文章编号　1004-924X(2026)04-0671-15

单头自注意力和频域-空域融合的水下目标检测

李大海，廖嘉伟*，王振东
(江西理工大学 信息工程学院，江西 赣州 341000)

摘要：水体折射散射、光照不均致目标纹理模糊，水生生物多为伪装密集小目标，且水下资源受限平台对轻量化实时性有约束，这些因素共同加剧了水下目标检测的难度。为此，本文提出一种单头自注意力和频域-空域融合的YOLOv8n改进模型，YOLOv8n-SD。首先，设计局部全局特征融合增强主干网络，通过动态划分通道比例的单头自注意力机制，高效提取部分通道的长程全局信息；并融合高效特征提取块提取的局部细节信息，实现局部与全局特征的互补增强。其次，构建频域空域高效融合颈部网络，设计基于Haar小波变换与空间到深度变换的下采样模块，融合浅层高分辨率特征中重要的高频与空间信息；同时，采用快速归一化加权策略，动态优化多尺度特征融合效率。在水下公开数据集URPC2020与RUOD上，YOLOv8n-SD的$mAP_{0.5:0.95}$与mAP_{50}指标分别达到51.2%，85.7%和50.6%，85.0%。同时，较基准模型参数量减少42.3%，计算负载降低17.2%。对比实验进一步验证，本文模型在多种复杂水下场景中均表现出良好的检测精度与鲁棒性。

关　键　词：水下目标检测；自注意力机制；Haar　小波变换；小目标检测
中图分类号：TP394.1；TH691.9　　**文献标识码**：A
doi：10.37188/OPE.20263404.0671　　**CSTR**：32169.14.OPE.20263404.0671

Underwater object detection based on single-head self-attention and frequency-domain & spatial-domain fusion

LI Dahai, LIAO Jiawei*, WANG Zhendong

(*School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou* 341000, *China*)
* *Corresponding author, E-mail*: 6720230861@*mail.jxust.edu.cn*

Abstract: Water refraction, scattering, and uneven illumination blur target textures. Aquatic organisms are mostly small, camouflaged, and dense. Resource-constrained underwater platforms demand lightweight, real-time models. These factors collectively exacerbate the difficulty of underwater object detection. Therefore, this paper proposed an improved YOLOv8n model based on single-head self-attention and frequency-domain & spatial-domain fusion, named YOLOv8n-SD. First, a backbone network enhanced by local-global feature fusion was designed. It used a single-head self-attention mechanism combined with dynamic channel ratio division to efficiently acquire long-range global information from partial channels, and further fused local detail information of efficient feature extraction blocks to realize comple-

收稿日期：2025-10-17；**修订日期**：2025-11-26.
基金项目：国家自然科学基金资助项目(No. 61563019, No. 61562037)；江西理工大学校级资助项目(No. 205200100013)

mentary enhancement of local and global features. Second, a neck network with efficient frequency-domain and spatial-domain fusion was constructed, and a downsampling module using Haar wavelet transform and space-to-depth transform was designed to fuse important high-frequency and spatial information of shallow high-resolution features. At the same time, a fast normalized weighting strategy was adopted to dynamically optimize the efficiency of multi-scale feature fusion. On the public underwater datasets URPC2020 and RUOD, the $mAP_{0.5:0.95}$ and mAP_{50} metrics of YOLOv8n-SD reach 51.2%, 85.7% and 50.6%, 85.0% respectively. Meanwhile, compared with the baseline, the number of parameters is reduced by 42.3% and the computational load is decreased by 17.2%. Comparative experiments further verify that the proposed model exhibits good detection accuracy and robustness in various complex underwater scenarios.

Key words: underwater object detection; self-attention mechanism; Haar wavelet transform; small object detection

1 引　言

近年来,随着海洋探索在科学研究与经济发展中的战略地位日益凸显,传统人工水下作业“效率低、成本高、风险大”的弊端愈发突出,水下目标检测技术由此逐渐成为主流,在海洋生物保护、海洋资源勘探、水产养殖等场景中得到广泛应用[1-2]。然而,水下环境具有特殊性,水体对光的吸收、悬浮颗粒引发的光散射,会导致目标边缘模糊、纹理信息丢失;加之水生生物的天然伪装特性与集群游动小目标的相互遮挡,共同构成了“弱特征、小尺度、强干扰”的复合检测难题。

传统水下目标检测算法多采用人工设计的特征提取方式,从图像中提取形状、纹理等特征后,结合机器学习算法完成检测任务。Hu等[3]提出基于多分类SVM(Support Vector Machine)的鱼类物种分类方法,通过融合颜色特征与纹理特征显著提升了分类精度。Chuang等[4]提出基于显著性引导动态标记的鱼类识别方法,借助分离性、适配性与判别性指标优化可变形组件学习,进一步提高了识别准确度。这类方法的检测性能高度依赖手工设计特征的表达能力,特征判别性不足会直接限制下游任务精度;且检测流程需人工设定边界框参数与固定阈值,使得其难以适配复杂水体环境的视觉检测任务。

针对上述的水下复合检测难题,卷积神经网络凭借高效的自动特征学习能力成为研究焦点,研究者围绕模型轻量化与检测精度提升展开系列优化。Feng等[5]提出CEH-YOLO模型,集成高阶可变形注意力模块聚焦关键区域,结合增强型空间金字塔池化模块,强化水下小目标颜色与纹理特征提取。Qu等[6]则提出YOLOv8-LA模型,通过选择性处理输入通道优化空间特征提取,引入轻量级上采样算子缓解目标信息丢失,实现精度与效率的平衡。何等[7]针对水下声呐弱特征,动态计算卷积核通道权重,强化关键通道关注度,显著提升了水下复杂环境下的弱特征表征能力。除直接优化检测模型外,图像增强作为预处理环节也成为性能提升的重要方法。例如,李等[8]提出基于颜色先验引导与注意力机制的水下图像增强方法,通过二者协同改善图像色彩失真与细节模糊问题。陶等[9]则基于Retinex理论设计可变注意力机制,针对低照度水下场景优化增强效果,为后续检测提供高质量预处理基础。

近年来,Transformer模型在陆上目标高精度检测中表现优异,研究人员开始尝试将自注意力机制引入水下目标检测方法,以提升检测精度。Gao等[10]通过路径增强模块与自适应点表示策略改进自注意力机制,使网络能自适应增强全局特征,对长条状目标的检测精度提升

显著。姚等[11]设计目标感知增强的双阶段检测头：一阶段通过增加区域提议生成网络深度与交并比分支获取目标先验信息；二阶段引入自注意力机制抑制背景干扰，并将一阶段先验信息融入分类求解过程，显著提升了特征判别能力。卷积神经网络能高效捕捉图像局部信息，因此诸多研究者将自注意力机制与卷积神经网络相结合。Liu 等[12]设计包含多头自注意力机制的主干网络，结合路径聚合网络融合深层语义特征与浅层细节特征，增强水下低对比度图像的特征描述能力。李等[13]将门控卷积网络与自注意力机制融合，引入结构重参数化技术降低推理计算负载，同时通过混合编码器实现浅层高频信息与深层语义信息的融合，平衡了检测精度与实时性。然而，这些融合方法未针对水下弱特征优化通道利用效率，仍存在通道冗余与计算资源浪费问题。

水下环境的复杂多变与光照条件的限制，会进一步导致频域与空域信息丢失，因此研究者们着手探索频域与空域信息的提取及利用方法，以期进一步提升水下目标检测精度。张等[14]提出频域注意力机制，通过离散余弦变换将图像映射至频域，还设计低频特征引导组件，捕获低频轮廓信息，提高了检测精度，有效验证了频域对水下弱特征的增强作用。韩等[15]则采用自适应空间分解模块替代传统步长卷积，从而保留更多空域信息，提升了模型对低分辨率图像及水下小目标的检测性能。由于现有算法的主干网络多依赖单一卷积或自注意力架构，在深层传播过程中易丢失浅层丰富的高频特征与空间细节，难以缓解深层网络的信息衰减问题。

现有研究表明，自注意力机制的全局建模能力、频域与空域方法的细节保留优势，是解决水下复合检测难题的核心，但单独应用或简单叠加难以适配水下复杂场景与资源受限平台的部署需求。为此，本文通过单头自注意力和频域-空域融合（Single-head self-attention and frequency-domain & spatial-domain fusion，SD）方法对 YOLOv8n[16]改进，提出 YOLOv8n-SD 模型。该模型针对低光照、高散射、小目标密集遮挡的复杂水下场景设计，能强化局部与全局信息互补，高效融合频域空域特征，适配小型水下机器人、便携式探测设备等资源受限平台；其在较基准模型参数量减少 42.3%，计算负载降低 17.2% 的前提下，于水下公开数据集 URPC2020 和 RUOD 上 $mAP_{0.5:0.95}$ 与 mAP_{50} 两个指标分别提升了 2.0%/2.1% 和 1.6%/1.7%，有效兼顾了检测精度与实时性需求。具体创新设计如下：

