

DOI:10.19651/j.cnki.emt.2417621

跨尺度特征融合的自适应水下目标检测算法^{*}

李海龙¹ 黄孙港^{1,2} 饶兴昌¹

(1.南京信息工程大学电子与信息工程学院 南京 210044;2.无锡学院江苏省集成电路可靠性技术及检测系统工程研究中心 无锡 214105)

摘要: 水下目标检测常面临复杂环境干扰、检测系统不稳定以及检测精度低等问题。为此,本文提出了一种基于自适应特征提取与跨尺度特征融合策略的轻量化目标检测算法 WAD-YOLOv8。首先,在主干网络中引入基于上下文信息的残差特征提取模块(CCRF),增强模型对全局和局部信息的综合能力。其次,采用可变大核卷积注意力机制引导的轻量化模块(ADFE),使网络在下采样阶段能够自适应调整采样特征,提高目标特征提取的精准性。最后,重构颈部网络特征融合策略,增加新的跨尺度特征融合连接,增强模型的抗干扰能力和多尺度目标检测性能。试验结果表明,WAD-YOLOv8在模型参数量和计算量均低于基准模型的情况下,检测精度提升了3.0%,平均检测精度 mAP50 提高了2.6%,推理速度达到64 FPS。与经典算法相比,WAD-YOLOv8在复杂水下场景中表现出更优的检测效果和更高的稳定性,为水下移动检测平台提供了一种高效、轻量化的目标检测解决方案。

关键词: 水下目标检测;YOLOv8;可变大核卷积;特征金字塔;自适应

中图分类号: TP391.4;TN914 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Adaptive cross-scale feature fusion for underwater object detection algorithm

Li Hailong¹ Huang Sungang^{1,2} Rao Xingchang¹

(1. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Jiangsu Province Engineering Research Center of Integrated Circuit Reliability Technology and Testing System, Wuxi University, Wuxi 214105, China)

Abstract: Underwater object detection often faces challenges such as complex environmental interference, unstable system performance, and low detection accuracy. To address these issues, this paper proposes WAD-YOLOv8, a lightweight object detection algorithm based on adaptive feature extraction and cross-scale feature fusion strategies. First, a context-aware residual feature extraction module (CCRF) is introduced in the backbone network, enabling the model to effectively integrate global and local information. Second, an adaptive down-sampling module (ADFE) guided by variable large kernel convolution attention mechanisms is employed to adjust sampling features dynamically, enhancing the network's adaptability. Finally, the neck network is restructured by incorporating new cross-scale feature fusion connections, significantly improving the model's robustness against environmental interference. Experimental results demonstrate that, compared to the baseline model, WAD-YOLOv8 achieves a 3.0% improvement in detection accuracy and a 2.6% increase in mAP50, while reducing model parameters and computation by substantial margins. The detection speed reaches 64 FPS, outperforming classical algorithms in both effectiveness and stability. These improvements highlight the model's capability to address the challenges of underwater object detection, offering a highly efficient and reliable solution for complex underwater environments.

Keywords: underwater object detection; YOLOv8; deformable convolution; feature pyramid; adaptive

0 引言

海洋是地球上最大的生态系统,占地表面积的70%以

上,蕴藏着丰富资源,其开发与利用对人类生存和经济发展至关重要。近年来,水下机器人平台凭借其成本低,运行高效的优点被广泛应用于海洋生态监测、水下救援、水产养殖

收稿日期:2024-12-12

^{*} 基金项目:国家青年自然科学基金(62204172)、江苏省高等学校自然科学基金(22KJB140016)资助

等领域。然而,为了在复杂的水下环境中有效执行这些任务,水下目标检测成为水下机器人的关键技术之一,直接影响水下作业的安全性和效率^[1]。

在实际应用场景中,水下目标检测面临诸多挑战。由于水下环境存在光线衰减、散射、浑浊和色彩偏移等现象,检测任务容易出现图像对比度降低、色彩偏移和边缘模糊等问题,尤其在水底沙尘、遮挡物增多的情况下,传统检测算法的精度和鲁棒性都会显著下降。此外,水下目标往往是小型目标,且目标形态复杂多变,检测算法需要具备更强的小目标识别能力和抗干扰能力^[2]。

近年来,深度学习技术的迅猛发展推动了卷积神经网络在水下目标检测领域的广泛应用。现有的目标检测方法通常分为两类:两阶段算法和单阶段算法^[3-4]。两阶段的经典算法如 Fast R-CNN^[5]和 Faster R-CNN^[6]等,通过生成候选区域并对其进行精确分类和定位,具备较高的检测精度,但由于生成候选区域的计算复杂度较高,难以满足水下机器人在实时动态环境中的轻量化部署需求。相比之下,单阶段检测算法如 SSD(single shot multibox detection)^[7]和 YOLO(you only look once)^[8-11]系列直接在输入图像上生成预测框,无需候选区域生成,具有更高的计算效率和实时性,尤其是 YOLO 系列模型因其良好的平衡性能在水下目标检测中广受关注。

尽管单阶段检测算法在一定程度上缓解了实时检测的需求,但现有研究仍然存在局限性。一方面,水下环境中复杂的光线变化和遮挡现象仍导致小目标检测效果不佳。例如,陶洋等^[12]提出的 LKCA-YOLOv5 算法通过空间融合模块提升了遮挡目标的检测能力,但忽略了对小目标的高效检测。周新等^[13]针对特征丢失问题提出暗通道-对比限制-光衰减算法用于丰富图像特征信息。另一方面,现有算法在增强特征表达能力的同时,往往伴随着模型参数数量和计算复杂度的显著增加,这不利于水下机器人平台的实际部署。例如,Bao 等^[14]基于 YOLOv5 提出了并行高分辨率水下目标检测网络,但因参数量增加降低了检测时效性。Yu 等^[15]基于 YOLOv7 设计了一种新的水下生物探测模型 U-YOLOv7 通过创建一个结合 CrossConv 和高效挤压激励模块的网络,增强了网络的特征融合,提高了模型在水下识别中的抗干扰能力。此外,针对水下图像的色彩偏移和噪声问题,Liu 等^[16]设计了图像增强模块,虽然有效改善了图像质量,但额外的预处理步骤增加了检测流程的复杂度。

基于上述分析,目前水下目标检测仍然面临以下挑战:

1) 光线变化和噪声导致的特征丢失问题,尤其在光线不足和浑水水体条件下,现有模型对图像的对比度变化和细节提取能力不足^[17]。

2) 小目标检测效果不佳,在海底生物和海洋垃圾等小目标场景中,现有算法无法有效提取小目标的细节特征。

3) 模型参数量过大,难以部署于轻量化平台,大多数现

有算法无法同时兼顾高精度检测和低计算复杂度的需求。

针对上述问题,本文提出了一种基于 YOLOv8 改进的水下目标检测模型 WAD-YOLO,旨在提升水下目标检测的精度、鲁棒性和轻量化性能。本文主要创新点包括:

