

基于细节增强与多尺度特征融合的水下目标检测^①



胡远江, 杨翼, 吴湘宁, 王梦雪, 潘志鹏

(中国地质大学(武汉)计算机学院, 武汉 430078)

通信作者: 吴湘宁, E-mail: wxning@cug.edu.cn

摘要: 针对水下目标检测任务中图像分辨率低、目标尺寸变化大、水体浑浊及目标遮挡而导致检测精度低的问题, 提出一种基于细节增强与多尺度特征融合的水下目标检测模型 DEMF-YOLO11n (YOLO11n underwater object detection model based on detail enhancement and multi-scale feature fusion). 模型以 YOLO11n 为主体, 为改善普通下采样对于低分辨率图像及小目标所造成的特征损失问题, 使用浅层与深层鲁棒下采样方法 RFD (robust feature downsampling) 分别替换模型浅层及深层步长卷积. 同时, 使用多核并行卷积模块 PKIModule (poly kernel inception module) 及上下文锚点注意力 CAA (context anchor attention) 对模型骨干中的 C3k2 进行重新设计, 以增强模型对遮挡目标及不同尺度目标的特征提取效果, 以及对复杂背景下目标的感知能力. 最后, 为解决水体浑浊等因素所导致的目标边缘细节模糊问题, 在头部网络中使用 CGAFusion (content guide attention fusion) 模块将深层特征与经 EEM (edge enhance module) 进行边缘增强后的浅层纹理特征进行自适应融合. 在 RUOD 数据集上的实验结果表明, DEMF-YOLO11n 较基准模型 mAP_{50} 提升 2.8%, mAP_{50-95} 提升 4.1%, 而参数量仅增加 0.53M.

关键词: 水下目标检测; YOLO11; 边缘增强; 多尺度特征融合; 注意力机制

引用格式: 胡远江, 杨翼, 吴湘宁, 王梦雪, 潘志鹏. 基于细节增强与多尺度特征融合的水下目标检测. 计算机系统应用, 2025, 34(9): 69-78. <http://www.c-s-a.org.cn/1003-3254/9925.html>

Underwater Object Detection Based on Detail Enhancement and Multi-scale Feature Fusion

HU Yuan-Jiang, YANG Yi, WU Xiang-Ning, WANG Meng-Xue, PAN Zhi-Peng

(School of Computer Science, China University of Geosciences, Wuhan 430078, China)

Abstract: To address the challenges of low image resolution, significant variation in target size, water turbidity, and occlusion that lead to poor detection accuracy in underwater object detection, a YOLO11n underwater object detection model based on detail enhancement and multi-scale feature fusion (DEMF-YOLO11n) is proposed. The model is built upon the YOLO11n architecture. To mitigate feature loss caused by standard downsampling in low-resolution images and small targets, the robust feature downsampling (RFD) method is adopted to replace the shallow and deep strided convolution layers. In addition, a module in the backbone is redesigned using the poly kernel inception module (PKIModule) and context anchor attention (CAA), enhancing feature extraction for occluded and multi-scale targets, as well as improving perception in complex underwater environments. To address the blurring of target edges caused by water turbidity and similar factors, shallow texture features are first refined using the edge enhancer module (EEM), and are subsequently fused with deep semantic features in the head network via the content guide attention fusion (CGAFusion) module. Experimental results on the RUOD dataset demonstrate that the proposed DEMF-YOLO11n achieves a 2.8% improvement in mAP_{50} and a 4.1% improvement in mAP_{50-95} compared to the baseline model, with an increase of only 0.53M in the number of parameters.

① 基金项目: 国家自然科学基金 (U21A2013); 湖北省自然科学基金 (2021CFB506)

收稿时间: 2025-01-09; 修改时间: 2025-02-17; 采用时间: 2025-02-24; csa 在线出版时间: 2025-07-14

CNKI 网络首发时间: 2025-07-15

Key words: underwater object detection; YOLO11; edge enhancement; multi-scale feature fusion; attention mechanism

1 引言

水下机器人可以代替人类深入水底完成资源开发、环境探测等任务^[1]。水下目标检测可以识别水下物体类别并确定位置,对水下机器人自动作业具有重要意义^[2]。然而,由于水下光线存在衰减现象,导致图像颜色严重失真,使目标与背景区分度较低^[3]。同时,浮游生物、泥沙等因素也会使水体变浑浊,呈现出雾霾现象,进一步降低了水下目标的对比度^[4]。此外,由于机位变化、目标遮挡及目标体型差异等因素,导致目标的尺度变化较大。这些不利因素使得现有的深度学习目标检测模型难以适应水下复杂场景,因此,研究高效、稳定的水下目标检测模型十分必要,将推动水下探测技术在海洋资源开发和环境保护中的高效应用。

目前目标检测主流研究主要使用两种模型:双阶段目标检测模型及单阶段目标检测模型。其中,双阶段目标检测模型以 Faster R-CNN^[5]为代表,采用候选框生成和精细分类回归的两步策略,第1阶段通过区域提议网络生成少量高质量候选框,第2阶段进一步优化检测结果。然而其推理速度较慢,模型训练成本较高。单阶段模型通过将图像划分为网格,直接回归目标的位置和类别,代表性方法为 YOLO 系列模型^[6],其结构简洁、计算效率高,在水下目标检测中主要使用该模型。

近年来,部分研究就水下复杂环境对 YOLO 系列模型进行了一系列针对性改进。赵永胜等^[7]在 YOLOv5 主干网络中引入多尺度注意力,以降低水下复杂背景的干扰,并在头部网络中使用自适应空间特征融合对多尺度特征进行有效融合。在高色偏和低对比度水下图像中,目标的边缘信息对检测至关重要,尤其是对于与背景区分度低的目标。Chi 等^[8]在 YOLOv5 主干网络中引入交叉注意力捕获全局上下文信息,并在颈部网络前引入伪装目标检测领域的纹理增强模块对目标边缘进行增强,从而提高对非显著目标的检测能力。陶洋等^[9]使用大核卷积注意力对 YOLOv5 主干网络重新设计,提高模型对全局特征的感知,从而实现重叠、遮挡目标的有效检测。辛世澳等^[10]在 YOLOv7 中采用 BiFPN 结构实现多尺度特征融合,并使用 BiFormer 注意力增强细节特征,以增强对低分辨率、模糊图像及

