文章编号：1674-5566(2025)03-0696-11 DOI：10. 12024/jsou. 20240404474

YOLO-U：基于结构重参数化和双重注意力机制的水下目标检测算法

李江川[1]，韩彦岭[1]，董传胜[2]，王 艳[1]，王 静[1]，张 云[1]，杨树瑚[1]

（1. 上海海洋大学 信息学院，上海 201306；2. 山东省国土测绘院，山东 济南 250102）

摘 要：水下目标检测算法的研究是实现水下机器人智能捕捞的前提。水下目标检测任务中存在的目标模糊、小目标众多以及相互遮挡等问题对实现精确的目标检测提出了挑战，本研究提出了一种基于YOLOv7-tiny的水下目标算法YOLO-U。该算法通过引入具有结构重参数化的RepViT骨干网络，融合ESE通道注意力机制，增强水下模糊目标的特征提取能力；同时，设计了浅层坐标信息特征融合的特征金字塔网络CAFPN，进一步增强检测模型对方向和位置信息的敏感度，融合不同尺度的特征信息以提高小目标的检测能力；最后，采用WIoUv2边界框损失函数有效降低了简单示例对损失值的贡献，使得模型能够聚焦于遮挡目标，进一步提高遮挡目标的检测精度。YOLO-U算法在URPC2021数据集上mAP_{50}取得了84.6%的检测效果，较YOLOv7-tiny、YOLOv5s和YOLOv8s分别提高了2.1%、5.2%和2.8%，检测结果显示，该算法可以有效提高水下目标检测精度，进一步改善水下模糊目标、小目标和遮挡目标的检测效果。

关键词：水下目标检测；YOLOv7-tiny；结构重参数化；注意力机制；损失函数

中图分类号：S 951.2；TP 391.4 **文献标志码：**A

探索海洋资源在人类生活中发挥着重要作用[1]。近年来，随着计算机视觉技术的蓬勃发展，基于深度学习的水下目标检测在海水养殖、海洋工程、海洋资源开发以及水下机器人等领域中得到了广泛的关注。然而，由于水下图像通常受到光线条件不均匀、对比度低、颜色偏差的影响，对水下目标检测的准确性和鲁棒性提出了巨大挑战。

现如今，越来越多的基于深度学习的水下目标检测算法被提出，已经形成了两种主流检测架构，分为两阶段和单阶段目标检测算法。Faster R-CNN算法是两阶段目标检测模型的典型代表。LIU等[2]提出了一种改进的基于Faster R-CNN算法，在ResNet101中嵌入自适应选择单元提高特征提取能力，适用于小而密集的场景。虽然检测精度较高，但是推理速度过慢。

与两阶段检测算法相比，单阶段检测算法无须对物体位置生成候选区域，从而大大降低了模型的计算成本，提高检测速度，因此，许多研究基于单阶段的水下目标检测算法。LI等[3]在YOLOv4算法的基础上，在残差块中结合通道注意力机制，调整最后一个卷积层的网格大小，从而能够检测更小的水下生物。HUA等[4]使用YOLOv5算法，提出了一种基于特征增强和渐进式动态聚合策略的水下目标检测算法，防止小物体被冲突信息淹没，提高小目标检测精度。

CHEN等[5]针对水下图像模糊导致细节特征缺失的问题，基于YOLOv7算法提出一种具有双分支特征提取结构的水下目标检测模型，以提高目标检测精度。然而，所提出的模型参数量和计算量分别达到80.4 M和278.4 G FLOPs，过大的模型和双分支结构无法满足实时检测需求以及移动终端部署的小尺寸模型需求，限制了其对水下场景的适用性。

收稿日期：2024-04-02 **修回日期：**2024-09-15

基金项目：国家自然科学基金面上项目(42176175，42271335)；十三五“蓝色粮仓科技创新”国家重点研发计划(2019YFD0900805)

作者简介：李江川(1999—)，男，硕士研究生，研究方向为基于深度学习的水下目标检测算法。E-mail：15879392839@163.com

通信作者：韩彦岭，E-mail：ylhan@shou.edu.cn

上述水下目标检测算法存在2点不足：(1)提高了小目标或模糊目标的检测精度，遮挡目标的检测效果并不理想。(2)模型较大，难以满足实时性和小尺寸模型部署需求。为了解决以上问题，本文基于YOLOv7-tiny[6]的基本模型结构提出了YOLO-U(YOLO Underwater)。在满足实时性的条件下，YOLO-U通过引入具有重参数化结构的骨干网络，提高模糊目标的特征提取能力；采用坐标注意力机制优化特征金字塔网络，多尺度特征融合捕获小目标特征和位置信息，提高小目标检测精度；优化损失函数，进一步提高遮挡目标的检测精度。

1　材料与方法

1.1　材料

1.1.1　实验环境

所有实验在配备Intel(R) Xeon(R) Silver 4210R CPU@ 2.40 GHz，64 GB RAM和NVIDIA GeForce RTX 3090 24 G的计算机上进行。软件配置环境为Python 3.9和Pytorch 1.13.1+cu117。本研究所有实验训练图像分辨率统一调整为640×640，batchSize设置为32，共训练300轮。实验结果取3次平均值。

1.1.2　评估指标

本研究使用平均精度(Mean average precision，mAP)、参数量(Params)、浮点计算量(Floating point operations，FLOPs)和检测速度(Frames per second，FPS)多个维度来评价模型的各项性能。mAP由精确率(Precision，P)和召回率(Recall，R)决定，精确率表示正确预测的正样本数目与所有预测为正样本的数目之间的比例。精确率如公式(1)所示：