1) 针对小目标检测中的特征提取困难问题,本文引入上下文残差连接思想,并设计了跨阶段上下文残差融合模块(cross-stage contextual residual fusion, CCRF)替代主干网络中的 C2f 特征提取模块,从而有效减少了模型参数量并提升了小目标检测精度。

2) 针对水下复杂场景中的不规则目标遮挡问题,本文引入了可变大核卷积注意力机制(deformable large kernel attention, DLKA),并设计自适应下采样特征提取模块(adaptive downsampling feature extraction, ADFE)替代颈部的下采样模块,以增强模型在复杂场景下的抗干扰能力和特征提取效果,同时降低了计算量。

3) 基于双向特征金字塔网络(bidirectional feature pyramid network, BiFPN),设计了新的跨尺度特征融合路径,改造网络的颈部结构,旨在增强模型的跨尺度特征融合能力,解决逐层特征融合时细节丢失的问题。

1 YOLOv8 结构

YOLOv8 主要分为主干网络(Backbone)、颈部网络(Neck)和检测头(Haed) 3 个部分,如图 1 所示。图像输入尺寸为 640×640 ,经过图像采样和特征融合后由 3 个检测头进行损失计算和目标识别。主干网络中采用 C2f、CBS 模块对图像进行降维处理,利用快速金字塔池化 SPPF 提取图像特征。颈部特征融合网络采用 PANet 结构,对不同维度的特征图像进行处理,一定程度上提高了模型的检测精度。YOLOv8 采用主流的解耦头作用,将分类和检测头分离,相比单一结构的检测头,提升了模型检测的准确性和效率。本文选取参数量最小的 YOLOv8n 作为基准模型。

2 WAD-YOLOv8 算法

2.1 跨阶段上下文残差融合模块

由于浑水水体的影响,水下成像条件相比空中更加复杂。光线在传播过程中不可避免地发生散射和水体吸收,导致图像细节大量丢失与遮盖。小目标聚集区域尤为明显,边界往往相互融合,难以有效区分。为了改善这一问题,提升模糊小目标的检测精度,本文提出了跨阶段上下文残差融合模块(CCRF)结构如图 2 所示,其设计灵感来源于人类视觉系统。与传统模型单一的特征提取方式相比,CCRF 通过融合目标周围的图像特征信息,从全局视角增强了网络对小目标特征的提取能力。

CCRF 模块设计结构轻量化,包括两部分:主干信息融合模块与上下文信息提取器(context guided block, CGB),其结构如图 3 所示。主干信息融合模块通过残差连接将目标的局部特征与提取到的上下文特征进行融合,而 CGB 专

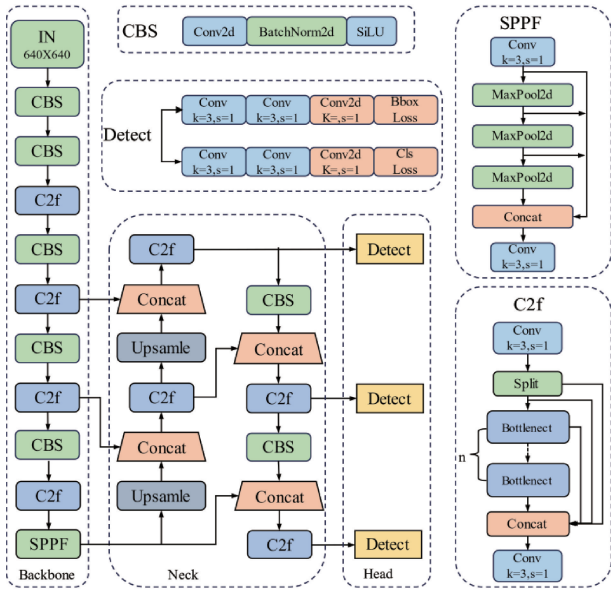


图1 YOLOv8 模型结构
Fig. 1 The structure of YOLOv8

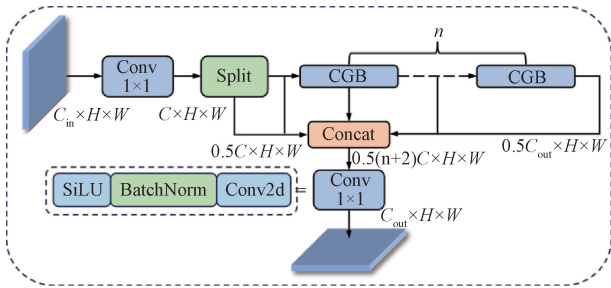


图2 CCRF 模块结构
Fig. 2 The structure of CCRF

门负责提取图像的上下文信息。CGB的核心设计思想是将图像的局部特征与其周围的上下文特征相结合,从而实现更丰富的特征表示。

CGB通过 1×1 卷积扩展输入图像的通道数并分为两部分,然后分别采用 3×3 标准卷积和 3×3 扩张卷积提取局部与上下文特征。随后,通过连接操作融合两部分特征,形成更完整的特征表示。融合后的特征经过批量归一化(batch normalization, BN)和参数化线性单元(PReLU)的处理后完成数据增强。最后,通过全局平均池化(global average pooling, GAP)和多层感知机(multilayer perceptron, MLP)整合全局信息,进一步剔除无关噪声,从而优化上下文特征。最终,经过CGB处理的上下文特征信息被送入主干信息融合模块与原始特征信息残差连接完成信息提取。CCRF模块的引入有效减少网络参数量,同时显著提升水下模糊小目标的特征提取能力。该模块为后续网络分级特征融合(Neck部分)提供了更加丰富的多尺度信息,进一步提升了检测精度。模块计算定义式如(1)~(5)所示。

