达到 53.2%、68.4% 和 67.8%。在两个数据集上，本文模型在 3 种尺度目标的检测任务中都达到了最高精度，表明本文方法在多尺度目标检测任务中有着良好的表现。

表 2　不同尺度目标检测平均精度均值对比

Tab. 2　Comparison of mAP of different scale objects　%

数据集	尺度	Faster R－CNN	Cascade R－CNN	Grid R－CNN	RepPoints	RetinaNet	FreeAnchor	FSAF	FCOS	ATSS	SSD	SSD－EMB	YOLO v5	YOLO v8	本文方法
RUOD	小	17.2	16.8	17.7	17.7	14.3	17.0	15.7	18.0	16.4	11.7	14.6	16.6	11.6	20.1
	中	40.9	42.2	41.2	43.3	39.2	42.9	38.0	40.0	41.1	31.6	38.7	44.2	42.7	48.8
	大	58.2	60.6	59.1	60.8	56.1	60.7	53.9	56.2	58.6	48.4	55.9	65.0	64.2	70.9
DUO	小	39.0	51.2	45.6	42.2	34.6	36.5	46.0	43.4	46.5	36.7	39.2	42.5	48.9	53.2
	中	55.2	57.5	57.1	57.3	52.8	56.5	57.1	55.4	60.3	52.1	58.6	64.7	65.9	68.4
	大	52.8	54.7	54.5	53.9	49.0	52.9	52.2	51.2	55.0	46.9	54.1	63.2	61.8	67.8

图 8 为不同模型与本文方法对不同尺度目标的检测效果。从图中可以明显看到 Faster R－CNN 未检测出重叠的鱼类，在目标尺度对比较大的情况下未能准确检测小尺度鱼类，同时对珊瑚和阴影产生了误检；Cascade R－CNN 和 Grid R－CNN 对尺度较小的鱼类存在漏检，错将单个珊瑚目标识别为多个目标，并且 Grid R－CNN 将阴影误检为鱼类；RepPoints 和 FreeAnchor 漏检了纹理特征不明显的小目标；RetinaNet 和 ATSS 将水下废弃物错分为潜水员，并漏检了较多的小目标；FSAF 和 FCOS 在较为复杂的水下环境中存在大量漏检的情况；SSD 漏检了重叠的鱼类和大量小目标；SSD－EMB 对小目标有着较好的检测效果，但未能识别重叠目标以及误检珊瑚和阴影；YOLO v5 和 YOLO v8 对重叠目标同样产生了漏检的情况，并且 YOLO v5 误检了珊瑚目标。从图中检测结果可以看出，在面对目标多尺度、目标遮挡以及目标特征模糊的情况时，本文方法均表现出良好的检测效果。

2.2.3　消融实验

通过消融实验进一步分析验证模型每部分有效性。表 3 展示了在 DUO 数据集上各个模块对实验结果的影响，首先基础模型精度为 61.1%，在添加 SACBS 模块后，精度达到 67.3%，中、大尺度目标精度分别提高 5.5、7.8 个百分点。在加入轻量级上采样算子后，精度提高 1.3 个百分点，小目标精度提高 4.1 个百分点。结合两种度量的损失函数相较于单一度量方式，小目标精度提高 11.7 个百分点，并且在其他尺度目标检测上都保持较高的精度，说明当数据集中各尺度目标样本不均衡时，该方法也能保持检测结果的准确性。本文模型在 DUO 数据集上精度最终达到 68.9%，较原模型提升 7.8 个百分点，并且对参数量的影响较小。从表 3 中可以清楚地发现本文方法中所添加的模块均提高了检测精度，并且针对不同尺度目标的检测精度均有着一定的提升，尤其是对小目标的检测精度提升 12.5 个百分点。

图 9 给出了通过 Grad－CAM[34] 得到的包含不同尺度目标的可视化热力图，通过对比图 9c、9d 可知，本文方法在面对多尺度检测任务时，检测更加全面，提高了对目标的定位能力以及小目标的检测效果，并且目标区域的显著性更高。

图 10 为在面对水下小目标任务时，原模型和改进后模型检测效果对比。由图 10 可知，YOLO v7 模型存在漏检情况，而本文方法识别出了改进前未识别出的海星、海胆等目标。在图 11 中（每幅图中均包含不同尺度目标），当目标尺度差异较大时，改进前的网络出现了大量小目标漏检问题，而本文方法则成功识别出大部分目标，展现出良好的检测效果。通过对比以上检测结果，本文提出的 UMS－YOLO v7 网络在面对小目标以及水下目标多尺度情