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

式中：TP(True positive)为正确预测为正样本的样本数；FP(False positive)为被错误预测为正样本的样本数。

召回率表示正确预测的正样本数目与所有实际为正样本的数目之间的比例。召回率如公式(2)所示：

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

式中：FN(False Negative)为被错误预测为负样本的样本数。

AP(Average precision)是精确率召回率曲线下的面积，mAP是所有类别的AP的均值。AP和mAP计算如下：

$$\mathrm{AP} = \int_0^1 P(R)\,\mathrm{d}R \tag{3}$$

$$\mathrm{mAP} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{AP}_i \tag{4}$$

mAP_{50}是指在计算mAP时，使用IoU(Intersection over Union)阈值为0.5。$\mathrm{mAP}_{50\sim95}$是指使用不同IoU阈值(0.5~0.95，步长0.05)的mAP。相比于mAP_{50}，$\mathrm{mAP}_{50\sim95}$考虑了更广泛的IoU范围，可以更全面地评估目标检测算法在不同重叠程度下的性能。

1.1.3　数据集

本实验使用URPC2021数据集进行验证，该数据集由8 200张水下真实图像组成。其中包含4类水下生物：海参、海胆、海星和扇贝。该数据集中图像存在模糊、严重色偏和低对比度等特点。水下目标以小目标为主，存在模糊、遮挡和背景相似等情况，尤其是海参在不同环境下形态各异，颜色和背景相似难以区分。本研究将数据集以6∶2∶2的比例随机分为4 920个训练样本、1 640个验证样本和1 640个测试样本。

1.2　整体网络结构

为了更好地识别复杂海底环境中的海洋生物，实现检测精度和速度的有效平衡，本研究对轻量级的YOLOv7-tiny模型进行了改进。YOLO-U整体网络结构如图1所示，由骨干网络(Backbone)、颈部网络(Neck)和检测头(Head)组成。YOLO-U引入RepViT[7]模型作为骨干网络，能够高效的提取水下模糊目标的特征，减少背景信息对后续融合结果的干扰。RepViT模型根据模型参数的大小分为RepViT-M1、RepViT-M2和RepViT-M3。为了平衡检测精度和速度，引入RepViT-M2作为骨干网络。RepViT模型中的4个Stage含有不同数量的RepViTBlock，输出4种不同尺度特征。颈部网络使用本研究提出的CAFPN(Coordinate attention feature pyramid network)结构提取骨干网络浅层坐标信息，融合4层尺度特征实现多尺度特征平衡，防止小目标信息被淹没。最后CAFPN输出3层特征图到IDetect中，完成对目标分类和定位。

