

混合注意力与动态特征引导网络的水下检测^{*}裴晓芳^{1,2} 杨继海² 周进² 徐永恒²

(1.无锡学院电子信息工程学院 无锡 214105; 2.南京信息工程大学电子与信息工程学院 南京 210044)

摘要: 针对背景干扰导致水下目标关键特征提取不充分,关注区域无法遍及整个目标的问题,提出了一种混合注意力与动态特征引导网络的水下目标检测算法。所提方法首先设计一种全维度动态提取模块(FDEM),充分理解目标整体性特征。其次,设计加权特征拼接模块(WFCM),以保留从浅层到深层的关键特征。将CARAFE算子用于特征上采样操作,从而使局部区域的关键特征得到更多关注。最后,构建混合注意力机制(HAM),有效结合目标的通道信息和像素点信息,保留关键特征。经实验验证,该方法能够充分提取关键特征,更有效地关注目标的整体性特征,从而减少水下背景干扰,进一步提升水下目标检测性能。

关键词: 特征提取;水下目标检测;注意力机制;YOLOv8

中图分类号: TP391.4;TN20 **文献标识码:** A **国家标准学科分类代码:** 520.20

Underwater detection with hybrid attention mechanism and
dynamic feature guidance networkPei Xiaofang^{1,2} Yang Jihai² Zhou Jin² Xu Yongheng²

(1. School of Electronic Information Engineering, Wuxi University, Wuxi 214105, China; 2. School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: In response to the problem that background interference leads to insufficient extraction of key features of underwater targets and the attention area cannot cover the entire target, an underwater target detection algorithm with mixed attention and dynamic feature guidance network was proposed. This method first designs a full-dimension dynamic extraction module (FDEM) to fully understand the overall features of the target. Secondly, a weighted feature concatenation module (WFCM) is designed to retain key features from the shallow to the deep layers. The CARAFE operator is used for the feature up-sampling operation, so that the key features of local areas can receive more attention. Finally, a hybrid attention mechanism (HAM) is constructed to effectively combine the channel information and pixel information of the target and retain the key features. Experimental verification shows that this method can fully extract key features, pay more attention to the overall features of the target, thereby reducing underwater background interference and further enhancing the underwater target detection performance.

Keywords: feature extraction; underwater target detection; attention mechanism; YOLOv8

0 引言

海洋约占地球总面积的71%,拥有丰富的生物以及矿产资源,能够为人类发展提供强大的资源支撑。随着陆地上的资源不断被消耗殆尽,探索海洋资源变得尤为重要。我国地处太平洋西岸,拥有1.8万公里的漫长海岸线,海洋资源丰富,发展空间广阔^[1]。据统计,我国2023年水产品总产量约为7116万吨,已连续35年位居世界第一,成为名副其实的“蓝色粮仓”。党的二十大报告中明确指出要大力

发展海洋经济,协同保护海洋生态环境。此外,十四五规划也明确强调要加快开发海洋牧场,建设绿色养殖和可持续性海洋渔业,深度参与全球海洋治理,着力推进海洋强国的建设,为进一步发展海洋、利用海洋、保护海洋建立系统性工程。

水下目标检测是海洋领域研究中的关键部分,在资源勘探、水产养殖、生态保护等方面发挥关键作用。然而,水下目标尺寸较小,常存在重叠遮挡、尺寸差异显著等问题^[2],导致现有检测算法误检率较高。研究学者也在不断

针对不同问题改进算法,以提升模型的检测性能。

针对目标尺寸小、尺度不一的问题,Yi等^[3]提出了一种基于改进YOLOv7的水下小目标检测算法,该算法以YOLOv7为基础网络,融入SENet注意力机制聚焦小目标关键特征,改进FPN网络拓扑并引入EIou损失函数,有效提升检测精度,降低漏检和误检率,在多种数据集上验证了算法的有效性和泛化能力。Chen等^[4]提出了基于改进YOLOv4神经网络的水下目标识别方法,该方法将YOLOv4的上采样模块替换为反卷积模块,并融入深度可分离卷积,同时改进Mosaic数据增强方法,有效提高了水下目标识别的精度和速度,降低了硬件性能需求。Liu等^[5]设计一种用于水下目标识别的网络,该网络由两个相同的一维卷积神经网络构成,用模拟数据训练,能有效识别水下目标辐射噪声的DEMON谱,实验显示其对不同多普勒频移、信噪比的小样本识别精度较高,具有良好的泛化能力。陈亮等^[6]基于SSD算法提出一种级联式融合结构,将高层级特征图经过上采样操作后得到的特征向量与低层级特征向量进行维度拼接操作,为低层特征图融入更多上下文信息,从而提升小目标的检测能力。李培坤等^[7]提出一种WCA-YOLOv8目标检测算法,该算法设计FFM模块提升对模糊和颜色偏移小目标的识别能力,加入FCA模块增强对小目标的特征提取能力,采用WIoUv3损失函数提高小目标检测精度,还设计DEM模块完善上下文信息,在多个数据集上检测效果良好。王书朋等^[8]提出一种ShapeLoss损失函数,充分结合Shape-IoU和Shape-NWD两类损失函数的优势,一方面减弱边界框自身形状和尺度对于回归的影响,另一方面提升小尺度目标边界框回归时的精准度。崔颖等^[9]设计一种深度特征金字塔模块,由空间注意力模块以及标准Transformer块共同构建而成,并将其嵌入到Deformable-DETR模型中,用以提高模型深层纹理信息的提取能力。刘侦龙等^[10]提出YOLOv10-MECAS模型,通过设计一种中值增强的通道和空间注意力模块MECAS降低图像噪声,提高海参图像特征的捕捉能力,同时引入了可切换空洞卷积SACConv模块,扩大了目标特征的感受野,有效提升复杂水下环境中小目标海参检测精度。

针对目标遮挡问题,罗偲等^[11]基于Faster R-CNN提出一种对抗样本生成网络,通过在特征空间上对前景目标的深层次特征创建遮挡,旨在获得足够遮挡样本的训练数据集,经实验分析,与传统检测算法相比,该改进算法具有更低的遮挡样本漏检率。谭鹤群等^[12]选择ResNeXt101网络替换Faster R-CNN原有的VGG16网络,采用横向堆叠的Bottleneck结构进行搭建,利用Shortcut Connection结构增加信息传递,增强图像中细节特征和深层次信息的感知能力。周昕等^[13]将可变形卷积加入主干网络,利用卷积核自适应形变的能力灵活调整其形状和采样位置,摆脱以往利用固定尺寸和形状的卷积核提取目标特征的局限性,提高模型在目标堆叠的水下环境的检测能力。陈亮等^[14]

在YOLOv4算法的基础上,联合真值排斥因子RepGt和预测排斥因子RepBox共同优化损失函数,以防止全部预测框陷入置信得分最高区域,同时扩大已匹配的目标框与预测目标框之间的界限,使预测框更加独立,从而提升遮挡目标的自适应检测能力。陶洋等^[15]在YOLOv5算法基础上提出大核卷积注意力特征提取模块,通过引入不同尺寸卷积核的卷积,获取到不同层次的语义信息,在不同感受野下提取遮挡物之间的差异性特征。Li等^[16]提出一种水下目标检测网络Underwater-YOLO,该网络引入双分支遮挡处理注意力机制,利用全局分支捕捉整体信息,而局部分支利用卷积增强对遮挡目标特征的提取,有效提升了对密集、遮挡目标的检测能力,在复杂水下场景检测实验中表现优异。周道先等^[17]构建一种水下目标检测网络MDLA-DCN,引入可形变卷积来取代传统的卷积,增强网络对不规则目标的提取能力,而多跳跃连接则聚合了不同分辨率的特征,增强了水下目标的空间信息。史朋飞等^[18]提出一种应用于YOLOv4的数据增强算法,通过线性加权的方式模拟目标的遮挡、重叠的水下环境,增强网络适应复杂环境的能力,有效增强模型的鲁棒性。