小目标的检测能力。袁红春等^[11]在 YOLOv8 中使用融合可变形卷积对卷积核进行自适应偏移采样,从而在不同尺度和方向上更全面提取特征,以应对水下目标重叠、遮挡、形状不规则问题。刘飞等^[12]使用大核卷积对 YOLOv8 中快速特征金字塔模块进行改进,提高模型感受野,从而一定程度上提升重叠目标及遮挡目标的检测能力。周昕等^[13]在 YOLOv8 主干中引入可变形卷积以增强对重叠遮挡目标的检测能力,并使用空洞卷积空间金字塔模块提取目标多尺度特征,以增强对远景近景中不同尺度目标的感知能力。

以上针对背景区分度低的目标、小目标及遮挡目标的改进方法取得了显著成效,为本文研究提供了进一步优化方向。然而,现有研究仍存在改进空间。第一,针对传统下采样操作对模糊目标边缘特征损失问题的考虑不足。第二,在多尺度特征提取中未综合利用目标上下文环境信息,而上下文环境信息对具有复杂背景的目标感知具有重要作用。第三,浅层特征代表了高频特征如边缘细节,深层特征代表了全局语义信息,简单的拼接并不能有效融合浅层与深层两个维度的特征信息。此外,当前的主流研究所使用的 URPC、DUO 等数据集仅包含 3、4 种目标类别,且所处环境相对单一,难以全面衡量水下目标检测模型性能。针对以上问题,提出一种基于细节增强及多尺度特征融合的水下目标检测模型 DEMF-YOLO11n,并使用包含有丰富环境及生物类别的 RUOD 数据集^[14]进行训练和评估。模型以 YOLO11n 为基础,对水下复杂环境的目标检测任务做出如下针对性改进。

(1) 为改善步长卷积对低分辨率图像及小目标特征细节完整性的损失,使用 RFD 模块^[15]替换步长卷积下采样操作。

(2) 为提高对尺度变化大的目标及复杂背景下目标的感知能力,使用结合多尺度特征提取及上下文感知策略的 PKIBlock^[16]对主干网络的 C3k2 模块进行了重新设计。

(3) 为充分利用浅层边缘纹理特征,并实现浅层深层特征的对齐融合,设计了 ECGAFusion 特征融合模块,在头部网络中将浅层边缘细节特征增强后与深层特征进行自适应融合。

2 YOLO11 结构

如图 1 所示, YOLO11 由骨干网络 Backbone、颈部网络 Neck 及头部网络 Head 构成, 分别用于特征提取、特征融合及目标分类与检测。

在骨干网络中, 对尺寸为 640×640 的输入图像, 首先进行两次步长卷积 Conv 进行 4 倍下采样, 将特征尺寸缩小为 160×160 。随后多次交替使用 C3k2 模块进行特征提取以及使用步长卷积 Conv 进行下采样。其中, C3k2 模块 (图 2) 为特征提取的核心, 对通道数为 c_1 的输入特征首先使用 1×1 卷积升维到 $2 \times c_1$; 随后将通道数均分为两部分, 其中一部分直接恒等连接, 而另一部分经过多个 BottleNeck 或 C3k 模块进行特征提取; 最后将每一层的输出特征按通道维度拼接后, 使用 1×1 卷积融合并降维得到输出特征。而 Conv 模块则将特征图尺寸减半并增加通道数, 以减少冗余特征并获得多层

次特征表示。最后使用空间池化金字塔 SPPF 及 C2PSA 结构进行特征融合与细化。

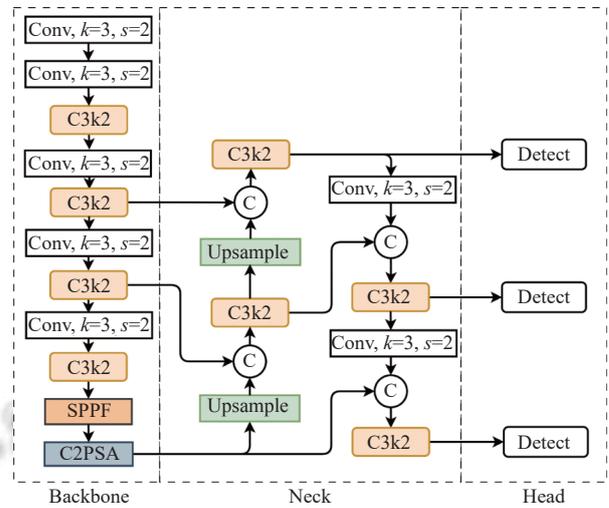


图 1 YOLO11 结构

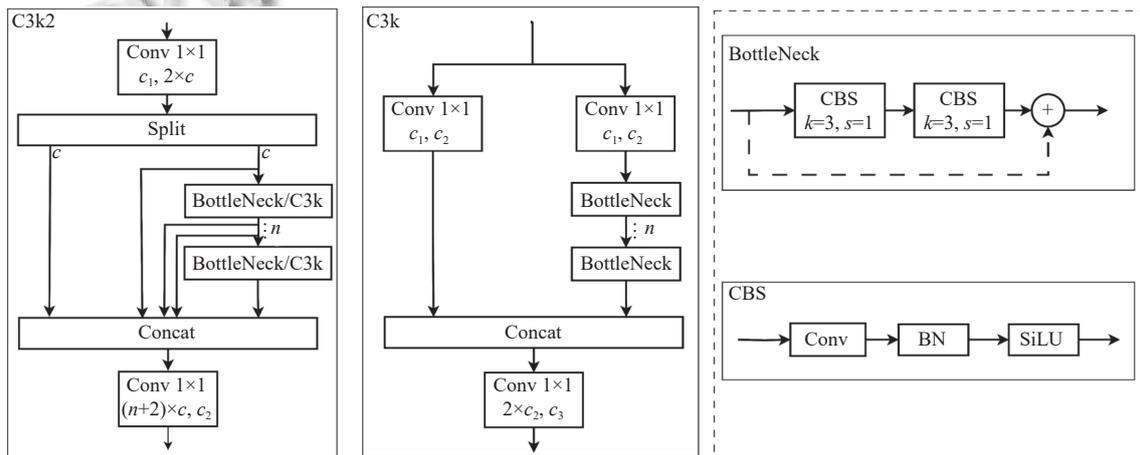


图 2 特征提取模块 C3k2 结构

颈部网络中的操作分为两个阶段, 第 1 阶段自底向上将骨干网络输出的特征进行上采样后, 与骨干网络中相同尺寸的浅层特征图拼接, 再使用 C3k2 提取特征; 第 2 阶段自顶向下将第 1 阶段输出的特征经下采样后, 与第 1 阶段中相同尺寸特征图进行拼接后再使用 C3k2 提取特征。由于浅层特征包含丰富的纹理细节, 而经过一系列下采样及特征提取后的深层特征包含语义信息, 在颈部网络中通过将浅层特征与深层进行特征融合, 可以增强浅层特征与深层特征的交互, 从而综合利用纹理细节与语义特征, 增强网络的目标感知能力。

