

# 基于改进 YOLOv8s 的水下目标检测方法

冯迎宾, 符珊

(沈阳理工大学 自动化与电气工程学院, 沈阳 110159)

**摘要:** 针对水下环境存在的色偏、细节模糊以及水下目标存在多尺度和遮挡等问题, 提出了一种改进 YOLOv8s 的水下目标检测算法 SLG-YOLOv8s。首先, 通过 Shallow-UWnet 网络对水下图像进行增强, 提高图像的对比度和清晰度; 其次, 提出了轻量化多尺度全局注意力 (lightweight multi-scale global attention, LMGA) 模块, 将此模块与 YOLOv8s 主干部分的 C2f 融合, 通过动态调整权重进行特征重标定, 加强特征表达能力, 减少计算量; 最后, 通过聚集和分发 (gather-and-distribute, GD) 机制采用跨层特征融合的方式增强中间层的信息融合能力, 从而提升模型对多尺度目标的检测效率。实验结果表明, SLG-YOLOv8s 算法的平均精度均值达到了 95.1%, 与 YOLOv8s 算法相比平均精度均值提升了 5.1%, 精确率和召回率分别提升了 4.5% 和 5.2%, IoU 为 0.5 时海参、海胆、海星、扇贝检测的平均精确率分别提升了 4.3%、5.3%、5.6%、5.5%, 研究结果可为水下机器人精准捕捞提供重要依据。

**关键词:** 水下目标检测; YOLOv8s; Shallow-UWnet; 跨层特征融合

**中图分类号:** TP391.41 **文献标志码:** A **DOI:** 10.3969/j.issn.1003-1251.2025.03.002

## Underwater Object Detection Method Based on Improved YOLOv8s

FENG Yingbin, FU Shan

(Shenyang Ligong University, Shenyang 110159, China)

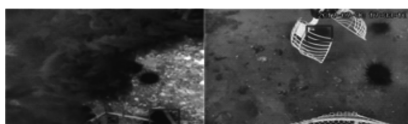
**Abstract:** To address the issues of color bias, blurring of details, and multi-scale and occlusion of targets in the underwater environment, an improved underwater target detection algorithm SLG-YOLOv8s is proposed, which improves YOLOv8s. Firstly, the underwater image is enhanced through the Shallow-UWnet network to improve the contrast and clarity of the image. Secondly, a lightweight multi-scale global attention (LMGA) module is proposed, which integrates this module with the C2f of the YOLOv8s backbone. By dynamically adjusting weights for feature recalibration, the feature expression ability is enhanced and the computational workload is reduced. Finally, the information fusion capability of the intermediate layer is enhanced through the Gather and Distribute (GD) mechanism using cross layer feature fusion, thereby improving the detection efficiency of the model for multi-scale targets. The experimental results show that the mAP of the SLG-YOLOv8s algorithm can reach 95.1%, which is 5.1% higher than the YOLOv8s algorithm. The precision and recall rates are improved by 4.5% and 5.2%, respectively. When the IoU of sea cucumber, sea urchin, sea star, and scallop is 0.5, the average precision increases by 4.3%, 5.3%, 5.6%, and 5.5%, respectively, and the results of the study can be much helpful for underwater robot precision fishing.

**Key words:** underwater object detection; YOLOv8s; Shallow-UWnet; cross-layer feature fusion

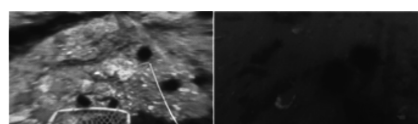
海底生物在经济和生态平衡中扮演着关键角色<sup>[1]</sup>。目前,捕捞海底生物主要依靠撒网和人工潜水等,存在捕捞效率低、成本高等缺点。研制自主抓取海底生物的机器人<sup>[2]</sup>是当前研究的热点问题之一,如何高效、准确识别海底生物<sup>[3]</sup>是实现机器人自主抓取<sup>[4]</sup>的关键。

传统水下目标检测算法依赖特征描述符<sup>[5]</sup>,其中纹理特征法<sup>[6]</sup>具有旋转不变性和较强的抗噪性,但计算量大、耗时长,张祥翔等<sup>[7]</sup>使用灰度共生矩阵提取纹理,效果显著但计算负担大。与上述两种传统方法相比,基于卷积神经网络(CNN)的目标检测方法准确率大幅提高,但水下环境的复杂性导致该方法在水下目标检测方面效果不佳。

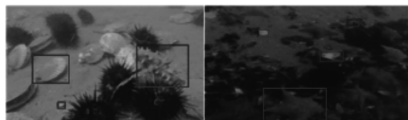
李辉等<sup>[8]</sup>通过在两阶段目标检测方法增加多尺度特征融合网络和注意力网络、优化样本比例以保证模型快速收敛,但检测速度较慢。Cai等<sup>[9]</sup>提出 Cascade R-CNN 算法,通过调整 IoU 阈值提升位置感知,但其使用三层级联检测器影响模型收敛和检测速度。李宝奇等<sup>[10]</sup>在 SSD 算法基础上对特征提取网络进行轻量化改进,但对于多尺度目标检测精度有待提高。此外,水下生物图像通常存在清晰度低和色彩失真等问题。郝琨等<sup>[11]</sup>将 ACE 算法应用于 MSRCR 算法增强水下图像,并改进 YOLOv3 算法,改善对遮挡目标的检测能力,但检测率仅为 78.7%。