(a) 原图 (b) 真值 (c) Faster R-CNN (d) Cascade R-CNN

(e) Grid R-CNN (f) RepPoints (g) RetinaNet (h) FreeAnchor

(i) FSAF (j) FCOS (k) ATSS (l) SSD

(m) SSD-EMB (n) YOLO v5 (o) YOLO v8 (p) 本文方法

图 8 不同算法检测结果对比

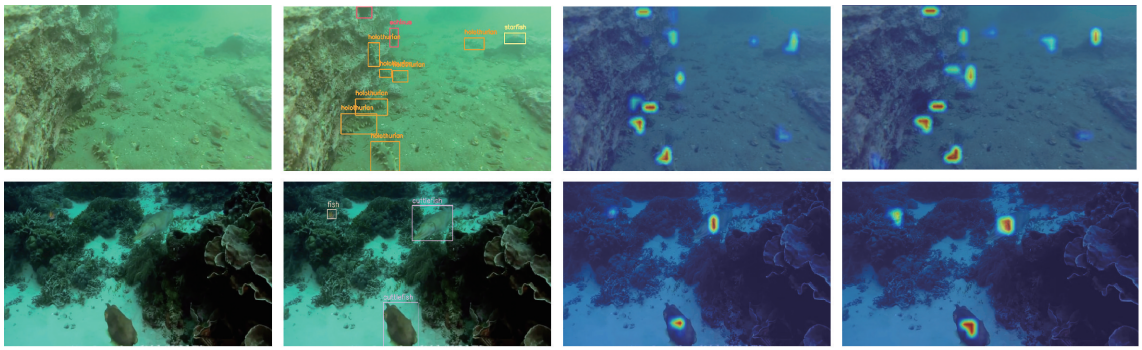
Fig. 8 Comparisons of detection results of different algorithms

表 3　DUO 数据集上消融实验结果

Tab. 3　Results of ablation experiments on DUO dataset

实验编号	SACBS	CARAFE	WIoU	NWD	mAP/%	mAP/%			参数量
						小尺度	中尺度	大尺度	
1	×	×	×	×	61.1	40.7	63.6	59.5	3.480×10^7
2	√	×	×	×	67.3	37.2	69.1	67.3	3.756×10^7
3	×	√	×	×	65.0	42.1	66.9	63.5	3.350×10^7
4	×	×	√	×	65.8	42.8	67.3	63.8	3.484×10^7
5	×	×	×	√	66.5	46.9	66.6	65.0	3.480×10^7
6	√	√	×	×	68.6	41.3	69.2	68.1	3.660×10^7
7	√	√	√	×	68.7	41.5	69.0	68.6	3.680×10^7
8	√	√	√	√	68.9	53.2	68.4	67.8	3.749×10^7

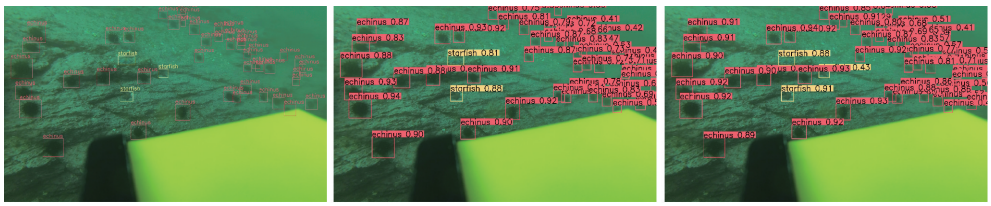
注：√表示使用该模块，×表示不使用该模块。



(a) 原图　(b) 真值　(c) YOLO v7　(d) 本文方法

图 9　特征可视化效果对比

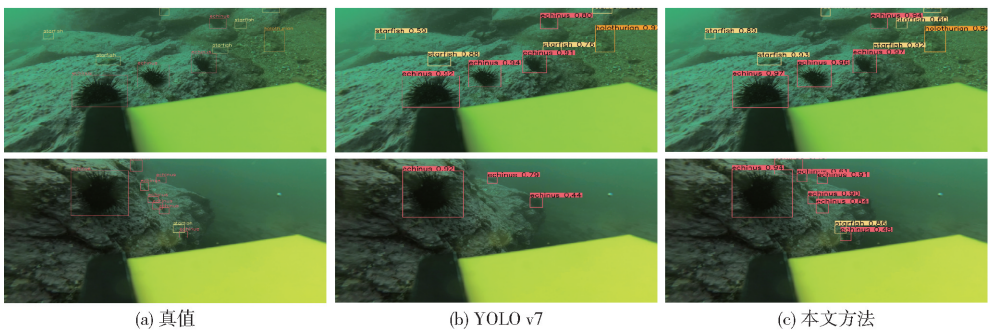
Fig. 9　Feature visualization effect comparison diagrams



