

基于改进 YOLO11 的水下目标检测模型<sup>\*</sup>方侦波<sup>1,2</sup> 高向阳<sup>3</sup> 张锬石<sup>3</sup> 程 俊<sup>3</sup> 杨梦杰<sup>4</sup>

(1. 桂林理工大学广西高校先进制造与自动化技术重点实验室 桂林 541006; 2. 桂林理工大学机械与控制工程学院 桂林 541006; 3. 中国科学院深圳先进技术研究院 深圳 518055; 4. 盛云科技有限公司 昆明 650000)

**摘 要:** 在复杂的水下环境中,针对传统 YOLO 目标检测方法检测性能差等问题,提出一种基于改进 YOLO11 的水下目标检测模型。首先,通过引入上下文引导模块 CGBD,采用多尺度特征提取器增强网络捕获能力;其次,为解决网络中特征冗余导致参数量过大的问题,设计轻量化高效聚合模块 RGCSPELAN 为模型减负;针对原有检测头定位识别能力不足且计算成本较高的问题,通过融合重参数化策略与细节增强卷积构建轻量高效的 DEC-Head 检测头。此外采用 Wise-Inner-MPD 损失函数提升模型的泛化能力并加速收敛。在 URPC 数据集上的实验结果表明,相较于基准模型 YOLO11,本文提出的方法在 mAP50 和 mAP50-90 平均精度均值上分别提升了 2.4% 和 2.1%。并且在 RUOD 数据集的实验结果中,本文所改进模型平均精度均值 mAP50 相比 YOLO11 提升了 1.3%,召回率 R 提升了 1.5%,较其他主流检测方法能够展现出更优的水下目标检测性能。

**关键词:** 上下文引导;轻量化高效聚合;检测头;损失函数

**中图分类号:** TP391; TN919.8 **文献标识码:** A **国家标准学科分类代码:** 510.4

## Underwater object detection model based on improved YOLOv11

Fang Zhenbo<sup>1,2</sup> Gao Xiangyang<sup>3</sup> Zhang Qieshi<sup>3</sup> Cheng Jun<sup>3</sup> Yang Mengjie<sup>4</sup>

(1. Key Laboratory of Advanced Manufacturing and Automation Technology, Guilin University of Technology, Guilin, 541006, China; 2. College of Mechanical and Control Engineering, Guilin University of Technology, Guilin 541006, China; 3. Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; 4. Shengyun Technology Company Limited, Kunming 650000, China)

**Abstract:** In the complex underwater environment, aiming at the poor detection performance of traditional YOLO target detection method, an underwater target detection model based on improved YOLO11 is proposed. Firstly, by introducing context guidance module CGBD, a multi-scale feature extractor is used to enhance the network capture capability. Secondly, in order to solve the problem that the number of parameters is too large due to feature redundancy in the network, the lightweight and efficient aggregation module RGCSPELAN is designed to reduce the burden of the model. To solve the problem that the localization and recognition ability of the original detection head is insufficient and the calculation cost is high, a lightweight and efficient DEC-Head detection head is constructed by combining the heavy parameterization strategy and detail enhancement convolution. In addition, Wise-Inner-MPD loss function is used to improve the generalization ability and accelerate the convergence of the model. The experimental results in URPC dataset show that compared with the benchmark model YOLO11, the proposed method improves the mean accuracy of mAP50 and MAP50-90 by 2.4% and 2.1% points respectively. Moreover, in the experimental results of RUOD dataset, Compared with YOLO11, the average accuracy of the improved model mAP50 increased by 1.3% and the recall rate R increased by 1.5%, showing better underwater target detection performance than other mainstream detection methods.

**Keywords:** context guidance; lightweight and efficient polymerization; detecting head; loss function

## 0 引 言

水下目标检测技术在海洋资源开发和维护国家海洋安

全与防御等方面发挥着重要作用。它可以用于勘探海底能源、如石油、天然气、矿产资源等,并能识别和跟踪水下潜在目标,从而支持海洋资源的有效开发和维护国家的海洋权

收稿日期:2025-01-11

<sup>\*</sup> 基金项目:国家自然科学基金—联合基金项目(U21A20487)、云南省科技人才与平台计划(院士专家工作站)(202305AF150152)、深圳市科技计划科技重大专项(KJZD20240903100000001)资助

益。此外,水下目标检测也广泛应用于海洋监测和军事领域,为科学研究和维护海洋领土完整提供了技术保障。因此对于水下目标检测技术的研究至关重要。

基于深度学习的水下目标检测方法主要分为两大类,一类是以 R-CNN<sup>[1]</sup> 系列为代表的双阶段目标检测方法,如 Fast-RCNN<sup>[2]</sup> 等,该方法首先找到图像特征的候选框,然后再对候选框区域进行分类和识别。双阶段方法检测精度较高,但是计算开销大,实时检测性能差,无法满足水下目标检测的要求。另一类则是以 YOLO<sup>[3-11]</sup> 系列和 SSD<sup>[12]</sup> 为代表的单阶段方法,该方法检测精度高、速度快、计算复杂度低,近年来在水下目标检测领域获得了广泛的研究和应用。