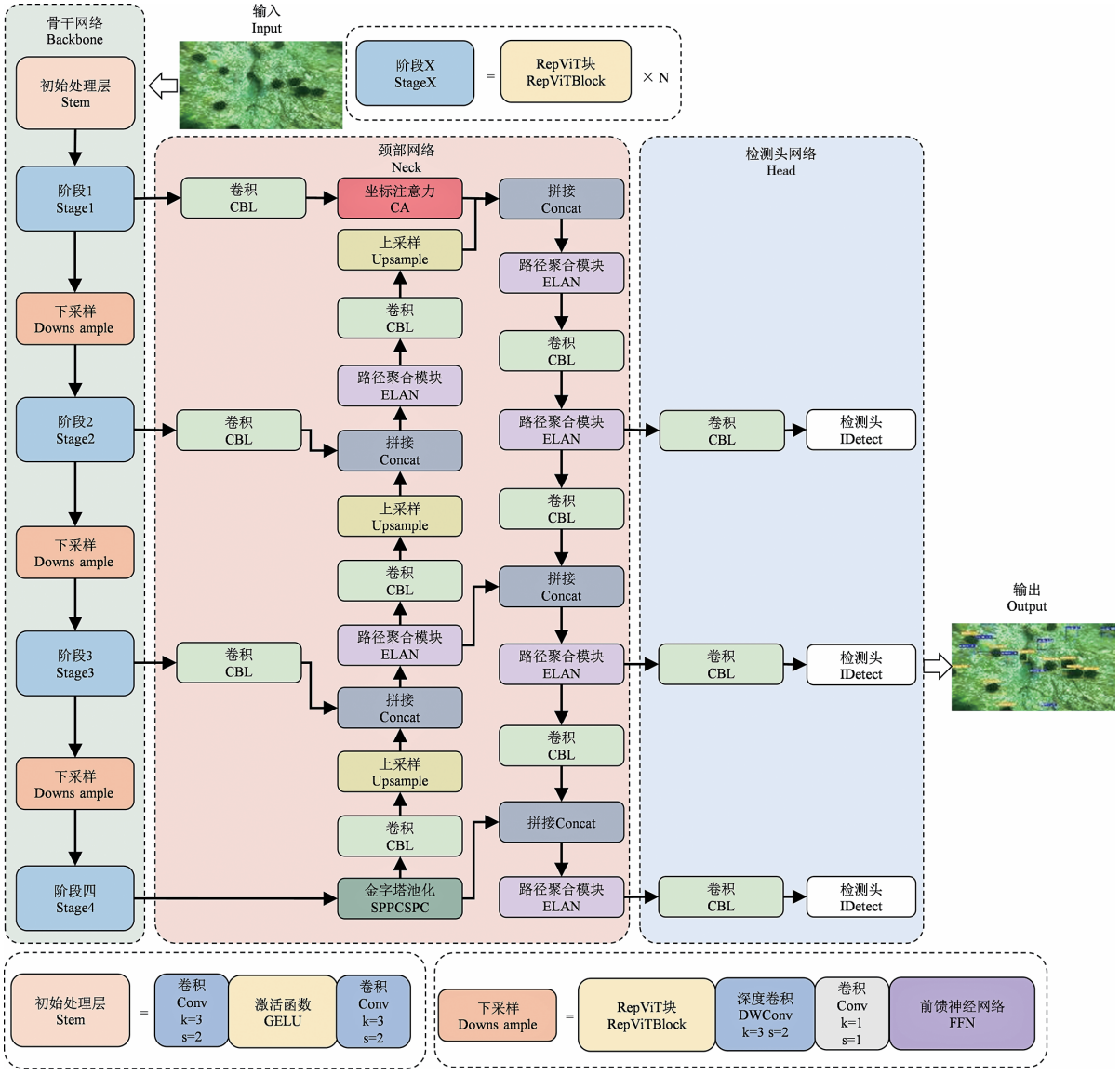


图1 整体网络结构图

Fig. 1 Overall network structure diagram

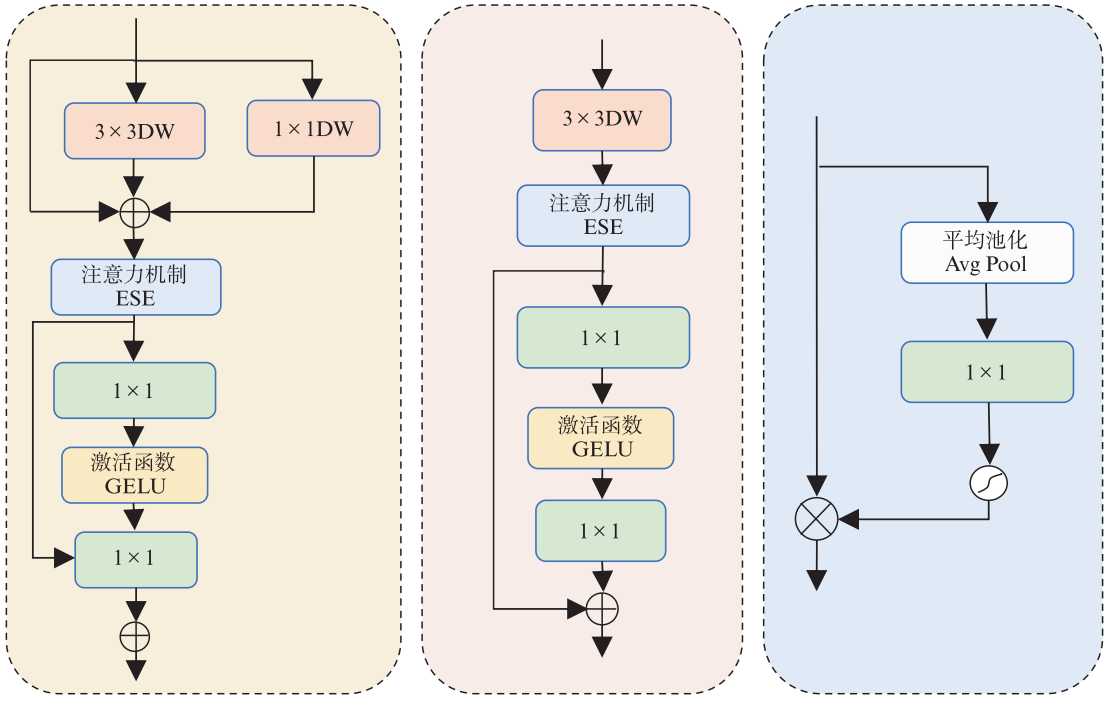
1.3 骨干网络的改进

在模糊的水下场景中，YOLOv7-tiny提取模糊目标特征和捕获全局上下文信息能力较弱，导致模糊目标存在较为严重的漏检情况。为了提高水下模糊目标的检测精度，引入具有结构重参数化的骨干网络RepViT。

结构重参数化方法的思想是借助多分支结构和单分支结构的优势：多分支结构训练更加稳定，易于收敛；单分支结构推理速度快，节省内存。如图2a所示，RepViTBlock在训练时为深度卷积引入多分支拓扑结构，能够提取更多水下目标的有效特征。在推理过程中，如图2b所示，深度卷积多分支结构可合并为单分支结构，消除多分支带来的额外计算和内存成本。此外，相比多分支结构，单分支结构更有利于部署在嵌入式设备、移动端设备以及边缘计算设备上。

RepViT在每个阶段的RepViTblock中交叉使用SE（Squeeze and excitation）[8]通道注意力机制捕捉全局上下文信息和通道权重。由于Excitation操作减少通道维度导致了通道信息的损失，引入ESE（Effective squeeze and extraction）[9]模块改进SE注意力机制，进一步提高模型的特征提取能力。如图2c所示，ESE模块首先通过全局平均池化生成一个1×1×C的向量，获取全局信息。为了加强通道间的信息交互，ESE使用一个1×1卷积层代替SE的两个全连接

层。随后，利用 Hard-sigmoid 激活函数得到特征图的归一化权重。最后，通过乘法逐通道加权到原始特征图的每一个通道上，完成对原始特征的重新标定。与 SE 模块相比，ESE 模块避免了通道信息的丢失，更有效地强调重要特征，抑制非重要特征，以增强整个网络的表达能力。