(a) 真值　(b) YOLO v7　(c) 本文方法

图 10　水下小目标检测结果对比

Fig. 10　Underwater small object detection results



(a) 真值　(b) YOLO v7　(c) 本文方法

图 11　水下不同尺度目标检测结果对比

Fig. 11　Detection results of underwater object at different scales

况时，均表现出较为出色的检测性能。

3　结论

（1）提出了一种水下生物多尺度目标检测模型。通过 SACBS 特征提取模块，提高了模型对水下多尺度目标的检测精度。通过引入 CARAFE 聚合上下文处理特征信息，提高了网络的特征表达能力。在实际应用中，本文模型更能适应水下环境中生物多尺度目标的特性。

（2）针对水下数据集分布不均衡问题，提出了结

合两种度量方式的边界框回归损失函数。解决了 IoU 度量的尺度敏感性和数据样本不均衡导致模型在训练过程中产生方向偏离的问题,平衡了模型对多种尺度目标的检测精度。对于在实际采集数据过程中造成的数据分布不均情况具有良好的泛化能力。

(3)与其他目标检测算法相比,UMS - YOLO v7 在检测精度上有着明显的优势,在 RUOD、DUO 公开水下数据集中都取得了良好的检测结果,mAP 分别达到 64.5%、68.9%,更适用于真实的复杂水下环境,能够为监测水下生物提供良好的技术支持。

参 考 文 献

[1] FOUAD M M M, ZAWBAA H M, EL-BENDARY N, et al. Automatic nile tilapia fish classification approach using machine learning techniques[C]//13th International Conference on Hybrid Intelligent Systems (HIS 2013), 2013: 173 - 178.

[2] ALSMADI M K, OMAR K B, NOAH S A, et al. Fish recognition based on robust features extraction from size and shape measurements using neural network[J]. Journal of Computer Science, 2010, 6(10): 1088 - 1094.

[3] 赵德安, 刘晓洋, 孙月平, 等. 基于机器视觉的水下河蟹识别方法[J]. 农业机械学报, 2019, 50(3): 151 - 158.
ZHAO Dean, LIU Xiaoyang, SUN Yueping, et al. Detection of underwater crabs based on machine vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(3): 151 - 158. (in Chinese)

[4] LI A, YU L, TIAN S. Underwater biological detection based on YOLOv4 combined with channel attention[J]. Journal of Marine Science and Engineering, 2022, 10(4): 469.

[5] WANG J, YU N. UTD - Yolov5: a real-time underwater targets detection method based on attention improved YOLOv5[J]. arXiv Preprint, arXiv: 2207.00837, 2021.

[6] CHEN G, MAO Z, WANG K, et al. HTDet: a hybrid transformer-based approach for underwater small object detection[J]. Remote Sensing, 2023, 15(4): 1076.

[7] LYU L, LIU Y, XU X, et al. EFP - YOLO: a quantitative detection algorithm for marine benthic organisms[J]. Ocean & Coastal Management, 2023, 243: 106770.

[8] JALAL A, SALMAN A, MIAN A, et al. Fish detection and species classification in underwater environments using deep learning with temporal information[J]. Ecological Informatics, 2020, 57: 101088.

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

[10] KARIMANZIRA D, RENKEWITZ H, SHEA D, et al. Object detection in sonar images[J]. Electronics, 2020, 9(7): 1180.

[11] PANG Y, WU C, WU H, et al. Over-sampling strategy-based class-imbalanced salient object detection and its application in underwater scene[J]. The Visual Computer, 2023, 39(5): 1959 - 1974.

[12] WANG H, XIAO N. Underwater object detection method based on improved Faster RCNN[J]. Applied Sciences, 2023, 13(4): 2746.

[13] LI Y, WANG B, LI Y, et al. Underwater object tracker: uostrack for marine organism grasping of underwater vehicles[J]. Ocean Engineering, 2023, 285: 115449.

[14] 刘世晶, 刘阳春, 钱程, 等. 基于 CycleGAN 和注意力增强迁移学习的小样本鱼类识别[J]. 农业机械学报, 2023, 54(增刊 1): 296 - 302.
LIU Shijing, LIU Yangchun, QIAN Cheng, et al. Recognition of small sample cultured fish based on CycleGAN and attention enhanced transfer learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(Supp. 1): 296 - 302. (in Chinese)

[15] WANG J, CHEN K, XU R, et al. CARAFE: content-aware reassembly of features[C]//International Conference on Computer Vision (ICCV), 2019: 3007 - 3016.

[16] TONG Z, CHEN Y, XU Z, et al. Wise - IoU: bounding box regression loss with dynamic focusing mechanism[J]. arXiv Preprint, arXiv: 2301.10051, 2023.

[17] WANG J, XU C, YANG W, et al. A normalized gaussian wasserstein distance for tiny object detection[J]. arXiv Preprint, arXiv: 2110.13389, 2021.