（1）提出局部全局特征融合增强主干网络：设计高效特征提取块（Efficient Feature Extraction Block，EFEB）提取局部纹理细节，构建部分单头自注意力（Partial Single-Head Self-Attention，PSSA）模块，对部分通道建模长程空间依赖，降低通道冗余；通过局部与全局特征互补融合，显著强化主干网络在水下模糊图像中的特征表征能力。

（2）构建频域空域高效融合颈部网络：设计频域空域信息融合下采样模块，采用 Haar 小波变换（Haar Wavelet Transform，HWT）提取高频纹理，通过空间到深度变换（Spatial-to-Depth Transform，SDT）提取空间结构，结合深度可分离卷积实现二者高效融合，缓解深层网络信息损耗；同时优化网络结构并引入快速归一化加权策略，提升多尺度特征融合效率，增强水下密集目标的特征判别能力。

2　研究方法

2.1　网络模型概述

本文构建的 YOLOv8n-SD 模型总体结构如图 1 所示。该模型以轻量级 YOLOv8n 为基础架构，在设计的局部全局融合特征增强主干网络中，采用高效特征聚合模块（Efficient Feature Aggregation Module，EFAM）替代主干网络 P4，P5 层的原始 C2f 模块。该模块先通过卷积提取局部细节特征，再按比例动态划分特征图通道并执行单头自注意力计算，降低通道冗余引发的计算负载，并高效建模长距离依赖关系。随后，局部细节特征与全局关联信息通过融合实现互补增强，显著提升主干网络对水下模糊图像的特征表征

能力。其次，构建频域空域高效融合颈部网络，设计了频域空域信息融合下采样模块（Frequency and Spatial Information Fusion Downsampling，FSIFD），直接对主干网络包含丰富小目标原始特征的P2层进行无损下采样并提取特征，并与P3层的特征图进行融合，有效缓解下采样过程中浅层细节易丢失的问题。同时，颈部网络还引入快速归一化加权策略Fusion，实现多尺度特征的自适应分配及高效融合提高模型对水下密集目标的特征判别能力。下文将对上述核心改进模块的结构设计与工作机制进行详细阐述。

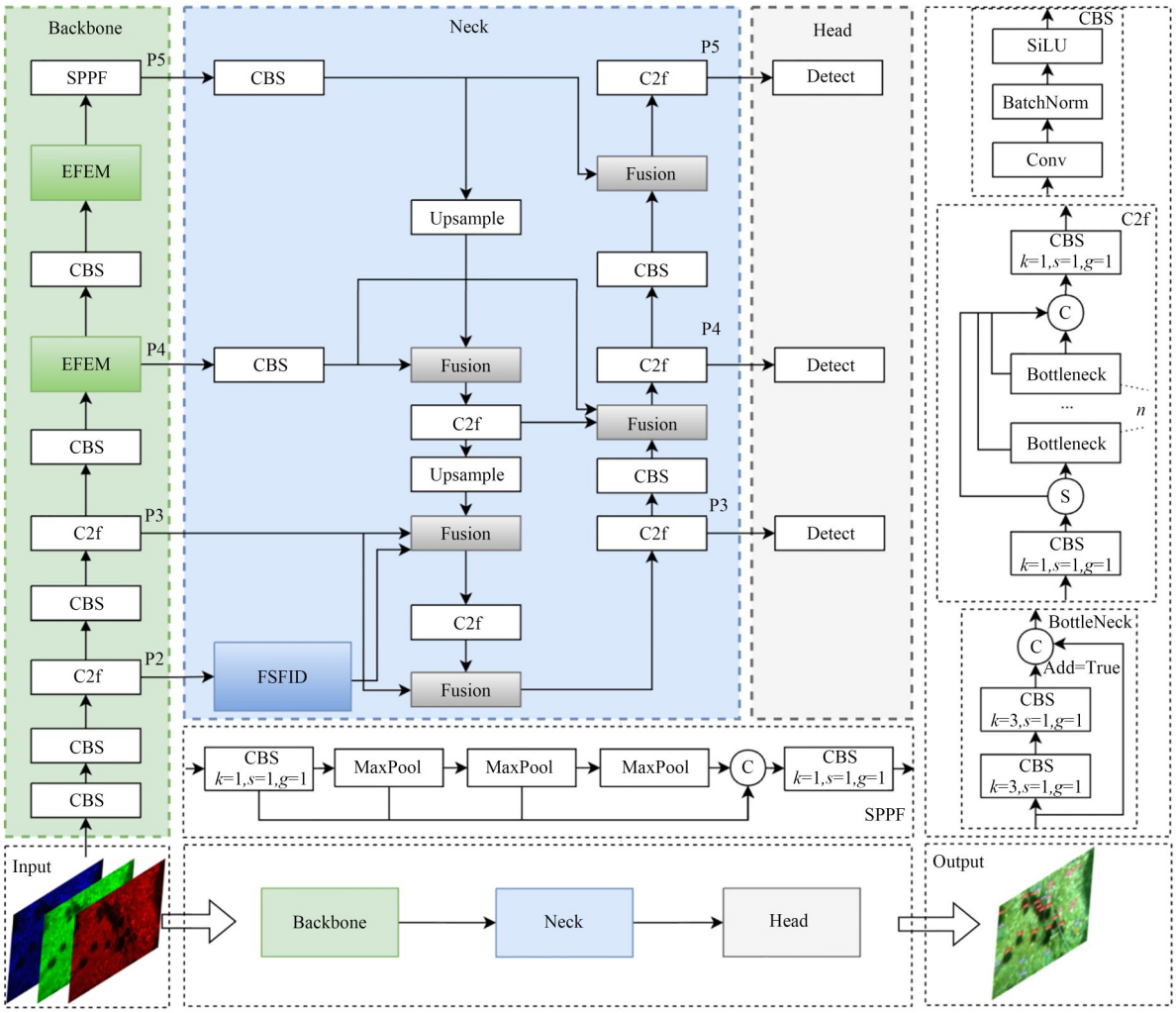


图1　YOLOv8n-SD 模型总体结构图

Fig. 1　Overall structure of the YOLOv8n-SD model

2.2　局部全局特征融合增强主干网络

受水体浑浊与光散射影响，水下图像普遍存在对比度低、纹理模糊等退化现象，而传统卷积由于感受野受限，难以有效提取全局特征。为此，本文提出局部全局融合特征增强主干网络，通过设计高效特征聚合模块替换原模型的C2f模块，实现局部卷积操作与全局自注意力计算的特征互补，减少计算量并显著提升特征表达能力。整体结构如图2所示，高效特征聚合模块主要由高效特征提取块与部分单头自注意力模块级联而成，其中k表示卷积核的大小，s表示卷积核滑动的步长，当$g=1$时表示该卷积为逐点卷积，当$g=C$时表示该卷积为深度卷积。

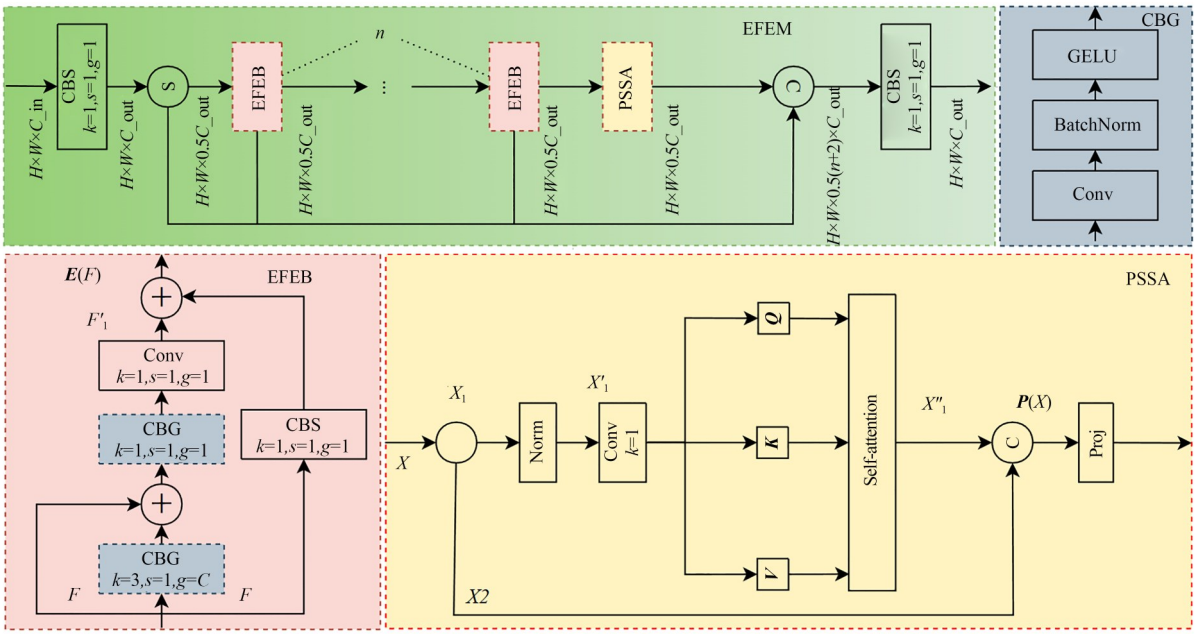


图 2　高效特征聚合模块结构图

Fig. 2　Structure of the efficient feature aggregation module

2.2.1　高效特征提取块

如图 1 所示，原有的 C2f 模块主要是由两个 3×3 卷积组成的多个 Bottleneck 瓶颈堆叠结构，因感受野固定而限制了主干网络的多尺度特征提取能力。为此，本文设计高效特征提取块，通过不同尺寸的卷积核，实现多尺度感受野融合与残差连接补偿，高效增强局部纹理特征信息。