然而,上述研究仅针对某类特殊水下场景,而水下环境复杂多变,同时水下目标受到背景干扰极其严重,容易使得目标关键特征提取不充分,模型关注区域无法全面覆盖目标,对于模型决策产生干扰,从而进一步影响检测性能。具体而言,水下环境中存在大量的悬浮物,如泥沙、藻类等,这些悬浮物会在图像中形成噪声,干扰对目标的识别。同时,复杂的海底地形以及各种自然或人工的背景物体,如礁石、沉船残骸、泥沙等,与目标物体的特征相互混淆。此外,不同深度的水体对光线的折射和散射程度不同,导致目标在不同深度下呈现出的特征差异较大,增加了目标与背景分离的难度,尤其在一些浑浊的水域,水下能见度极低,几乎无法通过视觉特征来区分目标和背景^[19]。这无疑对检测算法的性能提出了极高要求,给水下目标的精准检测带来一定的挑战。

为此,本文基于深度学习相关技术探寻可行的解决策略,提出一种混合注意力与动态特征引导网络的水下目标检测算法,旨在减少背景信息的干扰,进一步提升水下目标检测的精度。

1)设计全维度动态提取模块(full dimensional dynamic extraction module, FDEM),利用多分支瓶颈结构增加梯度信息的传递,借助全维度动态卷积(omni-dimensional dynamic convolution, ODCConv)^[20]动态调整卷积核权重,从多个维度提取重要特征,有助于捕获目标的整体性特征。

2)提出加权特征拼接模块(weighted feature concatenation module, WFCM),先对不同层级特征加权处理,以保留各层级的关键特征,再进行不同层级间的特征拼接操作,加强浅层级信息与深层级信息的融合,增强特征的表达能力。

3)采用 CARAFE 算子^[21]进行特征上采样,利用核预测模块生成上采样卷积核,借助内容感知重组模块加强局部信息的关注程度,重点把握关键特征的提取,减少背景信息的干扰。

4)提出混合注意力机制(hybrid attention mechanism, HAM),充分融合通道注意力和像素注意力机制的优势,有效保留关键的通道信息和像素级信息,进一步筛选重要特征,抑制冗余信息。

1 混合注意力与动态特征引导网络

本文以 YOLOv8s^[22] 目标检测模型为基线算法,提出一种混合注意力与动态特征引导网络的水下目标检测算法。所谓动态特征引导,是指在特征提取与融合过程中,基于输入图像的内容自适应调整特征提取的维度、权重分配及融合策略,通过特征交互与反馈机制,主动引导模型聚焦目标整体性特征,而非被动过滤冗余信息。与现有动态卷

积 DyConv 仅聚焦卷积核参数的单维度动态调整、传统注意力机制 CBAM 仅侧重通道或空间的单维度权重强化不同,该机制更有助于增强模型对图像关键信息的捕获能力。具体架构如图 1 所示,给定一张含有水下目标的大小为 640×640 的图像。首先,在主干网络利用 FDEM 对输入图像特征提取,利用动态瓶颈块增加特征传递,提取丰富的梯度信息,通过 ODCnv 生成多维度卷积核权重,捕获整体性特征。在颈部,利用 WFCM,将不同层级特征图加权后特征拼接,通过信息的双向流动与交互融合,使特征信息在不同层级间形成优势互补,从而丰富水下目标的特征多样性和捕获关键信息。接着,采用 CARAFE 作为上采样算子,利用核预测模块生成自适应上采样核,联合内容感知重组模块增强局部区域的关注程度,从而保留目标的关键特征。最后,构建混合注意力模块,充分融合关键通道特征和像素级区域特征,同时滤除干扰信息,增强模型对水下目标关键特征的提取。

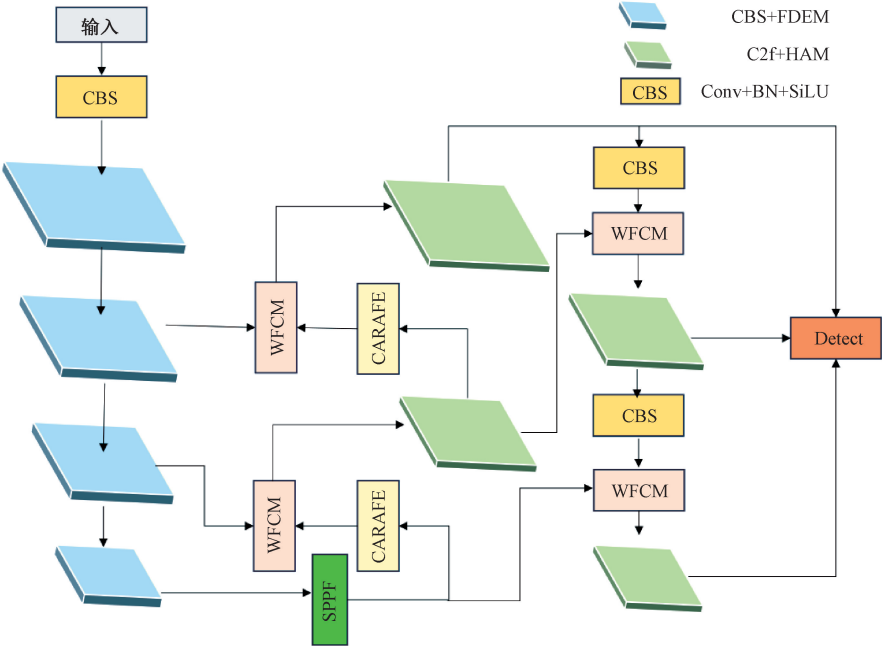


图 1 混合注意力与动态特征引导网络的整体框架

Fig. 1 The overall framework of the hybrid attention and dynamic feature guidance network

1.1 FDEM

特征提取部分往往通过加入更多分支结构,丰富梯度回传的支路,从而为模型带来更多梯度信息,有利于特征的提取。然而这种方式经过多个分支中的卷积操作,引入额外的噪声信息,势必会导致细粒度信息的丢失。此外,传统卷积的卷积核权重是固定的,在特征提取过程中缺乏对不同尺度目标的自适应能力,一般来说,对于大尺度目标,需要较大的卷积核来捕捉其整体特征,但对于小尺度目标,大卷积核可能会引入过多的无关信息,导致目标特征被淹没;而对于小目标有效的小卷积核,在处理大目标时又可能无法获取足够的上下文信息。而

ODConv 的特点在于,通过动态生成多维度卷积核权重,从而有效结合不同尺度特征,防止噪声信息的干扰,实现对输入特征的灵活自适应匹配。受此启发,本文构建一种 FDEM,如图 2 所示,设计动态瓶颈结构,自适应学习卷积内核参数特性,动态生成多维度卷积核权重,从而提取出各种不同类别图像的关键特征,提升模型的代表能力。

输入特征图首先进入一个 1×1 卷积层,用于调整输入通道数,提升模型的特征表达能力,接着进行特征分割,利用 Split 方法,将特征图在通道维度上平均分割成两个张量,一个保留原有信息用于后续特征拼接,另一个进入

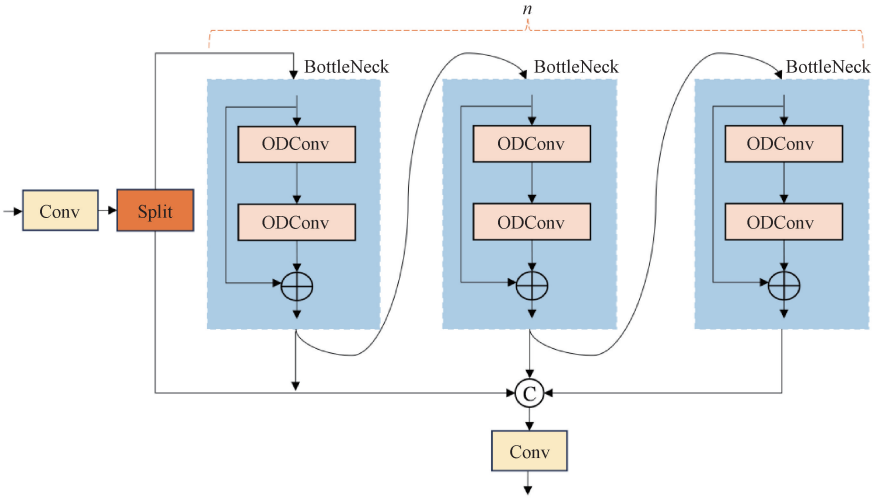


图2 全维度动态提取模块

Fig. 2 Full dimensional dynamic extraction module

BottleNeck 模块进行深层次特征提取。其中, BottleNeck 模块中利用 ODConv 卷积构建动态瓶颈块结构, 多层次动态提取特征信息, 保留输入图像的关键信息。为引导网络更加关注水下小尺度目标细节特征信息, 同时也能获取大尺度目标的其整体性特征, 增加动态瓶颈块的数量, 并在动态瓶颈块之间增加残差级连接, 丰富梯度流信息。接着, 将各个瓶颈块输出的动态特征信息与前面的输出进行特征拼接操作, 将该模块所有分支的输出特征进行融合, 丰富特征的表达能力, 使模型能够获取更全面的信息。最后, 拼接后的特征图进入第2个 1×1 卷积层特征图转换, 对特征进行整合和优化。