目前许多研究人员对水下目标检测领域展开了一系列研究。Ji 等<sup>[13]</sup> 提出一种特征增强和差分金字塔网络 (feature boosting and differential pyramid network, FBDPN),增强了不同尺度特征之间的上下文信息并实现了有效信息保留,但该网络的计算量和参数量较高。Dai 等<sup>[14]</sup> 提出一种门控跨域协作网络 (gated cross-domain collaborative network, GCC-Net),通过实时增强图像,并设计特征跨域交互和特征融合模块,提高了低对比度区域的可见度,但过度增强导致关键特征细节的丢失,在目标遮挡区域还是存在错检现象。Ma 等<sup>[15]</sup> 提出一种光照噪声模型,将最小加权熵误差准则 (MWME) 加入到 YOLO 网络的损失函数中,对网络参数进行训练优化,同时采用多误差处理策略对信息反向传播过程中的矢量误差进行处理,以加快收敛速度,实验表明该网络具有良好的检测性能。Zhou 等<sup>[16]</sup> 提出一种水下目标检测网络 (UODN),设计了跨阶段多分支模块 (CSMB) 和大核空间金字塔模块 (LSKP),将其组合作为检测网络的主干,增强了网络对各种水下物体尺度的提取能力。Gao 等<sup>[17]</sup> 设计了一种基于

自注意和远程关系捕获的大核卷积目标检测网络 (USF-Net),该网络首先提出一种混合扩展卷积注意机制 (HDLKA),其次将残差重构模块 (RConNeXt) 作为特征增强块,最后构建自适应空间特征融合检测头 (adaptively spatial feature fusion-head, ASFF-Head),该网络有效地减少了漏检和误检问题,大大提高了对小物体的识别能力。

上述研究为水下目标检测遇到的问题提供了良好的解决思路,但是水下目标检测面临的场景复杂,高噪声、图像模糊能见度较低等问题仍会影响模型检测性能。此外水下生物尺度大小不一,目标之间存在重叠遮挡导致错检漏检现象。因此本文对 YOLO11 模型进行改进,首先在主干和颈部网络中引入上下文引导模块 CGBD,通过学习局部特征和上下文联合特征,并进一步优化全局特征,提高了模型对场景中各种尺度特征的理解能力;其次针对卷积块特征冗余,参数和计算量大的问题,设计了一个轻量化高效聚合 RGCSPELAN 模块,减少了模型的参数量和计算量;然后提出了一种检测头 DEC-Head 代替原有检测头,增强检测头在水下场景中目标识别和定位的能力;最后将 Wise-Inner-MPD 三者进行结合作为模型的损失函数,通过动态聚焦机制、辅助边界框和最小点距离计算能够全面优化边界框回归,该损失函数更适合水下目标检测场景,能够有效提高检测精度。

1 YOLO11 网络及其改进模型

YOLO11 是由 Ultralytics 公司在 2024 年 9 月 30 日开源的最新版本,它是基于 YOLOv8 版本的又一次更新,实现了更高的精度、更快的速度和更少的参数量,支持图像检测、分割、分类识别等任务,该网络模型由主干网络 Backbone、颈部网络 Neck 和检测头 Detect Head 三部分组成,该网络结构如图 1 所示。

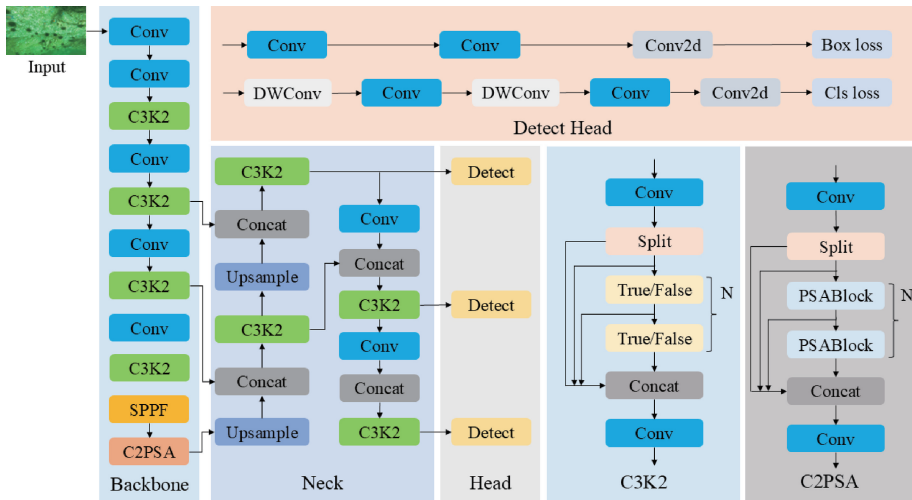


图 1 YOLO11 网络结构  
Fig. 1 YOLO11 network structure

在主干和颈部网络中, YOLO11 基于 C2f 改为 C3K2, 该模块使用两个卷积核, 并通过瓶颈层网络分割特征图, 细化了特征表示且加快了提取速度。另外在空间金字塔池化 SPPF 模块后面加入 C2PSA, C2PSA 基于 C2f 进行扩展, 引入 PSA 注意力机制, 通过多头注意力机制和前馈网络 FFN 来增强特征提取能力, 它可以选择性的添加残差块来优化梯度和训练效果, 同时使用 FFN 将特征映射到高维空间, 捕获特征的复杂非线性关系, 提升了模型对特征的捕捉能力。最后在分类检测头中加入两个深度可分离卷积 DWConv, 大大降低了计算冗余, 提升了计算效率。