头部网络包含 3 个检测头, 每个检测头同时实现了

目标分类与目标定位。由于网络层数越浅, 特征图尺寸越大, 而大尺寸特征图对小目标特征保留越完整, 因此自上往下分别是小目标、中等目标及大型目标检测头。

3 模型设计

为增强复杂水体环境下目标检测能力, 我们重新设计了 YOLO11n, 做出针对性改进, 并命名为 DEMF-YOLO11n, 其结构如图 3 所示。第一, 将步长卷积下采样 Conv 模块更换为 RFD 模块, 并在图片输入阶段使用 SRFD (shallow RFD) 进行 4 倍下采样, 抑制冗余像素特征; 而在后续阶段使用 DRFD (deep RFD) 进行 2 倍下采样。第二, 重新设计主干网络中的 C3k2 特征提

取模块,使用 PKIBlock 替换其内部的 BottleNeck,并将重新设计后的 C3k2 命名为 C3k2-PKI 特征提取模块.第三,在检测头前设计 ECGAFusion 特征融合模块,对浅层细节纹理特征进行边缘增强,并与深层语义特征进行特征自适应融合.

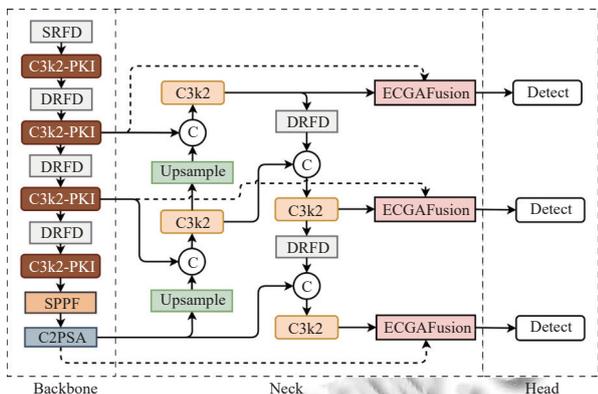


图3 DEMF-YOLO11n 结构

3.1 RFD 下采样模块

在目标检测模型中,下采样操作用于降低特征图的空间分辨率并提升通道维度.其中,降低空间分辨率通过缩减特征图尺寸实现,从而提高后续特征提取操作的计算效率,并扩大深层神经元的感受野;而提升通道维度则可以获得多层次特征表示.然而,作为特征提取的前序步骤,下采样需要在减少分辨率的同时保留重要的空间信息,否则可能导致目标关键特征的丢失,进而影响后续的特征提取和检测性能.YOLO11 采用步长为 2 的卷积操作实现下采样,在融合局部特征方面具有优势,但在包含低分辨率图像和小目标的水下环境中,该操作会导致目标关键边缘特征丢失.

为解决上述问题,使用 RFD 模块替换 YOLO11 中步长卷积下采样模块,以保留水下图像中目标关键细节.RFD 模块将输入特征图复制为 3 个副本,分别采用步长卷积 DWConv、切片下采样 CutD 和最大池化 MaxPooling 这 3 种方法进行下采样,并通过特征融合策略将 3 种下采样结果进行互补整合后得到最终下采样特征.通过以上操作,能够在压缩特征图尺寸的同时,保留更多的细节信息和纹理特征,为后续的特征提取提供更丰富的输入.其中,步长卷积下采样用于整合局部特征并提升模型的泛化能力,切片下采样在减少分辨率的同时能够尽可能地保留原始信息,而最大池化下采样则用于突出特征图关键细节,以避免特征信息的过多损失.切片下采样原理如图 4 所示,将特征图相

邻元素进行切片,按照奇数、偶数行与奇数、偶数列进行切分得到尺寸为原始特征图一半的 4 个切片特征图.并将切片特征图按通道进行拼接后执行 1×1 卷积进行特征融合,从而得到相较于输入特征图尺寸减半的切片下采样特征.

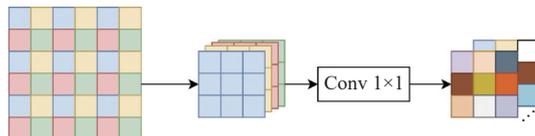


图4 切片下采样原理

为了适应不同阶段的下采样特性,RFD 针对模型不同深度设计了两种下采样操作:SRFD 和 DRFD,其结构如图 5 及图 6 所示.其中,SRFD 用于模型的浅层阶段,针对初始输入图像下采样.由于此时图像包含大量的冗余像素信息,SRFD 能够选择性地保留显著特征,同时抑制无关信息,提高后续特征的提取效率.而 DRFD 则用于模型的深层阶段,处理已经包含丰富语义信息的特征图.DRFD 通过精细的下采样策略,可避免丢失重要的语义特征.