对给定输入特征图 $F\in R^{H\times W\times C}$，$H$ 和 W 分别表示特征图的高度和宽度，C 为通道数。首先，进行初始特征压缩，再将通道平均划分为两条分支。主分支依次经过带有批量归一化和 GELU 激活函数增强的 3×3 深度卷积块 $\mathrm{CBG}_{3\times3}$ 和 1×1 逐点卷积块 $\mathrm{CBG}_{1\times1}$，共同组合成深度可分离卷积[17]。该设计通过先将通道数扩张至 $2C$ 再压缩回 C 的操作实现反瓶颈结构，能有效增强空间与通道之间的特征交互，并减少计算负载[18]。不同尺寸的卷积核在不同感受野范围下提取更多的局部细节，还能弥补水下目标因对比度低导致的特征区分度不足问题。随后，并行分支仅执行带有批量归一化和 SiLU 激活函数的 1×1 逐点卷积块 $\mathrm{CBS}_{1\times1}$，用于保留原始浅层细节。两条分支输出特征图逐元素相加并进行残差连接，得到局部增强特征图 $E(F)\in R^{H\times W\times C}$。可用公式表示为：

$$F_1'=\mathrm{PConv}_{1\times1}(\mathrm{CBG}_{1\times1}(\mathrm{CBG}_{3\times3}(F)+F))$$
$$E(F)=F_1'+\mathrm{CBS}_{1\times1}(F),\qquad(1)$$

其中，$\mathrm{PConv}_{1\times1}$ 为卷积核大小为 1×1 的逐点卷积。

2.2.2　部分单头自注意力模块

尽管多头自注意力机制能够建模全局依赖关系，但有研究指出该机制存在通道冗余现象，会大幅增加计算负载[19]。为此，本文设计部分单头自注意力模块建模长距离依赖，采用通道划分策略降低计算负载，实现高效的全局特征增强。

首先，对输入特征图 $E(F)\in R^{H\times W\times C}$ 的通道维度按比例 $r\in(0,1)$ 进行划分，得到特征图 $X_1\in R^{H\times W\times rC}$ 用于单头自注意力计算，未被划分的特征图 $X_2\in R^{H\times W\times(1-r)C}$ 用于保留局部增强后的特征信息。随后，对特征图 X_1 执行层归一化和 GELU 激活函数，得到预处理增强的特征图 X_1'。接着，通过 1×1 逐点卷积生成 Q，K 和 V 矩阵，并将空间维度 $H\times W$ 展平为序列维度，显著降低自注意力矩阵计算复杂度，得到矩阵 $Q,K\in R^{HW\times rC/2}$，$V\in R^{HW\times rC}$。随后，根据自注意力机制的公式，计算 Q 与 K 的相似度以生成自注意力权重，接着与 V 矩阵进行加权求和后，将维度空间重塑为 $H\times W$，得到全局特征增强后的特征图 $X_1''\in R^{H\times W\times rC}$。上述计算过程可用公式

所示：

$$\boldsymbol{X}_1'=\mathrm{GELU}\left(\mathrm{Norm}\left(\boldsymbol{X}_1\right)\right)$$

$$\boldsymbol{X}_1''=\mathrm{Reshape}\left(\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\cdot\boldsymbol{V}\right),\quad(2)$$

其中：d_k表示键向量维度。Reshape(.)表示维度重塑操作，将展平的序列特征重塑为原空间维度。接着，为了更好地将局部全局特征进行互补并增强融合效果，直接将保留局部特征增强后的特征图$\boldsymbol{X}_2$与全局特征增强后的特征图$\boldsymbol{X}_1''$沿通道维度拼接，并进入1×1逐点卷积投影层，输出局部与全局特征融合增强后的特征图$\boldsymbol{P}(X)\in\boldsymbol{R}^{H\times W\times C}$。以上操作可用公式表示为：

$$\boldsymbol{P}(X)=\mathrm{PConv}_{1\times1}\left(\mathrm{Concat}\left(\boldsymbol{X}_1'',\boldsymbol{X}_2\right)\right),\quad(3)$$

其中：Concat(.)是通道拼接操作。随后，在高效特征聚合模块中，连续采用了多个高效特征提取块与一个部分单头自注意力模块后，本文通过保留原C2f模块的瓶颈堆叠结构，将不同层级的特征图再次进行拼接融合输出，进一步融合多尺度特征。

总体而言，高效特征提取块采用反瓶颈结构和不同大小的卷积核，在不同空间感受野范围下提取了更多的局部纹理特征信息。部分单头自注意力模块，则利用通道比例划分策略，对部分通道进行单头自注意力求解，减少通道冗余现象并高效求取了长程全局信息，实现局部与全局的特征互补，缓解了水下弱特征难以有效捕捉的问题。因此，本文构建的主干网络，能够在水下复杂环境下高效提取有效特征，从而提高检测精度。

2.3　频域空域高效融合颈部网络

原YOLOv8n颈部的路径聚合网络，采用固定权重的邻层相加融合，致使浅层丰富的小目标特征信息在自顶向下传播时被深层语义逐步稀释，且缺乏对水下复杂背景的动态适应。为此，本文构建了频域空域高效融合颈部网络，如图1所示。

首先，颈部网络对于来自不同层级的输入特征图集，通过使用包含1×1逐点卷积的CBS块，将P4，P5层的通道数统一为P3层的通道数，实现跨越主干网络的横向连接；然后，对统一通道数的特征图与主干网络输出特征图，进行自下而上的特征融合。为缓解原YOLOv8n颈部浅层特征的稀释问题，本文采用快速归一化加权策略，通过可学习的动态权重在训练中自动调整特征贡献度，使重要特征图在多尺度融合中获得更高权重，从而实现高效融合。其公式可以表述为：

$$w_i'=\mathrm{ReLU}\left(\theta\left(w_i-\eta\cdot\nabla_\theta\mathcal{L}\right)\right)$$

$$O=\frac{\sum w_i'\cdot I_i}{\sum w_i'+\epsilon},\quad(4)$$

其中：θ为权重参数集合，η为学习率，$\nabla_\theta\mathcal{L}$为权重参数对应的损失函数梯度。w_i为原始特征权重，初始权重值区间为U(0，0.1)的均匀分布，避免过大导致部分特征主导融合过程，同时引导模型快速捕捉特征差异，提升训练收敛效率。w_i'为ReLU激活后的非负权重，通过反向传播与整体检测损失进行梯度更新，且ReLU激活可过滤负权重，确保权重的有效性。O为融合特征图。I_i为输入特征图。ε=0.000 1确保数值稳定性。

其次，本文在颈部网络对主干网络的浅层高分辨率P2层进行下采样，设计了频域空域信息融合下采样模块，保留并融合了水下目标检测关键的高频纹理和空间结构信息。最后，优化颈部网络结构，将频域空域信息融合下采样后的特征图直接加入P3层融合路径，还使用双向融合节点复用机制，以较小参数增量，提高多尺度特征融合效率，进而提高后续的小目标检测层的特征判别能力。

传统跨步卷积通过舍弃部分像素实现下采样，易导致高频分量混叠丢失与空间结构破坏，水下小目标本身特征微弱，信息丢失会直接导致检测精度下降。为此，本文设计频域空域信息融合下采样模块，将给定输入特征图$\boldsymbol{X}\in\boldsymbol{R}^{H\times W\times C}$拆分为两个子特征图$\boldsymbol{X}_1,\boldsymbol{X}_2\in\boldsymbol{R}^{H\times W\times C/2}$。其中，特征图$\boldsymbol{X}_1$利用Haar小波变换[20]提取频域信息，特征图$\boldsymbol{X}_2$则通过空间到深度变换[21]保留空间域信息。