本文提出的目标检测网络模型是基于 YOLO11 模型改进的, 首先在 Backbone 和 Neck 网络中加入上下文引导模块 CGBD, 通过学习上下文的联合特征并进一步优化; 然后设计轻量化高效聚合 RGCSPELAN 模块替换 C3K2; 其次提出一种检测头 DEC-Head 代替原有检测头, 减少参数的同时进一步增强检测头的识别和定位能力; 最后使用 Wise-Inner-MPDIoU 损失函数代替 YOLO11 的损失函数。改进的 YOLO11 网络结构如图 2 所示。

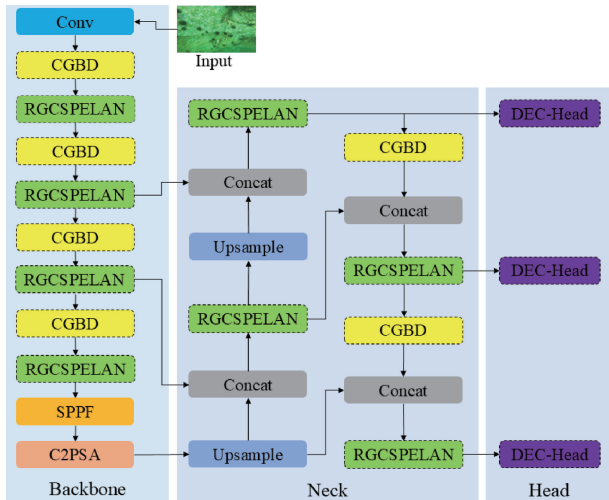


图 2 YOLO11 改进网络结构

Fig. 2 YOLO11 improved network structure

## 2 基于 YOLO11 的改进方法

### 2.1 上下文引导模块 CGBD

在 YOLO11 中, 传统的卷积通过设定步长来控制特征图的大小, 可以有效捕捉局部信息, 但是在保留细节特征方面存在局限性, 导致一些重要特征丢失。因此本文引入一种上下文引导模块 CGBD (context guide block down)<sup>[18]</sup>, 来弥补传统卷积块在水下目标检测中的不足。CGBD 模块包含局部特征提取器  $f_{loc}$ 、周围上下文提取器  $f_{sur}$ 、联合特征提取器  $f_{joi}$ 、全局上下文提取器  $f_{glo}$  四个部分。上下文引导模块网络结构如图 3 所示。图像经过  $1 \times 1$  的卷积处理后, 使用  $f_{loc}$  和  $f_{sur}$  来学习局部和周围上下文的特征,  $f_{loc}$  通过  $3 \times 3$  的标准卷积层实现, 能够更好捕捉

到局部变化信息, 在复杂的水下场景中区分出不同生物;  $f_{sur}$  使用  $3 \times 3$  的深度卷积来增加感受野的大小, 从而捕获更多的上下文信息;  $f_{joi}$  将局部特征提取器和周围上下文提取器提取的特征进行整合, 使用 Concat 连接层结合局部和周围上下文特征, 通过批量归一化 BN 和 PReLU 激活函数来增强特征表示能力;  $f_{glo}$  使用全局平均池化 GAP 聚合整个特征图的上下文信息, 然后利用两个全连接层 FC 进一步处理。上下文引导模块 CGBD 能够有效学习局部特征和周围环境的联合特征, 从而提高检测精度。

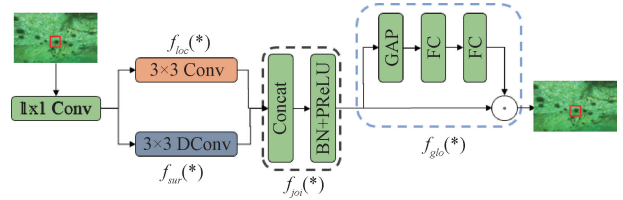


图 3 CGBD 原理图

Fig. 3 CGBD schematic diagram

### 2.2 检测头 DECHead

由于水下检测环境光照度低, 模糊难辨。YOLO 系列的传统检测头可能无法准确识别定位水下生物, 且传统的检测头参数量较大。因此本文设计了一种检测头 DECHead, 使用细节增强卷积 DEConv<sup>[19]</sup> 将先验信息整合到普通卷积层, 来增强表征和泛化能力, 然后通过重参数化技术, 将 DEConv 等效转化为普通卷积, 在保持精度的同时能够减少参数量和计算量, 检测头网络结构如图 4 所示。传统的 P3、P4、P5 检测层分别连接 3 个  $1 \times 1$  的分组归一化卷积层 Conv\_GN, Conv\_GN 将输入图像的特征维度进行分组, 并在每个组内进行归一化处理。然后将其送入 DEConv 捕捉更多图像细节, 融合后的特征图通过 Conv\_Reg 和 Conv\_Cls 层进行边界框的回归和类别的分类, scale 用于调整最终输出的尺度。

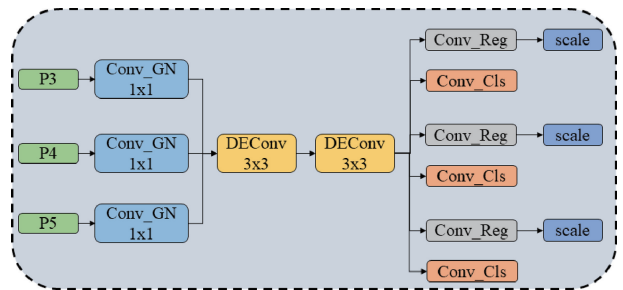


图 4 检测头原理图

Fig. 4 Schematic of the detection head

DEConv 原理如图 5 所示, 该卷积包含 5 个卷积层, 即普通卷积 (VC)、中心差分卷积 (CDC)、角差分卷积 (ADC)、水平差分卷积 (VDC)、垂直差分卷积 (VDC)。首先使用 CDC 和 ADC 将先验信息编码到卷积层中, 然后 VDC 和 HDC 将传统局部描述符集成到卷积层中, 通过学