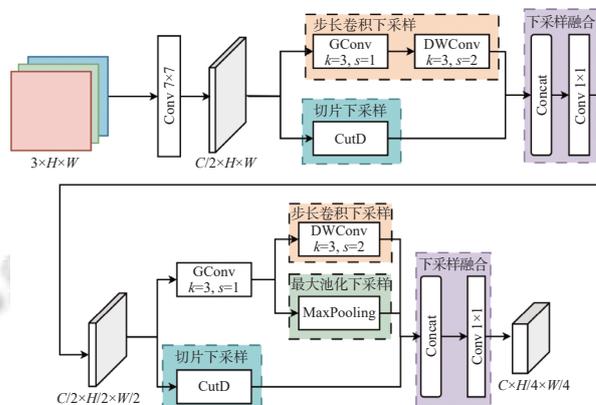


图5 SRFD 结构

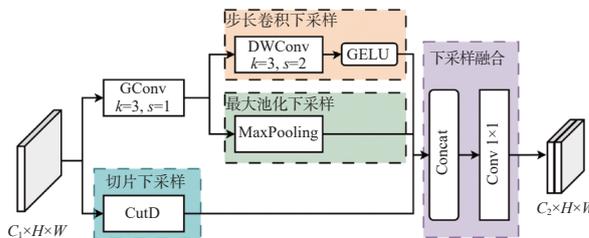


图6 DRFD 结构

3.2 C3k2-PKI 特征提取模块

在水下目标检测任务中,目标的部分遮挡、目标

属性的多样性以及摄像视角的变化,导致目标尺度差异显著,因此需要综合考虑不同尺度的目标特征提取。此外,水下图像因浑浊和偏色等问题,使得部分目标与背景在视觉上具有较高的相似性,进一步加大了目标检测的难度。在这一场景中,单纯依赖目标的外观特征难以实现高效识别,因而结合目标周围环境的上下文信息显得尤为关键。

针对此类问题,部分研究如 LSKNet^[17]通过融合大核卷积和膨胀卷积,有选择地扩大较大物体的空间感受野,以捕获更多的场景上下文信息,并提升模型对不同尺度目标的适应能力;空洞卷积空间金字塔池化 ASPP^[18]通过并行不同尺寸及膨胀率的卷积核进行多尺度特征提取。然而,大核卷积容易引入过多的背景噪声,在视线浑浊且分辨率低的水下图像中的使用效果

不佳,且大感受野对小目标的检测效果较差;而膨胀卷积虽然能扩展感受野,但其提取的特征往往较为稀疏,对感受野范围内的细粒度细节感知能力不足,从而可能会丢失部分与水下背景区分度较低的目标特征。这些方法在面对具有尺度多样性及细节敏感性的水下图像时存在明显局限。

为了解决上述问题,使用一种结合局部特征和远程上下文信息策略的 PKIBlock 结构对 YOLO11 中的 C3k2 模块进行重新设计,替换 C3k2 中的 BottleNeck 结构。PKIBlock 的结构如图 7(a) 所示,分别采用 PKI-Module 提取局部细节及上下文特征,利用 CAA 捕获远距离像素间的关联信息,并将两者融合,从而能够同时利用目标局部和全局上下文信息自适应提取目标特征,有助于提升模型在复杂水下场景中的目标检测能力。

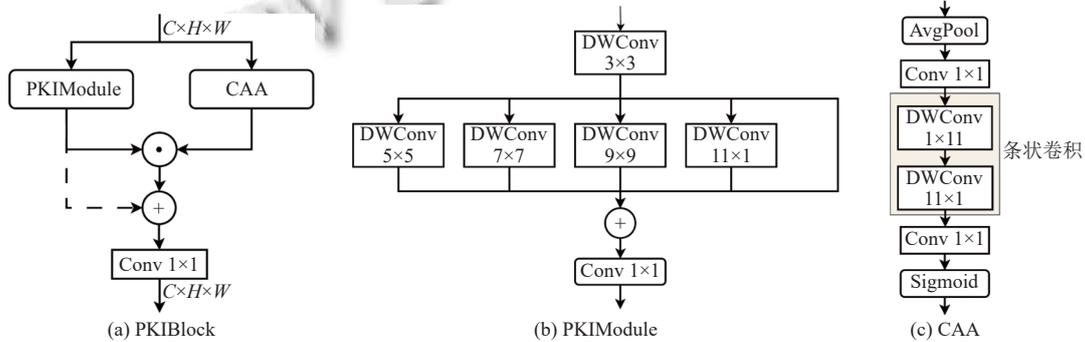


图7 PKIBlock 结构

PKI-Module 结构如图 7(b) 所示,首先使用 3×3 的小核深度卷积捕获目标的细粒度局部特征,如边缘与纹理信息。接着,通过并行不同感受野的深度卷积操作,进一步提取跨尺度的目标局部上下文信息,从而增强模型对不同目标尺度的适应能力。最终,采用 1×1 的点卷积将提取的不同感受野的纹理特征进行融合,生成局部特征上下文信息。CAA 模块结构如图 7(c) 所示,首先通过平均池化和 1×1 卷积对输入特征进行压缩,以提取关键的局部信息。随后,利用沿宽度方向与高度方向的一个一维条状深度卷积建立远程像素间的关系,并增强中心区域的特征。最后,通过 1×1 卷积和 Sigmoid 激活函数生成远程上下文权重向量,用于对全局上下文特征进行自适应增强。

3.3 ECGAFusion 特征融合模块

在水下目标检测任务中,浅层特征如边缘和轮廓在目标识别中具有重要地位,尤其是对于伪装于泥沙或礁石背景下的目标,但这些边缘特征在经过网络多

层次的传递后逐渐被削弱。针对这一问题,主流方法通过将浅层特征与深层特征进行跨层特征融合,以增强浅层到深层的特征传递。然而,浅层特征与深层特征在感受野和侧重点上存在显著差异,其中浅层特征的编码信息通常聚焦于细粒度的边缘与纹理,而深层特征更注重语义信息,其每个像素实际上综合了浅层特征中多个像素的信息。因此,简单的特征相加或串联操作无法很好地解决感受野不匹配所带来的信息丢失与冗余问题,需要使用合适的特征融合策略来改善特征对齐问题及实现自适应融合。

针对以上问题,设计一种基于自适应学习空间权重的特征融合方法 ECGAFusion,实现浅层特征与深层特征的有效融合,具体结构如图 8 所示。首先对浅层特征使用 EEM 边缘增强模块^[19]进一步增强边缘纹理细节特征,并与深层特征进行初步加权融合。随后,使用内容引导注意力 CGA (content guided attention)^[20]计算初步融合后特征每个通道特定的空间注意力权重,从

而为每个通道每个像素赋予不同的重要性.同时,自适应设置学习权重 W ,将空间注意力权重通过 W 及 $1-W$ 的权重分别作用于浅层特征与深层特征.最后,将注意力增强后的浅层与深层特征加性融合,并与增强之前的特征残差连接以减轻梯度消失问题,再通过 1×1 点卷积进行进一步特征融合,最终达到增强浅层特征