ODConv 工作原理如图3所示, 全局平均池化对输入进行压缩, 利用全连接层将特征向量映射至低维空间, 利用 ReLU 函数去除负值, 随后经过4个并行分支, 利用全连接层和非线性激活函数, 分别生成空间注意力权重 α_{si} 、输出通道权重 α_{fi} 、输入通道权重 α_{ci} 以及卷积核权重 α_{wi} , 用于对卷积核进行多维度调整, 使得卷积核能够根据输入特征图动态调整其参数特性。最后将4个分支的注意力权重与卷积核 W_i 执行乘法运算, 并与输入 x 执行卷积运算, 输出为:

$$y = (\alpha_{w1} \times \alpha_{f1} \times \alpha_{c1} \times \alpha_{s1} \times W_1 + \dots + \alpha_{wn} \times \alpha_{fn} \times \alpha_{cn} \times \alpha_{sn} \times W_n) * x \quad (1)$$

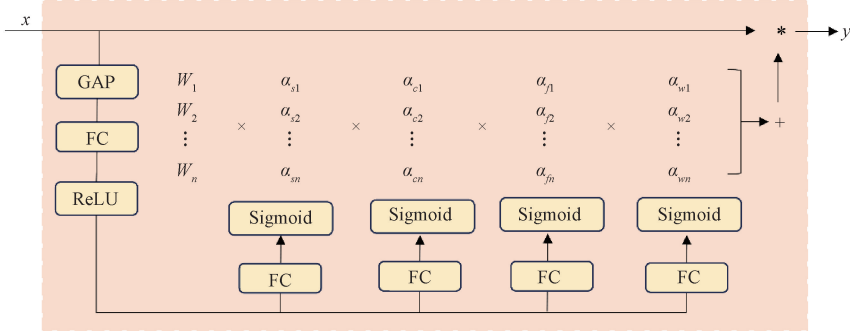


图3 ODConv 结构

Fig. 3 The structure of ODConv

1.2 WFCM

在卷积神经网络中, 为融合不同层级的特征信息, 往往通过简单的特征拼接操作, 以提高模型多尺度目标的检测能力。然而水下目标受到严重的背景干扰, 容易引入过多冗余信息, 从而在特征拼接操作过程中丢失不同层级的关键特征, 进一步影响模型的分析处理能力, 降低系统的检测性能。考虑到以上问题, 借鉴注意力机制的加权思想, 本文提出一种 WFCM, 依靠通道注意力生成的权重值,

与不同层级特征图进行加权融合, 保留从浅层到深层的重要特征, 以增强特征的表达能力和模型的性能, 如图4所示。

首先将来自不同层级的特征输入 F_L 和 F_H , 在通道维度上特征拼接, 得到:

$$F_1 = \text{Concat}(F_L, F_H) \quad (2)$$

式中: $\text{Concat}(\cdot)$ 表示特征拼接操作。对拼接后的特征分别进行最大池化和平均池化操作, 生成二维特征映射, 将两

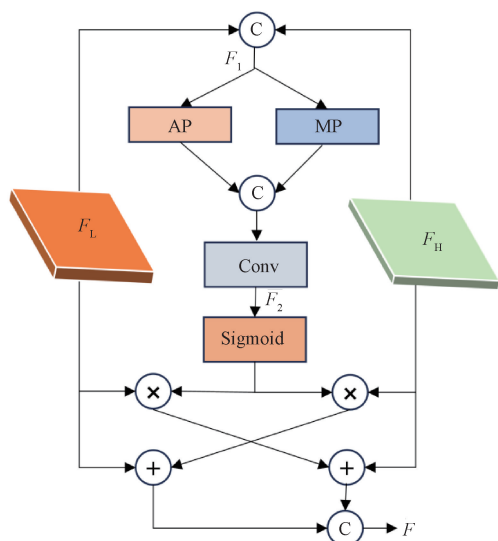


图 4 加权特征拼接模块

Fig. 4 Weighted feature concatenation module

个池化层分支输出在通道维度上特征拼接,通过卷积操作用于特征提取以及通道间信息的融合,得到特征图 F_2 :

$$F_2 = \text{Conv}(\text{Concat}(\text{AP}(F_1), \text{MP}(F_1))) \quad (3)$$

式中: $\text{Conv}(\cdot)$ 表示对特征进行卷积操作, $\text{AP}(\cdot)$ 表示平均池化, $\text{MP}(\cdot)$ 表示最大池化。

随后,经 Sigmoid 激活函数得到两个空间权重矩阵 W_L 和 W_H ,将权重与原始输入端不同层级的特征图相乘,能够灵活调整不同层级特征在融合过程中的重要程度,有效降低冗余信息的干扰,最大程度保留各层级的主要特征,得到加权后的特征图 F'_L 和 F'_H 。

$$W_L, W_H = \text{Split}(\text{Sigmoid}(F_2), [C_L, C_H]) \quad (4)$$

$$F'_L = F_L \times W_L \quad (5)$$

$$F'_H = F_H \times W_H \quad (6)$$

式中: $\text{Split}(\cdot)$ 表示对通道进行分割, $\text{Sigmoid}(\cdot)$ 表示 Sigmoid 激活函数,而 C_L 和 C_H 分别表示通道分割系数。

接着,将加权后的低层级特征图与高层级特征图逐元素相加输出为 F_{HL} ,使得保留细节信息的低层次特征有效整合高层次强语义信息,弥补低维特征信息中所缺失的整体性特征;同理,将加权后的高层级特征图与低层级特征图逐元素相加得到 F_{LH} ,使得保留强语义信息的高层次特征有效整合低层次细节信息,弥补高维特征信息中所缺失的纹理、边缘等特征。通过信息的双向流动与交互融合,使特征信息在不同层级间形成优势互补,从而丰富水下目标的特征多样性和捕获关键信息;利用特征加权操作通过生成通道注意力权重,对目标特征赋予重要性考量,保留关键特征,滤除噪声信息的干扰,同时避免因采样模块导致部分信息缺失。最后,将 F_{HL} 和 F_{LH} 的输出值在通道维度上进行特征拼接操作,得到 WFCM 的最终输出为 F 。

$$F_{HL} = F_H + F'_L \quad (7)$$

$$F_{LH} = F_L + F'_H \quad (8)$$

$$F = \text{Concat}(F_{HL} + F_{LH}) \quad (9)$$

1.3 CARAFE 上采样算子

在水下目标检测领域,研究者常通过上采样对图像或者特征图进行扩充放大,生成指定尺寸的特征图用于特征融合。例如最邻近元上采样,利用像素之间的空间距离指导下采样的过程,由于只考虑亚像素邻域,在进行密集检测任务时难以获取到丰富的语义信息,同时易产生锯齿状边缘;反卷积^[23]作为一种自适应的上采样方法,利用卷积的逆算子操作可以学习到更多的细节信息,但由于参数较多导致计算负载加重,容易产生过拟合现象。而 CARAFE 算子舍弃了常见的基于像素之间的位置距离,通过构建重组内核,增强了对于像素级区域内容的考量,而不会引入额外参数。基于此,本文采用 CARAFE 算子用于上采样操作,利用内容感知的动态核预测与局部特征重组实现高精度特征增强,有利于保留水下目标的细节信息,其基本结构如图 5 所示。

具体而言, CARAFE 算子主要由核预测模块和内容感知重组模块构成。核预测模块主要包含 3 个子模块:通道压缩器、内容编码器以及核归一化器。首先,对于输入大小为 $H \times W \times C$ 的特征图,由通道压缩器将输入特征图的通道数压缩为 C_m ,减少计算复杂度。其次,利用内容编码器生成不同位置上的上采样重组核 $\sigma H \times \sigma W \times K_{up} \times K_{up}$,其中 σ 为上采样倍率。为了获得更大的感受野,使用一个 $K_{encoder} \times K_{encoder}$ 卷积层,从而获得更大的上采样核。最后利用核归一化器,每个重组内核在空间上通过使用 Softmax 激活函数使卷积核权重和为 1,确保特征均值不变且避免缩放失真。