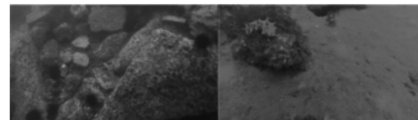
(a) 遮挡目标样例



(b) 不同光线的目标样例



(c) 多尺度目标样例



(d) 不同地形的目标样例

图1 UPRC2019 数据集样例

Fig. 1 UPRC2019 dataset example

## 1.2 研究方法

本文设计的网络结构主要由浅层网络图像增强、特征提取、特征融合和检测头检测四部分组成,具体技术路线如图2所示。输入的图片经过 Shallow-UWnet 网络进行图像增强,增强后的图片输入到 LMGA-CSPDrakNet-53 网络中,通过结合平均池化和最大池化缩小特征图的分辨率,逐层扩大感受野,获取更多目标信息。

综上所述,本文针对水下图像中目标尺寸多变、背景复杂、噪声大等问题,提出一种基于改进 YOLOv8s 的水下目标检测方法。首先,利用 Shallow-UWnet 算法增强水下图像,解决水下图像噪声大、颜色失真等问题;其次,提出多尺度全局注意力(lightweight multi-scale global attention, LMGA)与 YOLOv8s 特征提取网络的 C2f 模块融合,解决背景复杂导致的漏检;最后,采用聚集和分发(gather-and-distribute, GD)机制获取更多目标信息,解决目标尺度多变问题。

## 1 材料与方法

### 1.1 数据与来源

本研究采用国际水下机器人大赛开源数据集 UPRC2019<sup>[12]</sup>,此数据集由真实海底环境拍摄的视频截取所得。数据集包括 4 个类别,分别是海胆、海参、海星、扇贝,每个实例在整个数据集所占的比例为 10:3:4:3,共有 6 000 个实例。其中,光线亮与暗的图片比例为 1:4,礁石海底和沙子海底的比例为 3:7,数据集中的目标存在遮挡和多尺度等问题,本研究对数据集进行筛选,选取目标丰富的图片 2 000 张,数据样例如图 1 所示。采用 Labelimg 软件进行重标注,保存为 XML 格式,最后将数据集划分为训练集、验证集和测试集,划分比例为 8:1:1。

为了解决多尺度特征融合过程中信息损失问题,特征融合网络采用 GD 机制,其中低阶特征对齐模块(low-stage feature alignment module, Low-FAM)收集骨干网络提取的不同分辨率特征  $B_5$ 、 $B_4$ 、 $B_3$ 、 $B_2$ ,并通过上采样进行特征对齐,低阶信息融合模块(low-stage information fusion module, Low-IFM)融合对齐后的特征生成全局特征,信息注入模块采用注意力机制提取局部特征,并通过

切片操作分为两部分,将高层全局特征和局部特征融合分发给  $P3$  和  $P4$  层,从而增强分支的检测能力。 $P3$ 、 $P4$ 、 $B5$  层通过高阶特征对齐模块 (high-stage feature alignment module, High-FAM) 融合高层特征信息,通过高阶特征融合模块 (high-

stage information fusion module, High-IFM) 得到全局特征,然后信息注入模块采用注意力机制将信息注入不同层,将高层全局特征和局部特征融合送入  $N4$  和  $N5$  层,得到最终的特征图  $P3$ 、 $N4$ 、 $N5$ ,最后进行分类和边框回归。

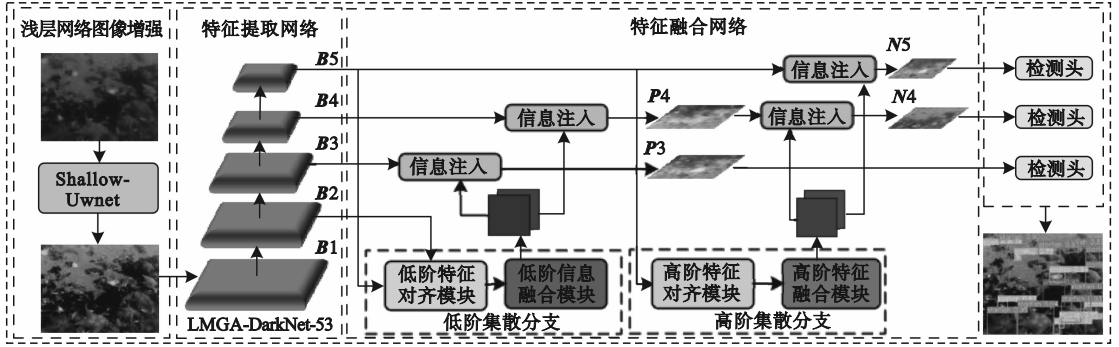


图 2 技术路线图

Fig. 2 Technology roadmap

1.2.1 浅层图像增强

传统的图像增强算法主要包括 HE 算法<sup>[13]</sup>、MSRCR 算法、DCP 算法<sup>[14]</sup>等。HE 算法处理耗时最少,但效果不佳;MSRCR 算法可以提高水下图像亮度,但颜色校正不佳。由于水下散射和吸收使图像偏向暗色调,限制了 DCP 算法通过暗通道估计透射率的能力,导致泛化能力弱。

基于深度学习的图像增强方法主要包括 Shallow-UWnet、FIVEAPLUS<sup>[15]</sup>、PUIE<sup>[16]</sup>等算法,其中 FIVEAPLUS 算法和 PUIE 算法训练起来难度相对较大。

为解决上述方法的缺点,本文采用基于 Shallow-UWnet 的水下图像增强算法,用于解决水下图像色偏和对比度低的问题,其结构如图 3 所示。Shallow-UWnet 网络是一个轻量型的全连接卷积神经网络,能够通过浅层神经网络实时恢复图像视觉感知质量,加强图像特征表达能力。相比于其他深度学习的图像增强方法,训练过程更高效;相比于 HE 算法和 MSRCR 算法,本文算法可有效提升图像对比度;相比于 DCP 算法,本文算法可以在不依赖环境先验知识的情况下,提高模型鲁棒性并抑制图像噪声。