(1) Haar小波变换下采样。该分支利用低通滤波器H_0对应系数向量$h=1/\sqrt{2}\ [1,-1]$，对相邻像素的加权平均提取水下图像特征图的低频轮廓信息。高通滤波器H_1则对应向量系数$l=1/\sqrt{2}[1,-1]$，通过相邻像素的差值运算捕捉高频细节信息。随后利用二维Haar小波变换将$\boldsymbol{X}_1$分解为多尺度的频域分量，保留轮廓、全局结构低频特征$\boldsymbol{D}_{ll}$和包含边缘、纹理特征的高频

特征 $\boldsymbol{D}_{lh}$，$\boldsymbol{D}_{hl}$，$\boldsymbol{D}_{hh}$。其变换公式如式(5)：

$$\boldsymbol{D}_{H_aH_b}=\left(\downarrow 2\right)\begin{bmatrix} H_b*(Z_2) \\ \left(\downarrow 2\right)\left(H_a*(Z_1)*X_1\right)\end{bmatrix}, \quad (5)$$

其中：$a,b\in\{0,1\}$。$*$表示卷积运算操作。Z_1，Z_2分别表示对行、列卷积运算。$\downarrow 2$为2倍下采样操作，实现频域分量分解。Haar小波变换下采样后，拼接各频域分量，得到保留频域信息的特征图 $\boldsymbol{X}_1'\in\boldsymbol{R}^{H/2\times W/2\times 2C}$。

(2) 空间到深度变换下采样。该分支则通过将空间坐标重新映射到通道维度来保持空间结构信息不变。坐标映射公式可以表示为：

$$f_{\text{out}}(i,j,c)=f_{\text{in}}\left(\left\lfloor\frac{i}{s}\right\rfloor\times s+\delta_i,\left\lfloor\frac{j}{s}\right\rfloor\times s+\delta_j,c\right), \quad (6)$$

其中：i和j表示空间坐标。s为缩放因子。c表示通道索引。δ_x和δ_y为偏移量。本文设置$s=2$时，特征图生成四个子块并沿通道维度拼接得到保留空间域信息的特征图 $\boldsymbol{X}_2'\in\boldsymbol{R}^{H/2\times W/2\times 2C}$。

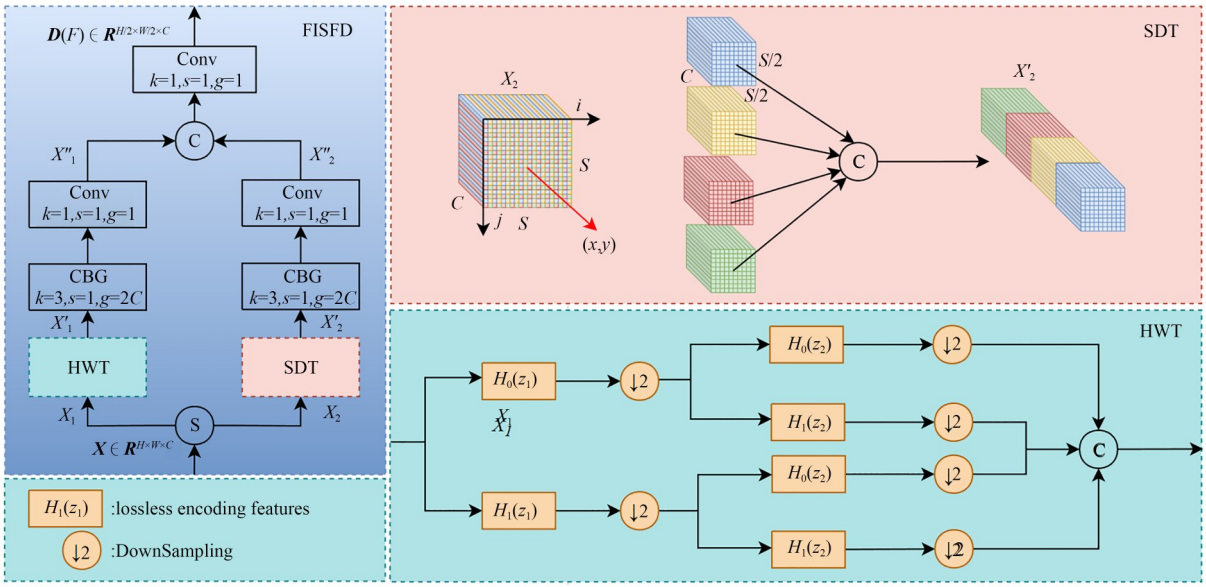


图3　频域空域信息融合下采样模块结构图

Fig. 3　Structure of the frequency-spatial information fusion downsampling module

为更好地融合频域和空域中重要的信息，本文在后续的卷积层中，对无损下采样后的特征图 $\boldsymbol{X}_1'$ 和 $\boldsymbol{X}_2'$ 依次经过包含3×3的深度卷积块 $\text{CBG}_{3\times3}$ 块和1×1逐点卷积，组成深度可分离卷积高效提取特征信息，并将通道数压缩至C，减少通道数增加至4倍所带来的计算负载。随后，沿通道维度拼接特征图 $\boldsymbol{X}_1''$ 和 $\boldsymbol{X}_2''$，通道扩展至$2C$，再经1×1卷积压缩至C，得到融合频域与空域信息的下采样特征图 $\boldsymbol{D}(X)\in\boldsymbol{R}^{H/2\times W/2\times C}$。以上操作可用公式表示为：

$$\boldsymbol{X}_i''=\text{PConv}_{1\times1}(\text{CBG}_{3\times3}(\boldsymbol{X}_i'))$$

$$\boldsymbol{D}(\boldsymbol{X})=\text{PConv}_{1\times1}(\text{Concat}(\boldsymbol{X}_1'',\boldsymbol{X}_2'')), \quad (7)$$

其中：$\boldsymbol{X}_i''$，$\boldsymbol{X}_i'$中的$i\in(1,2)$。综上所述，频域空域信息融合下采样模块直接对主干网络的高分辨率浅层特征图下采样，结合Haar小波频域分解和空间坐标重映射实现无损下采样，在后续非跨步卷积层中融合频域中的高频信息和空域中的空间结构信息。随后，再与构建的频域空域高效融合颈部网络的深层特征进行多尺度特征融合，为后续检测层提供更多复杂水下场景中所需的浅层细粒度细节，从而提高后续检测头对小目标检测层的特征判别能力。

3　实验结果与分析

本节首先介绍实验所选用的数据集、机器配置和参数设置，通过模块对比实验和消融实验，验证了网络模型中设计的子模块均对检测性能的提升有所贡献。最后，将本文算法与其他通用水下目标检测算法在选定数据集上进行实验对

比，证明了本文方法的有效性和高效性。

3.1　实验数据集与参数设定

本文选用URPC2020与RUOD两个公开水下数据集验证所提模型的有效性。其中，URPC2020数据集包含海参、海胆、海星、扇贝四类目标，共7 383幅图像，按7∶2∶1比例划分为训练集、验证集与测试集；为验证模型在复杂背景与尺度变化下的鲁棒性，本文对RUOD数据集进行筛选，筛除大目标类别后仅保留上述四类小目标，共4 268幅图像，同样按7∶2∶1比例划分数据集。

实验基于PyTorch 2.1.2深度学习框架和Windows 11 23H2系统。硬件配置为Intel i5-12490F CPU，32 GB内存及NVIDIA RTX 4070 Ti Super 16 GB显卡。所有模型均在相同环境下训练，输入图像尺寸为640×640，训练轮次为250，批量大小设为16，早停轮数设为20。优化器采用SGD，动量设为0.973，初始学习率0.01，权重衰减系数0.000 5。训练中采用Mosaic数据增强策略，并通过余弦学习率控制损失衰减。

为评估模型的检测性能，选取步长为0.05和交并比阈值从0.5到0.9的平均$mAP_{0.5:0.95}$、交并比阈值固定为0.5的平均精度均值mAP_{50}、计算负载GFLOPs、以百万为单位的参数量Param和帧率FPS作为关键评价指标。这些指标反映了模型在准确性、运算效率、实用性和部署可行性方面的综合表现。mAP表示不同召回率水平下的平均精度，其公式如下所示：

$$Precision=\frac{T_P}{T_P+F_P}, \tag{8}$$

$$Recall=\frac{T_P}{T_P+F_N}, \tag{9}$$

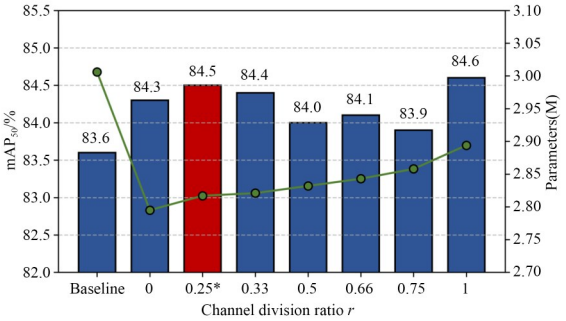
$$mAP_{0.5:0.95}=\frac{1}{N}\sum_{i=1}^{N}AP_i, \tag{10}$$

