

DOI:10.20033/j.1003-7241.(2026)04-0001-05

# 融合多尺度信息的水下目标检测算法研究

王宁生, 解志斌, 邵长斌  
(江苏科技大学, 江苏 镇江 212003)

**摘要:**水下目标检测在海洋探测中起着至关重要的作用。然而,传统的水下目标检测方法在复杂水下环境中面临诸多挑战,例如特征提取不准确、小目标检测效果差以及缺乏鲁棒性等问题。为了克服这些局限性,提出了一种适用于复杂水下环境的目标检测模型 EEN-YOLO。首先,所提模型在 YOLOv7 网络的基础上引入了 EVC 模块,通过并行学习,将底层特征与深层特征沿着通道维度相连接,扩大了感受野,从而解决特征提取精度差的问题。其次,在检测头中增添 Elan-Neck++ 结构,通过该结构中下采样模块提高了小目标检测精度。最后,采用 NWD 损失函数替换原模型中的 CIoU 损失函数,解决了原模型无法正确度量小目标边界框的问题,从而提高了模型训练精度和鲁棒性。在公开水下数据集 URPC 上进行的实验验证了所提模型的有效性。结果表明,改进后的 EEN-YOLO 模型平均准确率达到 88.3%,相较于原 YOLOv7、YOLOv8n、EfficientDet-d0 以及 SSD 模型,分别提升了 3%、4.5%、7.8% 和 13.1%。这充分证明, EEN-YOLO 在复杂水下场景中更准确、更稳健的目标检测性能。

**关键词:**水下目标检测; YOLOv7; 特征融合; 损失函数; Elan-Neck++; URPC

中图分类号: TP391.413

文献标志码: A

文章编号: 1003-7241(2026)04-0001-05

## Research on underwater object detection algorithm integrating multi-scale information

WANG Ningsheng, XIE Zhibin, SHAO Changbin

( Jiangsu University of Science and Technology, Zhenjiang 212003, Jiangsu, China )

**Abstract:** Traditional underwater target detection methods face many challenges in complex underwater environments, such as inaccurate feature extraction, poor small target detection performance, and lack of robustness. To overcome these limitations, this paper proposes an underwater target detection model EEN-YOLO suitable for complex underwater environments. Firstly, the proposed model introduces the EVC module on the basis of the YOLOv7 network, connecting the bottom features with the deep features along the channel dimension, expanding the receptive field. Secondly, the Elan-Neck++ structure is added to the detection head, which improves the accuracy of small target detection through the downsampling module in the structure. Finally, the NWD loss function is used to replace the original CIoU loss function in the model, solving the problem that the original model cannot correctly measure the bounding box of small targets. Experimental results on the public dataset URPC show that the mAP of the improved EEN-YOLO model reached 88.3%. Compared with YOLOv7, YOLOv8n, EfficientDet-d0, and SSD, it increased by 3%, 4.5%, 7.8%, and 13.1%, respectively.

**Keywords:** underwater target detection; YOLOv7; feature fusion; loss function; Elan-Neck++; URPC

为了更好地拓展海洋经济发展空间,需要对海洋的资源勘测、环境监测和生物识别等工作进行深入研究<sup>[1]</sup>。在这些研究中,水下目标检测技术起着至关重要的作用。

近年来,深度学习技术得到飞速发展,并在目标检测任务中取得了显著进展<sup>[2-4]</sup>。基于深度学习的目标检测算法可分为两种:两阶段目标检测和单阶段目标检测。两阶段目标检测主要包括 R-CNN<sup>[5]</sup>、Fast-R-CNN<sup>[6]</sup>、Faster-R-CNN<sup>[7]</sup>等。相比两阶段目标检测算法,单阶段目标检测算法检测速度更快<sup>[8-9]</sup>,主要包括 EfficientDet, SSD<sup>[10]</sup>, YOLO<sup>[11]</sup>系列等。