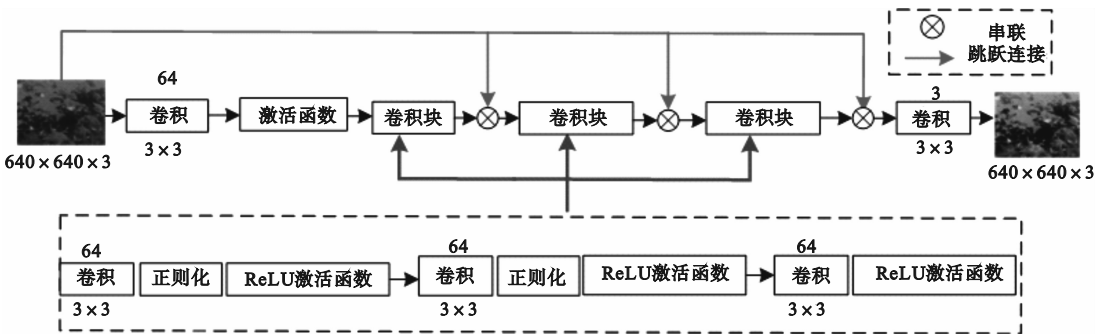


图 3 Shallow-UWnet 模型结构图

Fig. 3 Shallow-UWnet model structure diagram

图 3 所示模型中,首先采用带有  $3 \times 3$  卷积核的卷积层提取输入图像的边缘和纹理特征,并使用 ReLU 激活函数和批量归一化增加非线性。随后,卷积块进一步提取多样化特征,结合正则化,提高泛化能力。残差模块通过跳跃连接强化特征提取,并减轻梯度消失问题<sup>[17]</sup>。最后,模型通过

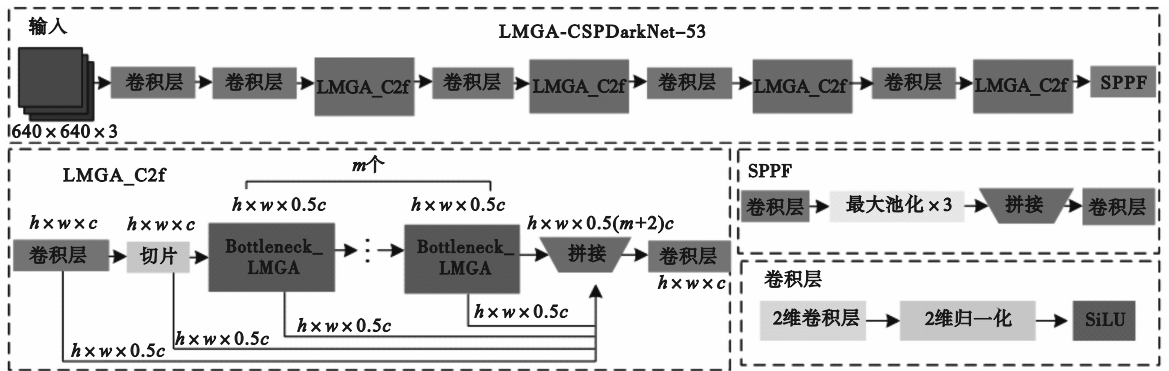
多任务损失函数训练,优化边缘清晰度和纹理结构。

1.2.2 引入轻量化多尺度全局特征提取模块

增强后的图片有利于检测网络提取更多特征,但随着网络层次不断加深,目标信息易遗漏。深层网络可以在不同尺度上扩大感受野,解决目

标漏检问题。对于复杂的水下环境,主干网络在提取遮挡目标的有效特征方面存在局限性。为此,本文将 YOLOv8 主干部分的 CSPDarkNet-53 网络替换为 LMGA-CSPDarkNet-53 网络,其结构如图 4 所示。输入图像首先通过两个连续卷积层模块,提取图像的基本特征,卷积层模块中包括 2 维卷积层、2 维归一化和 SiLU 非线性激活函数。其次,通过 LMGA\_C2f 模块进行切片处理,LM-

GA\_C2f 模块在原有 C2f 模块中融入 LMGA 注意力机制,利用  $m$  个 Bottleneck\_LMGA 瓶颈残差结构进行降维和升维,有效控制信息流并降低模型复杂度。SPPF 空间金字塔结构将不同尺度的特征图融合起来,然后对任意尺寸的特征图进行固定尺寸的最大池化操作,将每个池化得到的特征图拼接起来,形成一个固定长度的特征向量。最后,将该特征向量输入到全连接层。



注:图中  $h$  和  $w$  为特征图的宽和高,  $c$  为通道数。

图 4 LMGA-DarkNet-53 结构图

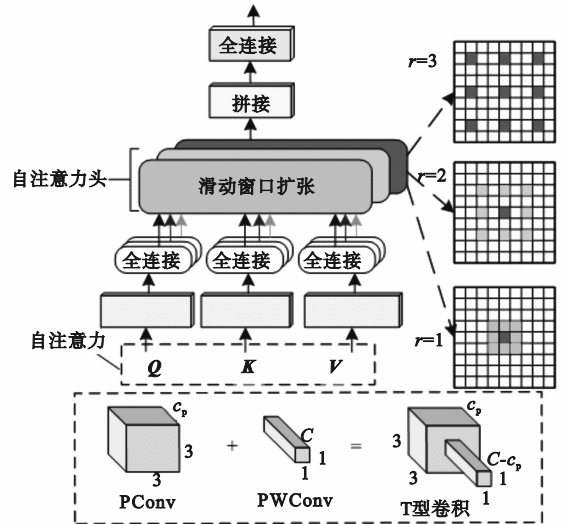
Fig. 4 LMGA-DarkNet-53 structure diagram