习梯度信息来增强表征和泛化能力。普通卷积和差分卷积用于获取水平强度和增强梯度信息,并将 5 个卷积层学习到的特征整合在一起。最后利用重参数化技术在获取丰富特征的同时不引入额外的计算成本,作为 DEConv 的输出,输出公式如式所示,  $F_{in}$  和  $F_{out}$  表示输入和输出,  $K_i$  分别表示 5 个卷积核,  $*$  表示卷积运算,  $K_{cat}$  表示将并行卷积组合在一起的转换核。

$$F_{out} = DEConv(F_{in}) = \sum_{i=1}^5 F_{in} * K_i = F_{in} * \left( \sum_{i=1}^5 K_i \right) = F_{in} * K_{cat} \quad (1)$$

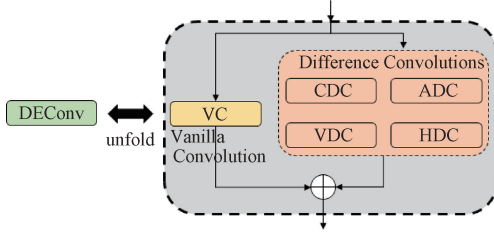


图 5 DEConv 原理图

Fig. 5 DEConv schematic diagram

### 2.3 轻量化高效聚合模块 RGCSPELAN

YOLO11 中的 C3K2 模块主要采用普通卷积对图像进行特征提取,不仅存在大量的冗余信息,而且增加了模型的计算量和参数,因此本文设计了 RGCSPELAN 模块来代替 C3K2 模块。该模块参考 GhostNet<sup>[20]</sup> 的思想,首先使用  $1 \times 1$  的标准卷积对输入图片进行通道压缩,然后分成两个分支,主分支使用重参数化卷积(reparameterizable convolution, RepConv)<sup>[21]</sup>, RepConv 在推理阶段将多分支卷积核和批量归一化 BN 层参数进行融合,生成一个等效的  $3 \times 3$  卷积核,然后将两个分支的特征整合在 Concat 层,最后利用  $1 \times 1$  的卷积进行输出,结构如图 6 所示。RGCSPELAN 模块利用 RepConv 增强特征提取和梯度流通能力,并使用缩放因子控制模块大小,使其同时兼顾大小模型。此外该模块显著减少了计算量和内存消耗,提高了推理速度。

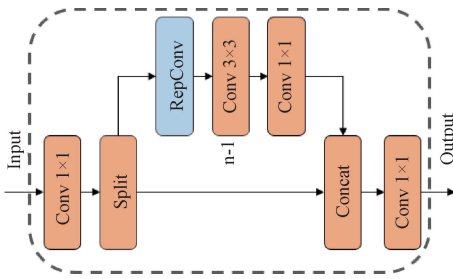


图 6 RGCSPELAN 原理图

Fig. 6 RGCSPELAN schematic diagram

### 2.4 损失函数改进

损失函数的设计与优化对于提高检测模型的精度至

关重要,通过合理选择和组合损失函数,可以有效地训练目标检测模型,提高其在复杂场景下的检测性能。YOLO11 默认采用 CIoU Loss 损失函数,CIoU 考虑了目标框之间的重叠面积、中心点距离以及长宽比,在多个几何因素上对预测框和真实框进行度量,有助于提高模型对目标的定位能力。但是在水下目标检测场景中,传统的 CIoU 损失函数泛化性能较差,对水下目标检测效果不佳,因此本文提出一种融合 WIoU-v3<sup>[22]</sup>、Inner-IoU<sup>[23]</sup> 和 MPD-IoU<sup>[24]</sup> 的复合损失函数(Wise-Inner-MPD),通过多层次几何约束优化边界框回归性能。

本文采用的 WIoU-v3 版本通过动态非单调聚焦机制评估锚框质量。首先定义离群值  $\beta$ ,  $\beta$  为  $L_{IoU}^*$  和  $\overline{L_{IoU}}$  的比值,其中  $L_{IoU}^*$  为锚框与目标框的损失函数,  $\overline{L_{IoU}}$  为  $L_{IoU}^*$  的指数移动平均值,然后基于  $\beta$  来构造非单调聚焦因子  $r$ ,其中  $\alpha$  和  $\beta$  为超参数, WIoU-v3 通过  $r$  动态调整损失权重,使得模型在训练过程中能够关注不同质量的样本。公式如下:

$$\begin{cases} L_{WIoU-v3} = r \cdot L_{WIoU-v1} \\ r = \frac{\beta}{\delta \alpha^{\beta-\delta}}; \beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty) \end{cases} \quad (2)$$

Inner-IoU 使用辅助边界框来计算 IoU 损失, Inner-IoU 通过缩小真实框和预测框的尺寸,计算缩小后框的交并比,从而提升边界框回归的精度。  $(x_c^{gt}, y_c^{gt})$  和  $(x_c, y_c)$  为真实框和预测框的中心点,  $w^{gt}, h^{gt}$  和  $w, h$  为真实框和预测框的宽高,  $s$  为比例因子,  $inter$  为两个框之间重叠区域的面积,  $union$  为两个框覆盖的区域面积。最后得到交并比。公式如下:

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} \cdot s}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} \cdot s}{2} \quad (3)$$

$$b_l^{gt} = y_c^{gt} - \frac{h^{gt} \cdot s}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} \cdot s}{2} \quad (4)$$

$$b_l = x_c - \frac{w \cdot s}{2}, b_r = x_c + \frac{w \cdot s}{2} \quad (5)$$

$$b_l = y_c - \frac{h \cdot s}{2}, b_b = y_c + \frac{h \cdot s}{2} \quad (6)$$

$$inter = \max(\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) \cdot \max(\min(b_b^{gt}, b_b) - \max(b_l^{gt}, b_l)) \quad (7)$$

$$union = (w^{gt} \cdot h^{gt}) \cdot s^2 + (w \cdot h) \cdot s^2 - inter \quad (8)$$

$$IoU_{inner} = \frac{inter}{union} \quad (9)$$

$$L_{Inner-IoU} = 1 - IoU_{inner} \quad (10)$$

MPD-IoU 通过最小化预测边界框与实际边界框之间的左上角和右下角点的距离来提高边界框回归的准确性和效率,考虑了重叠或非重叠面积、中心点距离以及宽度和高度的偏差,同时简化了计算过程。有助于改善边界框回归训练效果,提高收敛速度。  $w$  和  $h$  分别为真实框的宽度和高度,  $(x_1^{pred}, y_1^{pred})$  和  $(x_1^{gt}, y_1^{gt})$  分别为预测框和真实



框的左上角点坐标。 $(x_2^{pred}, y_2^{pred})$  和  $(x_2^{gt}, y_2^{gt})$  为预测框和真实框的右上角点坐标。公式如下:

$$L_{MPD-IoU} = 1 - MPD_{IoU} \quad (11)$$

$$MPD_{IoU} = \frac{inter}{union} + \frac{d_1^2}{w^2 + h^2} + \frac{d_2^2}{w^2 + h^2} \quad (12)$$

$$d_1^2 = (x_1^{pred} - x_1^{gt})^2 + (y_1^{pred} - y_1^{gt})^2 \quad (13)$$

$$d_2^2 = (x_2^{pred} - x_2^{gt})^2 + (y_2^{pred} - y_2^{gt})^2 \quad (14)$$

### 3 实验结果与分析

#### 3.1 数据集与实验参数设置

本文实验采用的数据集为开源的全国水下机器人专业比赛-光学赛事数据集 URPC,该数据集涵盖多种水下生物,如海星、海胆、扇贝、海参和海草等。通过对该数据集进行相关处理,移除干扰的样本和错误标注后,总共获得 7 383 张图片,该数据集包括海星(starfish)、海参(holothurian)、海胆(echinus)、扇贝(scallop)四种水下生物,其中各个类别数量分别为 1 785、1 227、5 792、2 766,并按照 8:2 的比例将数据集划分为训练集 5 907 张图像和测试集 1 476 张。

本文实验的硬件环境为 NVIDIA P40 显卡,24 G 显存,Ubuntu20.04 操作系统。软件开发环境为 PyTorch-2.2.0,CUDA11.8,python3.8。训练阶段使用 SGD 优化器训练模型,训练轮次为 100,输入图像分辨率大小为  $640 \times 640$ ,动量 0.937,权重衰减因子为 0.000 5,初始学习率 0.01,批次大小 32,线程数为 8。

#### 3.2 实验评价指标

本文实验通过精确度 P(precision)、召回率 R(recall)、

平均精度值 mAP(mean average precision)、模型推理速度(FPS)、浮点运算数(GFLOPS)、参数量(Parameters)等指标来对模型进行评估。计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (15)$$

$$R = \frac{TP}{TP + FN} \quad (16)$$

$$AP = \int_0^1 P(R) dR \quad (17)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP(i) \quad (18)$$

$$FPS = \frac{1\ 000}{t_{前处理} + t_{推理} + t_{后处理}} \quad (19)$$

其中,TP 代表实际为正样本且模型预测为正样本的数量;FP 代表实际为负样本但是模型预测为负样本的数量;FN 代表实际为正样本但被模型预测为负样本的数量。N 为数据集样本中的类别数量,AP 为 PR 曲线上精确度的平均值,mAP 代表对数据集中所有类别取平均得到的平均精度,该指标能够全面的评估模型的检测性能。FPS 为每秒处理多少帧图像,其中  $t_{前处理}$ 、 $t_{推理}$ 、 $t_{后处理}$  分别表示前处理耗时、推理耗时和后处理耗时。

#### 3.3 检测头对比实验

为了验证本文所提出检测头 DEC-Head 的有效性,在同样实验条件下,以 YOLO11 模型为基准,分别加入 DyHead<sup>[25]</sup>、SEAMHead<sup>[26]</sup>、MultiSEAMHead<sup>[26]</sup> 等检测头在 URPC 数据集上进行对比,对比实验结果如表 1 所示。

表 1 在 URPC 数据集上的对比实验  
Table 1 Comparative experiments on URPC datasets

检测头	P/%	R/%	mAP50/%	mAP50-95/%	Params/M	FLOPS/G	FPS
YOLO11	79.0	75.1	81.7	46.6	2.58	6.3	<b>71.6</b>
YOLO11+Dyhead	79.7	74.9	81.7	46.4	3.09	7.4	43.7
YOLO11+SEAMHead	79.7	74.3	81.4	46.0	2.49	6.1	52.8
YOLO11+MultiSEAMHead	<b>80.1</b>	75.3	81.9	46.1	4.59	6.0	37.9
YOLO11+DECHead	79.4	<b>76.6</b>	<b>82.5</b>	<b>47.3</b>	<b>2.26</b>	<b>6.0</b>	70.7