其中：T_P表示真正例，F_P表示假正例，F_N表示假负例。N为数据集中目标类别的总数。

3.2　主干网络模块对比实验

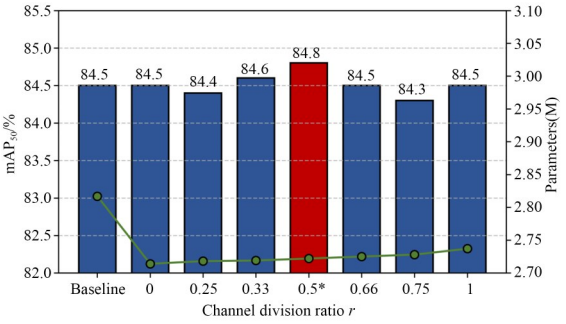
为系统评估局部全局特征融合主干网络的有效性，本文依次对主干网络中P4和P5特征层的高效特征聚合模块进行对比实验，验证不同通道比例r在URPC2020数据集上的性能表现。

实验结果如图4所示。当通道比例r=0时，表示不进行单头自注意力计算，仅使用高效特征提取块，参数量从基准模型的3.01 M降至2.85 M，还实现了0.7%的mAP_{50}精度提升。继续按通道比例动态划分，实验首先确定P4层的r=0.25时，mAP_{50}达84.5%，检测性能较优。再根据P4层的实验结果作为基准模型，测得P5层的r=0.5时，mAP_{50}达84.8%，检测性能较优。此时的改进主干网络mAP_{50}较基准模型提升了1.2%，同时参数量显著下降。通过观察其他通道比例r下的表现，可以发现当通道比例过大时，单头自注意力机制的通道冗余现象严重，模型计算量增加且泛化能力下降，导致模型精度下降；当通道比例过小时，则会丢失部分全局特征信息，导致精度提升不足。而不同层级的通道数不同，出现通道冗余的现象的通道数也不一致。因此，本文通过对不同层级进行动态的通道划分，可以有效减少自注意力机制带来的通道冗余现



(a) P4特征层不同通道比例r的对比实验

(a) Comparative experiments of different channel ratios r in the P4 feature layer



(b) P5特征层不同通道比例r的对比实验

(b) Comparative experiments of different channel ratios r in the P5 feature layer

图4　URPC2020数据集下的主干网络高效特征聚合模块对比实验

Fig. 4　Comparative experiments of the efficient feature aggregation module in the backbone network on the URPC2020 dataset

象，还能充分建模水下目标的长距离空间依赖关系。故本文设计的局部全局特征融合主干网络，可以充分提取局部和全局特征，在精度提升和计算效率之间能实现优秀平衡。

3.3　颈部网络模块对比实验

为系统评估高效颈部网络设计的有效性，本文在URPC2020水下目标检测数据集上开展消融实验。实验以YOLOv8n为基准模型，记为B，本文设计带有快速归一化加权策略的改进颈部网络结构，记为F。为进一步明确本文提出的频域空域信息融合下采样模块FIFSD模块有效性，实验将与使用传统跨步卷积下采样Conv，Haar小波变换下采样（Haar Wavelet Downsampling，HWD）与空间到深度变换下采样（Space-to-Depth，SPD）进行对比实验验证，其他实验配置保持一致。

表1　URPC2020数据集上的颈部模块对比实验

Tab. 1　Comparative experiments of neck modules on URPC2020 dataset

方法	URPC2020		GFLOPs	Param/M	FPS
	$mAP_{0.5:0.95}$/%	mAP_{50}/%			
B	49.2	83.6	8.1	3.01	258
F+Conv	49.7	84.3	**7.1**	**1.99**	222
F+HWD	**50.3**	**85.1**	7.8	2.05	204
F+SPD	49.9	84.8	7.8	2.05	214
F+FIFSD（本文）	**50.4**	**85.0**	**7.1**	**1.99**	217

由表1数据可知，在基础模型引入改进的颈部网络结构并针对P2层使用1×1传统跨步卷积下采样Conv后，$mAP_{0.5:0.95}$提升至49.7%，mAP_{50}提升至84.3%，同时计算量降至7.1 GFLOPs、参数量压缩至1.99 M，验证了颈部网络结构优化以及快速归一化加权策略，优化多尺度特征融合效率的优势。而在使用小波变换下采样模块HWD与空间到深度下采样模块SPD替换传统跨步卷积下采样的方法中，HWD使$mAP_{0.5:0.95}$进一步提高至50.3%，mAP_{50}达85.1%；SPD的$mAP_{0.5:0.95}$达49.9%，mAP_{50}达84.8%。这说明了水下目标检测中频域和空域的重要性，都显著增加模型的检测精度。但由于HWD与SPD下采样方法都将全部通道数增至4倍，计算量GFLOPs均从7.1显著增至7.8，参数量升至2.05 M，资源消耗增加且推理速度下降明显，不利于水下轻量化部署。而本文提出的FIFSD模块采用双分支融合设计，利用深度可分离卷积高效提取频域与空域特征并实现互补融合，进而提取有效的高频信息和纹理信息。该模块还通过逐步减低通道数的方式，使其保持7.1 GFLOPs计算量与1.99 M参数量的前提下，$mAP_{0.5:0.95}$达到了50.4%，mAP_{50}达到85.0%的优异性能，推理速度则维持在217 FPS。因而，本文构建的颈部网络能在保持轻量化设计的基础上，有效增强对水下目标的多尺度特征表达能力，进而提高后续检测头的特征判别能力，更加适配于水下复杂检测场景。

3.4　模型消融实验

为验证本文方法的局部全局特征融合主干网络与频域空域高效融合颈部网络两大核心模块对检测性能和效率的贡献，在URPC2020数据集上开展消融实验，以未引入任何改进的轻量级YOLOv8n为基线模型，所有实验保持相同硬件环境，不同方法的实验结果具体如表2所示。

基线模型YOLOv8n在URPC2020数据集上为$mAP_{0.5:0.95}$为49.2%，mAP_{50}为83.6%，计算量8.1 GFLOPs、参数量3.01 M、推理速度258 FPS。当仅引入局部全局特征融合主干网络时，该模块通过部分单头自注意力机制高效建模长程空间依赖、高效特征提取块提取局部纹理特征，实现全局与局部特征互补增强，其$mAP_{0.5:0.95}$提升至49.7%，mAP_{50}提升至84.8%。同时对比资源消耗优化，计算量降至7.6 GFLOPs，参数量降至2.72 M，推理速度为222 FPS，验证了该主干网络模块的有效性；当仅引入频域空域高效融合颈部网络时，$mAP_{0.5:0.95}$和mAP_{50}分别为50.4%，85.0%，且计算量降至7.1 GFLOPs，参

表 2　URPC2020 数据集的消融实验

Tab. 2　Ablation Experiments on the URPC2020 Dataset

局部全局特征融合主干网络	频域空域高效融合颈部网络	URPC2020		GFLOPs	Param/M	FPS
		$mAP_{0.5:0.95}$/%	mAP_{50}/%			
×	×	49.2	83.6	8.1	3.01	258
√	×	49.7	84.8	7.6	2.72	222
×	√	50.4	85.0	7.1	1.99	217
√	√	**51.2**	**85.7**	**6.7**	**1.73**	198

数量降至 1.99 M，推理速度为 217 FPS。当两大模块协同引入时，YOLOv8n-SD 实现了精度与效率的优秀平衡，$mAP_{0.5:0.95}$ 达 51.2%，mAP_{50} 达 85.7%，计算量进一步降至 6.7 GFLOPs，参数量降至 1.73 M，推理速度保持在 198 FPS，满足水下实时检测性能要求。综上所述，YOLOv8n-SD 中两个模块均能独立提升检测性能，且二者协同能进一步提升模型在水下场景中的检测精度。

图 5 展示了从 URPC2020 数据集选取的四张典型样本，采用 Grad-CAM[22] 方法对表 1 中四种消融实验模型，分别生成颈部网络末端的特征热力图。从可视化结果可见，仅引入改进颈部网络的热力图，特征聚焦能覆盖更多海胆且范围广，但对边缘模糊目标的特征捕捉不足；仅引入改进主干网络的模型热力图，较未改进模型对目标核心区域的聚焦更集中，但总体覆盖目标较少。而采用两项协同改进的本文模型 YOLOv8n-SD 的热力图清晰显示，其特征在密集遮挡区域能覆盖更多海胆和扇贝，且较完整勾勒出海星整体轮廓，特征响应更精准、抗干扰能力更