LMGA 原理如图 5 所示,为了有效捕获目标信息,将 PConv<sup>[18]</sup> 与 PWConv 结合成 T 型卷积结构,PConv 对输入通道进行分段处理,减少信息丢失。而 PWConv 采用  $1 \times 1$  的卷积核进行特征融合,进一步优化网络的计算效率。与传统卷积层处理所有输入通道相比,显著减少了计算量。为解决目标漏检问题,将 LMGA 集成在 Bottleneck 结构中,通过计算目标与背景显著性差异并调整检测权重,增强局部相关性。

由图 5 可知,输入的特征进行自注意力操作,  $Q$ 、 $K$ 、 $V$  分别代表查询、键和值向量,通过查询变量  $Q$  查询  $K$  中的关键信息,最后把关键信息整合到  $V$  中得到特征向量,其分量  $x_{ij}$  计算表达式为

$$x_{ij} = \text{Attention}(q_{ij}, K_r, V_r) = \text{Softmax}\left(\frac{q_{ij} K_r^T}{\sqrt{d_k}}\right) V_r \quad (1)$$

式中:  $x_{ij}$  代表输入  $i$  的各个部分对某时刻输入  $j$  映射的权重,  $1 \leq i \leq W, 1 \leq j \leq H$ ;  $\text{Attention}$  代表权重系数;  $(i, j)$  代表特征映射的原始查询位置;  $H$  和  $W$  分别代表特征映射的高度和宽度;  $K_r$  和  $V_r$  表示将特征映射到  $K$  和  $V$  中选择的键和值;  $d_k$  表示每个自注意力头的  $K$  向量维度;  $q_{ij}$  代表特征映射的查询变量;  $\text{Softmax}$  函数将向量转换为概率分布。



注:  $C$  为 PWConv 的通道数,  $c_p$  为 PConv 的通道数,  $r$  为膨胀率。

图 5 LMGA 原理图

Fig. 5 LMGA principle diagram

以上自注意力模型会带来特征冗余,因此使用 3 个维度为  $3 \times 3$  的 T 型卷积模块与零填充进行补丁嵌入,然后通过三个全连接层进行线性变换得到多个自注意力头。

为克服主干网络在提取被遮挡目标特征时的局限性,模型通过交替设置卷积核的步长调整输

出特征图的分辨率并扩大感受野,使用一个重叠核大小为 3、步长为 2 的卷积模块进行下采样,将特征映射的通道划分为  $n$  个,形成  $n$  个自注意力头,并以不同的扩张率进行滑动窗口扩张,计算表达式为

$$h_i = \text{SWDA}(Q_i, K_i, V_i, r_i), 1 \leq i \leq n \quad (2)$$

式中: $Q_i, K_i, V_i$  分别表示查询矩阵、键矩阵和值矩阵; $r_i$  代表膨胀率;SWDA 是指滑动窗口膨胀注意力; $h_i$  表示第  $i$  个特征映射通道的自注意力头。

随后通过一次线性变换得到与原始向量维度匹配的向量  $Y$ ,表达式为

$$Y = \text{Linear}(\text{Concat}[h_1, \dots, h_n]) \quad (3)$$

式中:Linear 表示线性变换;Concat 表示通道数合并; $h_1, \dots, h_n$  分别代表第 1 个到第  $n$  个特征映射通道的自注意力头,这里  $n$  设置为 8,通过计算 8 个维度的权重向量动态调整每个通道的权重,同时进行特征重标定,加强目标特征表示,多个头部并行处理信息,提高计算效率。

构建的 LMGA-CSPDarkNet-53 特征提取网络通过融合轻量化多尺度全局注意力机制,不仅聚焦于目标特征,还细致考虑位置信息,对水下目标进行准确分类和定位,从而解决复杂背景环境中

误判和遗漏问题。

### 1.2.3 跨层特征融合

输入图像通过特征提取网络的下采样和卷积操作得到特征图  $B_2, B_3, B_4, B_5$ 。特征融合部分将上述不同层次的特征图进行有策略的结合,融合过程可以是简单的特征叠加、特征加权融合或者使用更复杂的融合模块,如特征金字塔网络(FPN)、跨层连接网络等。

图 6 为 FPN、PANet、GD 结构简图。由图 6(a)可知,FPN 通过自顶向下的路径融合多尺度特征,易导致一些重要的细节信息丢失。为了克服 FPN 存在单向信息流的局限性,YOLOv8 使用 PANet 结构,如图 6(b)所示,引入自下而上的路径聚合网络,提高了信息利用率,但依旧未能解决多尺度信息融合能力不足问题。

为此,针对水下目标检测的目标尺寸多变问题,本文采用聚合-分发(GD)<sup>[19]</sup>的路径融合策略,如图 6(c)所示,通过建立有效的跨层连接,实现更为全面的多尺度信息整合,增强特征融合能力,使模型在不显著增加延时的情况下,提高信息利用率。GD<sup>[19]</sup>机制结构如图 7 所示。