具体到水下目标检测领域,为了应对水下环境的挑战,文献[12]提出了改进 YOLOv3-SPP 水下目标检测算法,利用水下对抗生成网络对水下原始图像进行恢复,并引入 CIoU 损失函数,同时将 K-Means 聚类算法更换成 K-Means++,达到提高网络精度的目的。但在复杂的水下环境中,该方法仍存在小目标漏检和检测速度慢的问题。文献以 YOLOv4 为基准算法,构建了新的主干网络,并采用可分离卷积替换普通网络卷积,以减少算法的体积。改进之后的算法检测速度得到了提升,但忽略了网络特征提取精度。

收稿日期:2025-02-02;录用日期:2025-04-17

基金项目:国家自然科学基金项目(62276117);高端外国专家引进计划(G2023014110)

作者简介:王宁生(2000—),男,硕士研究生,研究方向:目标检测。

引用本文:王宁生,解志斌,邵长斌.融合多尺度信息的水下目标检测算法研究[J].自动化技术与应用,2026,45(4):1-5.(WANG Ningsheng, XIE Zhibin, SHAO Changbin. Research on underwater object detection algorithm integrating multi-scale information [J]. Techniques of Automation and Applications, 2026,45(4):1-5.)

为解决水下小目标容易漏检,识别精度差等问题,本文选取 YOLOv7 网络作为基准算法,在网络中引入了视觉中心处理模块以增强特征提取,引入多尺度检测头来捕捉不同大小的目标,并优化了损失函数以提高对小目标的度量精度。通过在水下图像数据集上的实验验证,显示出改进后的算法在水下目标检测中的有效性。

## 1 YOLOv7 网络算法

YOLOv7 是一种实时目标检测算法,能够快速和准确的检测物体。其与高性能的 YOLOv5 相比,YOLOv7 采用了相似检测思路,但又集成了诸多策略,如 E-ELAN (extended efficient layer aggregation networks)、算法放缩以及算法重参数化等,以进一步提升性能和效果。

如图 1 所示,YOLOv7 由 4 个不同的模块组成:输入模块 (Input)、主干网络 (Backbone)、颈部 (Neck)、检测头 (Detection Head)。输入模块中,采用混合数据增强技术对输入图像进行预处理。主干网络包括 3 个主要组件: CBS 模块, ELAN 模块和 Downsampling 模块。颈部网络采用特征金字塔架构,分别包括 Sppcspc 模块, ELAN-Neck 模块, Upsample 模块以及 Downsampling-neck 模块。这里值得注意的是 Sppcspc 模块,该模块在空间金字塔结构中融合了卷积空间金字塔结构,在多个卷积操作中加入并行池化操作,改善网络提取特征能力。检测头部分采用 Repconv 模块调整头部网络输出,该结构可以在不牺牲其预测性能情况下降低网络复杂性。

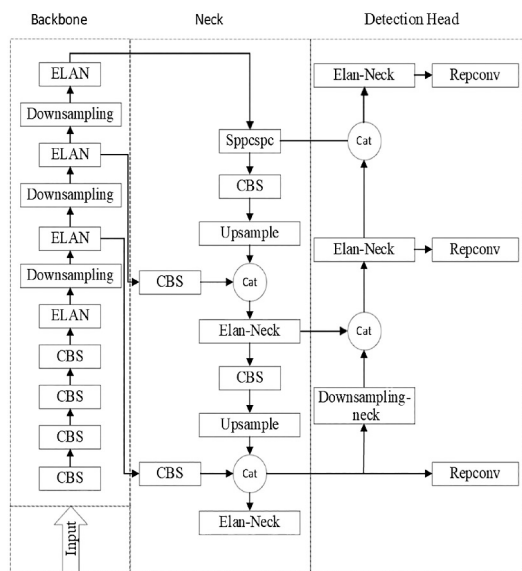


图 1 YOLOv7 网络架构

Fig. 1 YOLOv7 network architecture

## 2 水下多尺度检测算法

为了应对水下复杂环境中,目标与背景容易相互重叠而导致的特征提取不准确、小目标检测效果差及缺乏鲁棒性等实际问题,本文提出了 EEN-YOLO 目标检测算法,其结构如图 2 所示。