从表 1 中可以看出,与 YOLO11 基准模型相比,加入 DEC-Head 的 YOLO11 模型检测精度提升显著。相比于加入 YOLO11、Dyhead、SEAMHead 和 MultiSEAMHead 检测头,召回率 R 分别提升了 1.5%、1.7%、2.3% 和 1.3%,mAP50 分别提升了 0.8%、0.8%、1.1% 和 0.6%,参数量较其他检测头最低为 2.26 M,同时,DEC-Head 的计算复杂度较 YOLO11 减少了 0.3 G。因此加入 DEC-Head 检测头能够有效降低误检率,更有利于水下目标检

测任务。

#### 3.4 消融实验

为了进一步验证对 YOLO11 模型改进的有效性,对所改进的部分在 URPC 数据集上进行消融实验,首先加入 CGBD 模块,然后加入轻量化 RGCSPELAN 模块,其次添加检测头 DEC-Head 模块,最后使用 Wise-Inner-MPD IOU 损失函数。在同样的实验条件下来评估模型改进后的检测效果,实验结果如表 2 所示。

表 2 在 URPC 数据集上的消融实验  
Table 2 Ablation experiments on URPC datasets

实验	CGBD	RGCSPELAN	DEC-Head	损失函数	P/%	R/%	mAP50/%	mAP50-95/%	Params/M	FLOPS/G	FPS
1					79.0	75.1	81.7	46.6	2.58	6.3	71.6
2	✓				<b>81.5</b>	76.1	83.4	47.9	3.53	9.0	51.3
3		✓			80.6	74.3	82.4	47.1	<b>2.24</b>	6.2	<b>86.2</b>
4			✓		79.4	76.6	82.5	47.3	2.26	<b>6.0</b>	70.7
5				✓	80.1	74.8	82.0	46.5	2.58	6.3	73.2
6	✓	✓			80.4	77.2	83.3	48.2	3.19	9.0	56.5
7	✓	✓	✓		80.6	76.7	83.5	48.3	2.87	8.7	56.2
8	✓	✓	✓	✓	80.9	<b>77.6</b>	<b>84.1</b>	<b>48.7</b>	2.87	8.7	58.5

从表 2 中的结果来看,加入以上 4 个模块对 mAP50 指标均有提升,首先在 YOLO11 基准模型上加入 CGBD 模块,精确率 P 和召回率 R 分别提升了 2.5% 和 1.0%, mAP50 和 mAP50-95 指标分别提升了 1.7% 与 1.3%,说明该模块能够很好学习全局联合特征,提升了对水下生物中小目标的检测能力;将 C3K2 替换为轻量化聚合模块 RGCSPELAN 后,相较于基准模型召回率 R 下降了 0.9%,精确率 P 提升了 1.6%,但是参数量和计算量较 YOLO11 分别降低了 0.34 M 和 0.1 G,FPS 最高为 86.2;然后将原有检测头替换为 DECHead 模块,相比于 YOLO11 传统检测头精度有所提升,参数量减少了 0.32 M,并且计算量也降低了 0.3 G;将 Wise-Inner-MPD

损失函数应用于基准模型,平均检测精度 mAP50 提高了 0.3%;最后将 CGBD、RGCSPELAN、DECHead、损失函数四个模块同时加入 YOLO11,精确率 P 和召回率 R 分别提升了 1.9% 和 2.5%,且在 mAP50 和 mAP50-95 指标上较 YOLO11 分别提高了 2.4% 和 2.1%,所提改进方法相比于 YOLO11 有效地提升了检测精度。

3.5 对比实验

为了验证本文所提方法的有效性,将所改进的模型与当前比较主流常用的目标检测算法,如 Faster-R-CNN、Cascade R-CNN、YOLOv3-tiny<sup>[27]</sup>、YOLOv5、YOLOv8、YOLOv10<sup>[28]</sup>、YOLO11 在 URPC 数据集上进行对比,实验结果如表 3 所示。

表 3 不同算法对比实验  
Table 3 Comparative experiment of different algorithms

模型	P/%	R/%	mAP50/%	mAP50-95/%	Params/M	FLOPS/G	FPS
Faster R-CNN	75.2	73.8	79.3	42.4	28.2	158.7	17.9
Cascade R-CNN	78.8	71.9	80.6	44.5	56.1	186.4	11.8
YOLOv3-tiny	<b>81.4</b>	71.9	78.9	42.2	12.12	18.9	<b>173.6</b>
YOLOv5	80.1	74.1	81.5	46.4	<b>2.50</b>	7.1	88.8
YOLOv8	81.0	74.1	81.8	46.9	3.01	8.1	102.3
YOLOv10	77.7	73.9	79.6	45.1	2.69	8.2	71.7
YOLO11	79.0	75.1	81.7	46.6	2.58	<b>6.3</b>	71.6
本文算法	80.9	<b>77.6</b>	<b>84.1</b>	<b>48.7</b>	2.87	8.7	58.5

根据表中的数据分析,本文所提的改进方法相比于其他主流目标检测算法有更好的检测性能。在 mAP50 和 mAP50-95 指标上,本文所提方法明显优于其他模型,分别达到 84.1% 和 48.7%,模型参数量和计算量分别为 2.87 M 和 8.7 G,保持了精度和轻量的平衡,因此综合来看,本文方法在水下目标检测场景中具有一定的优势。