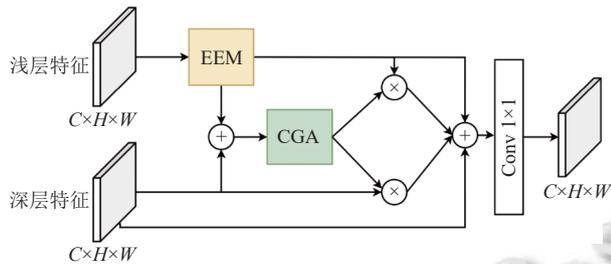


图8 ECGAFusion 结构

3.3.1 EEM 边缘增强模块

由于模型中存在下采样操作,虽已采用 PKI 及 RFD 进行改善,但仍会不可避免地丢失部分细节信息.为了进一步增强对模糊边界的水下目标的检测能力,在浅层特征与深层特征融合前,使用 EEM 边缘增强模块对浅层特征进行增强,模块结构如图 9 所示.首先,通过全局平均池化操作计算局部平滑特征图,并将原始特征图减去该平滑特征以提取显著边缘细节信息.随后,使用卷积操作进一步增强边缘信息,并将增强后的边缘特征经批归一化 BN 及 Sigmoid 激活后与原始特征相加,实现边缘细节的强化和特征恢复.增强后的浅层特征与深层特征进行融合后被送入检测头,可以提高对目标边界信息的感知能力,进而提升模型的整体检测性能.

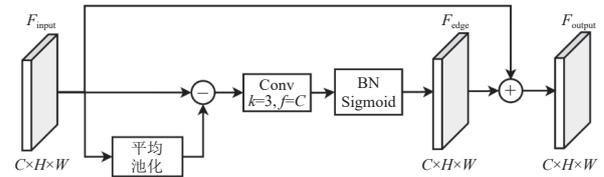


图9 EEM 边缘增强模块结构

3.3.2 CGA 内容引导注意力

传统空间通道注意力中,其通道注意力与空间注意力单独计算.对于尺寸为 $C \times W \times H$ 的输入,通道注意力会生成 $C \times 1 \times 1$ 的通道权重,每个通道对应一个标量权重,从而自适应地赋予不同通道重要性,但是其对于同一通道中不同像素没有区分性.而空间注意力会生成 $1 \times W \times H$ 的空间权重,为特征图每个像素点赋予一个标量权重,从而自适应地赋予特征空间中不同像素点的重要性.然而所有通道的同一个坐标像素共享同一个权重,忽略了不同通道间的差异.

针对上述不足,使用 CGA 内容引导注意力生成针对不同通道的空间注意力权重,以提高模型泛化性. CGA 结构如图 10 所示,首先对输入特征 F_{input} 计算 $1 \times W \times H$ 尺寸的空间注意力 F_{s2} 及 $C \times 1 \times 1$ 尺寸的通道注意力 F_{c1} .随后,这两种注意力权重通过 PyTorch 的广播机制加权,生成初步融合权重 F_{w1} .为进一步提升融合权重对每个通道不同的表达能力,使用 F_{input} 作为指导,与 F_{w1} 在通道维度上进行拼接,并采用通道重排操作使两者的每个通道交替排列.最后,经过 7×7 分组卷积及 Sigmoid 激活函数处理后,生成精细化的通道特异性空间注意力权重 F_{output} .通过以上操作,可以对各个通道不同空间位置自适应地赋予独立的权重,以增强特征图中重要信息的表达能力.

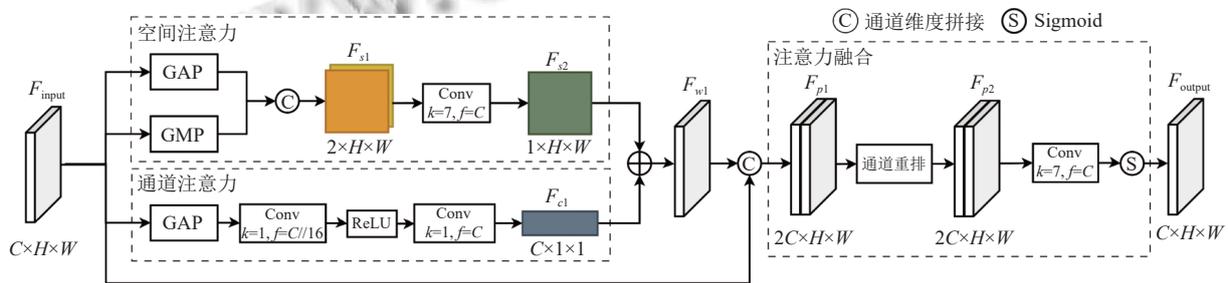


图10 CGA 注意力机制原理

4 实验结果与分析

4.1 实验数据集

实验数据集为现实世界水下环境数据集 RUOD,

包括 14000 张图像及 74903 个标记目标,其中训练集有 9800 张图像,测试集有 4200 张图像.数据集中涵盖了复杂的水下图像场景,包括光干扰、雾霾效果、偏色

等,并包含 10 种目标类别: 鱼类 fish、潜水员 driver、海星 starfish、珊瑚 corals、海龟 turtle、海胆 echinus、海参 holothurian、扇贝 scallop、墨鱼 cuttlefish 及水母 jellyfish.

4.2 实验环境参数

实验设备 CPU 为 Intel(R) Xeon(R) W-2223 @ 3.60 GHz, 内存大小为 32 GB, GPU 为 NVIDIA RTX A2000. 模型迭代次数 epoch 为 150, 批次大小为 32. 使用 3.10 版本 Python 下 2.2.2 版本 PyTorch 进行开发, CUDA 版本为 12.1.

4.3 评价指标

为综合评估模型精度及计算量, 使用如下评价指标: P 、 R 、 mAP_{50} 、 mAP_{50-95} 及参数量, 计算公式如式 (1)–(4) 所示. P 为精确率, 表示预测的正样本中实际的正样本比例; R 为召回率, 表示实际的正样本中预测正确的样本比例; mAP_{50} 表示在交并比 IoU 阈值为 0.5 时平均精度 AP 的均值; mAP_{50-95} 表示在 IoU 阈值为 0.5–0.95 且步长为 0.05 的多个阈值下 AP 的均值; 参数量表示模型的总参数量, 用于衡量模型复杂度.