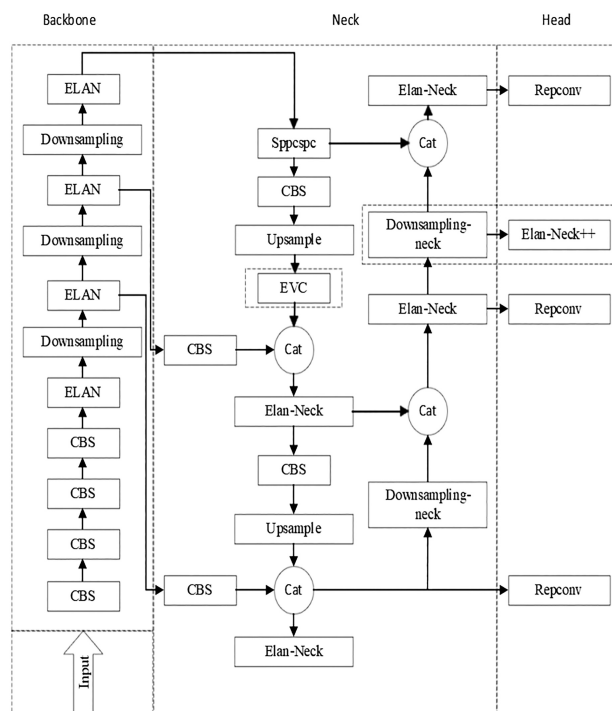


图 2 EEN-YOLO 网络架构

Fig. 2 EEN-YOLO network architecture

首先,所提算法在网络颈部引入 EVC (explicit visual center) 模块<sup>[14]</sup>,该模块通过有效的并行训练策略来提高训练效率,同时引入动态卷积网络和多层感知器来减少计算量,并行学习策略一方面既能捕捉到网络不同特征层信息,另一方面又能捕捉到层内信息,这使得网络保留了更多有效信息,从而提高特征提取的准确性。其次,设计了创新性的检测头,在算法本身 ELAN-Neck 结构中增加池化结构,让网络能够适应不同输入对象的大小以及维度,从而进一步提升目标检测的性能。另外,由于 CIOW 损失函数容易忽略微小目标,所以采用 NWD (normalized wasserstein distance)<sup>[15]</sup> 损失函数。NWD 损失函数首先对目标边界框和预测边界框进行二维高斯分布变换,然后使用最优传输距离公式度量这两个边界框。

### 2.1 融合位置信息的特征提取网络

在训练过程中,特征提取是至关重要的一步,在这个过程中,不仅希望能够注意到层与层之间的特征交互,还需要关注到层内特征之间的联系。以往一些方法使用了注意力机制来学习层内特征之间的联系,但忽略了对于小目标检测任务非常重要的局部区域。因此,为了解决这些问题,引入了具有多层信息捕捉优势的 EVC 模块,降低了原始算法出现漏检的可能性。

如图 3 所示,EVC 模块由 Conv (卷积)层、BN (批归一化)层、ReLU 激活层和两个并行连接 LM 模块和 LVC 模块组成。其中 LM 模块使用轻量级多层感知器来捕捉全局特征之间的关系,而 LVC 模块采用可学习的视觉中心机制,用于捕捉局部区域特征之间的关系。最后,这两个模块的结果特征图会沿着通道维度连接在一起,作为 EVC 模块的输出。具体而言,EVC 模块为了提升训练稳定性和鲁棒性,采

用了通道数为 256 的 7×7 卷积层,批归一化层和 ReLU 激活层对数据进行预处理,如式(1)所示。

$$X_{in} = \sigma_{ReLU}(BN(Conv_{7 \times 7}(X))) \quad (1)$$

其中  $Conv_{7 \times 7}(\cdot)$  表示步长为 1,通道数为 256 的 7×7 卷积层。 $BN(\cdot)$  表示批归一化层,  $\sigma_{ReLU}(\cdot)$  表示 ReLU 激活函数。