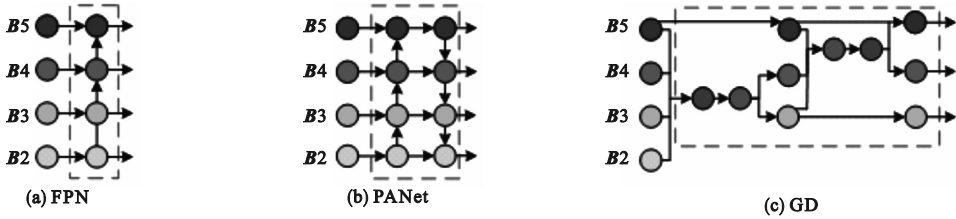


图 6 FPN、PANet、GD 结构简图

Fig. 6 Structure diagrams of FPN, PANet and GD

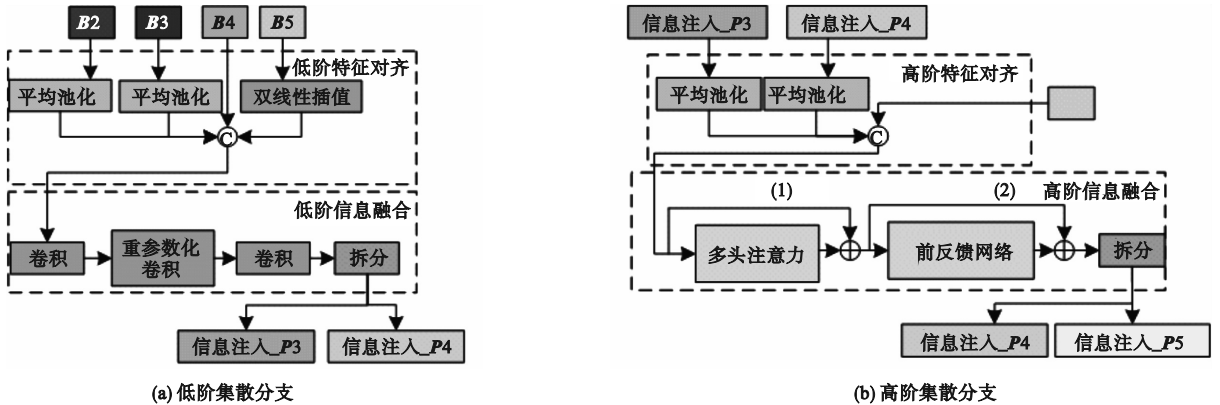


图 7 GD 机制结构图

Fig. 7 GD mechanism structure diagram

在 GD 机制的低阶集散分支中,通过选取主干网络的输出层( $B_2, B_3, B_4, B_5$ ),并利用低阶特

征对齐进行平均池化,将特征图统一为  $40 \times 40$  像素,随后通过低阶信息融合模块的重参数化卷积

块调整特征表达,并通过拆分与重组操作精细融合不同层级的特征,提高信息整合效率。在高阶集散分支中,使用平均池化对输入特征进行降维,有利于高阶信息聚合,其中高阶信息融合模块包含多头注意力、前馈网络和拆分操作,用于处理和融合高阶特征。通过图7中(1)、(2)两个残差连接,提升网络对多尺度信息的融合能力。

为平衡速度与精度,引入轻量级邻层融合

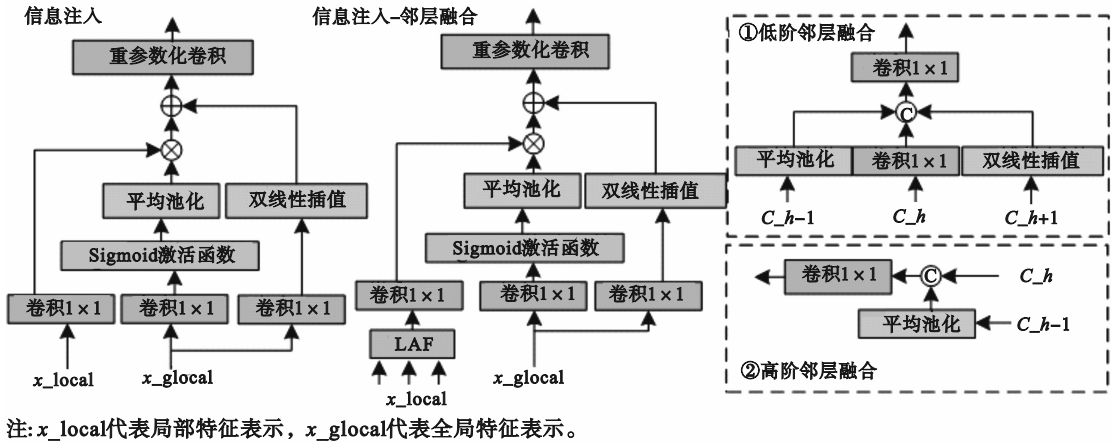


图8 轻量级邻层融合模块结构图

Fig. 8 Lightweight adjacent layer fusion module structure diagrams

## 2 实验环境及评价指标

### 2.1 实验平台与参数设置

本实验硬件平台采用 Intel (R) Core (TM) i5 - 12500H 2.50 GHz CPU, NVIDIA GeForce RTX 3050 Laptop GPU, 采用的软件平台为 Ubuntu20.04 操作系统, 采用 PyTorch 作为深度学习框架, CUDA 和 CUDNN 版本分别为 12.2 和 9.0, python 版本为 3.8。本文实验输入图像尺寸为  $640 \times 640$  像素, 实验批次为 4, 训练 200 个轮次, 初始学习率设置为 0.001, 经过多次实验得出, 学习率设为 0.001 时模型收敛最好。

### 2.2 评价指标

本文实验结果评价指标有精确率、召回率、所有类别的平均精度均值 (mAP)、每张图片检测时间  $T_{All}$ 。 $T_{All}$  用于评估模型的处理速度, 其计算公式为

$$T_{All} = T_{pr} + T_{in} + T_{NMS} \quad (4)$$

式中:  $T_{pr}$  为每张图像预处理时间;  $T_{in}$  为每张图片的推理时间;  $T_{NMS}$  为每张图片的非极大值抑制处

(LAF) 模块, 其结构如图 8 所示。为了确保来自不同级别的特征图与目标大小对齐, 两个邻层融合模块仅使用三个算子:  $C_{h-1}$  通过双线性插值对太小的特征进行上采样,  $C_{h+1}$  通过平均池化对太大的特征进行下采样,  $C_h$  通过  $1 \times 1$  卷积调整与目标通道数不同的特征。邻层融合模块与信息注入模块的结合有效平衡了准确性和速度。