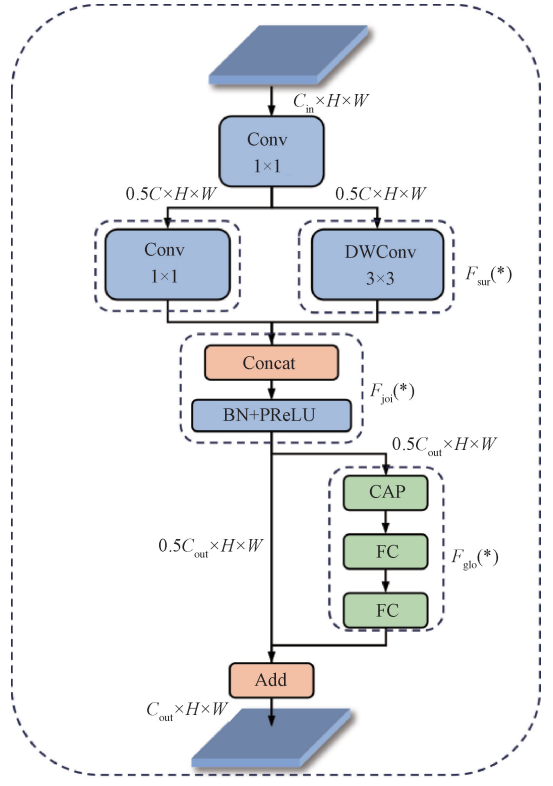


图3 CGB 模块结构
Fig. 3 The structure of CGB

$$f_{loc} (*) = Conv_{1 \times 1}(Conv_{1 \times 1}(*)) \quad (1)$$

$$f_{sur} (*) = DConv_{3 \times 3}(Conv_{1 \times 1}(*)) \quad (2)$$

$$f_{joi} (*) = PReLU(BN(f_{loc} (*) + f_{sur} (*))) \quad (3)$$

$$f_{glo} (*) = MLP(GAP(f_{joi} (*))) \quad (4)$$

$$f_{out} (*) = f_{glo} (*) \odot f_{joi} (*) \quad (5)$$

其中, $f_{loc} (*)$ 为局部特征, $f_{sur} (*)$ 为上下文特征, \odot 代表元素逐个相乘, $DConv$ 为扩张卷积。

2.2 自适应下采样特征提取模块

YOLOv8在特征融合网络中分别使用两次普通卷积完成图像下采样,但对于水下大量复杂干扰信息和光线明暗变化,在下采样的过程中会造成网络保留过多冗余信息,在增加计算开支的情况下还会导致特征提取不够精确容易丢失目标的细节信息。为了减小下采样模块的计算复杂度,提高网络对细节特征的提取能力,本文设计了一个融合分组卷积模块(Adown)^[18]和可变大核注意力机制(DLKA)^[19]的自适应下采样特征提取模块(ADFE)。如图4所示,输入图像经过分组卷积融合降低图像维度,再由可变大核注意力机制对其进行特征提取,最终完成图像下采样过程。

图像首先通过平均池化操作减小特征图尺寸,并提取全局特征。然后,图像在通过分组卷积模块时被分成多个组,每个组通过不同的卷积核进行卷积操作,从而减少计算量并有效保留细节特征。在两个支路上分别进行最大

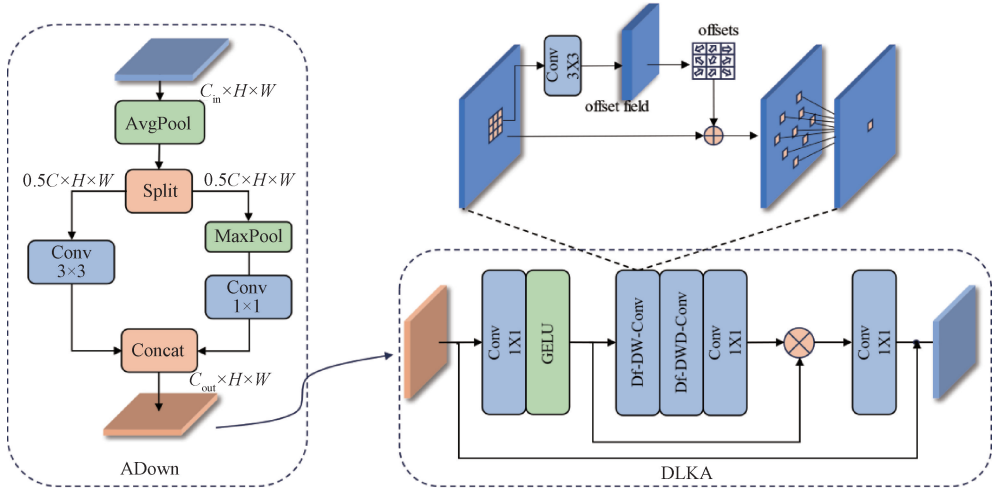


图 4 ADFE 模块结构

Fig. 4 The structure of ADFE

池化和卷积操作,最后将输出特征进行拼接,送入后续的注意力机制模块。这种采用混合池化策略的下采样方式可以有效捕获图像中的不同尺度的特征,从而在捕捉背景信息的同时更好的保留细节特征。

在特征提取阶段,DLKA 模块结合了可变形深度可分离卷积(Deform-DW-Conv)和可变形深度可分离空洞卷积(Deform-DWD-Conv)。其中,可变形卷积通过在特征图上学习偏移量来动态调整卷积核的采样位置,从而生成自适应的卷积核,这些卷积核能够根据输入特征自适应地变化,精确提取目标的变形和边界特征。相比传统卷积操作,可变形卷积显著减少了计算量和参数量,同时通过自适应调整,增强了网络的感受野并更好地融合了全局和局部特征信息。结合空洞卷积后,DLKA 模块不仅扩大了感受野,还能自适应地调整采样点,从而更精准地捕捉细节信息,尤其适用于处理水下复杂光照和背景条件下的目标检测任务。相较于普通卷积,可变形卷积不仅可以显著减少计算量和网络参数量还可以借助自适应的优势来有效融合全局和局部特征信息,增强网络感受野。式(6)和(7)分别是参数量和计算量的比值:

$$\frac{Deform - DW - Conv}{Conv} = \frac{MD_k^2 + MN}{MND_k^2} = \frac{1}{N} + \frac{1}{D_k^2} \quad (6)$$

$$\frac{Deform - DW - Conv}{Conv} = \frac{MD_k^2 D_F^2 + MND_F^2}{MND_k^2 D_F^2} = \frac{1}{N} + \frac{1}{D_k^2} \quad (7)$$

其中, D_k^2 为深度卷积核大小, D_F^2 为逐点卷积核大小, M, N 为卷积核个数。以 3×3 大小的卷积核为例,卷积核个数 N 较大时, $D_k^2 = 1/9$ 可变形深度可分离卷的参数量和计算量将下降到普通卷积的 $1/9$ 。

ADFE 模块进行特征提取的计算过程如式(8)~(10)所示。

$$F' = GELU(Conv(F)) \quad (8)$$

$$Attention = Conv_{1 \times 1}(DDWDConv(DDWConv(F'))) \quad (9)$$

$$Output = Conv_{1 \times 1}(Attention \otimes F') + F \quad (10)$$

其中, F 为经过降维处理的特征图像, \otimes 表示逐元素的乘积运算,图像首先经过普通卷积改变通道数,由 $GELU$ 激活函数非线性处理成 F' 送入注意力机制,最后与原始特征图像 F 相加完成加权特征的输出。值得注意的是,相比于其他自注意力机制,DLKA 并未使用额外的归一化函数,如 Sigmoid 或 Softmax,这些归一化函数往往会忽略高频信息,从而降低自注意力机制的性能。