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \int_0^1 P(R) dR \quad (3)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (4)$$

其中, TP 表示模型预测正确的正样本数; FP 表示模型预测为正样本的样本中, 实际为负样本的样本数; FN 为模型预测为负样本的样本中, 实际为正样本的样本数; n 代表样本数, $AP(i)$ 代表第 i 类样本的 AP .

4.4 消融实验

为了验证 DEMF-YOLO11n 对于 YOLO11n 各针对性改进在水下复杂环境下的有效性, 设计消融实验如表 1 所示. 其中, 基准模型选取 YOLO11n, 并分别添加 RFD 下采样模块、C3k2-PKI 特征提取模块、ECGA-Fusion 特征融合模块的不同组合, “√”表示在基准模型上添加了此改进模块.

表 1 YOLO11n 消融实验结果

实验编号	RFD	C3k2-PKI	ECGAFusion	P (%)	R (%)	mAP_{50} (%)	mAP_{50-95} (%)	参数量 (M)
1	—	—	—	83.8	74.6	82.4	57.2	2.58
2	√	—	—	83.8	76	83.2	57.9	2.56
3	—	√	—	84.7	77.1	84.1	59.8	2.59
4	—	—	√	83.9	76.9	83.8	59.2	3.07
5	√	√	—	84.6	77.5	84.4	60.2	2.56
6	√	—	√	85	76.8	84.3	59.9	3.1
7	—	√	√	85	77.1	84.3	60.2	3.08
8	√	√	√	85.4	77.5	85.2	61.3	3.11

从表 1 中实验 2–4 的数据可以看出, 对 YOLO11n 分别单独添加以上 3 个改进模块后, 各精确度指标相较于实验 1 均有显著提升. 其中, 添加 RFD 下采样模块有助于保留弱边界水下目标的边缘细节, 相较于 YOLO11n, 其 mAP_{50} 提升了 0.8%, mAP_{50-95} 提升了 0.7%, 且参数量减少了 0.02M; 添加 C3k2-PKI 特征提取模块增强了模型对多尺度变化及存在遮挡的水下目标的感知能力, 相较于原 YOLO11n 模型, 其 mAP_{50} 提升 1.7%, mAP_{50-95} 提升 2.6%, 参数量仅增加 0.01M, 综合性能在这 3 种改进模块中效果最好; ECGAFusion 特征融合模块对浅层边缘细节特征进行强化, 并与深层语义特征进行自适应深度融合, 能够有效提升模型对复杂水下环境的全局特征及局部细节感知能力. 相较

于 YOLO11n 模型, mAP_{50} 提升了 1.4%, mAP_{50-95} 提升了 2.0%.

此外, 表 1 中实验 5–7 展示了各改进模块在不同组合下对模型性能的影响. 由实验数据可知, 相较于单独添加各改进模块, 对 3 种改进模块的任意组合均能进一步提升模型精度, 且对模型参数量影响较小. 最终, 由实验 8 可知, 对 YOLO11n 添加所有改进模块后, 其精度相较于 YOLO11n 取得更大幅度提升, 其中 P 提升了 1.6%, R 提升了 2.9%, mAP_{50} 提升了 2.8%, mAP_{50-95} 提升了 4.1%, 而参数量仅增加了 0.53M, 能够满足水下目标检测中模型部署要求. 由此可知, 所提模型针对水下复杂环境的目标检测任务进行针对性改进, 取得一定性能提升, 并在精确率和参数量上取得高效折中.

4.5 对比实验

为了验证所提模型相较于其他模型的优势, 将 DEMF-YOLO11n 与模型参数量相近的主流目标检测模型在相同参数设置下进行实验对比, 包括 YOLOv5n、YOLOv8n、YOLOv9t、YOLOv10n 及 YOLO11n, 实验结果如表 2 所示。

表 2 对比实验结果

模型名称	P (%)	R (%)	mAP ₅₀ (%)	mAP ₅₀₋₉₅ (%)	参数量 (M)
YOLOv5n	83.7	74.2	81.9	56.3	2.5
YOLOv8n	83.9	75	82.5	57.3	3.01
YOLOv9t	83.3	74.1	82.2	56.5	1.97
YOLOv10n	82.2	74	81.1	56.1	2.70
YOLO11n	83.8	74.6	82.4	57.2	2.58
DEMF-YOLOv8n	84.6	77.9	84.9	60.5	4
DEMF-YOLO11n	85.4	77.5	85.2	61.3	3.11

从表 2 可以得知, 各基准对比模型的 mAP₅₀ 均在 81.1%–82.5% 之间. 其中, YOLOv9t 参数量最少, 但 mAP₅₀ 低于 YOLO11n. YOLOv10n 在该水下数据集上的表现不足, 仅取得了 81.1% 的 mAP₅₀. YOLOv5n 与 YOLOv8n 作为水下目标检测领域主流的骨干网络, 其效果仍有改进的空间. 其中, 虽然 YOLOv8n 相较于 YOLO11n 其 mAP₅₀ 稍高 0.1%, 然而其参数量却增加

了 0.43M; YOLOv5n 的参数量虽然比 YOLO11n 少 0.03M, 但其 mAP₅₀ 却降低了 0.5%. 最后, 选用精度最佳的 YOLOv8n 和 YOLO11n 作为主干模型, 对于水下的复杂环境进行针对性改进, 使用 RFD 下采样模块、C3k2-PKI 特征提取模块及 ECGAFusion 特征融合模块对其进行重新设计, 得到 DEMF-YOLOv8n 和 DEMF-YOLO11n 模型. 其中, DEMF-YOLOv8n 的 mAP₅₀ 达到 84.9%, 参数量却达到 4M; 而 DEMF-YOLO11n 的 mAP₅₀ 达到 85.2%, 优于 DEMF-YOLOv8n, 且参数量仅为 3.11M, 相较于 DEMF-YOLOv8n 少 0.89M. 因此, DEMF-YOLO11n 在检测精度与模型参数量两个度量上在对比模型中取得最佳性能, 能够有效满足水下目标检测任务需求。