同时,为了降低计算复杂度,采用深度卷积模块如式(2)所示。

$$\tilde{X}_{in} = DConv(GN(\tilde{X}_{in})) + \tilde{X}_{in} \quad (2)$$

这里  $\tilde{X}_{in}$  是经过深度卷积模块的输出,  $GN(\cdot)$  表示归一化,  $DConv(\cdot)$  表示卷积核大小为 1×1 的深度卷积。

为了进一步降低计算复杂度,在 LM 模块中采用多层感知器 MLP 而非卷积神经网络完成特征的提取,如式(3)所示。

$$MLP(\tilde{X}_{in}) = CMLP(GN(\tilde{X}_{in})) + \tilde{X}_{in} \quad (3)$$

其中  $CMLP(\cdot)$  是指多层感知器 MLP 模块。

为了更好地捕捉局部区域特征,LVC 模块采用卷积网络和编码器组合运算的方式实现了可学习的视觉中心机制,如式(4)所示。

$$LVC(X_{in}) = X_{in} \oplus X_{in} \otimes (\delta(Conv_{1 \times 1}(\sum_{k=1}^K \phi(e_k)))) \quad (4)$$

其中  $X_{in}$  指 LVC 模块的输入,  $e_k$  指经过 codebook 模块的输出向量,  $\phi(\cdot)$  包含 ReLU 激活函、BN 层和 Mean 层,  $Conv_{1 \times 1}(\cdot)$  表示卷积核大小为 1×1 的卷积模块,  $\delta(\cdot)$  指比例系数。

由此可知,基于 EVC 模块的低复杂度并行学习策略,加强了网络在全局和局部两个方面对小目标特征的学习能力,从而改善了原算法的特征提取性能。

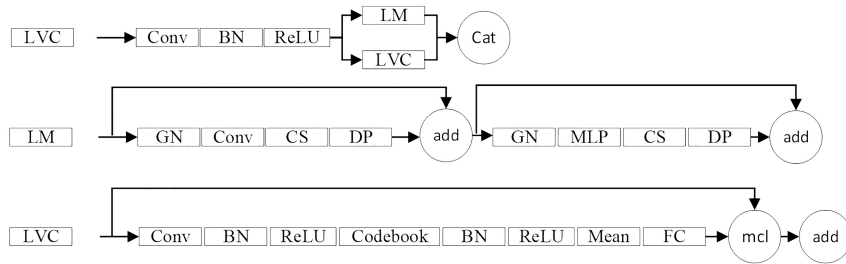


图3 EVC 模块

Fig. 3 EVC module

### 2.2 融合多尺度信息的检测头

在 YOLOv7 网络中,检测头首先对网络中间特征图进行提取,然后通过回归任务预测物体的边界框,并对边界框内的目标进行分类,以确定物体所属的类别。然而,在 YOLOv7 的原始结构中,仅有的检测头无法精确检测到输入图像中不同大小的对象,导致网络无法有效地识别小目标任务,影响了检测精度。

为了解决这个问题,本文增加了一个全新的检测头,专门用于小目标检测。其结构如图 4 虚线框所示。新增加的检测头首先使用图 4 虚线框中 Downsampling-neck 模块处理来自上一层的输出特征图,通过减小特征图的尺寸来降低模型的复杂度,避免网络过拟合的风险。接下来,采用图 4 虚线框中 Elan-Neck++ 模块,在小幅提升网络复杂度的情况下提升特征提取能力。Elan-Neck++ 模块首先通过 Elan-Neck 模块增强网络对特征的学习能力,然后再次使用 Downsampling-neck 模块降低模型复杂度,最后通过 Repconv 模块输出类别信息和位置信息以及置信度。这种设计使得检测头能够更准确地捕捉到小目标的细节,从而显著提示网络的整体性能。

### 2.3 针对小尺度目标的损失函数

由于 YOLOv7 损失函数惩罚项的不稳定性,导致了 CIUO 损失函数对小尺度目标非常敏感,较小的目标反而会受到较大的损失,这导致了算法在小目标检测上表现不佳。