内容感知重组模块将输出特征图中的每个位置映射回输入特征图后,为了使局部区域中的某些特征得到更多关注,取出每个位置以 $K_{up} \times K_{up}$ 为中心的区域并与其预测出的该点的上采样重组核做点积运算,得到输出特征图,使之获得更多的语义信息。这种机制突破了传统方法的像素级局部性限制,能够有效融合长程上下文信息。

1.4 FDEM

水下环境繁杂,不仅受到不同程度光照的干扰,还存在噪声、悬浮物等干扰因素,将会对模型的决策产生影响。而注意力机制往往通过给与检测目标大的权重信息,同时给予干扰信息较小的目标权重,最终使得模型聚焦目标区域,从而忽略不重要的信息,以此来提升模型的检测效果。为了进一步提升模型的检测性能以及抗干扰能力,本文充分结合通道注意力机制和像素注意力机制的特点,构建了 HAM,如图 6 所示。

水下目标在图像中所占区域有限,其完整特征主要分散在某些通道中,而通道注意力机制能实现跨通道的信息交互与整合,将不同通道中与目标相关的信息进行融合,形成更全面、更具代表性的特征表示。通道注意力机制能

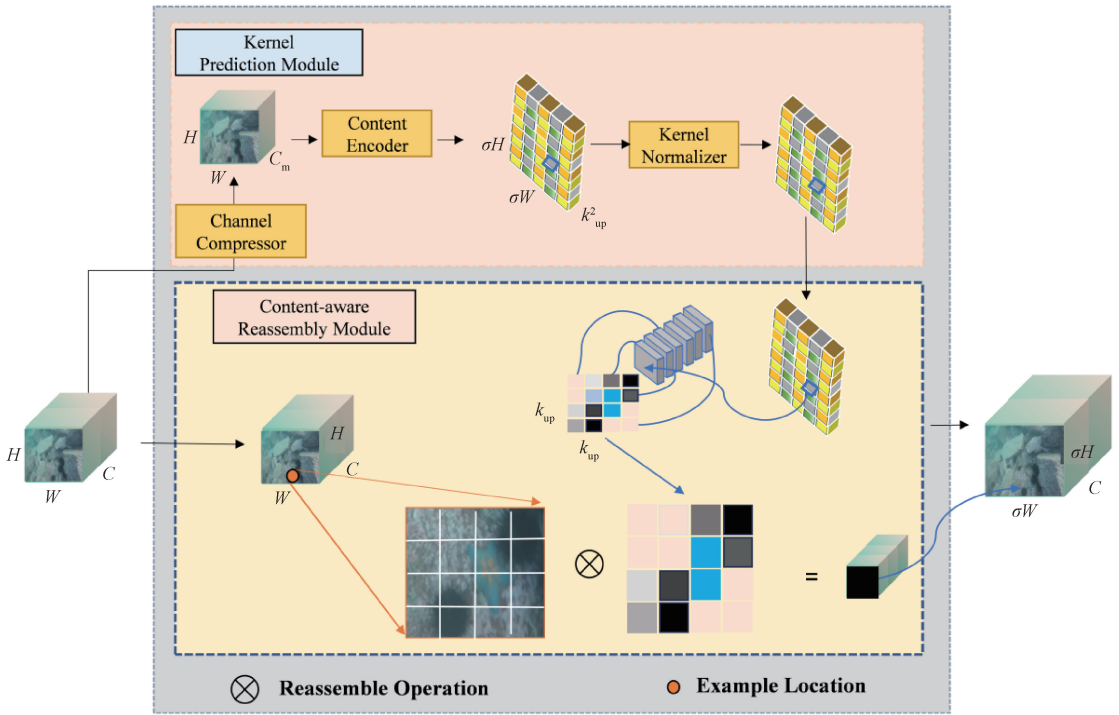


图5 CARAFE 的基本结构

Fig. 5 The basic structure of CARAFE

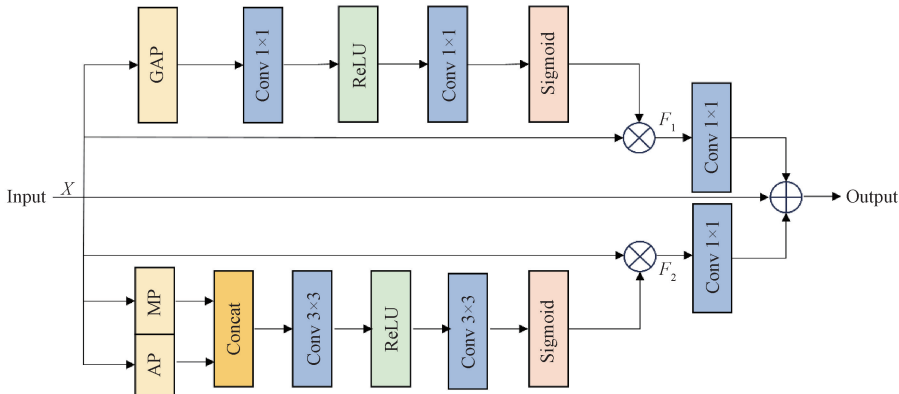


图6 混合注意力机制

Fig. 6 Hybrid attention mechanism

自动学习每个通道的重要性权重,着重突出包含目标关键语义信息的通道,通过强化这些通道可使目标特征更明显。首先对输入的特征图进行全局平均池化操作,将每个通道的二维特征压缩为一个具有全局信息的实数,该实数可以在一定程度上代表该通道的整体特征强度。然后,通过一个或多个全连接层来学习这些实数之间的权重关系。这些全连接层能够挖掘通道之间的重要性差异。最后,把学习到的权重再重新映射到原始特征图的各个通道上,从而增强那些被判定为重要通道的特征,同时抑制相对不重要通道的特征,滤除杂乱信息的干扰,以此来突出对任务更关键的通道信息。对于输入,输出为 F_1 :

$$F_1 = \delta_s(f_{1 \times 1}(\delta_R(f_{1 \times 1}(\text{GAP}(X)))) \otimes X \quad (10)$$

式中: $\delta_s(\cdot)$ 指 Sigmoid 函数, $f_{1 \times 1}(\cdot)$ 指 1×1 大小的卷积, $\delta_R(\cdot)$ 指 ReLU 激活函数, $\text{GAP}(\cdot)$ 指全局平均池化。

像素注意力机制重点关注图像中的每个像素点信息,通过像素级自适应权重分配和并行池化结构的设计,在目标检测任务中展现出显著优势。对于目标所在的像素区域,赋予更高的权重,突出水下目标的像素特征;而对于背景的像素区域,赋予更低的权重,有效滤除背景所带来的干扰,以保留对任务更关键的像素信息。对于输入 X ,输出为 F_2 。最后,将 F_1 和 F_2 分别通过卷积调整通道参数后与原输入累加得到最终输出。

$$F_2 = \delta_s \left(f_{3 \times 3} \left(\delta_R \left(f_{3 \times 3} \left(\text{Concat} \left(\begin{matrix} f_{AP}(X) \\ f_{MP}(X) \end{matrix} \right) \right) \right) \right) \right) \otimes X \tag{11}$$

$$\text{Output} = f_{1 \times 1}(F_1) + f_{1 \times 1}(F_2) + X \tag{12}$$

式中： $f_{3 \times 3}(\cdot)$ 指 3×3 大小的卷积， $f_{AP}(\cdot)$ 指平均池化， $f_{MP}(\cdot)$ 指最大池化。

2 实验结果与分析

2.1 实验环境与数据集

1) 实验数据集

本文使用的水下光学图像数据集有两种，一种是来自

全国水下机器人大赛公布的开源数据集 URPC，手动剔除掉海草这一干扰标签后，共包含 4 752 张图像，主要有海参、海胆、扇贝和海星 4 类生物，其中训练集 3 802 张图像，测试集包含 950 张图像。另一种是涵盖多种场景的水下数据集 RUOD，共 14 000 张图像，涵盖了多种海洋生物，包括海参、海胆、扇贝、海星、潜水员、鱼、珊瑚、海龟、鱿鱼、水母 10 个类别，其中训练集包含 9 800 张图像，测试集包含 4 200 张图像。两类数据集中部分图像如图 7 所示。