4.6 实验可视化结果分析

为直观展示 DEMF-YOLO11n 对复杂水下环境针对性改进的有效性, 选取部分代表性测试集实验结果与 YOLO11n 进行对比. 在图 11 中, 每一行为一种代表性场景下实验结果, 每一列分别为原图、YOLO11n 检测结果、DEMF-YOLO11n 检测结果, 并在图中显著区别处使用箭头与序号标注。

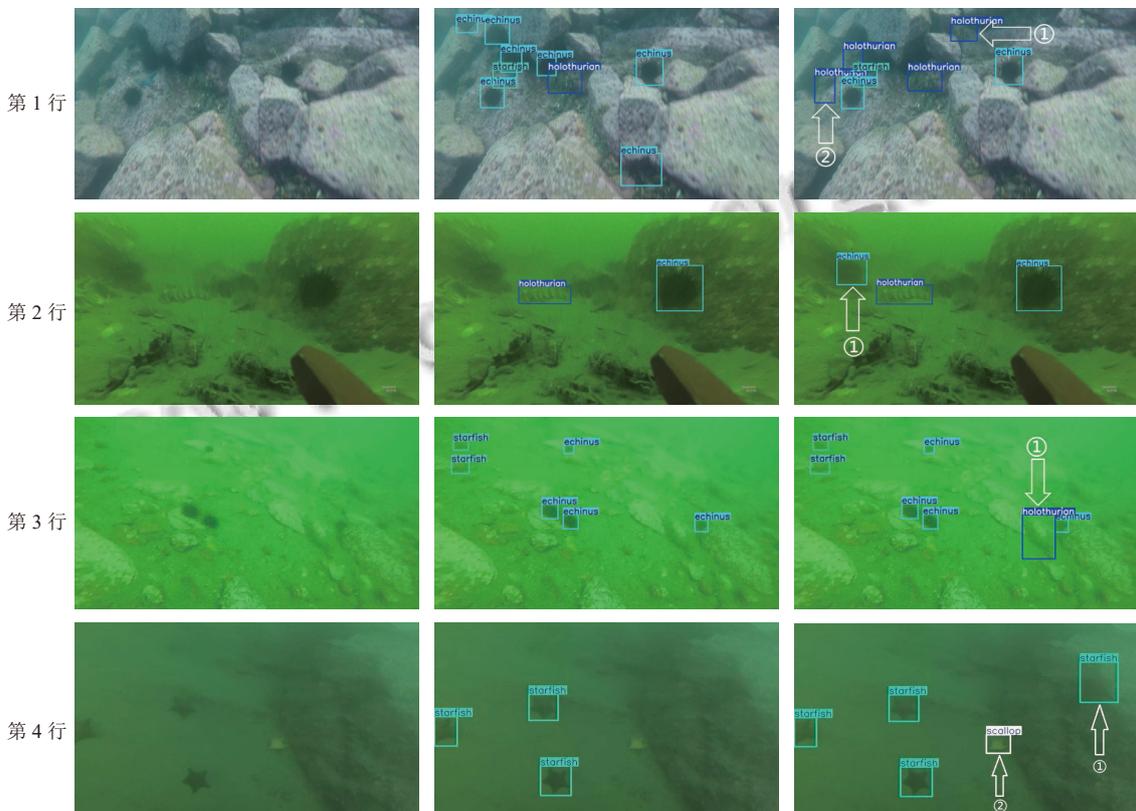


图 11 水下目标检测可视化结果

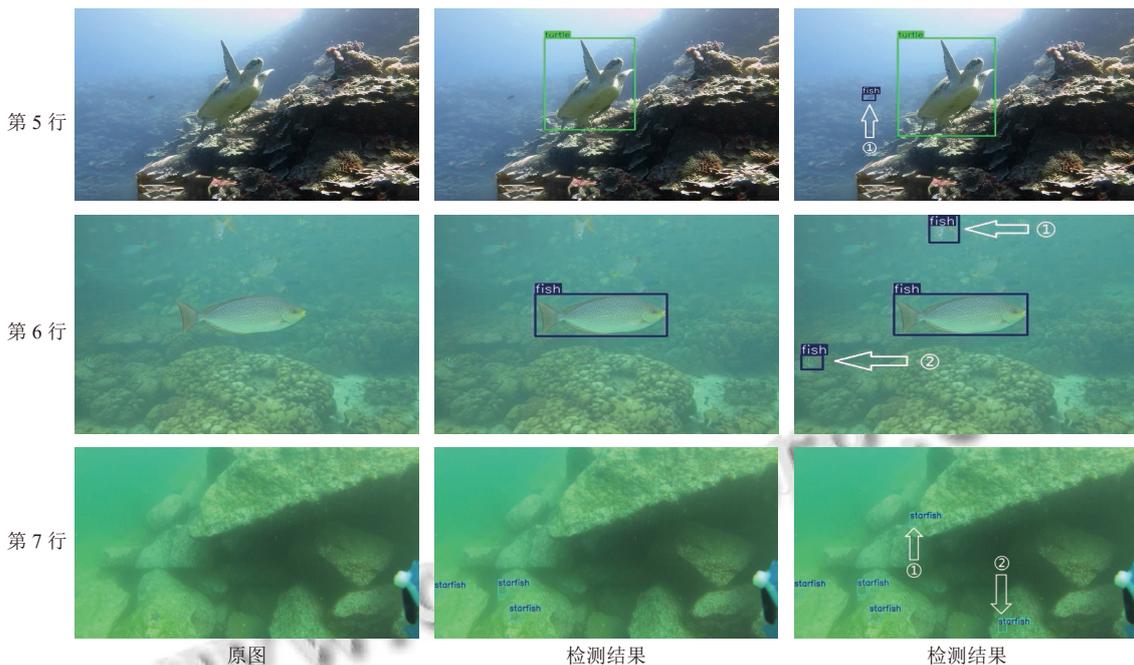


图 11 水下目标检测可视化结果(续)