[18] QIAO S, CHEN L C, YUILLE A. DetectoRS: detecting objects with recursive feature pyramid and switchable atrous convolution[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 10208 - 10219.

[19] FU C, LIU R, FAN X, et al. Rethinking general underwater object detection: datasets, challenges, and solutions[J]. Neurocomputing, 2023, 517: 243 - 256.

[20] LIU C, LI H, WANG S, et al. A dataset and benchmark of underwater object detection for robot picking[C]//2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2021: 1 - 6.

[21] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: common objects in context[C]//Computer Vision-ECCV 2014: 13th European Conference, 2014: 740 - 755.

[22] 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.

(下转第 409 页)

[25] 韦宝婧, 苏杰, 胡希军, 等. 基于“HY－LM”的生态廊道与生态节点综合识别研究[J]. 生态学报, 2022, 42(7): 2995－3009.
WEI Baojing, SU Jie, HU Xijun, et al. Comprehensive identification of eco-corridors and eco-nodes based on “HY－LM”[J]. Acta Ecologica Sinica, 2022, 42(7): 2995－3009. (in Chinese)
[26] 俞孔坚, 王思思, 李迪华, 等. 北京市生态安全格局及城市增长预景[J]. 生态学报, 2009, 29(3): 1189－1204.
YU Kongjian, WANG Sisi, LI Dihua, et al. The function of ecological security patterns as an urban growth framework in Beijing City[J]. Acta Ecologica Sinica, 2009, 29(3): 1189－1204. (in Chinese)
[27] 宋宏利, 王悦悦, 张启斌, 等. 邯郸西部生态空间网络结构时空演变及其与固碳功能相关性研究[J]. 农业机械学报, 2024, 55(9): 226－237, 248.
SONG Hongli, WANG Yueyue, ZHANG Qibin, et al. Spatial-temporal evolution of ecological spatial network structure in western Handan region and its correlation with carbon sequestration function[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(9): 226－237, 248. (in Chinese)
[28] 侯宏冰, 郭红琼, 于强, 等. 基于 LMBA 策略的鄂尔多斯市生态空间网络优化[J]. 农业机械学报, 2021, 52(1): 219－227.
HOU Hongbing, GUO Hongqiong, YU Qiang, et al. Ecospatial network optimization in Ordos based on LMBA strategy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(1): 219－227. (in Chinese)
[29] 汪昊, 吴正平, 程洁, 等. 基于复杂网络的电网结构静态鲁棒性分析[J]. 通信电源技术, 2018, 35(1): 33－35, 38.
WANG Hao, WU Zhengping, CHENG Jie, et al. Static robustness analysis of power network structure based on complex network[J]. Telecom Power Technology, 2018, 35(1): 33－35, 38. (in Chinese)
[30] 朱强, 俞孔坚, 李迪华. 景观规划中的生态廊道宽度[J]. 生态学报, 2005, 25(9): 2406－2412.
ZHU Qiang, YU Kongjian, LI Dihua. The width of ecological corridor in landscape planning[J]. Acta Ecologica Sinica, 2005, 25(9): 2406－2412. (in Chinese)
[31] 高贵兵, 岳文辉, 欧文初. 制造系统局域世界网络演化模型及其复杂性度量[J]. 计算机集成制造系统, 2019, 25(8): 2016－2024.
GAO Guibing, YUE Wenhui, OU Wenchu. Multi-local-world network modeling and complexity measure of manufacturing system [J]. Computer Integrated Manufacturing Systems, 2019, 25(8): 2016－2024. (in Chinese)

(上接第 396 页)

[23] LABAO A B, NAVAL P C. Cascaded deep network systems with linked ensemble components for underwater fish detection in the wild[J]. Ecological Informatics, 2019, 52: 103－121.
[24] LU X, LI B, YUE Y, et al. Grid R－CNN[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 7355－7364.
[25] YANG Z, LIU S, HU H, et al. RepPoints: point set representation for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 9656－9665.
[26] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 99: 2999－3007.
[27] ZHANG X, WAN F, LIU C, et al. FreeAnchor: learning to match anchors for visual object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 3096－3109.
[28] ZHU C, HE Y, SAVVIDES M. Feature selective anchor-free module for single-shot object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 840－849.
[29] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 9626－9635.
[30] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 9756－9765.
[31] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the Computer Vision—ECCV 2016, 2016: 21－37.
[32] CHOI H T, LEE H J, KANG H, et al. SSD－EMB: an improved SSD using enhanced feature map block for object detection [J]. Sensors, 2021, 21(8): 2842.
[33] REIS D, KUPEC J, HONG J, et al. Real-time flying object detection with YOLOv8[J]. arXiv Preprint, arXiv: 2305.09972, 2023.
[34] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad－CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336－359.