(a) RepViTBlock训练结构
RepViTBlock training structure

(b) RepViTblock推理结构
RepViTESEBlock inference structure

(c) ESE结构
ESE structure

图 2　RepViTBlock 结构图

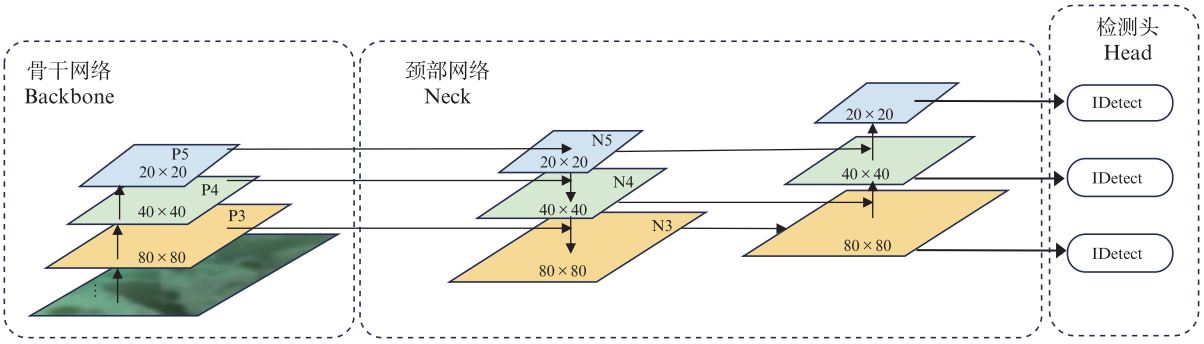
Fig. 2　RepViTBlock structure diagram

1.4　颈部网络的改进

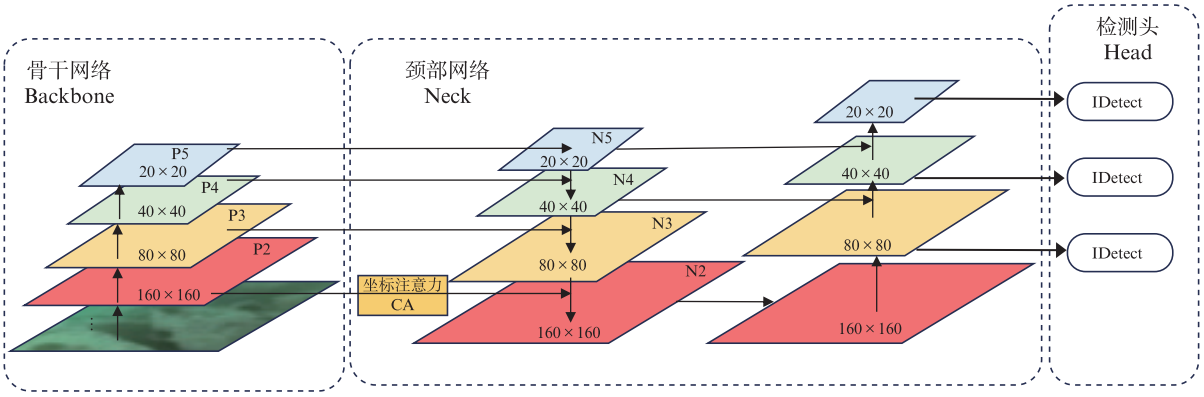
目前 YOLOv7 的 Neck 模块中使用的特征融合结构是 FPN（Feature pyramid network）[10] 和 PANet（Path aggregation network）[11]，通过自上而下和自下而上的聚合路径实现了多尺度特征融合。在水下场景中，通常会存在大量的小目标，特征信息有限。YOLOv7 中 FPN 融合骨干网络最深三层特征，没有充分利用浅层纹理信息不利于检测小目标。如图 3a 所示，FPN 主要提取骨干网络深层的 P3、P4 和 P5 层特征，然后再输入到 PANet 中。经过多次下采样之后，水下小目标的大部分特征已经丢失。

为了增强水下小目标的检测能力，针对 YOLOv7 目标检测算法中的特征金字塔网络多尺度特征融合进行优化。如图 3b 所示，本研究建议的方法 CAFPN 通过提取骨干网络中具有较高分辨率的 P2 层，应用 CA（Coordinate attention）[12] 模块捕捉小物体的细节信息和位置信息，以增强模型的表达能力。接下来，将 N2 层特征和具有位置信息的 P2 层特征拼接，进一步融合高层语义信息和底层位置敏感信息。多尺度特征融合能够综合利用不同尺度和分辨率的信息，提供更全面、更丰富的特征表示，有助于捕捉水下小目标的细节和上下文信息。相比于其他水下模型[13-14]，本研究提出的模型并未同时引入高分辨率的 N2 层检测头。增加的检测头会导致特征信息过多，导致另外 3 个检测头的有效特征信息主导地位被削弱，还会大大增加模型的计算量和内存开销。YOLO-U 维持 3 个尺度的检测头，其特征图尺度分别为 80×80、40×40 和 20×20。

CAFPN 中 CA 模块是一种嵌入位置信息的注意力机制。CA 模块如图 4 所示，将通道注意力分为两个一维特征编码过程，并沿两个空间方向聚合特征。这种设计使得 CA 模块能够捕获沿一个空间方向的远程依赖关系，还可以沿另一个空间方向保留精确的位置信息。通过将位置信息嵌入到通道注意力中，CA 模块能够更好地利用位置信息，提高模型对物体的识别性能。



(a) YOLOv7特征金字塔网络 YOLOv7 feature fusion network



(b) CAFPN结构 CAFPN structure

图3 两种不同特征融合网络

Fig. 3 Two different feature fusion networks

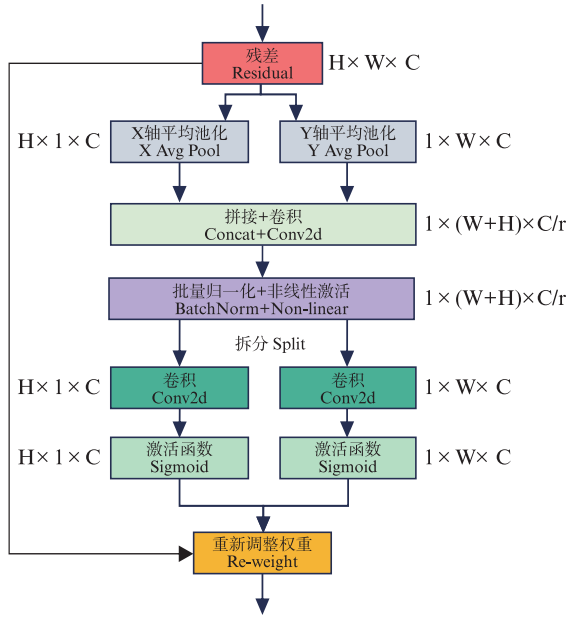


图4 CA结构图

Fig. 4 CA structure diagram

1.5 损失函数的改进

YOLOv7 模型采用 CIoU (Complete intersection over union)[15]作为边界框损失函数，CIoU中纵横比的几何度量会加剧对低质量示例的惩罚从而使模型的泛化性能下降，造成水下遮挡目标漏检以及边界框位置不准确的问题。为了提高水下遮挡目标的检测效果，引入WIoUv2 (Wise-IoU v2)[16]损失函数作为边界框回归损失。WIoUv2是在WIoUv1基础上通过构造梯度增益的计算方法来附加聚焦机制。其中，WIoUv1能在锚框与目标框较好地重合时削弱几何度量的惩罚，不过多地干预训练将使模型有更好的泛化能力。在此基础上，根据中心点距离度量构建了距离注意力，得到了具有两层注意力机制的WIoUv1，如公式(5)和(6)所示：

$$\mathcal{L}_{\mathrm{WIoUv1}} = R_{\mathrm{WIoU}} \times \mathcal{L}_{\mathrm{IoU}} \tag{5}$$

$$R_{\mathrm{WIoU}} = \exp\left[\frac{(x - x_{\mathrm{gt}})^2 + (y - y_{\mathrm{gt}})^2}{(W_{\mathrm{g}}^2 + H_{\mathrm{g}}^2)^*}\right] \tag{6}$$