因此,本文使用更加灵活的 NWD 损失函数替换原网络中边界框损失函数。首先对真实框和预测框采用二维高斯分布变换,然后使用最优传输距离公式来测量变换之

后边界框的相似性。由于不同尺寸的边界框经过二维高斯分布变换之后可以被更好地度量,这使得网络的鲁棒性得到了很大的提升。

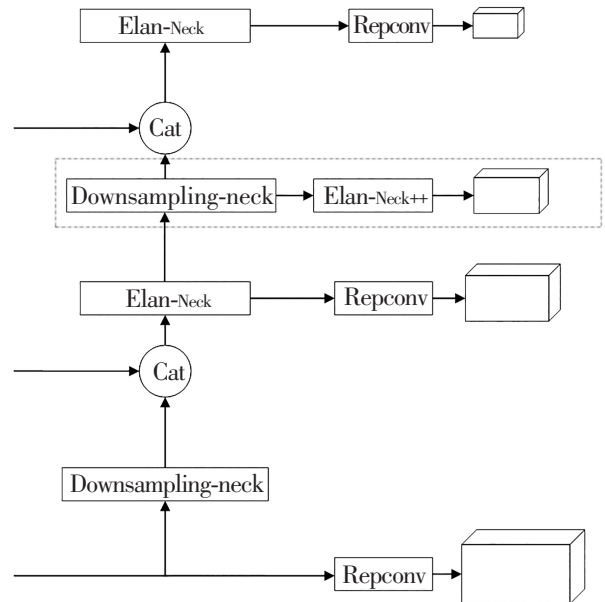


图4 改进后的检测头

Fig. 4 Detectionhead after improvement

将边界框  $R = (c_x, c_y, w, h)$  建模为二维高斯分布  $N(\mu, \Sigma)$ , 如式(5)所示。

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \Sigma = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \quad (5)$$

式中,  $N$  表示服从均值为  $\boldsymbol{\mu}$  和方差为  $\boldsymbol{\Sigma}$  的二维高斯分布,  $(c_x, c_y)$  表示预测框中心坐标位置,  $w$  和  $h$  表示预测框宽度和高。

由式(5)可知,将预测框和真实框分别进行二维高斯分布变换,可得  $N_a$  和  $N_b$ 。将其代入 Wasserstein 距离公式,如式(6)所示为

$$\text{NWD}(N_a, N_b) = \exp \left( - \frac{\sqrt{\left\| \left[ \begin{matrix} c_{x_a}, c_{y_a}, \frac{w_a}{2}, \frac{h_a}{2} \end{matrix} \right]^T, \left[ \begin{matrix} c_{x_b}, c_{y_b}, \frac{w_b}{2}, \frac{h_b}{2} \end{matrix} \right]^T \right\|_2^2}}{C} \right) \quad (6)$$

式中,  $T$  表示矩阵转置,  $(c_{x_a}, c_{y_a})$  为预测框中心位置坐标;  $(w_a, h_a)$  为预测框宽和高;  $(c_{x_b}, c_{y_b})$  为真实框中心位置坐标;  $(w_b, h_b)$  为真实框宽和高;  $C$  为自定义的常数,与数据集密切相关。

由上述公式,可以观察到, NWD 损失函数去除了影响 CIOW 损失函数的惩罚项,使得其对边界框比例并不敏感,并且采用的二维高斯分布变换可以满足小尺度目标的需求,在后续实验中也能证明这一点。

## 3 实验结果分析

### 3.1 数据集

为了对 EEN-YOLO 的有效性进行验证,采用了 URPC 数据集作为评估基准。该数据集是全国水下机器人专业竞赛数据集,包含了 5 543 张水下光学图像,其中尺寸小于  $32 \times 32$  像素的物体的图片为 2 708 张。数据集中的目标生物涵盖了 4 类:海参、海胆、扇贝和海星。将数据集随机划分为训练集和测试集,划分比例为 8 : 2。