理时间。

## 3 结果与分析

### 3.1 消融实验结果

为验证所提算法的有效性, 本文共设计 3 组消融实验, 实验结果如表 1 所示。实验 1 在 YOLOv8s 网络的基础上加入 Shallow-UWnet 网络, 对原始图像进行预处理, 相比于 YOLOv8s 算法 mAP 值提升 1.9%, 说明图像增强的效果明显, 使网络能够提取更多的特征。实验 2 在实验 1 的基础上在 C2f 模块中融合构建的 LMGGA 模块, 相比于实验 1 的 mAP 值提升了 3.0%, 说明 LMGGA 注意力机制有利于网络引导模型关注重要的特征, 提高了目标检测的准确性。实验 3 在实验 2 的基础上加入 GD 模块, mAP 值进一步提升了 0.2%, 证明采用跨层特征融合方法可以充分利用浅层和深层特征图的通道信息, 增强特征融合能力。此外, 虽然本文方法检测时间仅增加了 0.002 s, 但检测精度大幅度提升。

表 1 消融实验结果  
Table 1 Ablation study results

模型	mAP/%	检测时间/s
YOLOv8	90.0	0.027
YOLOv8 + Shallow-UWnet	91.9	0.027
YOLOv8 + Shallow-UWnet + LMGA_C2f	94.9	0.028
YOLOv8 + Shallow-UWnet + LMGA_C2f + GD	95.1	0.029

### 3.2 不同注意力机制对比实验

为验证不同注意力机制对模型性能的影响,在 YOLOv8 的 C2f 模块融合 LMGA 注意力机制与分别融入 SE<sup>[20]</sup>、CBAM<sup>[21]</sup>、MSDA<sup>[22]</sup> 注意力机制进行比较,共设计 4 组对比实验,结果如表 2 所示。实验 4 与实验 1 相比,检测精度、召回率、

mAP 分别提高了 2.0%、4.8%、5.2%;实验 4 与实验 2 相比,检测精度、召回率、mAP 分别提高了 1.4%、2.6%、4.4%;实验 4 与实验 3 相比,检测精度、召回率、mAP 分别提高了 1.8%、3.8%、4.9%。通过对比可知本文算法准确率最高,并且检测时间最短。

表 2 C2f 融合不同注意力机制性能对比  
Table 2 Performance comparison of C2f fusion with different attention mechanisms

序号	C2f 融合不同注意力模块	精确率/%	召回率/%	mAP@50/%	检测时间/s
1	SE_C2f	86.8	90.3	89.7	0.034
2	CBAM_C2f	87.4	92.5	90.5	0.031
3	MSDA_C2f	87.0	91.3	90.0	0.033
4	LMGA_C2f(本文算法)	88.8	95.1	94.9	0.028

### 3.3 不同图像增强算法结果对比

本文将 Shallow-UWnet 算法与其他图像增强算法进行对比,实验结果如图 9 所示。通过比较可以看出,本文算法更清晰并且真实,HE 算法无

法解决颜色失真问题,MSRCR 算法颜色校正不佳,呈黄色调,UCM 算法呈绿色调,FIVEAPLUS 算法输出的图片偏灰色调,PUIE 算法输出的图片偏绿色调。

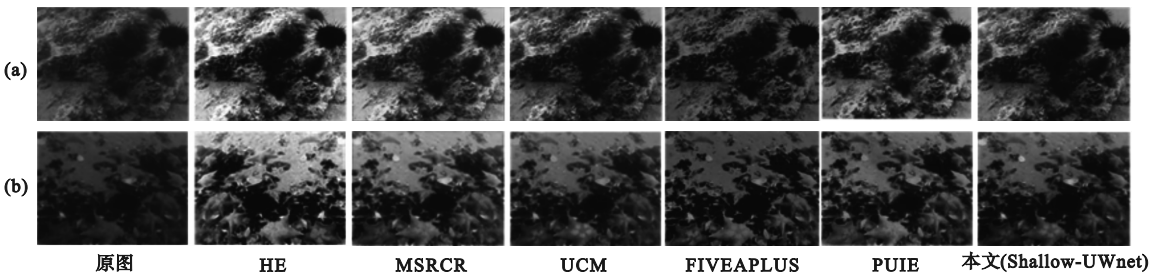


图 9 图像增强结果对比图

Fig. 9 Comparison diagrams of image enhancement result

### 3.4 不同检测方法的性能对比

为检测本文算法与主流算法的性能差异,在经过图像增强后的数据集上开展对比实验,实验结果如表 3 所示。本文算法同 Faster R-CNN<sup>[23]</sup>、Cascade R-CNN、SSD、YOLOv5<sup>[24]</sup>、YOLOv8<sup>[25]</sup> 等

经典水下目标检测算法相比,平均精度均值分别提升了 4.5%、3.6%、11.4%、9.0%、5.1%,精确率和召回率也都有提升,四种生物检测的平均精确率均最高,但检测时间相比 YOLOv8s 算法延迟了 0.002 s。