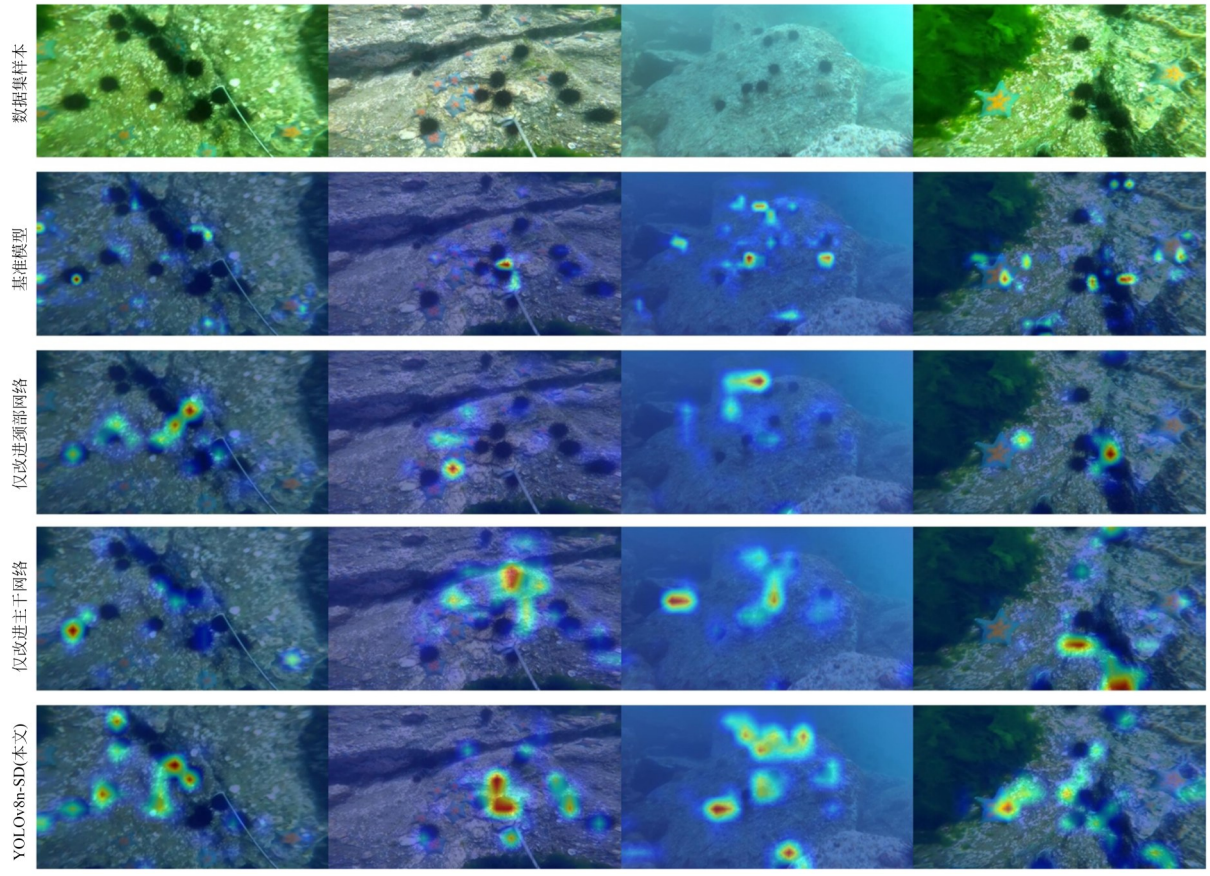


图 5　URPC2020 数据集下模型的 Grad-CAM 可视化对比图

Fig. 5　Grad-CAM visualization of models on the URPC2020 dataset

强。这足以说明本文方法能有效图像色彩偏移、密集遮挡目标等复杂水下场景。

3.5　不同模型对比实验

为验证本文方法的水下目标检测效果，选取经典检测模型 Faster-RCNN[23]，DETR-DC5[24]，SSD[25]，PANet[26]；主流 YOLO 系列检测模型 YOLOv3tiny[27]，YOLOv5n[28]，YOLOv8n/s/m，YOLOv10n[29]，YOLOv11n[30]；以及其他 YOLO 改进模型 CEH-YOLO[5]，YOLOv8-LA[6]进行对比实验，其余实验设置与环境保持一致。

图 6 展示了 URPC2020 数据集上各模型的 mAP_{50} 训练曲线，可直观呈现不同模型在训练过程中的性能差异。在训练轮次小于 50 时，本文方法 YOLOv8n-SD 模型收敛速度较快。当训练轮次大于 150 时，YOLOv8n-SD 较基线模型 YOLOv8n 的 mAP_{50} 高 1.5%～2.1%。当 YOLOv8n-SD 完全收敛时，其 mAP_{50} 接近于中、大模型版本 YOLOv8m 和 YOLOv8s，初步验证了本文模型的高效性。

对比实验结果如表 3 所示。可以进一步清晰地观察不同模型在 URPC2020 和 RUOD 验证集上的检测精度、计算成本与推理速度间的差异。本文方法 YOLOv8n-SD 在轻量级模型中表现突出，URPC2020 数据集上 $mAP_{0.5:0.95}$ 和 mAP_{50} 分别达到 51.2% 和 85.7%，在 RUOD 数据集上 $mAP_{0.5:0.95}$ 和 mAP_{50} 分别为 50.6% 和 85.0%。

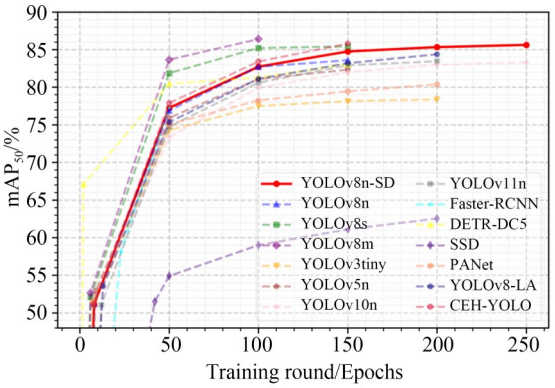


图 6　URPC2020 数据集上模型的 mAP_{50} 训练曲线对比图

Fig. 6　Comparison of mAP_{50} training curves of different models on URPC2020 dataset

表 3　在 URPC2020 和 RUOD 数据集上的不同模型对比实验

Tab. 3　Comparative experiments of different models on URPC2020 and RUOD datasets

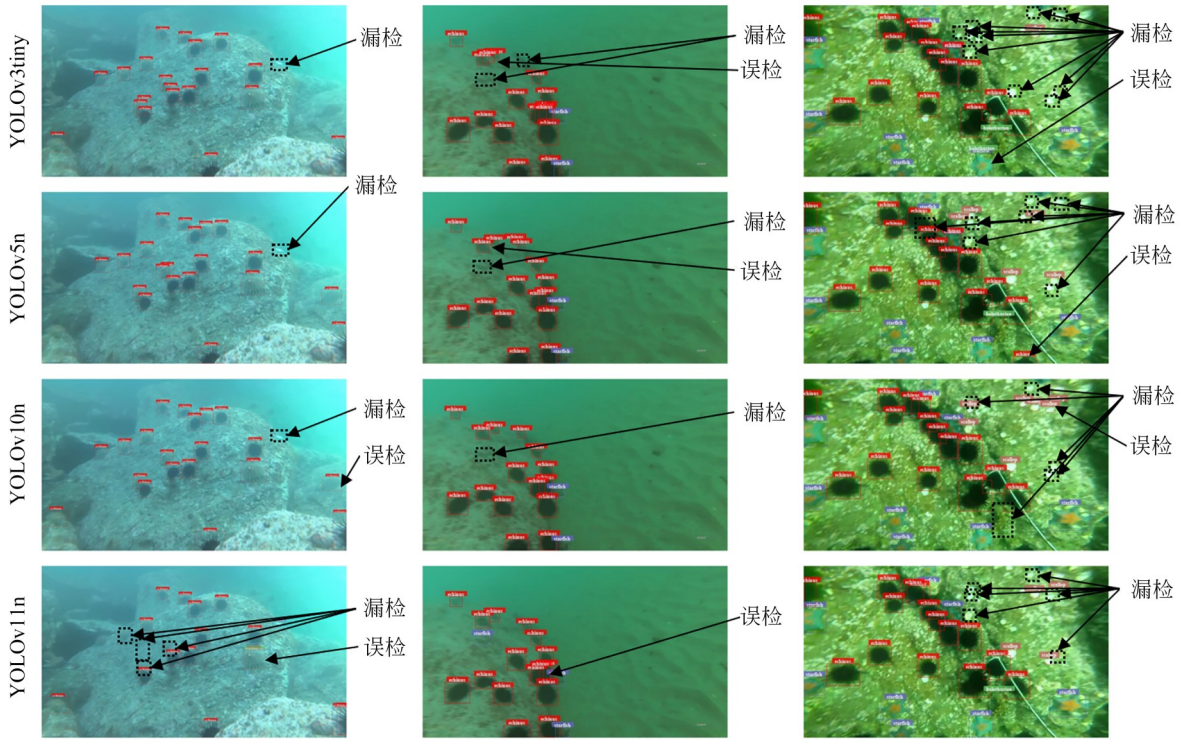
方法	URPC2020		RUOD		GFLOPs	Param/M	FPS
	$mAP_{0.5:0.95}$/%	mAP_{50}/%	$mAP_{0.5:0.95}$/%	mAP_{50}/%			
Faster-RCNN	39.2	74.6	38.4	70.4	369.8	136.73	60
DETR-DC5	45.7	83.4	44.6	79.2	225.0	60.22	58
SSD	34.4	62.3	32.0	60.1	61.0	24.01	98
PANet	48.6	80.5	48.2	78.5	6.5	31.63	125
YOLOv8-LA	50.2	84.7	49.6	84.5	7.5	**2.42**	179
CEH-YOLO	**51.5**	**86.3**	**51.0**	**85.3**	11.6	4.40	154
YOLOv3tiny	39.5	79.4	38.5	76.6	12.9	8.67	156
YOLOv5n	47.0	82.2	46.6	80.0	**4.1**	1.76	236
YOLOv10n	50.2	83.4	49.1	82.5	**6.5**	2.27	196
YOLOv11n	49.4	83.2	48.8	83.4	**6.3**	2.58	188
YOLOv8m	**51.9**	**87.5**	**51.4**	**86.5**	78.7	25.84	88
YOLOv8s	**51.1**	**86.2**	**50.8**	**85.8**	11.2	11.13	155
YOLOv8n	49.2	83.6	48.9	83.4	8.1	3.01	258
YOLOv8n-SD(本文)	**51.2**	**85.7**	**50.6**	**85.0**	**6.7**	**1.73**	198

