

文章编号 1004-924X(2025)08-1303-10

感知增强混合网络的水下目标检测

姚婷婷*, 李 宁, 张 煜

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

摘要:水下目标检测技术在海洋资源勘探、环境保护等领域发挥着重要作用。然而,水下环境存在成像模糊、目标尺度多变等问题给检测任务带来了难度。为提高水下目标检测精度,提出一种感知增强混合网络。首先,构建全局-局部混合增强特征提取主干网络,利用自注意力机制求取图像中的长程全局信息,进一步构建卷积注意力增强模块提取更丰富的图像局部细节信息。通过在特征提取过程中更好地捕获图像全局和局部特征关联,增强各尺度特征描述的判别性;随后,为提高成像模糊对比低时的目标检测精度,构建目标感知增强双阶段检测头。通过增加一阶段区域提议生成网络的深度,在低质图像中提取更多目标语义信息,并通过在二阶段引入自注意力机制,抑制背景或次要信息干扰。进一步引入目标检测交并比分支,将一阶段目标先验信息融入二阶段分类求解中,提高不同种类目标的检测精度。所提方法在水下目标检测数据集 TrashCan 与 WPBB 上的 $mAP_{0.5:0.95}$ 和 AP_{50} 分别达到了 37.8%, 61.8% 和 82.0%, 98.9%。定性和定量的对比实验结果表明,模型对不同种类水下目标均具有良好的检测准确性和鲁棒性。

关键词:水下目标检测;特征增强;自注意力机制;混合网络

中图分类号:TP391 文献标识码:A

doi:10.37188/OPE.20253308.1303

CSTR:32169.14.OPE.20253308.1303

Perception enhanced hybrid network for underwater object detection

YAO Tingting*, LI Ning, ZHANG Yu

(Information Science and Technology College, Dalian Maritime University, Dalian 116026, China)

* Corresponding author, E-mail: ytt1030@dlmu.edu.cn

Abstract: Underwater object detection technology plays an important role in areas of marine resource exploration and environmental protection. However, the problems of blurred imaging and variable object scales in underwater environments pose difficulties for detection tasks. As a result, it is challenging for accurate underwater object feature extraction, which influences the detection performance of existing methods. To solve the above-mentioned problem, a feature enhanced hybrid network was proposed to improve the detection accuracy of underwater objects. Firstly, a global-local hybrid feature enhancement network was constructed. The long-range global information in the image was extracted via self-attention mechanisms, and the richer local detailed information was further calculated through the devised convolutional attention enhancement module. The global and local relationships in the images could be better established, hence the multiscale feature representation powers of the network were enhanced. Subsequently, in order

收稿日期:2024-10-24;修订日期:2024-12-24.

基金项目:国家自然科学基金(No. 62001078);中央高校基本科研业务费(No. 3132025245)

to suppress the interference of imaging blurriness and low contrast in underwater environments, a two-stage object perception enhanced detection head was constructed. The depth of the first-stage region proposal network was increased, thus more semantic information of underwater objects could be extracted. Besides, the self-attention mechanism was introduced in the second stage to suppress the interference from background noise. Moreover, an intersection-over-union branch was added to further integrate the prior information of objects obtained from the first stage into the second stage. The proposed method achieves 37.8%, 61.8%, and 82.0%, 98.9% of $mAP_{0.5:0.95}$ and AP_{50} on the TrashCan and WPBB datasets respectively. The qualitative and quantitative comparison experimental results demonstrate that this method could obtain superior detection results for various underwater objects.

Key words: underwater object detection; feature enhancement; self-attention; hybrid network

1 引言

利用水下目标检测技术可以自动定位出水中感兴趣目标在图像中的位置,并给出目标的类别信息。该技术在海洋资源勘探、海洋生态系统检测、环境保护和渔业管理等方面有着广泛的应用^[1-2]。

传统水下目标检测算法通常使用霍夫变换、边缘检测算子、灰度共生矩阵等方法提取图像特征,进而确定目标的位置和类别。由于水下环境中成像图像存在光照不均匀、水质模糊干扰等特殊挑战,这些方法虽然计算相对简单,但是提取的目标特征描述力不足,因此检测精度往往较低。