为进一步直观展示改进算法的检测效果和优势,将本文所改进的算法与目标检测算法 YOLOv5、YOLOv8、YOLOv10、YOLO11 进行检测效果对比,检测效果如图 7

所示。从图 7 中可看出,所对比的主流检测模型在水下场景中都存在漏检、误检、重复检测现象,但是本文所提方法对误检和漏检有明显改善,证明了本文方法的有效性,能够更好应用于水下目标检测。

3.6 RUOD 数据集对比

为进一步验证本文所提方法在复杂场景下的水下目标检测性能,在 RUOD 数据集上进行实验,RUOD 数据集包含多种水下生物,包括海参、扇贝、海胆、海星、珊瑚、鱼、潜水员、墨鱼、海龟、水母共 10 种类别,其中训练集图像为 9 800 张,测试集图像为 4 200 张。本文算法与当下主流检

测模型进行对比,对比数据如表 4 所示。从表 4 可以看出,本文所改进算法相较于 YOLO11 基准模型,精确度 P、召回率 R、mAP50 等指标分别提升了 0.6%、1.5%、1.3%。

本文所改进模型的检测精度优于其他主流目标检测算法,而且参数量和计算量适中,验证了本文改进模型在复杂水下环境下良好的检测性能。

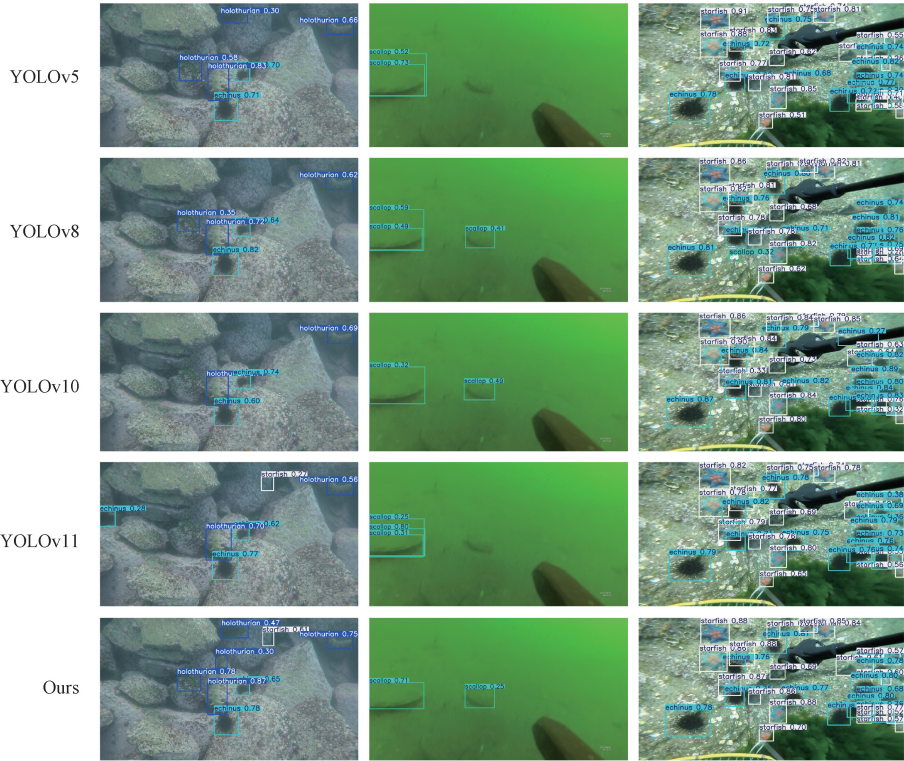


图 7 检测效果对比  
Fig. 7 Comparison of detection effect

表 4 RUOD 对比实验

Table 4 Comparative experiment of RUOD

模型	P/%	R/%	mAP50/%	mAP50-95/%	Params/M	FLOPS/G	FPS
Faster R-CNN	81.2	72.1	80.3	47.0	28.3	150.5	19.3
Cascade R-CNN	80.6	72.0	79.6	49.7	56.2	179.2	12.3
YOLOv3-tiny	84.0	73.8	80.6	52.7	12.12	18.9	<b>166.9</b>
YOLOv5	83.1	74.1	81.5	55.9	<b>2.50</b>	7.1	90.7
YOLOv8	84.0	75.7	82.8	<b>57.8</b>	3.01	8.1	97.0
YOLO10	81.7	74.0	81.0	56.4	2.69	8.2	72.3
YOLO11	83.9	75.5	82.6	57.7	2.58	<b>6.3</b>	74.9
本文算法	<b>84.5</b>	<b>77.0</b>	<b>83.9</b>	<b>57.8</b>	2.87	8.7	56.8

4 结 论

在水下目标检测任务中,由于检测环境复杂,受干扰影响大,导致检测精度低,更易出现漏检错检的现象,因此本文提出一种改进 YOLO11 的水下目标检测模型。首先在主干和颈部网络中引入 CGBD 模块增强了上下文特征的学习能力;其次在网络中加入轻量化高效聚合模块 RGCSPELAN,在保持精度基本不变的情况下减少参数量

和计算量;然后设计轻量检测头 DEC-Head 增强检测头在水下的定位识别能力,并减少参数量;最后应用 Wise-Inner-MPD 损失函数平衡不同检测场景的训练结果,使模型快速收敛回归。实验结果表明本文提出的检测模型具有良好的检测性能。未来将继续优化算法,进一步加强模型的泛化性,提高模型的实时检测性能,并部署到水下航行设备上。