本文将 YOLOv8 颈部(Neck)的两个 CBS 下采样模块全部替换为 ADFE 下采样模块,轻量化设计的同时可以使模型捕获更重要的特征信息,增强模型的抗干扰能力,进一步提高水下复杂光照背景中的目标检测精度。

2.3 基于改进双向金字塔的多尺度特征融合网络

YOLOv8 的颈部采用的是双路径聚合网络(PAN)(如图 5(b)所示),与传统融合网络(FPN)(如图 5(a)所示)不同的是,FPN 采用的是自顶向下的路径来融合不同层的多尺度特征,图中为第 2~5 层(P2~P5),但是单向信息流的路径融合方式并不充分,相距较远的高层和低层信息在融合时会丢失细节特征。为了改善这类问题,YOLOv8 引入 PANet,在原有基础上增加自底向上的额外路径,对自顶向下路径进行扩充解决了单向信息流的细节丢失问题。

随着网络层数增加,特征图像维度变化较大时,这种逐层直接拼接的特征融合方式会增加网络的参数量和计算量,不利于水下实时检测部署要求,且会出现特征信息冗余的问题。为了提高特征融合效率和网络性能,本文根据双向特征金字塔网络(BiFPN)^[20](如图 5(c)所示)改进颈部网络,BiFPN 允许特征信息同时在自顶向下和自底向上两个路径上流动和融合,在无需增加额外计算量的情况下使网络更高效。同时,对处于同一层级的节点添加额外

的残差连接,优化跨尺度特征的融合效果。BiFPN还引入了加权融合机制,为不同尺度的特征添加权重,目的使网络可以根据贡献程度高效完成特征层的融合,防止信息冗余和细节特征丢失。

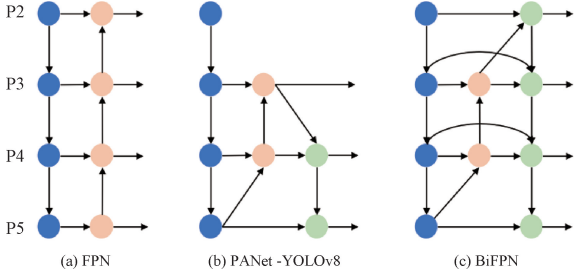


图5 不同特征融合网络

Fig. 5 The structure of different feature fusion network

以 P4 为例,加权特征融合过程如下:

$$\omega'_i = \text{ReLU}(\omega_i) \quad (11)$$

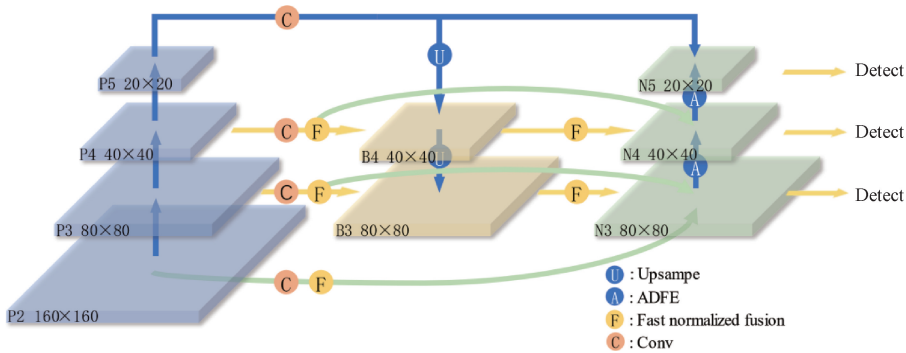


图6 改进 F-BiFPN 结构

Fig. 6 The structure of improved F-BiFPN

2.4 改进的网络结构 WAD-YOLOv8

改进后的模型结构如图7所示,在YOLOv8的基础上综合了本文所提出的多个创新点,为水下复杂场景的目标检测提供了一种高效且轻量化的解决方案。首先使用CCRF模块替换主干网络中的C2f模块,接着将颈部网络中的下采样模块替换为ADFE模块,减少下采样过程中的计算开支并融入自适应机制。最后基于BiFPN对网络融合策略改进,提升颈部网络整体的特征提取效率和准确性。

3 实验设计与结果分析

3.1 实验环境与训练参数配置

本文实验的硬件配置使用NVIDIA GeForce RTX 4090显卡,核心处理器为Intel(R) Xeon(R) Gold 6430。软件环境为ubuntu20.04操作系统,深度学习加速框架为PyTorch 1.10.0,Cuda 11.3,编程开发环境为Python 3.8。训练时,输入图像尺寸统一为 640×640 ,Epoch设置为

$$P_4^{id} = \text{Conv} \left(\frac{\omega_1 P_4^{in} + \omega_2 \text{Resize}(P_5^{in})}{\omega_1 + \omega_2 + \epsilon} \right) \quad (12)$$

$$P_4^{out} = \text{Conv} \left(\frac{\omega'_1 P_4^{in} + \omega'_2 P_4^{id} + \omega'_3 \text{Resize}(P_3^{in})}{\omega'_1 + \omega'_2 + \omega'_3 + \epsilon} \right) \quad (13)$$

其中, i 表示层数, ω_i 为可学习的权重, P_i^{in} 表示第*i*层输入的特征, P_4^{id} 表示自上而下路径第*i*层的中间特征, P_4^{out} 表示第*i*层输出特征,Resize表示采样过程。

改进的F-BiFPN结构如图6所示,在原网络的基础上分别在B3和N3,B4和N4层级添加了新的跳跃连接,用于优化跨尺度融合效果。为了弥补低层在采样过程中信息的丢失,创新性的在输出端新增N3层直接与骨干网络中P2层级连接,深度采样后的特征图与浅层特征图融合可以有效增强网络的互补能力,在水下目标检测中,避免因水下图像模糊导致的深度采样后的细节特征丢失。

300,Batchsize设置为32,初始学习率为0.01,使用SGD作为优化器。

3.2 数据集

为了综合评估本文模型在复杂水下场景对小目标的检测能力,选取DUO数据集^[21]进行消融和对比实验,样本如图8所示。DUO数据集是对历年水下机器人大赛所使用的URPC数据集^[22]图片的筛选和整理,相比于单一年份URPC数据集,图像数据覆盖面更广,场景更加多样化。DUO数据集共有7782张图片,一共有4类目标:海参(holothurian)、海胆(echinus)、扇贝(scallop)和海星(starfish)。

选取TrashCan数据集^[23]进行模型泛化性验证,该数据集为水下机器人采集的真实数据,包含多种水下垃圾并对生物做了更细致的分类,更加贴近真实的水下状态,可以更好的用来评估模型的鲁棒性。