在 URPC2020 数据集上，二阶段检测模型 Faster-RCNN 的 $mAP_{0.5:0.95}$ 和 mAP_{50} 分别达到了 39.2% 和 74.6%，但由于其区域提议网络的冗余候选生成机制，导致计算成本高达 369.8 GFLOPs，推理速度仅 60 FPS；基于 Transformer 进行全局建模的检测模型 DETR-DC5，受限于多

头自注意力机制带来的高计算复杂度，推理速度也仅 58 FPS，较 YOLOv8n-SD 在 $mAP_{0.5:0.95}$ 和 mAP_{50} 指标上分别低出 5.5% 和 2.3%。作为经典单阶段检测模型的 SSD，其 $mAP_{0.5:0.95}$ 和 mAP_{50} 仅分别为 34.4% 和 62.3%，精度表现较差。PA-Net 以 6.5 GFLOPs 达到 80.5% 的 mAP_{50}，但因其采用固定权重融合方法，多尺度特征融合效率不足，难以适应水下复杂场景。近年来的水下目标检测改进模型中，YOLOv8-LA 通过选择性通道处理与轻量级上采样算子平衡精度与效率，其在 URPC2020 数据集上的 $mAP_{0.5:0.95}$ 和 mAP_{50} 分别为 50.2% 和 84.7%，但参数量达 2.42 M 较 YOLOv8n-SD 多 40%，计算量 7.5 GFLOPs 高出 11.9%，推理速度 179 FPS 慢于本文模型，整体检测效率与精度不及 YOLOv8n-SD；CEH-YOLO 模型则通过高阶可变形注意力与增强型空间金字塔池化模块强化小目标特征提取，其 $mAP_{0.5:0.95}$ 51.5% 和 mAP_{50} 86.3% 略高于本文模型，但计算量激增至 11.6 GFLOPs，参数量达 4.40 M，推理速度仅 154 FPS，牺牲较多实时检测性能，难以适配水下嵌入式设备需求。主流 YOLO 系列模型中，YOLOv8n-SD 新推出的 YOLOv10n 和嵌入轻量级多头自注意力机制的 YOLOv11n，参数量和计算负载相近，但 $mAP_{0.5:0.95}$ 分别高出了 1.0% 和 1.8%，mAP_{50} 分别高出了 2.3% 和 2.5%，在轻量化模型中表现优异；与 YOLOv8 的中大模型版本的 YOLOv8s 和 YOLOv8m 比较，YOLOv8n-SD 综合表现优异，以较小的参数量和计算量达到较为接近的检测精度表现，推理速度则显著优于二者。

在 RUOD 数据集上，YOLOv8n-SD 的优势进一步凸显，其检测精度与资源效率的平衡优势尤为突出。相较于 Faster-RCNN 的 $mAP_{0.5:0.95}$ 38.4%，mAP_{50} 70.4%，YOLOv8n-SD 两项指标分别高出 12.2%，14.6%；较 DETR-DC5 的 $mAP_{0.5:0.95}$ 44.6%，mAP_{50} 79.2%，则分别高出 6.0%，5.8%；与水下场景优化的改进模型 YOLOv8-LA 比较，在 $mAP_{0.5:0.95}$ 和 mAP_{50} 指标精度分别高 1.0%，0.5%；较 CEH-YOLO $mAP_{0.5:0.95}$ 51.0%，mAP_{50} 85.3%，精度略低 0.4%，0.3%；较基准模型 YOLOv8n，mAP0.5:0.95 和 mAP50 分别提升 1.7%，1.6%。综合而言，本文方法 YOLOv8n-SD 通过两大核心模块的协同作用，实现了更优秀的检测精度与效率平衡表现，在不同数据集上具有良好的鲁棒性以及对水下嵌入式设备部署需求的适配性。

图 7 则展示了不同模型在 RUOD 数据集下的复杂水下场景检测结果，标注出漏检和误检情



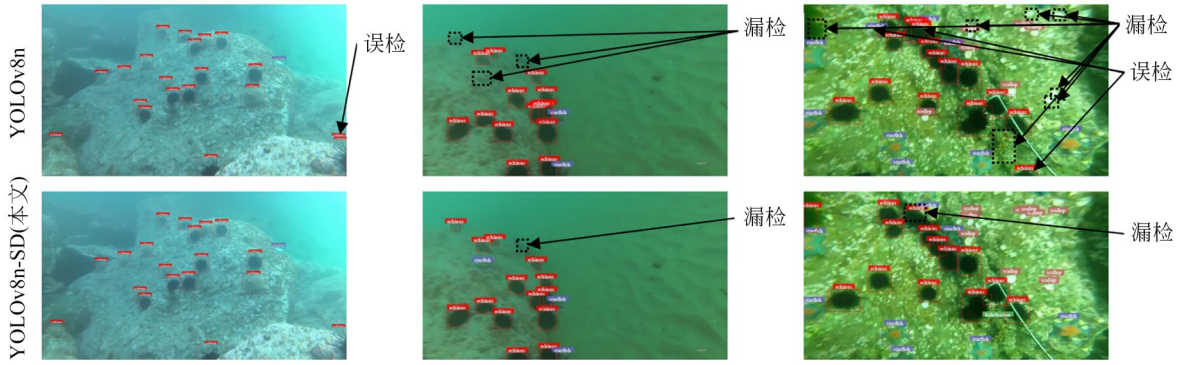


图 7　不同方法在 RUOD 数据集上的检测结果对比图

Fig. 7　Detection result comparison of different methods on RUOD dataset

况。在主流 YOLO 系列轻量级模型中，如 YOLOv3tiny，YOLOv5n 由于主干网络特征提取能力较弱，在第三列的水体散射导致边界模糊、小目标密集遮挡场景时，未能检测出小尺度扇贝与背景岩石的区别，导致出现大量漏检误检；较新的 YOLOv10n，YOLOv11n 则在第一列的颜色偏移中未能识别隐藏在边缘的海参，在第二列的模糊、多尺度密集场景中则未能有效检出远处模糊的海胆，说明其颈部网络在水下复杂环境中的多尺度有效特征提取能力较弱，导致漏检误检情况严重。而本文方法 YOLOv8n-SD 可以清晰地看出在水下密集目标的边界框定位更精准，且无明显误检，仅存在少量漏检情况。综上所述，YOLOv8n-SD 通过构建局部全局特征融合增强主干网络与频域空域高效融合颈部网络，能显著提升颜色偏移、图像模糊等复杂水下场景的正确检出率，进一步验证了本方法的有效性和鲁棒性。