式中：(x,y)为预测框中心点坐标；$(x_{\mathrm{gt}},y_{\mathrm{gt}})$为真实框中心点坐标；$W_{\mathrm{g}}$和$H_{\mathrm{g}}$分别为预测框和真实框的最小封闭框的宽和高。为了防止R_{WIoU}产生阻碍收敛的梯度，将W_{g}和H_{g}从计算图中分离(*表示此操作)，因为它有效地消除了阻碍收敛的因素。R_{WIoU}将显著放大普通质量锚框的$\mathcal{L}_{\mathrm{IoU}}$，$\mathcal{L}_{\mathrm{IoU}}$将

显著降低高质量锚框的R_{WIoU}，并在锚框与目标框重合较好的情况下显著降低其对中心点距离的关注。

WIoUv2 借鉴了 Focal Loss[17]在 WIoUv1 基础上设计了一种针对交叉熵的单调聚焦机制，有效降低了简单示例对损失值的贡献，使得模型能够聚焦于遮挡目标，如公式(7)所示：

$$\mathcal{L}_{\mathrm{WIoUv2}} = \mathcal{L}_{\mathrm{IoU}}^{\gamma^*}\mathcal{L}_{\mathrm{WIoUv1}},\ \gamma > 0 \tag{7}$$

在模型训练过程中，梯度增益$\mathcal{L}_{\mathrm{IoU}}^{\gamma^*}$随着$\mathcal{L}_{\mathrm{IoU}}$的减小而减小，导致训练后期收敛速度较慢。因此，引入$\mathcal{L}_{\mathrm{IoU}}$的滑动平均值作为归一化因子，如公式(8)所示：

$$\mathcal{L}_{\mathrm{WIoUv2}} = \left(\frac{\mathcal{L}_{\mathrm{IoU}}^{*}}{\overline{\mathcal{L}_{\mathrm{IoU}}}}\right)^{\gamma}\mathcal{L}_{\mathrm{WIoUv1}} \tag{8}$$

动态更新归一化因子使梯度增益$\left(\frac{\mathcal{L}_{\mathrm{IoU}}^{*}}{\overline{\mathcal{L}_{\mathrm{IoU}}}}\right)^{\gamma}$整体保持在较高水平，解决了训练后期收敛速度慢的问题。

2　结果

2.1　消融实验

为了验证 YOLO-U 模型中不同模块的有效性，本文设计了 4 个消融实验，分别对 RepViT-M2 模块、CAFPN 模块和 WIoUv2 损失函数进行评估。

首先，将 YOLOv7-tiny 模型作为基准模型（记为 Baseline）。第二，将 YOLOv7-tiny 的 Backbone 进行改进，引入融合 ESE 注意力机制的 RepViT-M2（记为 Model 1）。第三，基于 Model 1 对 FPN 进行优化，加入浅层坐标信息融合的 CAFPN 模块（记为 Model 2）。第四，基于 Model 3，将 WIoUv2 替换原始的 CIoU（记为 Model 3）。表 1 显示了基于 YOLOv7-tiny 的消融实验的结果。

表 1　消融实验结果
Tab. 1　Results of ablation experiment

模型 Models	RepViT-M2	CAFPN	WIoUv2	mAP_{50}/%	$mAP_{50\sim95}$/%	参数量 Params/M	计算量 FLOPs/G	帧数 FPS/(帧/s)
Baseline				82.6	47.4	7.0	13.0	45
Model 1	√			83.8	48.3	11.6	27.9	35
Model 2	√	√		84.3	48.7	11.7	29.4	30
Model 3	√	√	√	84.7	49.0	11.7	29.4	30

基线模型 YOLOv7-tiny 的mAP_{50}为 82.6%，$mAP_{50\sim95}$为 47.4%，检测速度达到 45 FPS。Model1 相比 YOLOv7-tiny，mAP_{50}提高了 1.2%，$mAP_{50\sim95}$增加了 0.9%。结果表明 RepViT-M2 能够有效的提取水下模糊目标的特征，增强了水下模糊目标的检测能力。在检测效率方面，由于 RepViT-M2 加深了模型的深度使得检测速度略有下降，检测速度保持在 35FPS。

与 Model 1 相比，Model 2 的mAP_{50}提高了 0.5%，$mAP_{50\sim95}$提高了 0.4%。结果表明：CAFPN 比 FPN 有更好的特征融合效果，保留小目标必要的纹理和位置信息。在检测效率方面，引入上采样等操作使模型检测速度稍有下降，检测速度保持在 30FPS。

与 Model 2 相比，Model 3 的mAP_{50}提高了 0.4%，$mAP_{50\sim95}$提高了 0.3%。结果表明 WIoUv2 有效降低了简单示例对损失值的贡献，使得模型能够聚焦于遮挡目标，进一步带来性能的提升，检测速度保持在 30FPS。

相比 YOLOv7-tiny 算法，最终模型的mAP_{50}提高了 2.1%，$mAP_{50\sim95}$提高了 1.6%。虽然检测速度有所下降，但仍能满足实时检测需求 30FPS，实现了检测精度和速度的平衡。

从表 2 中可以看出，RepViT-M2，CAFPN 和 WIoUv2 的引入使得海参、海星和扇贝的检测精度均有明显提升。海参的AP_{50}提高了 4.3%，$AP_{50\sim95}$提高了 2.3%；海胆的AP_{50}提升了 0.2%，$AP_{50\sim95}$提高了 0.8%；海星的AP_{50}提高了 0.9%，$AP_{50\sim95}$提高了 1.1%；扇贝的AP_{50}提高了 2.9%，$AP_{50\sim95}$提高了 2.3%。实验结果证明在满足水下实时检测的前提条件下，本研究提出的 3 个策略都是有效的。