2) 实现细节

本文实验环境如表 1 所示。

此外，本文根据大量实验所得经验，在训练时采用 SGD 优化器对模型进行优化，动量设置为 0.937，迭代批次

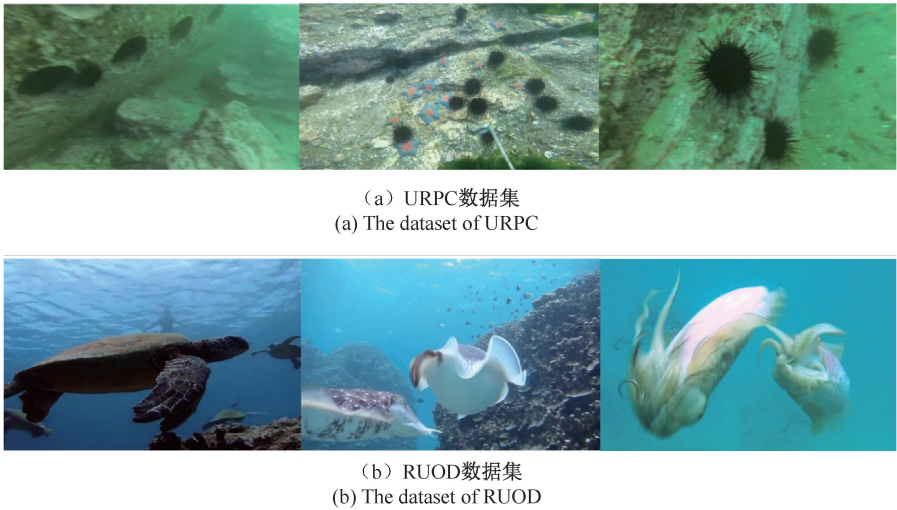


图 7 水下光学图像数据集

Fig. 7 Underwater optical image dataset

表 1 实验环境配置信息

Table 1 The experimental environment configuration information

配置名称	配置信息
操作系统	Windows11
深度学习框架	Pytorch1.11.0
CPU	13th Intel(R)Core(TM)i7-13700KF@3.40 GHz
GPU	NVIDIA GeForce RTX4070
集成开发环境	PyCharm2022.3.2
开发语言	Python3.8.5
虚拟环境	Anaconda3

设为 16，初始学习率设为 0.01，权重衰减设为 0.000 5，输入图像分辨率设为 640×640 ，共训练 300 个周期。

2.2 评价指标

本文主要利用精确率 (precision, P)，召回率 (recall,

R)，平均精度均值 (mean average precision, mAP) 来评估水下目标检测模型的性能。

1) 精确率

精确率是指在预测为正样本的所有样本中，实际为正样本的比例。它衡量的是模型预测为正样本的准确性，公式如式(13)所示。

$$P = \frac{TP}{TP + FP} \tag{13}$$

式中：TP 表示实际为正样本且预测为正样本的个数；FP 表示实际为负样本且被分类器预测为正样本的个数。

2) 召回率

召回率是指在实际为正样本的样本中，被预测为正样本的比例。它衡量的是模型能够正确识别出正样本的能力，公式如式(14)所示。

$$R = \frac{TP}{TP + FN} \tag{14}$$

式中：FN 表示实际为正样本而被分类器错误地预测为负样本的个数。

3)平均精度均值

平均精度均值是通过计算所有目标类别的平均精度 (average precision, AP) 的平均值得到,能够更全面地评估模型对不同水下目标类别的检测性能。公式如式(15)、(16)所示。

$$AP = \int_0^1 P(R) dR \tag{15}$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \tag{16}$$

式中: N 表示水下目标的类别数量。 P - R 曲线描述了精确率与召回率之间的关系,而 AP 值通过计算 P - R 曲线下的面积得到,反映为模型对某一类水下目标的检测性能。

2.3 消融实验

为了验证本文改进算法中每个关键模块的有效性,有助于提升模型检测水下小尺度目标的性能,本文采用 YOLOv8s 作为基准模型,并在 URPC 数据集上进行消融实验。表 2 为各个模块的检测结果,其中“√”代表加入此模块。

表 2 消融实验
Table 2 The ablation experiment (%)

实验	基准模型	FDEM	WFCM	CARAFE	HAM	P	R	mAP
1	√					79.1	72.6	79.7
2	√	√				79.8	73.2	80.9
3	√		√			80.4	72.9	81.2
4	√			√		79.5	73.0	80.6
5	√				√	79.2	72.8	80.4
6	√	√	√			80.7	73.4	81.6
7	√	√	√	√		81.3	73.7	82.3
8	√	√	√	√	√	82.0	73.9	82.8

1)FDEM 的有效性

从表 2 可以观察到,实验 1 为基准模型,检测结果分别为精确率 79.1%、召回率 72.6%、平均准确率 79.7%,将全动态特征提取模块 FDEM 加入基准模型后,精确率、召回率以及 mAP 分别增加 0.7%、0.6%、1.2%,得益于 FDEM 中融入多个动态瓶颈块结构,通过扩展各分支输出特征之间的融合,增加了梯度流信息的传递,丰富特征的表达能力,同时,该模块中通过增加残差连接以及 ODConv 层,使得模型能够动态调整卷积核参数特性,自适应不同尺度特征,增强对目标细节特征的关注程度,略除杂乱背景信息的干扰,提升对于不同尺度目标的特征提取能力。

2)WFCM 的有效性

由表 2 可知,当不使用 WFCM 模块时,模型在各项评价指标上均出现了较为明显的下降。基准模型未加入该模块与加入该模块相比,精确率、召回率以及 mAP 均降低 1.3%、0.3%、1.5%,同时通过对比实验 2 和 6 的数据,未加入该模块后,指标均降低 0.9%、0.2%、0.7%。这是由于 WFCM 模块通过加权操作,为目标信息分配不同权重值,对不同层级特征图加权处理,有助于保留低层级小目标的重要信息,滤除杂乱特征的干扰,此外,加权后的特征图与各层级交互融合,增加了水下目标高层次强语义信息的融合,使得检测效果大大提升。

3)CARAFE 的有效性

在颈部,本文将特征融合中的最邻近上采样方法替换为 CARAFE 算子,其他结构保持不变,修改 CARAFE 上

采样算子中 K_{up} 以及 $K_{encoder}$ 的值。其中, K_{up} 是用于上采样的核,它通过对低分辨率特征图进行上采样操作,生成与高分辨率特征图相同大小的特征图;而 $K_{encoder}$ 用于编码特征,它通过对输入特征进行编码操作,生成位置特定的核,这些核可以用于后续的特征重组。在基准模型 YOLOv8s 上,对 CARAFE 中两个参数修改后进行实验,以调整到适合该模型的参数,实验结果如表 3 所示。

表 3 CARAFE 算子参数设置
Table 3 The CARAFE operator parameter settings

$K_{encoder}$	K_{up}	$P/\%$	$R/\%$	mAP/%
—	—	79.1	72.6	79.7
3	5	79.3	71.1	79.5
3	7	78.9	71.4	79.3
5	5	79.5	73.0	80.6
5	7	78.8	72.6	80.0
7	7	79.6	71.8	80.5

经过实验分析可知,当参数 $K_{encoder} = 5, K_{up} = 5$ 时,模型取得较好的检测效果,精确率、召回率、平均精度均值较于采用最邻近元采样算子的基准模型分别提升 0.4%、0.4%、0.9%。由于最邻近元采样算法只选择与目标点最近的纹理像素点作为采样点,因此在面临存在噪声或者纹理复杂的水下环境时检测的准确性不高;而 CARAFE 算子,在特征图的每个位置上预测上采样重组内核,并在预定义的附近区域内重组特征,通过内容感知处理,利用预

先生成的自适应内核实时适应特定实例的内容,加强对局部重要区域的关注程度,在复杂的水下环境中有助于提升模型的性能。

4)混合注意力的有效性

最后,将混合注意力模块加入模型中,模型的各参数指标均有所提升,具体而言,模型在增加 FDEM、WFCM 以及 CARAFE 的基础上增加 HAM 模块,精确度从 81.3%增加为 82%,召回率从 73.7%增加为 73.9%,平均精确度从 82.3%增加为 82.8%,而实验 1 和 5 相比,指标分别提高多少 0.1%、0.2%、0.7%,这充分说明 HAM 能够重点关注水下目标的有用信息,忽略背景信息的干扰,这得益于利用通道注意力跨通道交互信息,利用像素注意