表3 不同检测算法性能对比

Table 3 Comparison of performance metrics of different detection algorithms

方法	精确率/%	召回率/%	mAP/%	检测时间/s	AP@50/%			
					海参	海胆	海星	扇贝
Faster R-CNN	86.9	90.8	90.6	1.250	90.1	92.5	89.5	90.3
Cascade R-CNN	87.2	92.3	91.5	1.320	91.9	92.4	90.3	91.4
SSD	80.0	87.3	83.7	0.061	82.2	84.7	83.9	84.0
YOLOv5	83.2	88.8	86.1	0.030	86.3	84.4	83.6	90.1
YOLOv8s	84.4	89.9	90.0	0.027	91.3	89.2	88.7	90.6
本文算法	88.9	95.1	95.1	0.029	95.6	94.5	94.3	96.1

为了直观地展示本文算法的检测效果,实验中随机抽取图片进行对比,本文算法与 Faster R-CNN、Cascade R-CNN、SSD、YOLOv5、YOLOv8 等主流算法进行比较,为保证实验的公平性和有效性,所有实验都在同一台设备上,所有算法的输入图像都是经过 Shallow-UWnet 算法图像增强后得到的图片,对比结果如图 10 所示。在图 10 (a) 中 SSD 算法未能检测出扇贝和右下方的海星, YOLOv5 算法优于 SSD 算法但仍未检测出扇贝, Faster R-CNN 和 YOLOv8s 算法对扇贝有漏

检;在图 10 (b) 中 SSD 算法和 Cascade R-CNN 算法将背景误检为海星, YOLOv5 算法对海参有漏检, YOLOv8s 算法将背景误检为海参, Faster R-CNN 算法未能检测出下方海星;在图 10 (c) 中 SSD 算法对海参的检测效果差, YOLOv5、Faster R-CNN 和 YOLOv8s 算法对图像上方海参均有漏检,且 YOLOv8s 算法将背景误检为海参。由此可见,本文算法适应复杂环境的能力较强,对遮挡目标检测的准确率较高。

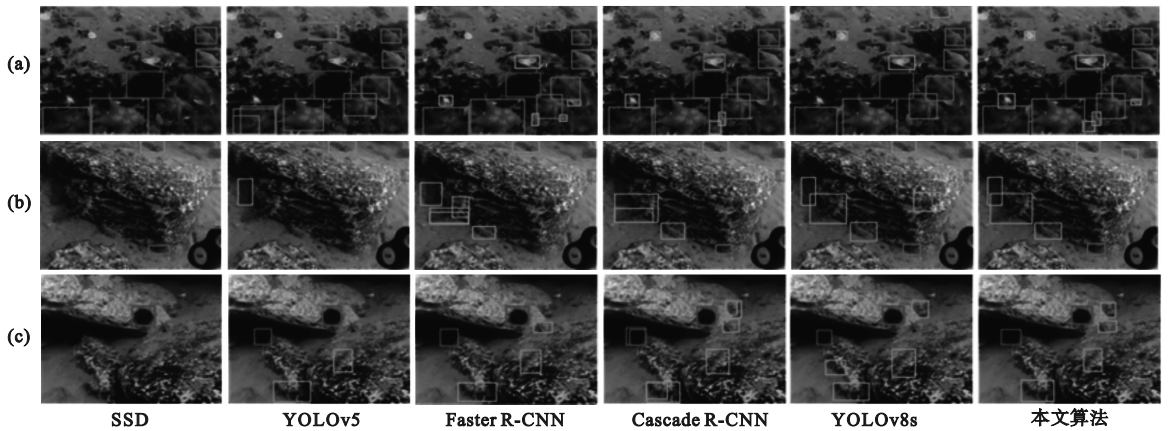


图 10 不同算法检测结果对比图

Fig. 10 Comparison diagrams of different algorithm results

## 4 结论

由于水下环境光照不稳定,导致水下图像存在噪声、对比度低的问题。本文引入 SLG-YOLOv8s 算法解决此问题,主要结论如下:

1) 针对噪声大、对比度低、颜色失真等问题,通过加入 Shallow-UWnet 图像增强算法进行优化,与原图相比,模型在增强后的图像上能够更准确地捕捉目标特征信息;

2) 为充分提取目标特征信息,本文设计了

LMGA 注意力机制,使模型在训练过程中自适应地调整权重,集中于从复杂环境中提取目标信息。与 C2f 未引入 LMGA 相比,将 mAP 再次提升了 3.0%,有效增强了网络特征提取能力;

3) 为解决多尺度特征融合不充分的问题,本文引入了 GD 机制进行特征融合,运用跨层连接,使高层语义信息和低层空间信息充分交换。与 PANet 相比,mAP 再次提升了 0.2%,证明了聚集分发机制的有效性。

本文方法在水下环境中表现出良好的性能,能够在水下复杂场景中实现高质量的目标检测。

同时,平均检测时间为 0.029 s,虽然符合实时检测要求,适用于水下机器人智能捕捞任务,但落后于 YOLOv8s 算法。未来可进一步优化算法,利用参数剪枝、知识蒸馏等方法在提高响应速度的同时,确保结果的准确性。