2.2 对比实验

将本研究所提出的YOLO-U与一些主流的实时目标检测模型进行了比较，即YOLOv5s, YOLOv5m,YOLOv7-tiny,YOLOv8n和YOLOv8s。不同模型的评估指标的比较结果如表3所示。

表2 4类生物的消融实验结果

Tab. 2 Ablation results of 4 types of organisms %

模型 Models	AP_{50}				$AP_{50\sim95}$			
	海参 Holothurian	海胆 Echinus	海星 Starfish	扇贝 Scallop	海参 Holothurian	海胆 Echinus	海星 Starfish	扇贝 Scallop
Baseline	73.1	90.6	88.0	78.7	40.1	49.5	52.7	47.1
Model 1	76.3	90.1	88.7	79.9	42.1	49.7	53.2	48.0
Model 2	77.1	90.3	88.6	81.1	42.5	49.3	53.6	49.3
Model 3	77.4	90.8	88.9	81.6	42.4	50.3	53.8	49.4

表3 不同水下目标检测模型的实验结果

Tab. 3 Experimental results of different underwater object detection models

模型 Models	mAP_{50}/%	$mAP_{50\sim95}$/%	参数量 Params/M	计算量 FLOPs/G	帧数 FPS/(帧/s)
YOLOv5s	79.5	45.8	7.0	15.8	36
YOLOv5m	81.6	48.8	20.8	47.9	25
YOLOv7-tiny	82.6	47.4	7.0	13.0	45
YOLOv8n	80.3	47.0	3.0	8.1	52
YOLOv8s	81.9	48.8	11.1	28.4	40
YOLO-U	84.7	49.0	11.7	29.4	30

通过对比mAP50可以发现，本研究提出的模型要优于YOLOv5s, YOLOv5m, YOLOv7-tiny, YOLOv8n和YOLOv8s。与前5种模型相比，mAP_{50}分别提高了5.2%, 3.1%, 2.1%, 4.4%和2.8%。YOLO-U在主流模型中取得了最高的检测精度，同时满足实时检测的需求，有效地在检测精度和速度之间取得了更好的平衡。值得注意的是，与为边缘设备和移动设备等低算力平台设计的轻量化模型YOLOv8s相比，YOLO-U的参数量和计算量与其接近，满足了YOLO-U在移动终端部署的小尺寸需求。

从表4我们可以看出，与其他主流模型相比，YOLO-U在海参、海胆、海星和扇贝都取得了最高的AP_{50}。实验结果表明模型在满足实时检测需求的前提下，具有更强的泛化能力，更适合水下检测任务。

表4 各模型不同水下目标的检测结果

Tab. 4 Detection results of different underwater objects for each model %

模型 Models	AP_{50}			
	海参 Holothurian	海胆 Echinus	海星 Starfish	扇贝 Scallop
YOLOv5s	70.1	87.2	84.6	75.9
YOLOv5m	72.2	88.9	86.1	79.1
YOLOv7-tiny	73.1	90.6	88.0	78.7
YOLOv8n	68.8	89.3	86.4	76.7
YOLOv8s	71.1	89.6	87.6	79.3
YOLO-U	77.4	90.8	88.9	81.6

3　讨论

3.1　消融实验可视化分析

本研究使用XGrad-CAM[18]定位模型感兴趣的区域，通过热力图可视化结果验证RepViT-M2、CAFPN和WIoUv2方法分别在检测模糊目标、小目标和遮挡目标上具有优越性。图5展示了消融实验中4种模型在3种不同场景下表现的热力图结果。