### 3.2 实验环境

本次实验采用 Pytorch 深度学习框架。实验环境具体配置:操作系统是 Ubuntu22.04, GPU 是两块英伟达 RTX3080,其他软件包括 python3.8、cuda11.0、torch1.7.0、torchvision0.8.1。其他超参数设置如下:批量大小为 16,训练轮数设置为 300,优化器采用 SGD 优化器,初始学习率为 0.001,学习率下降方式为余弦退火。

### 3.3 实验结果分析

如图 5 所示,将二者算法的可视化检测结果进行了对比。图片中包含了预测框,所属类别和置信度分数。

其中,图 5(a)和图 5(c)为 YOLOv7 的检测结果,图 5(b)和图 5(d)为 EEN-YOLO 的检测结果。对比图 5(a)和图 5(b)可知,改进后的算法检测出了原算法中检测不出的小目标“海参”和“扇贝”并且检测精度明显上升;由图 5(c)和图 5(d)可知,改进后的算法提取特征能力加强,检

测到相同物体的置信度明显提高,充分说明了本文方法的有效性。

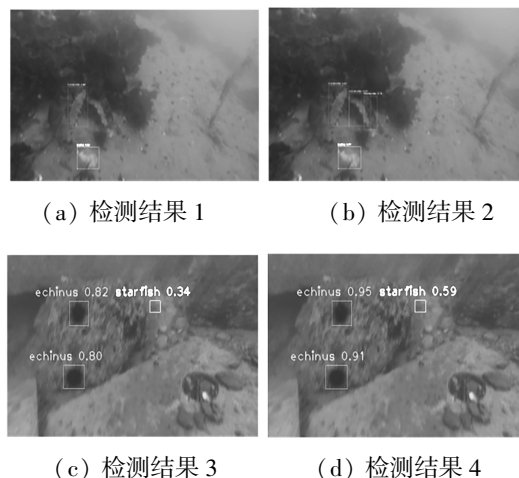


图 5 可视化检测结果对比

Fig. 5 Comparison of visual detection results

### 3.4 不同网络对比实验

为探索 EEN-YOLO 的有效性,在采用相同电脑配置和数据集的情况下,将其与近几年经典算法进行了对比实验,实验结果如表 1 所示。

从表 1 中可以看出改进之后的网络平均准确率比基础算法 YOLOv7 高 3%,分别比当前流行算法 SSD, EfficientDet-d0, YOLOv5s 和 YOLOv8n 的 mAP 高 13.1%, 7.8%, 3.4% 和 4.5%。这些实验结果充分说明了本文改进算法的有效性。

表 1 不同网络算法实验结果对比

Tab. 1 Comparison of experimental results from different network algorithms

算法	精确率/%	召回率/%	mAP/%
SSD	74.3	69.8	75.2
EfficientDet-d0	82.3	74.6	80.5
YOLOv5s	85.2	78.8	84.9
YOLOv7	82.2	80.3	85.3
YOLOv8n	82.1	78.5	83.8
EEN-YOLO	85.3	82.4	88.3

### 3.5 消融实验

为了验证所提出的各部分优化算法都是切实有效的,分别做了 4 组消融实验,实验结果如表 2 所示。

由表 2 可知,未改进算法时,水下目标检测平均精度为 85.3%,当依次采取添加 EVC 模块,检测头增加 Elan-Neck++ 结构以及改进损失函数后, EEN-YOLO 得到了最高的 mAP 值,其为 88.3%,结果表明改进措施有效。

表 2 消融实验

Tab. 2 Ablation experiments

序号	EVC	Elan- Neck++	NWD	精确率/%	召回率/%	mAP/%	参数量/ $10^6$
1	×	×	×	82.2	80.3	85.3	37.21
2	√	×	×	84.8	80.5	86.3	41.50
3	√	√	×	83.8	83.3	87.6	46.56
4	√	√	√	85.3	82.4	88.3	46.56