### 参考文献 (References):

- [1] YUE Z X, YAN B, LIU H Z, et al. An effective method for underwater biological multi-target detection using mask region-based convolutional neural network[J]. *Water*, 2023, 15 (19): 3507.
- [2] HUO J L, JIANG Q. IG-YOLOv5-based underwater biological recognition and detection for marine protection[J]. *Open Geosciences*, 2023, 15 (1): 2391 - 5447.
- [3] YANG Q, MENG H J, GAO Y C, et al. A real-time object detection method for underwater complex environments based on FasterNet-YOLOv7[J]. *Journal of Real-Time Image Processing*, 2023, 21 (1): 8.
- [4] 许裕良, 杜江辉, 雷泽宇, 等. 水下机器人在渔业中的应用现状与关键技术综述[J]. *机器人*, 2023, 45 (1): 110 - 128.  
XU Y L, DU J H, LEI Z Y, et al. Review: application status and key technologies of underwater robots in fishery[J]. *Robot*, 2023, 45 (1): 110 - 128. (in Chinese)
- [5] 王慧敏, 霍冠英, 周亚琴, 等. 基于偏振成像和显著区域自补偿的水下显著目标检测[J]. *计算机应用研究*, 2022, 39 (7): 2210 - 2216.  
WANG H M, HUO G Y, ZHOU Y Q, et al. Underwater salient target detection based on polarization imaging and salient region self-compensation[J]. *Application Research of Computers*, 2022, 39 (7): 2210 - 2216. (in Chinese)
- [6] 要紫丹, 黄小华, 李根, 等. 基于水面红外图像的深海网箱鱼群夜间智能监测方法研究[J]. *南方水产科学*, 2024, 20 (1): 81 - 88.  
YAO Z D, HUANG X H, LI G, et al. Research on nighttime intelligent monitoring method for deep-sea cage fish school based on water surface infrared images[J]. *South China Fisheries Science*, 2024, 20 (1): 81 - 88. (in Chinese)
- [7] 张祥翔, 陈永和, 傅雨田. 基于改进曲波变换的水面弱纹理提取方法[J]. *光学学报*, 2021, 41 (9): 0910001.  
ZHANG X X, CHEN Y H, FU Y T. Extraction method of water surface weak texture based on improved curvelet transformation[J]. *Acta Optica Sinica*, 2021, 41 (9): 0910001. (in Chinese)
- [8] 李辉, 王晓宇, 刘云, 等. 融合多尺度特征和多重阳力的水下目标检测[J]. *农业工程学报*, 2022, 38 (20): 129 - 139.  
LI H, WANG X Y, LIU Y, et al. Detection underwater object using multi-scale features fusion and multiple attention [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38 (20): 129 - 139. (in Chinese)
- [9] CAI Z W, VASCONCELOS N. Cascade R-CNN: high quality object detection and instance segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43 (5): 1483 - 1498.
- [10] 李宝奇, 黄海宁, 刘纪元, 等. 基于改进 SSD 的水下光学图像感兴趣目标检测算法研究[J]. *电子与信息学报*, 2022, 44 (10): 3372 - 3378.
- LI B Q, HUANG H N, LIU J Y, et al. Underwater optical image interested object detection model based on improved SSD [J]. *Journal of Electronics & Information Technology*, 2022, 44 (10): 3372 - 3378. (in Chinese)
- [11] 郝琨, 王阔, 赵璐, 等. 基于图像增强与改进 YOLOv3 的水下生物检测算法[J]. *吉林大学学报(工学版)*, 2022, 52 (5): 1088 - 1097.  
HAO K, WANG K, ZHAO L, et al. Underwater biological detection algorithm based on image enhancement and improved YOLOv3 [J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2022, 52 (5): 1088 - 1097. (in Chinese)
- [12] CHENG C S, HOU X J, WEN X N, et al. Small-sample underwater target detection: a joint approach utilizing diffusion and YOLOv7 model [J]. *Remote Sensing*, 2023, 15 (19): 4772.
- [13] WANG X, HU S L, LI J C. Improved Retinex algorithm for low illumination image enhancement in the chemical plant area[J]. *Scientific Reports*, 2023, 13 (1): 21932.
- [14] 汤忠强, 周波, 戴先中, 等. 基于改进 DCP 算法的水下机器人视觉增强[J]. *机器人*, 2018, 40 (2): 222 - 230.  
TANG Z Q, ZHOU B, DAI X Z, et al. Underwater robot visual enhancements based on the improved DCP algorithm[J]. *Robot*, 2018, 40 (2): 222 - 230. (in Chinese)
- [15] JIANG J X, YE T, BAI J B, et al. Five A + network: you only need 9K parameters for underwater image enhancement[C]//*British Machine Vision Conference*. Aberdeen, UK: Springer, 2023: 149 - 151.
- [16] FU Z Q, WANG W, HUANG Y, et al. Uncertainty inspired underwater image enhancement [C]//*European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022: 465 - 482.
- [17] XING Z Y, CAI M, LI J X. Improved Shallow-U2net for underwater image enhancement [C]//*2022 IEEE International Conference on Unmanned Systems (ICUS)*. Guangzhou, China: IEEE, 2022: 1191 - 1196.
- [18] 高宏伟, 王雨桐. 基于颈颈模块与单解耦头的航拍车辆检测方法[J]. *沈阳理工大学学报*, 2024, 43 (2): 1 - 6.  
GAO H W, WANG Y T. Aerial vehicle detection method based on fast neck module and single decoupling head [J]. *Journal of Shenyang University of Technology*, 2024, 43 (2): 1 - 6. (in Chinese)
- [19] REN H, ZHANG Y J, CHEN J T, et al. A method for detecting surface defects in hot-rolled strip steel based on deep learning[J]. *Metallurgija*, 2024, 63 (3 - 4): 423 - 426.
- [20] 张艳, 李星汕, 孙叶美, 等. 基于通道注意力与特征融合的水下目标检测算法[J]. *西北工业大学学报*, 2022, 40 (2): 433 - 441.  
ZHANG Y, LI X S, SUN Y M, et al. Underwater object detection algorithm based on channel attention and feature fusion [J]. *Journal of Northwestern Polytechnical University*, 2022, 40 (2): 433 - 441. (in Chinese)
- [21] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module [C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. Munich, Germany: Springer, 2018: 3 - 19.
- [22] JIAO J, TANG Y M, LIN K Y, et al. Dilateformer: multi-scale dilated transformer for visual recognition[J]. *IEEE Transactions on Multimedia*, 2023, 25: 8906 - 8919.