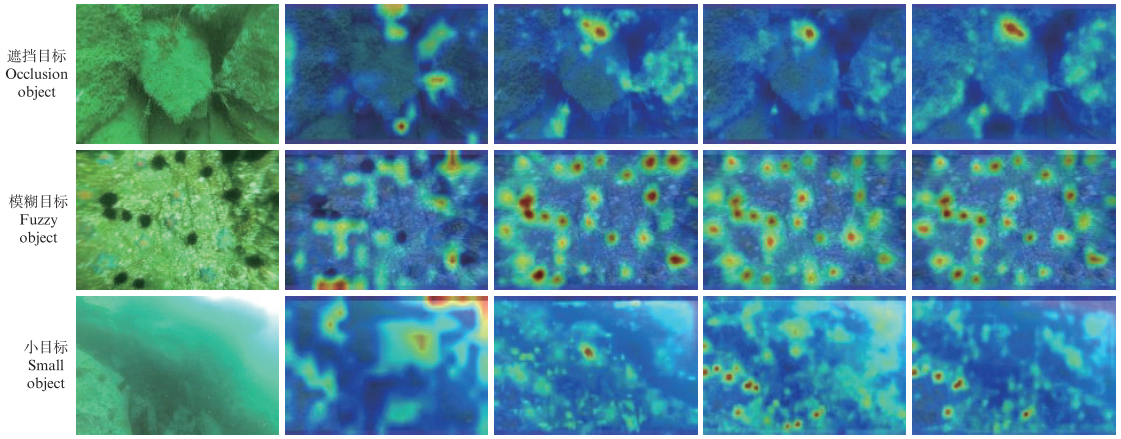


图5　4种模型的热力图

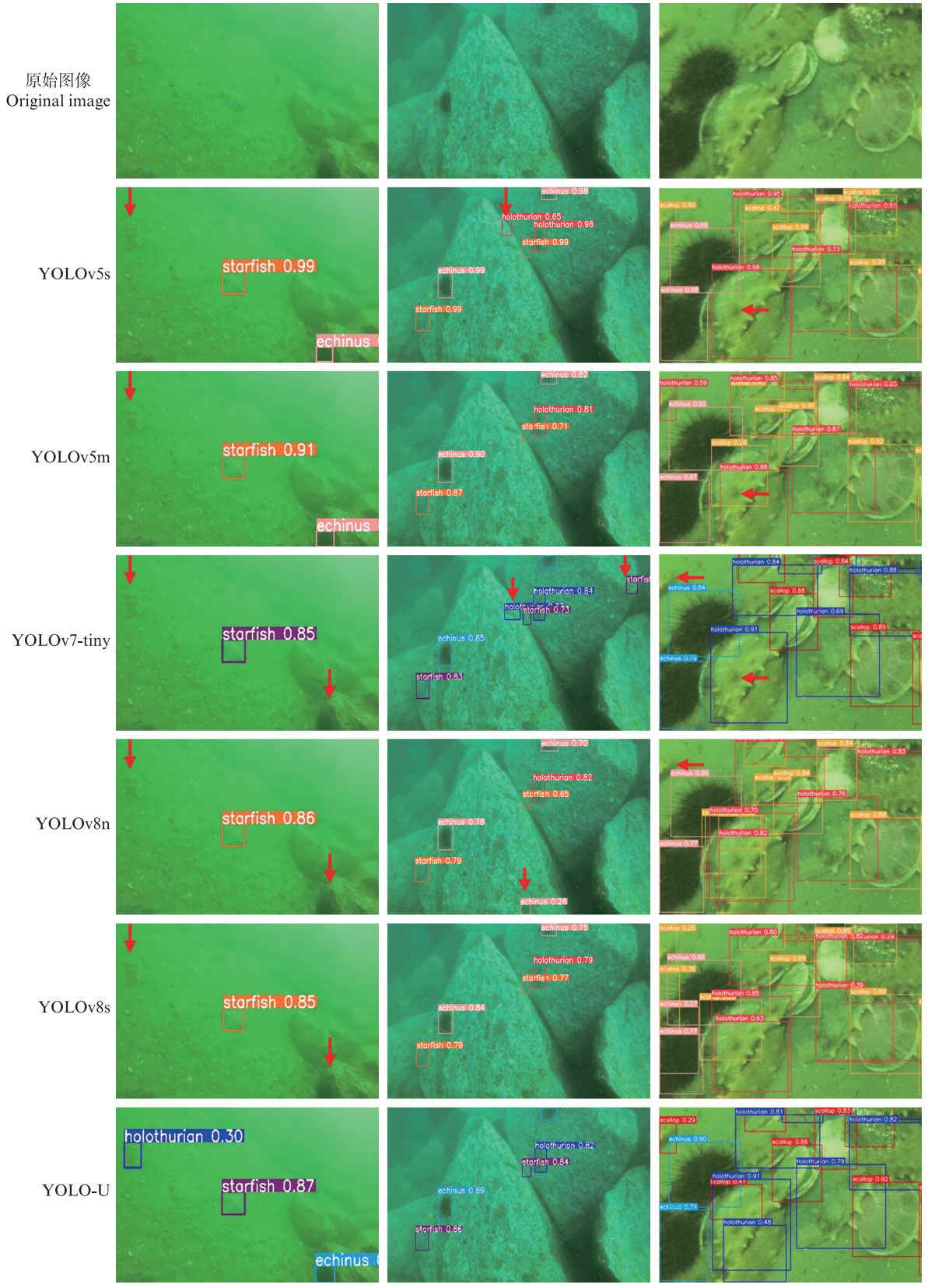
Fig. 5　Heatmaps of four models

YOLOv7-tiny关注了更多的背景信息，可能会忽略目标物体的特征，导致目标物体未被正确地检测和定位（图5b）。如图5c所示，引入RepViT-M2骨干网络后，模型大大减少了对背景干扰信息的关注，更关注于目标本身。如图5d所示，添加CAFPN后，模型能关注到模糊目标和小目标上，然而定位不够准确。如图5e所示，引入WIoUv2后，模型能更准确的关注到目标而非背景，对目标定位也更加精准。在目标遮挡场景中，对于被遮挡的海参，YOLOv7-tiny只能关注到部分特征，造成定位不准确的问题。引入WIoUv2损失函数优化后，YOLO-U能关注到完整特征，从而能给出更准确的检测框定位。四种模型的热力图效果对比验证了我们的结论：RepViT-M2、CAFPN和WIoUv2三种方法分别在模糊目标、小目标和遮挡目标上的检测效果有较大改善。

3.2　对比实验可视化分析

选择严重色偏的水下图像，其中包含模糊目标、小目标、遮挡目标的场景来比较模型检测效果。与其他主流方法相比，YOLO-U在3种场景下显示出显著优势（图6）。

如图6a所示，图像整体质量较低，存在严重的色偏和模糊的问题，其余算法表现出不同程度的遗漏。YOLOv5s和YOLOv5m都检测出了其中的海星和海胆，然而未能检测出与背景相似的海参。YOLOv8s，YOLOv8n和YOLOv7-tiny在模糊图像中仅仅检测出了海星。本文提出的模型成功检测出了所有目标，略微不足的是海参置信度较低。图6b中的小目标与背景中的物体容易混淆，一些模型出现了误检情况。YOLOv5s将红色箭头处石块纹理错误的识别为海参；YOLOv8n将石块纹理错误的识别为海胆；YOLOv7-tiny错误的将两处石块纹理识别为海参以及海星；YOLO-U准确区分了小物体和背景，可以更好地提取有关小物体的信息。从图6c中可以看到，大量的海洋生物堆积在一块。所有的主流检测模型对于大部分目标都能正确检测，但是在图像左下角箭头处，YOLOv5s，YOLOv5m和YOLOv7-tiny都未能正确检测出两个重叠在一块的海参，而是错误的当作一个海参。YOLO-U、YOLOv8n和YOLOv8s正确的识别了两个海参，这表明了模型在遮挡情况下的检测能力有一定的改善。从图6a和图6b中我们可以清晰地了解到改进后的模型对于检测目标与背景相似的情况也有更好的鲁棒性。



(a) 模糊目标 Fuzzy object (b) 小目标 Small object (c) 遮挡目标 Occlusion object

红色箭头表示漏检或错检。

The red arrow indicates missed or incorrect detection.