3.3 水下复杂场景模拟

DUO数据集是通过有主动光照的水下机器人拍摄获

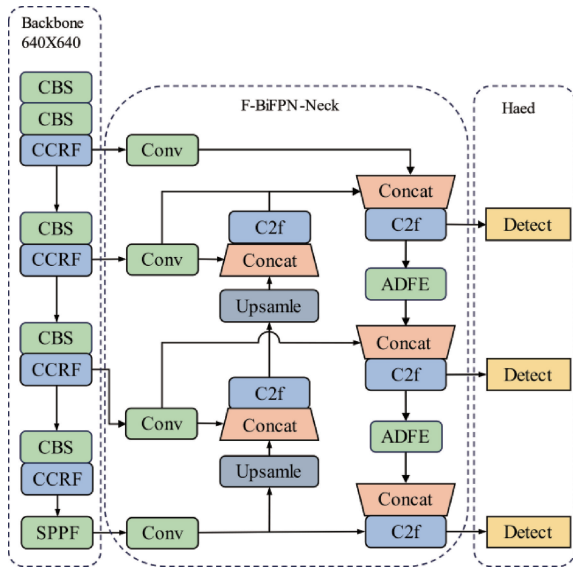


图 7 WAD-YOLOv8 模型结构

Fig. 7 The structure of WAD-YOLOv8

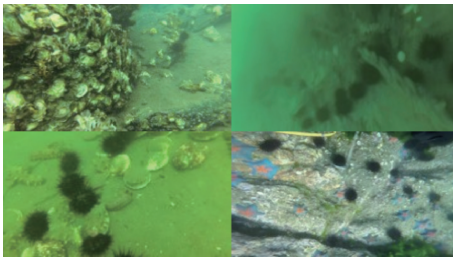


图 8 数据集样本

Fig. 8 Dataset sample

得,水下光照条件和水体透光率较为稳定,为了使模型覆盖到水下不同的图像采集场景,本文选择对 DUO 数据集进行图像处理,模拟不同的光照条件和水体透光率,生成一个综合数据集 Mix-DUO。

模拟算法参考图像加雾原理^[24],使用指数衰减函数来模拟水体透光率和亮度的变化,如式(14)~(16)所示。

$$t(d) = e^{-\beta d}, l(d) = A e^{-\alpha d} \quad (14)$$

$$I' = I l(d) \quad (15)$$

$$I_f = I' t(d) + (1 - t(d)) \quad (16)$$

其中, I 表示原始图像像素值, A 为全局亮度参数,数值为 1 时图像保持原始状态, d 表示当前像素点与雾化中心的距离, α, β 为控制因子控制亮度和浓度的变化,数值为 0 时图像保持原始状态, I_f 表示处理过后的图像像素值。

将原始数据集中的 7 782 张图像按照低光照图像、低透光率图像和原始图像类别,以 2:3:5 的比例随机分组,生成新的数据集 Mix-DUO,如表 1 所示。

在新数据集中,各类别图像分别占总数据集的 20%、30%和 50%。随后,将 Mix-DUO 数据集按 8:2 的比例划分为训练集和测试集,用于模型训练和性能验证,处理实

列如图 9 所示。

表 1 Mix-DUO 数据集组成

Table 1 Composition of Mix-DUO dataset

| 数据集 | 类别 | 图像数量 | 比例/% |
|---------|------|-------|------|
| Mix-DUO | 低光照 | 1 556 | 20 |
| | 低透光率 | 2 334 | 30 |
| | 原始图像 | 3 892 | 50 |
| 数据集划分 | 训练集 | 6 225 | 80 |
| | 测试集 | 1 557 | 20 |

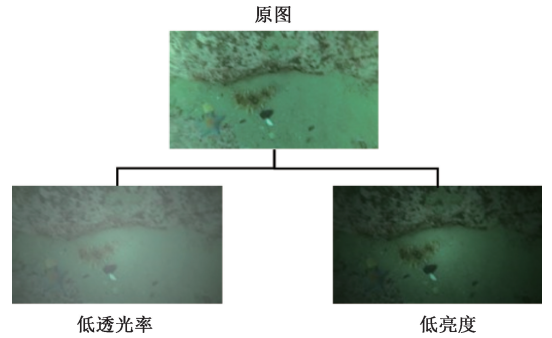


图 9 图像增强

Fig. 9 Image enhancement

3.4 评价指标

为了准确评估模型性能,采用精度(precision, P),平均精度(average precision, AP),平均精度均值(mean average precision, mAP),召回率(recall, R)作为模型改进效果的指标。本文中 mAP 选择两个阈值,分别是 mAP@ (0.5 : 0.95)、mAP@0.5。使用参数量(parameters, params),和浮点运算量(GFLOPs)作为模型轻量化的衡量标准。

3.5 消融实验

为了验证模型改进的有效性,本文在基准模型 YOLOv8n 的基础上,分别对不同改进点进行分组对比,量化分析其对模型性能的具体影响,实验结果如表 2 所示。

实验 1 为原始的 YOLOv8n 基准模型,未进行任何改进。实验 2 和 3 主要用于验证提出模块的轻量化效果。结果表明,主干网络中引入的上下文残差特征提取模块(CCRF)和改进的双向特征金字塔(F-BiFPN)颈部网络,均能显著降低模型的参数量和运算量,同时使 mAP50 分别提升了 0.5%和 0.8%。实验 4 结合了 CCRF 模块和 F-BiFPN 的改进策略。结果显示,相较于基准模型,mAP50 提高了 0.9%,参数量和计算量分别降低了 25.8%和 17.8%。这表明 CCRF 模块能够有效缓解特征提取阶段的细节信息丢失,而 F-BiFPN 模块则提升了颈部网络的融合效率,并增强了多尺度特征的捕捉能力。实验 5 将颈部网络中的下采样模块替换为融合可变核注意力机制的 ADFE 模块。结果表明,尽管模型的参数量和计算量有所增加,但网络对关键信息的提取能力和复杂环境的抗干扰

能力显著增强,mAP50提升了1.8%。实验6采用改进后的WAD-YOLOv8模型,综合了多种有效改进策略。实验结果显示,与基准模型相比,mAP50提升了2.6%,mAP50:95提升了3.0%,参数量和运算量分别减少了13.0%和6.1%。此外,模型在推理过程中保持了64 fps的高检测速度,展现出稳定可靠的水下检测能力,满足水下检测平台的部署与应用要求。

3.6 对比实验

1) 不同模型对比

为了进一步验证本文所提模型的有效性,表3展示了其与其他6种现有经典模型的检测对比结果,实验基于Mix-DUO数据集,并在统一的实验平台上进行。纵向对比表明,单阶段算法整体优于两阶段算法。YOLO系列模型在参数量和运算量上显著小于Faster-RCNN,同时在检测精度(P)和召回率(R)方面也表现出显著优势,更适合水下移动检测平台的部署需求。横向对比显示,YOLOv8s的平均精度值mAP50达到了86.3%,比YOLOv8n提升了2.1%,召回率也优于YOLOv5系列。然而,YOLOv8s主要通过增加网络深度和宽度来提升检测精度,这种方法