力增强关键像素点区域特征,从而形成全面的特征表示,有利于捕获水下目标的关键特征。

2.4 对比实验

1)模型对比

为了探究所提模型的有效性,选取部分现阶段水下目标检测模型并在 URPC 数据集与 RUOD 数据集上进行对比试验,从表 4 可以看到,改进模型在 URPC 数据集上的评价指标更高,这说明模型能够有效检测水下目标,降低误检率,从而提高模型的检测效能;在 RUOD 数据集上,改进模型所取得的检测效果更为优异,这表明改进模型能够适应多种水下场景,并能进一步验证其具有较强的泛化能力。

表 4 在 URPC、RUOD 数据集上对比试验

Table 4 The comparative experiments on the URPC and RUOD dataset (%)

模型	年份	主干	URPC			RUOD		
			P	R	mAP	P	R	mAP
SSD	2021	ResNet-50	64.4	56.5	63.4	71.5	61.7	72.0
YOLOv5s	2020	CSPDarknet-53	79.3	64.3	76.7	80.8	68.7	80.5
YOLOX	2021	CSPDarknet-53	74.5	70.6	75.6	79.5	74.90	79.8
Cascade R-CNN	2022	Swin-Large	80.6	73.2	81.1	84.3	74.6	83.8
Underwater-YCC	2022	CSPDarkNet53	80.2	71.9	80.3	81.9	72.3	82.5
GFocal V2 ^[24]	2023	Res2Net-101-DCN	79.3	71.2	79.8	80.9	72.5	81.7
CenterNet V2 ^[25]	2022	R2-101-DCN-BiFPN	74.2	65.3	74.5	76.1	66.4	77.6
本文	—	CSPDarknet-53	82.0	73.9	82.8	84.9	76.1	86.7

2)可视化对比

为了更加直观感受改进模型的检测效果,在 URPC 数据集和 RUOD 数据集上取出一部分测试集进行检测。其中,图 8、9 所示为 URPC 数据集上的检测效果对比,图 8

为背景光干扰情况,图 9 为背景颜色干扰情况。可以看到,SSD、Underwater-YCC、YOLOv5s 等模型虽然可以检测到目标,但由于受到背景干扰导致检测精度不高,尤其受到强光干扰下容易产生漏检情况;相比较而言,改进模

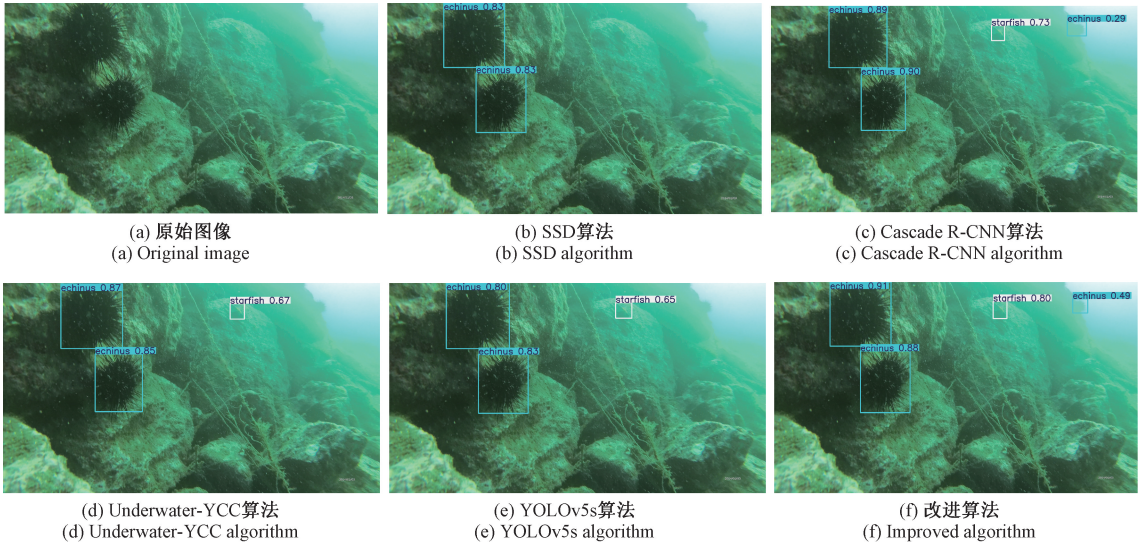


图 8 背景光干扰(URPC 数据集)

Fig. 8 The background light interference(URPC dataset)

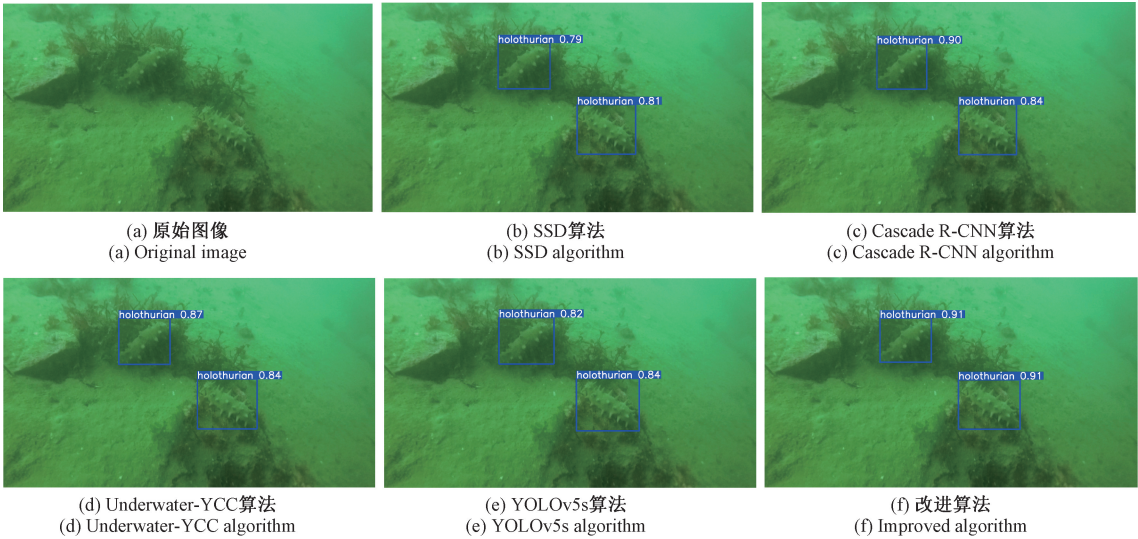


图 9 背景色干扰(URPC 数据集)

Fig. 9 The background color interference(URPC dataset)

型能够有效改善背景带来的干扰,检测到所有目标并提升检测的精确度。图 10、11 所示为在 RUOD 数据集上的检测效果对比,图 10 为背景光干扰情况,图 11 为背景颜色干扰情况。通过检测效果可知,改进后的模型在背景干扰下不仅准确检测到目标,减少错检和漏检情况的发生,而且精确度明显提升,从而进一步验证改进模型的有效性,能较好适应于复杂的水下环境。

2.5 热力图效果对比

由引言部分可知,模型在水下环境受到较为严重的背景信息干扰,例如不均匀的光照干扰、背景色干扰等,直接或者间接地对模型决策产生影响。为进一步验证本文改

进模型在水下具有较强的抗干扰性,能够适应复杂的水下环境,本文选择在 URPC 以及 RUOD 数据集中两个具有代表性的背景干扰场景(第 1 行为背景色干扰,第 2 行为背景光干扰)进行热力图可视化测试,测试结果分别如图 12、13 所示。其中,第 1 列为原始图片,第 2 列为改进前基准模型(YOLOv8s)的检测效果,而第 3 列为本文改进模型(混合注意力与动态特征引导网络)的检测效果。从图中可以看到,基准网络在检测时由于受到背景信息的干扰,热力区域难以覆盖整个目标,而且边缘界限不明显,难以有效判定其大致形状,而且热力图中心红色关键区域的面积较小,这说明模型对于目标判断的决策能力相对较弱。