近年来,为了更好地提高水下复杂成像环境中的目标检测精度,研究人员利用深度神经网络模型对水下图像提取更多的特征描述,一定程度上提高了水下目标的检测性能。Shen等^[3]在YOLO网络中引入选择性注意力模块,通过获取全局信息并高效地实现多尺度特征交互融合,提高了水下目标的检测精度。Chen等^[4]提出基于卷积运算的编码器和解码器。利用编码器提取更多包含丰富语义信息的特征描述,进一步在解码器中自适应地融合不同通道的特征,获取了更加有效的水下目标多尺度特征表达。Yu等^[5]提出一种路径聚合网络结构,通过在特征融合网络中增加自下而上的下采样路径,提高特征提取能力。进一步将注意力机制融合到网络中,提高海洋目标的检测准确性。Dai等^[6]在主干网络中引入边缘注意力模块提取目标的边界信息,从而捕获更多的判别性特征。同时,在特征融合网络中引入特征聚合模块有效地融合多尺度特征,提高

了水下目标的定位和检测精度。

近年来Transformer模型在陆上目标检测任务中取得了令人瞩目的检测精度,因此学者们尝试利用自注意力机制实现水下目标检测。如Liu等^[7]利用包含多头注意力机制的Swin Transformer作为骨干网络进行特征提取,并在特征金字塔中加入路径聚合网络,通过更好地结合富含语义信息的深层特征和富含细节信息的浅层特征,增强网络对水下成像质量较低图像的特征描述力。Dai等^[8]提出了一种以自注意力机制为中心的门控跨域协作网络,用于水下目标检测。该网络结合了自注意力机制,利用其对不同空间区域的关联性进行建模,从而提高了水下目标的检测准确性和鲁棒性。陶洋等^[9]利用RepViTS作为特征提取网络,并通过结构重参数化提升水下目标的检测精度和模型推理速度。此外,引入了空间和通道重构模块,增强了网络对水下模糊目标的定位能力。

由于卷积神经网络可以很好地捕捉图像中的局部信息,而自注意力求解可以在全局范围内建模像素之间的关联,因此已有研究人员尝试将自注意力机制和卷积神经网络相结合,同时提高网络对目标的全局和局部特征描述力。如Liang等^[10]提出将检测头得到的初始特征输入到自注意力模块中,再分别经过线性层和卷积层获得最终的分类分数和边界框坐标。通过有效地学习不同水下目标之间地关联性,提高了检测的准确性和鲁棒性。Xu等^[11]在特征融合网络中引入自注意力模块。在上采样操作后加入自注意力机制来获得输入特征图的静态和动态上下文信息,提高了网络对小尺度密集海洋

生物的检测能力。Zhang等^[12]提出在特征金字塔中引入多头自注意力机制,对已有特征进行非局部增强,并在多头自注意力机制后加入过滤特征集成机制来综合浅层特征描述和深层语义信息,提高了模型在海洋复杂多变环境下的目标检测能力。

通过将卷积神经网络与Transformer网络进行融合,有效捕获了目标更多的局部和全局信息,进而提高了水下目标的检测精度。然而,现有算法的特征提取主干网络大多单纯依靠卷积或自注意力求解,在特征融合阶段才引入混合模型结构。由于主干网络的结构单一性在一定程度上影响了网络所提取的特征描述判别性。此外,当水下成像环境对比度较低时,现有算法的检测性能也会受到一定程度的影响。基于此,本文提出一种感知增强混合网络的水下目标检测方法,通过构建全局-局部混合增强特征提取骨干网络和目标感知增强双阶段检测头,更好地在水下复杂水质环境中提取出目标的特征描述,从而提高模型的检测性能。本文主要贡献如下:

(1) 在特征提取主干网络中构建了全局-局部混合特征增强网络,利用自注意力机制捕获长程全局信息,并进一步构建卷积注意力增强模块在水下图像中提取详细的局部信息。通过在特征提取过程中更好地捕获图像中全局和局部特征关联,增强网络的多尺度特征判别性。

(2) 为了进一步提高水下成像模糊对比度低时的目标检测精度,构建了目标感知增强双阶段检测头。通过增加一阶段区域提议生成网络的深度,从低质量图像中获取更多水下目标的语义信息。同时,在二阶段引入自注意力机制抑制背景噪声的干扰。进一步引入目标检测交并比分支,在二阶段分类求解中融入一阶段目标先验信息,提高了不同种类目标的检测精度。

(3) 所提方法在水下目标检测数据集 Trash-Can 与 WPBB 上的 $mAP_{0.5:0.95}$ 和 AP_{50} 分别达到了 37.8%, 61.8% 和 82.0%, 98.9%。定性和定量的对比实验结果表明,模型对不同的水下目标均具有良好的检测准确性和鲁棒性。