4　结　论

本文针对水下复合检测难题与水下探测设备资源受限导致的检测性能及效率不佳问题，提出融合单头自注意力与频域-空域特征的改进模型 YOLOv8n-SD。首先，设计局部全局特征融合增强主干网络，通过部分单头自注意力建模长程空间信息，结合高效特征提取块提取的局部纹理细节，实现了局部与全局特征的互补融合，显著强化主干网络在复杂水下场景的特征表征能力。此外，构建频域空域高效融合颈部网络，设计频域空域信息融合下采样模块，利用 Haar 小波变换和空间到深度变换，对浅层特征图进行无损下采样，并融合提取重要的高频与空域信息。还通过快速归一化加权策略，优化多尺度融合效率，进一步提高了水下模糊、低光照场景下的检测精度。实验结果表明，YOLOv8n-SD 在公开水下数据集 URPC2020 和 RUOD 上表现优异，保持 198 FPS 实时检测速度的同时，$mAP_{0.5:0.95}$ 和 mAP_{50}，分别达到了 51.2%，85.7% 和 50.6%，85.0%。此外，参数量仅 173 万，计算量仅 6.7 GFLOPs。这说明 YOLOv8n-SD 模型足以适配小型水下机器人、便携式探测设备等资源受限平台，为水下资源勘探与生态保护提供了兼顾实用性与部署性的新方法。最后，为进一步降低漏检率，后续研究将采用更敏感的损失函数，优化水下目标边框定位精度，持续提升模型在极端复杂环境下的稳定性。

作者贡献声明：

李大海：方法的整体构思和设计，；

廖嘉伟：论文撰写和实验实施；

王振东：提供实验条件和审核论文。

参考文献：

[1] COSTELLO M J. Biodiversity: the known, unknown, and rates of extinction[J]. *Current Biology*, 2015, 25(9): R368-R371.

[2] HALPERN B S, LONGO C, HARDY D, *et al*. An index to assess the health and benefits of the global ocean [J]. *Nature*, 2012, 488 (7413):

615-620.

[3] HU J, LI D L, DUAN Q L, *et al*. Fish species classification by color, texture and multi-class support vector machine using computer vision[J]. *Computers and Electronics in Agriculture*, 2012, 88: 133-140.

[4] CHUANG M C, HWANG J N, WILLIAMS K. A feature learning and object recognition framework for underwater fish images[J]. *IEEE Transactions on Image Processing*, 2016, 25(4): 1862-1872.

[5] FENG J F, JIN T. CEH-YOLO: a composite enhanced YOLO-based model for underwater object detection[J]. *Ecological Informatics*, 2024, 82: 102758.

[6] QU S M, CUI C, DUAN J L, *et al*. Underwater small target detection under YOLOv8-LA model [J]. *Scientific Reports*, 2024, 14: 16108.

[7] 何梦云, 何自芬, 张印辉, 等. 用于水下声呐目标检测的弱特征共焦通道调控方法[J]. 中国光学(中英文), 2024, 17(6): 1281-1296.
HE M Y, HE Z F, ZHANG Y H, *et al*. Weak feature confocal channel regulation for underwater sonar target detection[J]. *Chinese Optics*, 2024, 17(6): 1281-1296. (in Chinese)

[8] 李文彪, 陶洋, 董源, 等. 基于颜色先验引导和注意力机制的水下图像增强[J]. 液晶与显示, 2025, 40(8): 1163-1176.
LI W B, TAO Y, DONG Y, *et al*. Underwater image enhancement based on color prior guidance and attention mechanism[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(8): 1163-1176. (in Chinese)

[9] 陶洋, 龚霁霆, 周立群. 基于Retinex的可变注意力低照度水下图像增强[J]. 液晶与显示, 2025, 40(3): 481-492.
TAO Y, GONG J T, ZHOU L Q. Variable attention low illumination underwater image enhancement based on Retinex[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(3): 481-492. (in Chinese)

[10] GAO J X, ZHANG Y H, GENG X, *et al*. PE-Transformer: Path enhanced transformer for improving underwater object detection[J]. *Expert Systems with Applications*, 2024, 246: 123253.

[11] 姚婷婷, 李宁, 张煜. 感知增强混合网络的水下目标检测[J]. 光学 精密工程, 2025, 33(8): 1303-1312.
YAO T T, LI N, ZHANG Y. Perception enhanced hybrid network for underwater object detection[J]. *Opt. Precision Eng.*, 2025, 33(8): 1303-1312. (in Chinese)

[12] LIU J, LIU S, XU S J, *et al*. Two-stage underwater object detection network using swin transformer[J]. *IEEE Access*, 2022, 10: 117235-117247.

[13] 李瑜辉, 崔慧霞, 李耀敏, 等. 基于轻量化门控卷积网络的实时Transformer水下目标检测方法[J]. 水下无人系统学报, 2025, 33(2): 229-237.
LI Y H, CUI H X, LI Y M, *et al*. Real-time transformer detection of underwater objects based on lightweight gated convolutional network[J]. *Journal of Unmanned Undersea Systems*, 2025, 33(2): 229-237. (in Chinese)

[14] 张天, 温显斌, 薛彦兵, 等. 基于频域注意力的水下目标检测算法研究[J]. 光电子·激光, 2024, 35(6): 604-611.
ZHANG T, WEN X B, XUE Y B, *et al*. Research on detection algorithm for underwater object based on frequency domain attention[J]. *Journal of Optoelectronics·Laser*, 2024, 35(6): 604-611. (in Chinese)

[15] 韩丽, 马春海, 林志浩, 等. 一种用于低分辨率小目标的水下垃圾检测算法[J]. 科学技术与工程, 2024, 24(35): 15126-15136.
HAN L, MA C H, LIN Z H, *et al*. Underwater trash detection algorithm for low-resolution small targets[J]. *Science Technology and Engineering*, 2024, 24(35): 15126-15136. (in Chinese)

[16] JOCHER G, CHAURASIA A, QIU J. Ultralytics YOLO (Version 8.0.0)[CP]. 2023[2024-10-01]. https://github.com/ultralytics/ultralytics.

[17] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *July* 21-26, 2017, *Honolulu, HI, USA*. IEEE, 2017: 1800-1807.

[18] SANDLER M, HOWARD A, ZHU M L, *et al*. MobileNetV2: inverted residuals and linear bottlenecks[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. June* 18-23, 2018, *Salt Lake City, UT, USA*. IEEE, 2018: 4510-4520.

[19] VASWANI A, SHAZEER N, PARMAR N, *et al*. Attention is all you need[C]. *Advances in*

Neural Information Processing Systems. 2017, 30.

[20] XU G P, LIAO W T, ZHANG X, *et al*. Haar wavelet downsampling: a simple but effective downsampling module for semantic segmentation [J]. *Pattern Recognition*, 2023, 143: 109819.

[21] SUNKARA R, LUO T. No more strided convolutions orpooling: a new CNN building block forlow-resolution images andsmall objects [C]. *Machine Learning and Knowledge Discovery in Databases. Cham*: *Springer*, 2023: 443-459.

[22] SELVARAJU R R, COGSWELL M, DAS A, *et al*. Grad-CAM: visual explanations from deep networks *via* gradient-based localization [J]. *International Journal of Computer Vision*, 2020, 128 (2): 336-359.

[23] REN S 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*, 2017, 39(6): 1137-1149.

[24] ZHAO Y A, LV W Y, XU S L, *et al*. DETRs Beat YOLOs on real-time object detection [C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 16-22, 2024, *Seattle*, *WA*, *USA*. IEEE, 2024: 16965-16974.

[25] LIU W, ANGUELOV D, ERHAN D, *et al*. SSD: single shot multibox detector [C]. *Computer Vision - ECCV* 2016. *Cham*: Springer, 2016: 21-37.

[26] LIU S, QI L, QIN H F, *et al*. Path aggregation network for instance segmentation [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. *June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 8759-8768.

[27] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. 2018: *arXiv*: 1804.02767. https://arxiv.org/abs/1804.02767

[28] JOCHER G, CHAURASIA A, STOKEN A, *et al*. Ultralytics/YOLOv5: v7.0-YOLOv5 SOTA realtime instance segmentation [DB/OL]. *Zenodo*, 2022.

[29] CHEN H, CHEN K, DING G G, *et al*. YOLOv10: Real-Time end-to-end object detection [C]. *Advances in Neural Information Processing Systems* 37. *December* 10-15, 2024. *Vancouver*, *BC*, *Canada*. *Neural Information Processing Systems Foundation*, *Inc*. (*NeurIPS*), 2024: 107984-108011.

[30] KHANAM R, HUSSAIN M. YOLOv11: an Overview of The Key Architectural Enhancements [EB/OL]. 2024: *arXiv*: 2410.17725. https://arxiv.org/abs/2410.17725

作者简介：

李大海(1975—)，男，山东乳山人，副教授，硕导，博士，CCF 会员(85595M)，主要研究方向为智能优化算法、强化学习算法及应用、深度学习的图像处理等。E-mail：9120130107@ mail. jxust. edu. cn

通讯作者：

廖嘉伟(2001—)，男，江西赣州人，硕士研究生，主要从事深度学习、图像识别的研究。E-mail:6720230861@mail.jxust. edu. cn