未优化网络结构,同时显著增加了参数量和运算量,难以满足轻量化部署的实际需求。本文提出的模型基于YOLOv8n进行了深度优化,通过提升主干网络对全局信息的敏感度,显著降低了漏检和误检率。同时,颈部网络引入注意力机制,增强了模型对水下复杂环境干扰的抵抗能力,使召回率(P)提高了4.2%,进一步减少了漏检和误检问题。得益于轻量化的改进策略,本文模型在保持检测精度的同时,参数量和运算量仅为YOLOv8s的24.1%和27.7%,且检测速度保持在较高水平。相比文献[12]提出的模型,本文模型的参数量显著减少,同时mAP50提升了1.1%。此外,本文新增了与Transformer系列先进目标检测模型RT-DETR-R18^[25]的对比。RT-DETR-R18的精度达到了89.6%,略高于本文模型。然而,其作为RT-DETR系列中规模最小的模型,参数量为20.1 M,远大于本文模型,FLOPs为60.2 G,约为本文模型8倍。尽管精度有所提升,但其计算资源需求高于轻量化部署的要求。相比之下,本文模型在保证检测精度的同时,具备显著的轻量化优势,更适合复杂水下环境中的实际部署需求。

表2 消融实验指标比较

Table 2 Comparison of ablation experiment metrics

| Model | YOLOv8n | CCRF | ADFE | F-BiFPN | mAP50/% | mAP50:95/% | Params/M | FLOPs/G | FPS |
|-------|---------|------|------|---------|-------------|-------------|------------|------------|-----------|
| 1 | ✓ | | | | 84.2 | 64.5 | 3.1 | 8.4 | 62 |
| 2 | ✓ | | | ✓ | 85.0 | 64.8 | 2.8 | 8.2 | 66 |
| 3 | ✓ | ✓ | | | 84.7 | 64.4 | 2.5 | 6.8 | 71 |
| 4 | ✓ | ✓ | | ✓ | 85.1 | 65.7 | 2.3 | 6.9 | 73 |
| 5 | ✓ | | ✓ | | 86.0 | 66.5 | 4.0 | 10.1 | 58 |
| 6 | ✓ | ✓ | ✓ | ✓ | 86.8 | 67.5 | 2.7 | 7.9 | 64 |

表3 不同模型对比实验

Table 3 Comparison experiment of different models

| Model | P/% | R/% | mAP50/ % | Params/ M | FLOPs/ G | FPS |
|--------------|-------------|-------------|-------------|--------------|-------------|-----------|
| Faster-R-CNN | 64.2 | 58.2 | 67.1 | 42.1 | 66.6 | 22 |
| SSD | 69.2 | 54.3 | 62.6 | 24.4 | 42.8 | 58 |
| YOLOv10n | 84.6 | 77.1 | 85.2 | 2.5 | 7.2 | 71 |
| YOLOv8n | 83.1 | 75.0 | 84.2 | 3.1 | 8.4 | 61 |
| YOLOv8s | 85.7 | 78.5 | 86.3 | 11.2 | 28.5 | 56 |
| YOLOv5n | 76.8 | 68.5 | 77.1 | 2.6 | 7.1 | 73 |
| RT-DETR | 89.6 | 82.4 | 90.3 | 20.1 | 60.2 | 45 |
| 文献[12] | / | / | 85.7 | 12.4 | / | 55 |
| Ours | 86.1 | 79.2 | 86.8 | 2.7 | 7.9 | 64 |

图10所示分别为基准模型和本文所提改进模型检测过程中的注意力分布热力图,其中颜色由浅至深表明网络对目标关注程度由低到高的变化。由图像可知,改进后的

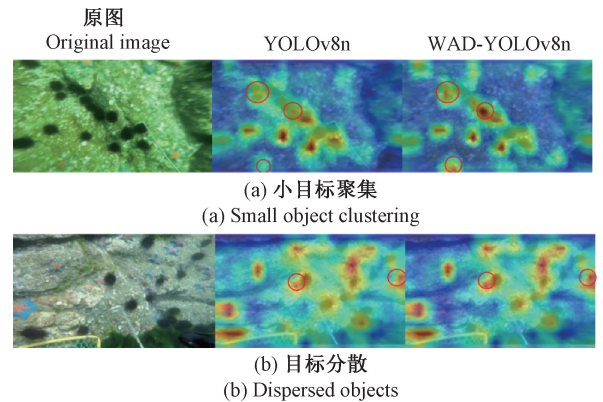


图10 热力图可视化

Fig. 10 Heatmap visualization

模型有效抑制了目标周围岩石信息的干扰,将注意力收缩到检测目标上,表现出了良好的鲁棒性。在面对多个海胆目标聚集时,能够将目标有效分开并检测出来,降低了重叠目标的漏检率。

2) 模型复杂度对比

为了进一步评估本文模型在水下目标检测任务中的部署可行性,本文从时间复杂度和空间复杂度两个方面对改进模型进行了分析,并与其他主流目标检测算法进行了对比。时间复杂度直接影响模型的推理速度,通常由网络深度、特征图分辨率和计算量(FLOPs)决定。从表 3 的数据可见,本文模型的 FLOPs 为 7.9 G,低于 RT-DETR 的 60.2 G,同时推理速度达到 64 fps,优于 Faster R-CNN 的 22 FPS 和 RT-DETR 的 45 fps。

空间复杂度主要由模型的参数量和存储需求决定。本文模型的参数量仅为 2.7 M,远低于 RT-DETR 的 20.1 M 和 YOLOv8s 的 11.2 M,体现了其轻量化设计优势。通过优化特征提取和跨尺度融合模块,本文模型进一步减少了内存占用,为水下机器人平台提供了更好的部署条件。

综上所述,本文模型在时间复杂度和空间复杂度的综合性能上具有显著优势,既能满足水下目标检测的实时性需求,又能适应水下设备对轻量化和低功耗的要求。与 RT-DETR 等复杂模型相比,本文模型以较低的计算量和存储需求实现了接近的检测精度,验证了其在实际应用中的可行性与可靠性。

3) 不同下采样模块对比

为验证本文提出的下采样模块 ADFE 的检测效果,在改进后的模型基础上,通过调整颈部网络的不同下采样策略进行对比实验。实验中,将 ADFE 模块中的注意力机制分别更换为多尺度扩张注意力机制(MSDA)^[26]和高效多尺度注意力机制(EMA)^[27],其中 MSDA 与 DLKA 属于空间注意力机制,EMA 属于通道注意力机制。