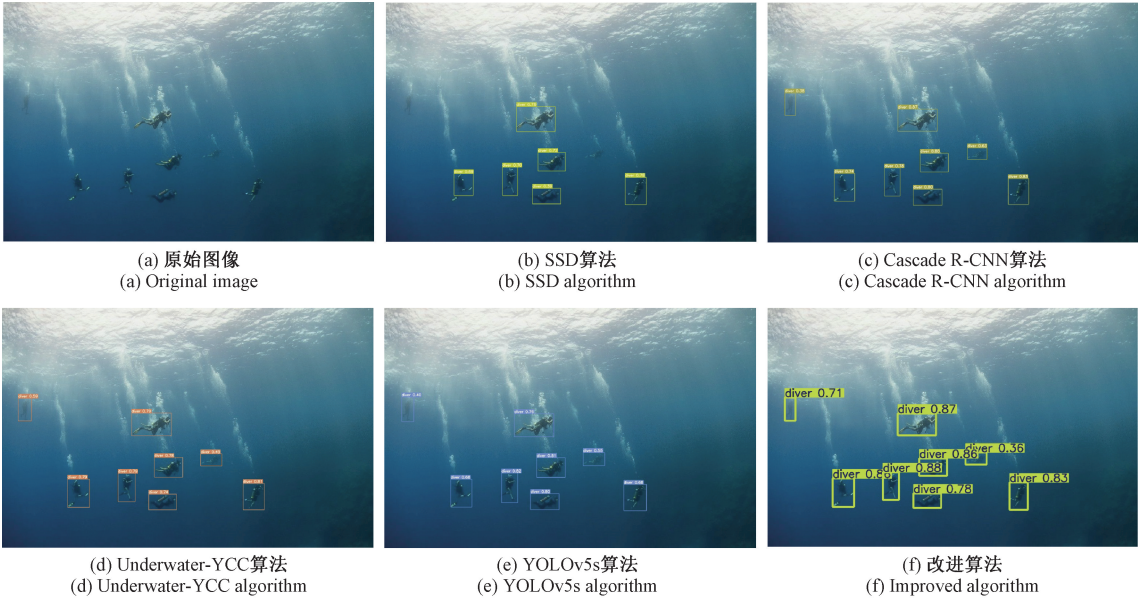


图 10 背景光干扰(RUOD 数据集)

Fig. 10 The background light interference(RUOD dataset)

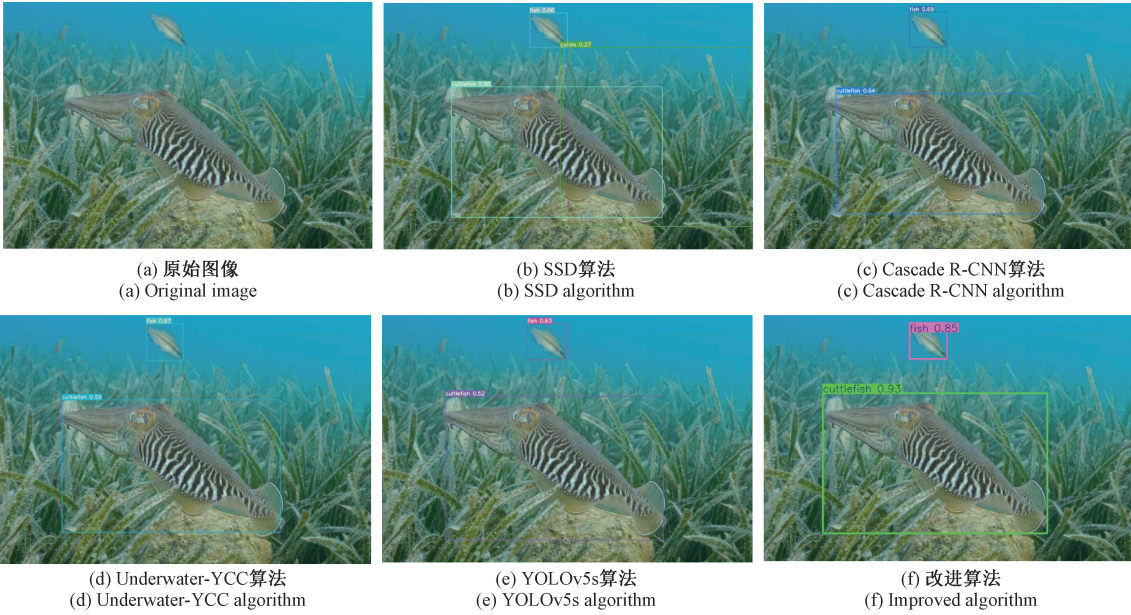


图 11 背景色干扰(RUOD 数据集)

Fig. 11 The background color interference(RUOD dataset)

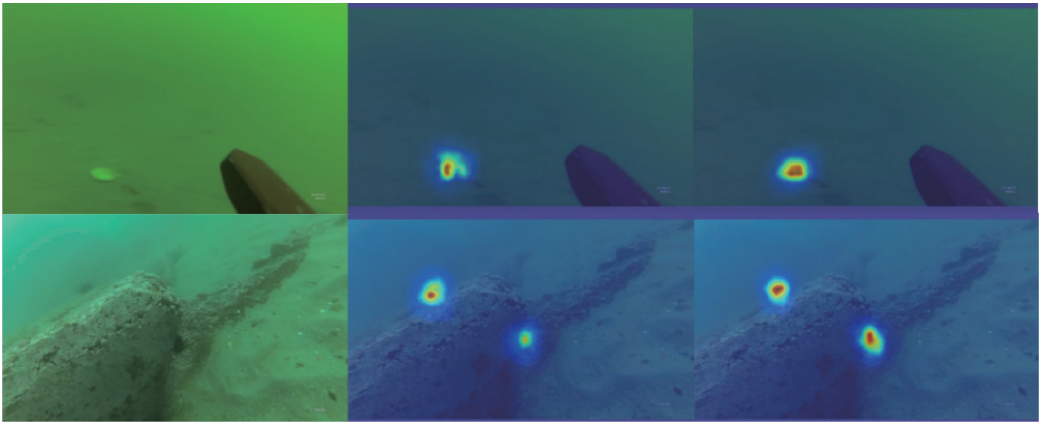


图 12 URPC 数据集下热力图可视化

Fig. 12 The heatmap visualization of the URPC dataset

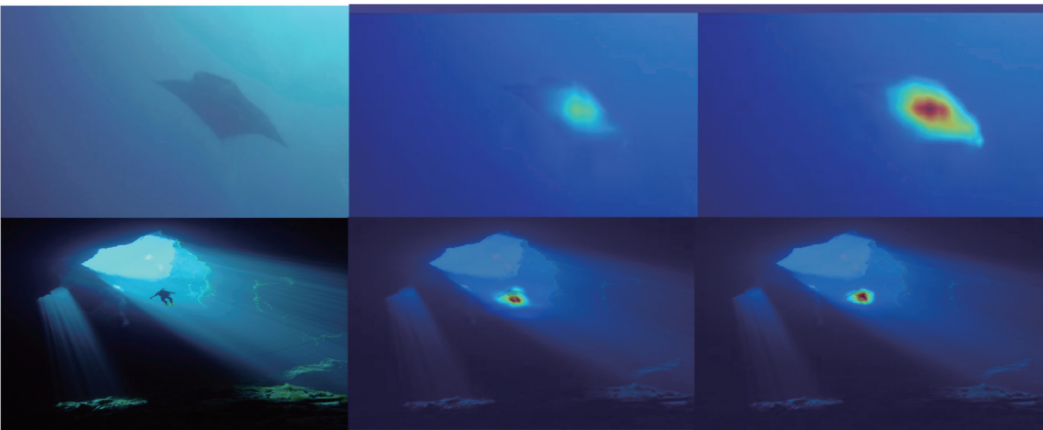


图 13 RUOD 数据集下热力图可视化

Fig. 13 The heatmap visualization of the RUOD dataset

相对而言,本文改进算法目标红色中心关键区域更为集中,面积也较大,不同热力区域过渡平滑且覆盖目标区域,边缘界限明显,能更好反映出目标的范围和整体形状,这充分说明本文改进算法能较好地适应复杂的水下环境,更加关注目标的关键区域特征,同时精准提取到目标整体性特征,使得目标边界框回归更为精准,改善模型的决策能力,提升模型在水下的应用范围。