由图 11 可以分析, 第 1 行中, ①号与②号海参与背景石块颜色相近, 区分度低, YOLO11n 发生漏检现象, 并且将石块阴影误检为海胆. 第 2 行中, ①号海参与背景融为一体, 颜色相近且边缘不清晰, YOLO11n 发生漏检. 第 3 行中, 画面整体分辨率低, 导致①号海参与边缘特征不明显, 因此 YOLO11n 发生漏检. 第 4 行中, 画面整体分辨率低且光照不足, 同时存在大量噪点, 导致①号海星边缘几乎模糊难以分辨, 与背景石块几乎融为一体, YOLO11n 发生漏检. 第 5 行中, 海龟与①号鱼类呈现多尺度变化现象, 海龟尺度大, 占据画面中大部分. 相比之下鱼类为典型小目标, 而 YOLO11n 未能正确处理该尺度变化强烈的场景, 因而发生漏检. 第 6 行中, 存在多个鱼类类别, 且鱼之间体型与外观差异较大, 此时 YOLO11n 发生漏检. 第 7 行中存在多个小型海星目标, 且该场景下受水体折射影响, 海星颜色与背景石头相似, 导致 YOLO11n 未能识别出①号与②号海星, 甚至肉眼也难以辨别. 然而受益于细节增强、多尺度特征融合及上下文锚点注意力等设计, DEMF-YOLO11n 能够准确识别出这些复杂环境下普通目标检测模型难以感知的目标, 具有显著优势.

5 结束语

为了实现色偏严重、水体浑浊、目标尺度变化大

等复杂条件下的水下目标检测, 提出了一种基于细节增强和多尺度特征融合的水下目标检测模型 DEMF-YOLO11n. 首先, 使用 RFD 下采样模块改善普通下采样对小目标及低分辨率图像带来的特征损失. 其次, 结合多尺度特征提取模块 PKIModule 与上下文注意力 CAA 对 YOLO11n 主干网络中特征提取模块 C3k2 重新设计, 以增强对尺度变化大及遮挡目标的检测能力. 最后, 在检测头前对浅层特征使用 EEM 进行边缘增强, 并与深层特征基于 CGAFusion 特征融合模块进行深度融合, 以提升目标边界的感知能力. 在 RUOD 数据集上的实验评价指标及可视化结果表明, DEMF-YOLO11n 能有效提升对水下复杂环境下目标的检测能力, 且对参数量影响较小, 对水下机器人任务具有一定指导意义.

参考文献

- 张路, 魏本昌, 魏鸿奥, 等. 基于改进 RT-DETR 的水下目标检测. 计算机系统应用, 2024, 33(12): 131-140. [doi: 10.15888/j.cnki.csa.009684]
- 袁明阳, 宋亚林, 张潮, 等. 基于 GA-RetinaNet 的水下目标检测. 计算机系统应用, 2023, 32(6): 80-90. [doi: 10.15888/j.cnki.csa.009134]
- 刘祎恒, 邓箴. 降质感知的小波变换水下图像增强网络. 计算机系统应用, 2024, 33(9): 201-207. [doi: 10.15888/j.cnki.

- csa.009616]
- 4 Liu RS, Fan X, Zhu M, *et al.* Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(12): 4861–4875. [doi: [10.1109/TCSVT.2019.2963772](https://doi.org/10.1109/TCSVT.2019.2963772)]
 - 5 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
 - 6 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 779–788.
 - 7 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. *电子测量技术*, 2024, 47(13): 183–190.
 - 8 Chi JN, Zheng L, Miao J. Underwater object detection algorithm based on improved YOLOv5. In: Qu Y, Gu MC, Niu YF, *et al.* eds. *Proceedings of the 3rd International Conference on Autonomous Unmanned Systems*. Singapore: Springer, 2024. 260–269.
 - 9 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. *小型微型计算机系统*, 2024, 45(11): 2688–2694.
 - 10 辛世澳, 葛海波, 袁昊, 等. 改进 YOLOv7 的轻量化水下目标检测算法. *计算机工程与应用*, 2024, 60(3): 88–99. [doi: [10.3778/j.issn.1002-8331.2308-0333](https://doi.org/10.3778/j.issn.1002-8331.2308-0333)]
 - 11 袁红春, 李春桥. 基于改进可变形卷积的 FDC-YOLO v8 水下生物目标检测方法研究. *农业机械学报*, 2024, 55(11): 140–146. [doi: [10.6041/j.issn.1000-1298.2024.11.015](https://doi.org/10.6041/j.issn.1000-1298.2024.11.015)]
 - 12 刘飞, 杨德刚, 章鑫, 等. 基于 YOLOv8 改进的水下目标检测算法. *计算机与现代化*, 2025(1): 113–119. [doi: [10.3969/j.issn.1006-2475.2025.01.018](https://doi.org/10.3969/j.issn.1006-2475.2025.01.018)]
 - 13 周昕, 李远祿, 吴明轩, 等. 改进 YOLOv8 的水下目标检测. *计算机系统应用*, 2024, 33(11): 177–185. [doi: [10.15888/j.cnki.csa.009680](https://doi.org/10.15888/j.cnki.csa.009680)]
 - 14 Fu CP, Liu RS, Fan X, *et al.* Rethinking general underwater object detection: Datasets, challenges, and solutions. *Neurocomputing*, 2023, 517: 243–256. [doi: [10.1016/j.neucom.2022.10.039](https://doi.org/10.1016/j.neucom.2022.10.039)]
 - 15 Lu W, Chen SB, Tang J, *et al.* A robust feature downsampling module for remote-sensing visual tasks. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 4404312.
 - 16 Cai XH, Lai QX, Wang YW, *et al.* Poly kernel inception network for remote sensing detection. *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2024. 27706–27716.
 - 17 Li YX, Hou QB, Zheng ZH, *et al.* Large selective kernel network for remote sensing object detection. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE, 2023. 16748–16759.
 - 18 Chen LC, Zhu YK, Papandreou G, *et al.* Encoder-decoder with atrous separable convolution for semantic image segmentation. *Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018. 833–851.
 - 19 Gao SX, Zhang PP, Yan TY, *et al.* Multi-scale and detail-enhanced segment anything model for salient object detection. *Proceedings of the 32nd ACM International Conference on Multimedia*. Melbourne: ACM, 2024. 9894–9903.
 - 20 Chen ZX, He ZW, Lu ZM. DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention. *IEEE Transactions on Image Processing*, 2024, 33: 1002–1015. [doi: [10.1109/TIP.2024.3354108](https://doi.org/10.1109/TIP.2024.3354108)]

(校对责编: 张重毅)