对比结果如表 4 所示,空间注意力机制相比通道注意力机制在水下目标检测场景中表现更优,检测精度显著提升。对于受水体浑浊和弱光照影响的小目标聚集区域,其边界通常较为模糊,而空间注意力能够帮助网络更精准地捕捉关键区域的边缘信息。相比之下,通道注意力主要对全局特征通道进行加权处理,忽视了局部特征的细节达。

表 4 不同下采样模块对比

Table 4 Comparison of different down-sampling modules

| 模块 | mAP50/ mAP50:95/ | | Params/ FLOPs/ | |
|-------------------------------|------------------|------|----------------|-----|
| | % | % | M | G |
| Conv | 85.1 | 65.7 | 2.3 | 6.9 |
| ADFE(A _{down} +DLKA) | 86.8 | 67.5 | 2.7 | 7.9 |
| A _{down} +MSDA | 86.5 | 67.3 | 3.1 | 8.3 |
| A _{down} +EMA | 85.6 | 66.1 | 2.4 | 7.1 |

3.7 可视化分析

为了直观展示 WAD-YOLOv8 模型在复杂水下场景的检测效果,从测试集中取了具有典型特征的图像进行推理测试,包括小目标聚集重叠,光照环境不足,水体浑浊和图像偏色 4 种,检测结果如图 11 所示。图 11(a)和(b)展

示了小目标聚集和遮挡场景。在图 11(a)图像中,由于多个小目标海胆聚集在海参周围产生干扰,YOLOv8n 模型未能正确识别出海参,发生了漏检现象。

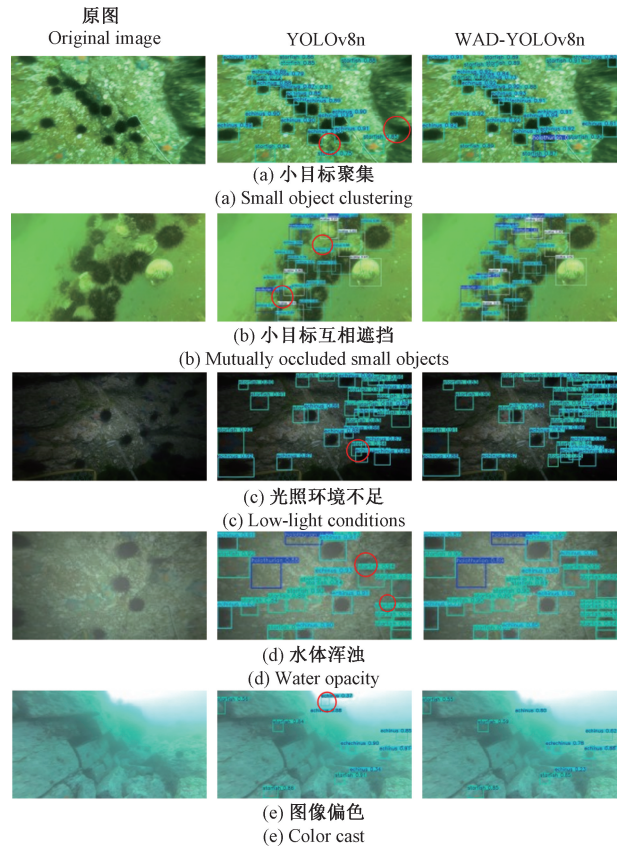


图 11 检测结果对比

Fig. 11 Comparison of detection results

相比之下,WAD-YOLOv8 模型通过引入可变大核卷积注意力机制,能够灵活区分目标和背景信息,避免了聚集目标的干扰,在复杂场景中成功检测出了海参。对于图 11(b)组图像,由于海胆之间的相互重叠,目标边界模糊,YOLOv8n 未能有效区分两个重叠目标,而改进后的 WAD-YOLOv8 模型通过更高效的目标感知能力成功识别出了两个目标,展现出其在目标遮挡条件下的优越性能。图 11(c)和(d)展示了光照不足和水体浑浊影响下的场景,两个模型均出现一定精度下降。但由于缺乏多尺度特征融合能力和敏感的目标感知,YOLOv8n 模型在这些场景中发生漏检。相比之下,WAD-YOLOv8 模型通过引入跨连接双向金字塔融合策略,增强了低维和高维特征的加权处理能力,降低了信息冗余,实现了对模糊目标的精准检测,表现出较强的鲁棒性。图 11(e)展示了光照变化引起图像偏色的场景。YOLOv8n 模型将礁石误检为海胆,而 WAD-YOLOv8 模型通过改进的注意力机制有效避免了此类问题。总体而言,WAD-YOLOv8 模型在处理漏检和误检问题上表现出色,并增强了对小目标的识别能力,特别是在水下复杂环境中,展现了较好的稳定性和检测准确性。

3.8 泛化性验证

为验证 WAD-YOLOv8 模型在不同数据集上的泛化能力,选用 TrashCan 数据集进行实验。泛化实验结果如表 5 所示,相较基准模型,改进模型的精确率(P)、召回率(R)和平均检测精度(mAP50)分别提高了 0.7%、2.2%和 1.6%。同时,其 FPS 达到 76,满足实时检测需求。

图 12(a)和(b)分别展示了较差水质和较清水质下的检测结果。在较低水质条件下,图像较为模糊,但 WAD-YOLOv8 模型依然能够稳定地检测到水底的塑料垃圾,并且给出了较高的置信度,反映出其在不同水质条件下的良好鲁棒性。这一结果表明,WAD-YOLOv8 模型不仅在标准数据集上展现了卓越性能,也能够具有挑战性的水下环境中表现出较好的迁移能力和泛化性能。

表 5 基于 TrashCan 数据集的泛化实验

Table 5 Generalization experiments based on the TrashCan dataset

| 模块 | P/% | R/% | mAP50/% | FPS |
|-------------|-------------|-------------|-------------|-----------|
| YOLOv5n | 87.2 | 71.4 | 87.9 | 82 |
| YOLOv8n | 90.6 | 77.1 | 91.1 | 68 |
| Ours | 91.3 | 79.3 | 92.7 | 76 |