## 参考文献

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [2] REN SH Q, HE K M, 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.
- [3] REDMON J. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:779-788.
- [4] LI CH Y, LI L L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976,2022.
- [5] 周道先,张吟龙,徐高飞,等.基于形变卷积和深层聚合网络的水下文物检测[J].仪器仪表学报,2023,44(11):185-195.  
ZHOU D X, ZHANG Y L, XU G F, et al. Underwater cultural relics detection based on deformation convolution and deep aggregation network[J]. Chinese Journal of Scientific Instrument, 2023,44(11):185-195.
- [6] 赵永胜,严志远,毛瑞霞,等. MEAS-YOLO: 改进 YOLOv5 的水下目标智能检测算法[J]. 电子测量技术,2024,47(13):183-190.  
ZHAO Y SH, YAN ZH Y, MAO R X, et al. MEAS-YOLO: An improved underwater target intelligent detection algorithm for YOLOv5 [J]. Electronic Measurement Technology, 2024,47(13):183-190.
- [7] 杨继海,裴晓芳.多尺度与跨空间信息聚合网络的水下目标检测[J].激光与光电子学进展,2024,61(24):374-383.  
YANG J H, PEI X F. Underwater target detection in multi-scale and cross-spatial information aggregation networks [J]. Laser & Optoelectronics Progress, 2019, 61(24):374-383.
- [8] FU CH P, LIU R SH, FAN X, et al. Rethinking general underwater object detection: Datasets, challenges, and solutions[J]. Neurocomputing, 2023, 517: 243-256.
- [9] 周志耀,马常霞,杨丽莎,等.基于轻量级改进的 YOLOv8 水下目标检测模型[J]. 电子测量技术, 2024,47(19):181-189.  
ZHOU ZH Y, MA CH X, YANG L SH, et al. Underwater object detection model based on lightweight improved YOLOv8 [J]. Electronic Measurement Technology, 2024,47(19):181-189.
- [10] 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.
- [11] ZHAO Z J,谈宇光,刘鹏,等.基于激光选通成像技术的水下远距离目标智能识别系统研究[J].集成技术, 2023,12(2):39-52.  
ZHAO ZH J, TAN Y G, LIU P, et al. Research on underwater long-range target intelligent identification system based on laser range-gated imaging technology[J]. Journal of Integration Technology, 2023, 12(2):39-52.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [13] JI X, CHEN SH J, HAO L Y, et al. FBDPN: CNN-Transformer hybrid feature boosting and differential pyramid network for underwater object detection[J]. Expert Systems with Applications, 2024, 256: 124978.
- [14] DAI L H, LIU H, SONG P H, et al. A gated cross-domain collaborative network for underwater object detection [J]. Pattern Recognition, 2024, 149: 110222.
- [15] MA H P, ZHANG Y J, SUN SH Y, et al. Weighted multi-error information entropy based you only look once network for underwater object detection [J]. Engineering Applications of Artificial Intelligence, 2024, 130: 107766.
- [16] ZHOU H, KONG M W, YUAN H X, et al. Real-time underwater object detection technology for complex underwater environments based on deep learning [J]. Ecological Informatics, 2024, 82: 102680.
- [17] GAO Z R, SHI Y L, LI SH. Self-attention and long-range relationship capture network for underwater object detection[J]. Journal of King Saud University-Computer and Information Sciences, 2024, 36(2): 101971.
- [18] WU T Y, TANG SH, ZHANG R, et al. CGNet: A light-weight context guided network for semantic segmentation [J]. IEEE Transactions on Image Processing, 2020, 30: 1169-1179.
- [19] CHEN Z X, HE Z W, LU ZH M. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024,33:1002-1015.
- [20] HAN K, WANG Y H, TIAN Q, et al. GhostNet:



- More features from cheap operations[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580-1589.
- [21] DING X H, ZHANG X Y, MA N N, et al. RepVGG: Making VGG-style convnets great again[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13733-13742.
- [22] 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.
- [23] ZHANG H, XU C, ZHANG SH J. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[J]. ArXiv preprint arXiv:2311.02877, 2023.
- [24] MA S L, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression[J]. ArXiv preprint arXiv:2307.07662, 2023.
- [25] DAI X Y, CHEN Y P, XIAO B, et al. Dynamic head: Unifying object detection heads with attentions [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7373-7382.
- [26] WANG C Y, YE H I, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]. European conference on computer vision. Cham: Springer Nature Switzerland, 2024: 1-21.
- [27] FARHADI A, REDMON J. YOLOv3: An incremental improvement[C]. Computer Vision and Pattern Recognition. Berlin/Heidelberg, Germany: Springer, 2018.
- [28] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [J]. ArXiv preprint arXiv:2405.14458, 2024.

## 作者简介

**方侦波**, 硕士研究生, 主要研究方向为移动机器人。

E-mail: 3195291955@qq.com

**高向阳**, 硕士, 高级工程师, 主要研究方向为嵌入式系统、智能机器人。

**张镔石**(通信作者), 博士, 副教授, 主要研究方向为图像增强与理解, 智能机器人。

E-mail: qs.zhang@siat.ac.cn

**程俊**, 博士, 研究员, 主要研究方向为机器视觉、智能机器人、人工智能。

**杨梦杰**, 本科, 高级工程师, 主要研究方向为人工智能、智慧城市、智慧园区。