### 3.6 鲁棒性实验

为了进一步验证提出的 EEN-YOLO 的鲁棒性,将其余改进算法与 EEN-YOLO 在 RUOD 数据集上进行实验。RUOD 数据集收集了 14 000 张不同水下目标照片。数据集中的目标生物涵盖了 10 个类别:鱼、潜水员、海星、珊瑚、海龟、海胆、海参、扇贝、鱿鱼、水母。将上述数据集按照 8:2 的比例划分训练集和测试集实验结果如表 3 所示。

表 3 在 RUOD 数据集上的鲁棒性实验

Table 3 Robustness study on the RUOD dataset

算法	精确率/%	召回率/%	mAP/%
SSD	75.3	68.9	74.3
EfficientDet-d0	83.8	73.2	81.5
YOLOv5s	84.5	80.0	85.6
YOLOv7	86.0	78.3	85.8
YOLOv8n	84.0	74.8	82.5
EEN-YOLO	87.0	82.0	88.5

从表 3 实验结果可以看出,在 RUOD 数据集上。改进之后的网络平均准确率比基础算法 YOLOv7 高 2.7%,分别比当前流行算法 SSD, EfficientDet-d0, YOLOv5s 和 YOLOv8n 的平均准确率高 14.2%,7%,2.9%和 6%。

## 4 结论

本文提出了一种改进的水下目标检测算法。首先,引入的 EVC 模块采用高效的并行学习策略,分别对特征图的局部信息和全局信息加以提取并融合,从而提升算法的特征提取能力。同时,还对检测头结构进行改进,通过引入的 Elan-Neck++ 模块有效提升了算法检测不同尺度目标的准确性。最后,采用 NWD 损失函数,合理地解决了计算小目标边界框损失不准确的问题,使网络在小目标检测性能上得到进一步提升。为验证改进后算法性能,在 URPC 数据集上进行了实验,并将结果与近几年流行的目标检测算法进行了对比。实验结果表明,EEN-YOLO 算法适用于复杂的水下目标检测环境,相比原算法性能得到了有效的提升,为水下目标检测任务提供了一种有效的解决方案。

## 参考文献

- [1] LIU Y, ANDERLINI E, WANG S, et al. Ocean explorations using autonomy: technologies, strategies and applications [J]. Offshore Robotics, 2022,1(1):35-58.
- [2] ZHANG W, SUN W. Research on small moving target detection algorithm based on complex scene [J]. Journal of Physics: Conference Series, 2021, 1738(1):012093.
- [3] 周宏威, 王文博, 王伟光, 等. 基于改进 YOLOv7 算法的驾驶分心行为检测模型 [J]. 自动化技术与应用, 2025, 44(11):70-74.
- [4] 王敬启. 改进 YOLOv7-Tiny 的电力现场作业安全装备实时检测算法 [J]. 自动化技术与应用, 2025, 44(9):89-94.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Columbus: IEEE, 2014: 580-587.
- [6] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015:1440-1448.
- [7] 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, 2015(39):1137-1149.
- [8] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020:10781-10790.
- [9] 马赛, 解志斌, 邵长斌. 融合位置信息和上下文的水面目标检测方法 [J]. 小型微型计算机系统, 2024, 45(9):2221-2227.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]//Proceedings of the European Conference on Computer Vision (ECCV). Amsterdam Natherlanys: IEEE, 2016:21-37.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016:779-788.
- [12] 叶赵兵, 段先华, 赵楚. 改进 YOLOv3-SPP 水下目标检测研究 [J]. 计算机工程与应用, 2023, 59(6):231-240.
- [13] 史先鹏, 王宏娟. 基于 YOLOv4 改进的轻量级水下目标检测网络 [J]. 哈尔滨工程大学学报, 2023, 44(1):154-160.
- [14] QUAN Y, ZHANG D, ZHANG L, et al. Centralized feature pyramid for object detection [J]. IEEE Transactions on Image Processing, 2023(32):4341-4354.
- [15] WANG J, XU C, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection [J/OL]. (2021-10-26). <https://arxiv.org/abs/2110.13389>.