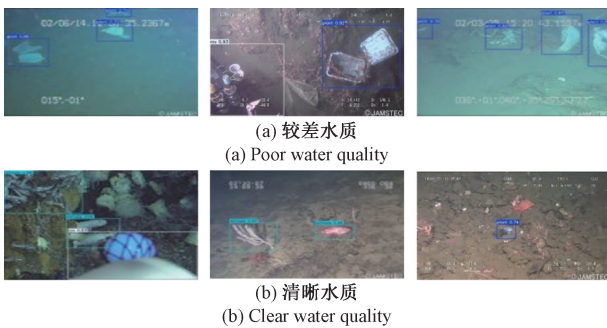


图 12 TrashCan 数据集检测结果对比

Fig. 12 Comparison of detection results of TrashCan datasets

4 结 论

本文针对水下复杂场景提出了一个高效且轻量化的目标检测模型 WAD-YOLOv8,旨在解决因水下成像条件差、小目标聚集以及噪声干扰等因素导致的检测精度低和检测不稳定问题。首先,将主干中的 C2f 特征提取模块替换为可以结合上下文特征信息进行引导的轻量化模块 CCRF,提高模型对小目标的检测能力;其次,重构颈部特征提取网络,引入双向特征金字塔融合策略,并增加新的跨尺度融合层,优化模型在逐层特征融合时的信息丢失问题,显著提高模型检测效率;最后,采用全新下采样模块 ADFE,结合轻量化卷积和可变大核卷积注意力机制,扩大模型的感受视野,进一步减少漏检和误检现象。实验结果表明,在 Mix-DUO 数据集上 WAD-YOLOv8 模型的

mAP50 可以达到 86.8%,本文所提模型与基准模型相比,参数量和计算量分别降低了 13.1%和 6.1%。在不同类型的水下复杂场景中,WAD-YOLOv8 均能实现对目标的精准检测,表现出显著优于基准模型的稳定性和检测性能,为水下移动检测平台提供了一种高效、轻量化的解决方案。未来的研究将进一步探索模型的轻量化优化,并尝试将模型应用于更广泛的海洋场景和其他实际需求,以拓展其潜在应用范围,助力水下目标检测领域的发展。

参考文献

- [1] 方俊,邱春荣. 水下图像目标检测技术研究[J]. 科技资讯, 2022, 20(12): 26-29.
FANG J, QIU CH R. Research on underwater image target detection technology[J]. Science & Technology Information, 2022, 20(12): 26-29.
- [2] 郭银辉,张春堂,樊春玲. 基于生成对抗网络的深海图像增强算法[J]. 电子测量技术, 2024, 47(12): 173-181.
GUO Y H, ZHANG CH T, FAN CH L. Deep-sea image enhancement algorithm based on generative adversarial network [J]. Electronic Measurement Technology, 2024, 47(12): 173-181.
- [3] 罗飞,王润峰. 基于 YOLOv5 水下目标检测算法研究与改进[J]. 通信与信息技术, 2024(1): 34-40.
LUO F, WANG R F. Research and improvement of underwater target detection algorithm based on YOLOv5 [J]. Communication and Information Technology, 2024(1): 34-40.
- [4] XU S, ZHANG M, SONG W, et al. A systematic review and analysis of deep learning-based underwater object detection [J]. Neurocomputing, 2023, 527: 204-232.
- [5] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015:1440-1448.
- [6] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards realtime object detection with region proposal networks [C] 28th International Conference on Neural Information Processing Systems, 2015: 91-99.
- [7] LYU Z, JIN H, ZHEN T, et al. Small object recognition algorithm of grain pests based on ssd feature fusion[J]. IEEE Access, 2021, 9: 43202-43213.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [9] LI C Y, LI L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. ArXiv preprint arXiv:

- 2209.02976, 2022.
- [10] 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]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023; 7464-7475.
- [11] WANG C Y, YE H I H, LIAO H Y M. Yolov9: Learning what you want to learn using programmable gradient information [J]. ArXiv preprint arXiv: 2402.13616, 2024.
- [12] 陶洋, 朱腾, 钟邦乾. RepViTS-YOLOX: 水下模糊及遮挡目标检测方法[J]. 计算机工程与应用, 2024, 60(13): 200-208.
- TAO Y, ZHU T, ZHONG B Q. RepViTS-YOLOX: A method for underwater blurred and occluded object detection [J]. Computer Engineering and Applications, 2024, 60(13): 200-208.
- [13] 周新, 张春堂, 樊春玲. 基于 YOLOv5_PGS 的轻量级水下生物识别目标检测[J]. 电子测量技术, 2023, 46(21): 168-175.
- ZHOU X, ZHANG CH T, FAN CH L. Lightweight YOLOv5-PGS based objective detection for underwater biological identification [J]. Electronic Measurement Technology, 2023, 46(21): 168-175.
- [14] BAO Z, GUO Y, WANG J, et al. Underwater target detection based on parallel high-resolution networks[J]. Sensors, 2023, 23(17): 7337.
- [15] YU G, CAI R, SU J, et al. U-YOLOv7: A network for underwater organism detection [J]. Ecological Informatics, 2023, 75: 102108.
- [16] LIU Z, WANG B, LI Y, et al. UnitModule: A light weight joint image enhancement module for underwater object detection[J]. Pattern Recognition, 2024, 151: 110435.
- [17] 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.
- [18] WANG C Y, YE H I H, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]. European Conference on Computer Vision. Cham: Springer, 2025: 1-21.
- [19] AZAD R, NIGGEMEIER L, HTEMANN M, et al. Beyond self-attention: Deformable large kernel attention for medical image segmentation[C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 1287-1297.
- [20] REDDY M V, REDDY K A, GOUND M S S, et al. EfficientDet: Scalable and efficient object detection[J]. Neuro Quantology, 2022, 20(19): 5559.
- [21] 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.
- [22] CHEN L, LIU Z, TONG L, et al. Underwater object detection using invert multi-class adaboost with deep learning[C]. 2020 International Joint Conference on Neural Networks (IJCNN), 2020: 1-8.
- [23] HONG J, FULTON M, SATTAR J. TrashCan: A semantically segmented dataset towards visual detection of marine debris [J]. ArXiv preprint arXiv: 2007.08097, 2020.
- [24] 尹胜楠, 崔学荣, 李娟, 等. 基于快速 ACE 算法的视觉里程计图像增强方法[J]. 电子测量与仪器学报, 2021, 35(6): 27-33.
- YIN SH N, CUI X R, LI J, et al. Visual odometry image enhancement method based on fast ACE algorithm[J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(6): 27-33.
- [25] YU C, SHIN Y. MCG-RTDETR: Multi-convolution and context-guided network with cascaded group attention for object detection in unmanned aerial vehicle imagery[J]. Remote Sensing, 2024, 16(17): 3169.
- [26] JIAO J, TANG Y M, LIN K Y, et al. Dilateformer: Multiscale dilated transformer for visual recognition[J]. IEEE Transactions on Multimedia, 2023, 25: 8906-8919.
- [27] OUYANG D, HE S, ZHANG G, et al. Efficient multiscale attention module with cross-spatial learning[C]. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023: 1-5.

作者简介

李海龙, 硕士研究生, 主要研究方向为深度学习、目标检测。

E-mail: 202312490398@nuist.edu.cn

黄孙港(通信作者), 副教授, 主要研究方向为计算机视觉、光电检测仪器与系统。

E-mail: hsg@cwku.edu.cn

饶兴昌, 硕士研究生, 主要研究方向为深度学习、目标检测。

E-mail: raoxingchang2001@163.com