图 6 3 种场景下不同模型的预测结果

Fig. 6 Prediction results of different models in three scenarios

总之，与其他模型相比，在目标模糊、体积小和密集遮挡的水下场景中，YOLO-U对4种不同的水下生物有更低的误检和漏检情况，综合性能方面优于其他轻量化模型。相比于YOLOv7-tiny，YOLO-U的mAP_{50}提高了2.1%，$mAP_{50\sim95}$提高了1.6%。在检测速度方面，YOLO-U虽然有所下降，但仍满足水下实时检测的需求，实现检测精度和速度的平衡。

4　结论

实现自动化、智能化捕捞的前提是准确识别水下目标。为了在复杂海底环境中能较好的检测目标，针对水下目标模糊、体积小和密集情况遮挡严重等特点，本文提出一种改进的YOLOv7-tiny水下目标检测算法。首先，本研究采用了一种具有结构重参数化的模型RepViT结合ESE通道注意力机制重构骨干网络，增强了水下模糊目标的检测能力；其次，提出CAFPN优化特征融合，改进后的特征金字塔网络兼顾深层的语义特征和浅层的纹理以及位置特征，有效提升了小目标的检测精度；最后，引入WIoUv2损失函数，模型能够聚焦于遮挡目标，进一步降低遮挡目标的漏检率的同时能更好的定位边界框。实验表明，所提出的模型在URPC2021数据集上实现了84.7%的mAP50，在综合性能方面优于其他所列算法，在复杂的水下场景中具有更强的鲁棒性。

本研究提出的方法仍有改进的空间，在未来的工作中将致力于进一步优化模型结构和算法，在不牺牲检测性能的情况下提高检测速度；引入合适的图像增强算法，提高海参等的检测精度；将提出的模型部署到移动设备，结合实际检测结果进一步优化。

作者声明本文无利益冲突。

参考文献：

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

[2] LIU Y, WANG S N. A quantitative detection algorithm based on improved faster R-CNN for marine benthos[J]. Ecological Informatics, 2021, 61: 101228.

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

[4] HUA X, CUI X P, XU X H, et al. Underwater object detection algorithm based on feature enhancement and progressive dynamic aggregation strategy [J]. Pattern Recognition, 2023, 139: 109511.

[5] CHEN X, YUAN M J H, FAN C Y, et al. Research on an Underwater Object Detection Network based on dual-branch feature extraction [J]. Electronics, 2023, 12(16): 3413.

[6] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 7464-7475.

[7] WANG A, CHEN H, LIN Z J, et al. Rep ViT: revisiting mobile CNN from ViT perspective [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2024: 15909-15920.

[8] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[9] LEE Y, PARK J. CenterMask: real-time anchor-free instance segmentation [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 13903-13912.

[10] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 936-944.

[11] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE, 2018: 8759-8768.

[12] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 13708-13717.

[13] XUAN K, DENG L M, XIAO Y, et al. SO-YOLOv5: small object recognition algorithm for sea cucumber in complex seabed environment [J]. Fisheries Research, 2023, 264: 106710.

[14] 陶洋，钟邦乾，赵文博，等. 融合显示视觉中心与注意力机制的水下目标检测算法[J]. 激光与光电子学进展，2024，61(12)：1237009.
TAO Y, ZHONG B Q, ZHAO W B, et al. Underwater object detection algorithm integrating explicit visual center and attention mechanism [J]. Laser & Optoelectronics Progress, 2024, 61(12): 1237009.

[15] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box

regression［C］//Proceedings of the AAAI Conference on Artificial Intelligence. New York：AAAI Press，2020：12993-13000.

［16］ TONG Z J，CHEN Y H，XU Z W，et al. Wise-IoU：bounding box regression loss with dynamic focusing mechanism［J］. arXiv preprint arXiv：2301. 10051，2023.

［17］ LIN T Y，GOYAL P，GIRSHICK R，et al. Focal loss for dense object detection［C］//Proceedings of the IEEE International Conference on Computer Vision. Venice：IEEE，2017：2999-3007.

［18］ SUNDARARAJAN M，TALY A，YAN Q Q. Axiomatic attribution for deep networks［C］// Proceedings of the 34th International Conference on Machine Learning. Sydney：JMLR. org，2017：3319-3328.

YOLO-U：An underwater object detection algorithm based on structural reparameterization and dual attention mechanism

LI Jiangchuan[1]，HAN Yanling[1]，DONG Chuansheng[2]，WANG Yan[1]，WANG Jing[1]，ZHANG Yun[1]，YANG Shuhu[1]

（1. *College of Information Technology*，*Shanghai Ocean University*，*Shanghai* 201306，*China*；2. *Shandong Provincial Institute of Land Surveying and mapping*，*Jinan* 250102，*Shandong*，*China*）

Abstract: The research on underwater object detection algorithms is a prerequisite for achieving intelligent fishing with underwater robots. The problems of fuzzy object，numerous small objects and mutual occlusion in underwater object detection pose challenges to the realization of accurate object detection. This paper proposes a YOLO-U algorithm for underwater object detection based on YOLOv7-tiny. The algorithm introduces a RepViT backbone network with structural reparameterization and fuses an ESE channel attention mechanism to enhance the feature extraction capability for underwater fuzzy objects. Additionally，a feature pyramid network CAFPN with shallow coordinate information feature fusion is designed to further enhance the sensitivity of the detection model to directional and positional information，and integrate feature information of different scales to improve the detection ability of small objects. Furthermore，the WIoUv2 bounding box loss function is employed to effectively reduce the contribution of easy examples to the loss value. This allows the model to focus on occluded objects and further improve the detection accuracy for occluded objects. The YOLO-U algorithm achieves a mAP50 of 84.6% on the URPC2021 dataset，which is an improvement of 2.1%，5.2%，and 2.8% compared to YOLOv7-tiny，YOLOv5s，and YOLOv8s，respectively. The detection results show that the algorithm can effectively improve the detection accuracy of underwater objects and further improve the detection performance of underwater fuzzy objects，small objects，and occluded objects.

Key words: underwater object detection；YOLOv7-tiny；structural reparameterization；attention mechanism；loss function