3 结 论

水下目标极易受到背景信息的干扰,使模型难以有效聚焦关键特征、关注区域难以有效覆盖整个目标,进一步限制模型的检测性能。为解决上述问题,本文提出一种混合注意力与动态特征引导网络的水下目标检测算法。在两个数据集下进行广泛实验,并与多个主流模型进行对比实验,通过数据分析以及算法可视化结果可知本文算法的检测效果优于多个主流模型。同时,基准模型与改进后模型在背景光干扰、背景色干扰等情况下进行热力图可视化实验可知,本文算法能够有效捕捉目标的关键特征,最大程度关注目标的整体性特征,有效降低背景所带来的干扰,进一步提升模型的检测性能。然而,目前含有水下目标的光学图像数据集比较少,标注的种类也比较有限,导致模型的泛化性还不够理想,未来工作将融合多源数据,除光学图像,水下声呐数据也蕴含丰富的目标信息,通过融合多模态数据,可提供互补信息,进一步扩展应用场景,有助于在复杂水下环境中更精准地识别和定位目标,提升模型的鲁棒性。

参考文献

- [1] 王旭,张敏学. 中国海洋牧场建设:战略意涵、风险挑战、发展路径[J]. 太平洋学报, 2024, 32(12): 67-80.
Wang Xu, Zhang Minxue. China's marine ranching construction: strategic implications, risks and challenges, and development paths [J]. Pacific Journal, 2024, 32(12): 67-80.
- [2] Shen L H, Reda M, Zhang X, et al. Polarization-driven solution for mitigating scattering and uneven illumination in underwater imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 4202615.
- [3] Yi W G, Wang B. Research on underwater small target detection algorithm based on improved YOLOv7[J]. IEEE Access, 2023, 11: 66818-66827.
- [4] Chen L Y, Zheng M C, Duan S Q, et al. Underwater target recognition based on improved YOLOv4 neural network[J]. Electronics, 2021, 10(14): 1634.
- [5] Liu D L, Shen W H, Cao W J, et al. Design of siamese network for underwater target recognition with small sample size[J]. Applied Sciences, 2022, 12(20): 10659.
- [6] 陈亮,黄代琴,时慧晶,等. 基于多级特征混叠融合的水下小目标检测算法[J]. 探测与控制学报, 2022, 44(5): 77-82.
Chen Liang, Huang Daiqin, Shi Huijing, et al. An multilevel feature hybrid fusion algorithm of underwater small object detection [J]. Journal of Detection & Control, 2022, 44(5): 77-82.
- [7] 李培坤,李锋,葛忠显,等. 基于改进 YOLOv8n 的水下目标检测算法[J]. 电子测量技术, 2025, 48(3): 172-179.
Li Peikun, Li Feng, Ge Zhongxian, et al. Underwater target detection algorithm based on improved YOLOv8n[J]. Electronic Measurement Technology, 2025, 48(3): 172-179.
- [8] 王书朋,李凡. 抗混叠与多尺度特征融合的水下目标检测算法[J]. 计算机工程与应用, 2025, 61(18): 209-217.
Wang Shupeng, Li Fan. Underwater object detection algorithm with anti-aliasing and multi-scale feature fusion[J]. Computer Engineering and Applications, 2025, 61(18): 209-217.
- [9] 崔颖,韩佳成,高山,等. 基于改进 Deformable-DETR 的水下图像目标检测方法[J]. 应用科技, 2024, 51(1): 30-36,91.
Cui Ying, Han Jiacheng, Gao Shan, et al. An object detection method of underwater image based on improved Deformable-DETR[J]. Applied Science and Technology, 2024, 51(1): 30-36,91.
- [10] 刘侦龙,王骥,麦仁贵. 基于改进 YOLOv10s 的海洋牧场水下海参检测方法[J]. 农业工程学报, 2025, 41(10): 186-194c.
Liu Zhenlong, Wang Ji, Mai Rengui. Detecting underwater sea cucumber in marine ranching using improved YOLOv10s[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(10): 186-194.
- [11] 罗偲,李凯扬,吴吉花,等. 基于对抗注意力机制的水下遮挡目标检测算法[J]. 计算机工程, 2024, 50(10): 313-321.
Luo Cai, Li Kaiyang, Wu Jihua, et al. Underwater occlusion target detection algorithm based on adversarial attention mechanism [J]. Computer Engineering, 2024, 50(10): 313-321.
- [12] 谭鹤群,李玉祥,朱明,等. 通过图像增强与改进 Faster-RCNN 网络的重叠鱼群尾数检测[J]. 农业工程学报, 2022, 38(13): 167-176.
Tan Hequn, Li Yuxiang, Zhu Ming, et al. Detecting

- overlapping fish population using image enhancement and improved Faster-RCNN networks [J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(13): 167-176.
- [13] 周昕, 李远禄, 吴明轩, 等. 改进 YOLOv8 的水下目标检测[J]. 计算机系统应用, 2024, 33(11): 177-185.
Zhou Xin, Li Yuanlu, Wu Mingxuan, et al. Underwater target detection via improved YOLOv8[J]. Computer Systems & Applications, 2024, 33(11): 177-185.
- [14] 陈亮, 杨羽翼, 张剑, 等. 基于特征增强与损失优化的水下遮挡目标检测算法[J]. 探测与控制学报, 2023, 45(3): 109-115.
Chen Liang, Yang Yuyi, Zhang Jian, et al. Underwater occlusion target detection algorithm of feature enhancement and loss optimization[J]. Journal of Detection & Control, 2023, 45(3): 109-115.
- [15] 陶洋, 赵文博, 钟邦乾, 等. 融合大卷积核注意力机制的水下目标检测算法[J]. 小型微型计算机系统, 2024, 45(11): 2688-2694.
Tao Yang, Zhao Wenbo, Zhong Bangqian, et al. Underwater object detection algorithm with large kernel convolutional attention mechanism[J]. Journal of Chinese Computer Systems, 2024, 45(11): 2688-2694.
- [16] Li Z M, Zheng B, Chao D, et al. Underwater-Yolo: Underwater object detection network with dilated deformable convolutions and dual-branch occlusion attention mechanism[J]. Journal of Marine Science and Engineering, 2024, 12(12): 2291.
- [17] 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 Daoxian, Zhang Yinlong, Xu Gaofei, et al. Underwater cultural artifact detection based on deformable convolution and deep layer aggregation network[J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 185-195.
- [18] 史朋飞, 韩松, 倪建军, 等. 结合数据增强和改进 YOLOv4 的水下目标检测算法[J]. 电子测量与仪器学报, 2022, 36(3): 113-121.
Shi Pengfei, Han Song, Ni Jianjun, et al. Underwater object detection algorithm combining data enhancement and improved YOLOv4 [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 113-121.
- [19] Zhuang H W, Liu W S. Underwater biological target detection algorithm and research based on YOLOv7 algorithm[J]. IAENG International Journal of Computer Science, 2024, 51(6): 594-601.
- [20] Li C, Zhou A J, Yao A B. Omni-dimensional dynamic convolution [PP/OL]. V1. (2022-09-16). <https://arxiv.org/abs/2209.07947>.
- [21] Wang J Q, Chen K, Xu R, et al. Carafe: Content-aware reassembly of features [C]//IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.
- [22] Reis D, Kupec J, Hong J, et al. Real-time flying object detection with YOLOv8[PP/OL]. V1. (2023-05-17). <https://arxiv.org/abs/2305.09972>.
- [23] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation [C]//IEEE International Conference on Computer Vision, 2015: 1520-1528.
- [24] Liu Y, Zhang K H, Zhao Y Q, et al. Bi-RRNet: Bi-level recurrent refinement network for camouflaged object detection[J]. Pattern Recognition, 2023, 139: 109514. DOI:10.1016/j.patcog.2023.109514.
- [25] Xu L Q, Ji Z P, Zhou X D, et al. Prohibited items detection on X-ray images with multi-task learning [C]//8th International Conference on Computing and Artificial Intelligence, 2022: 558-565.

作者简介

裴晓芳(通信作者), 副教授, 硕士研究生导师, 主要研究方向为智能信号处理、嵌入式系统设计与计算机视觉等。

E-mail: xiaofangpei@cxwu.edu.cn

杨继海, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 2642792538@qq.com

周进, 硕士研究生, 主要研究方向为图像处理。

E-mail: 2067969710@qq.com

徐永恒, 硕士研究生, 主要研究方向为缺陷检测。

E-mail: 1150444587@qq.com