2 研究方法

2.1 网络模型概述

本文构建的网络模型总体结构如图1所示。模型在PVT^[13]基础上,为了提高每个尺度特征描述的判别性,构建了全局-局部混合特征增强主干网络。利用自注意力机制求取图像中的长程信息,对特征描述进行全局增强。进一步设计了卷积注意力增强模块,提取更加丰富的图像局部细节信息。接着,采用特征金字塔网络^[14]进一步对多尺度特征进行融合后,送入构建的目标感知增强双阶段检测头中。在一阶段区域提议生成网络中,通过增加卷积层提高网络对成像模糊、

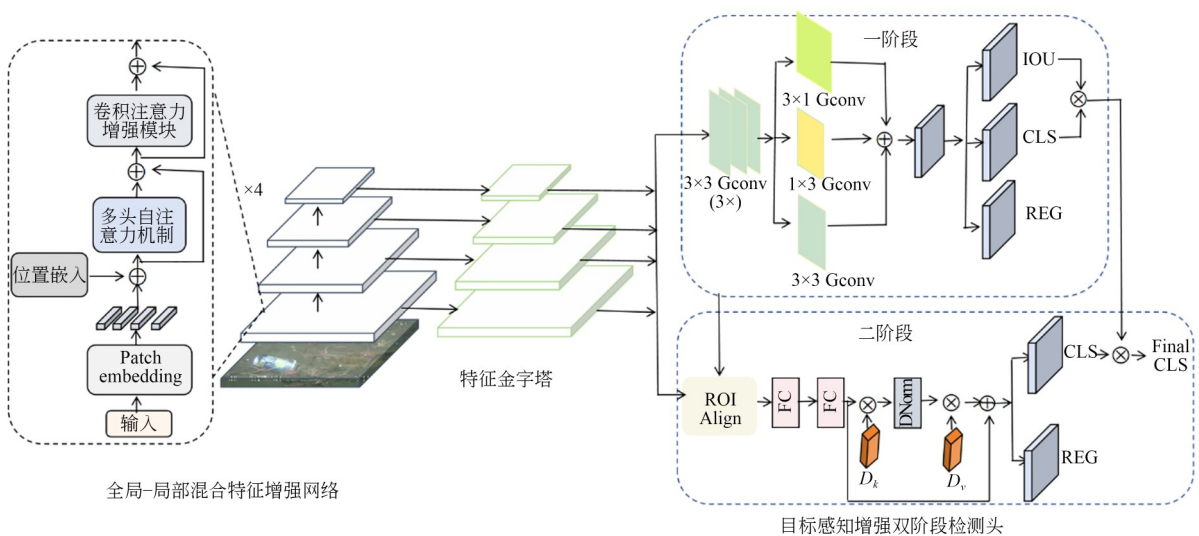


图1 网络模型总体结构图

Fig. 1 Overall structure of network

对比度低的水下图像目标语义特征的描述力。在二阶段,通过引入自注意力机制^[15]抑制背景或次要信息的干扰。接着将得到的特征分别经过分类子网络、回归子网络和目标检测交并比分支子网络求取目标先验信息和候选框^[16]。随后利用目标先验和候选框提高二阶段目标分类和检测精度。下文将对模型中的每个模块进行详细介绍。

2.2 全局-局部混合特征增强网络

由于水下光线不足,并且水中的悬浮物较多使得水下图像质量较差,因此利用一般的卷积神经网络难以有效地提取图像特征。本文通过构建全局-局部混合特征增强模块,综合全局自注意力求解和局部卷积操作,从而提高主干网络提取特征的判别性。

本文构建的全局-局部混合特征增强模块结构如图2所示。首先,将网络输入特征图分割成 N 个不重叠的特征块 $N=HW/B^2$, H, W 表示特征图的高度和宽度, B 表示每个特征块的大小。由于主干网络每层输出不同尺度大小的特征图,参考论文^[13],本文为四个尺度特征图分别设置 B 的大小为4,2,2,2。接着将每个特征块经过线性投影,并利用正余弦函数计算对应的位置编码,得到特征映射 $X \in R^{hw \times C}$, $h=H/B$, $w=W/B$, C 表示通道数量。随后将特征 $X \in R^{hw \times C}$ 经过线性变换求取对应的 $Q, V, K \in R^{hw \times C}$ 矩阵,并利用卷积操作将 V, K 的宽高分别降低为原来的 S 倍得到 $V' \in R^{hwC/S^2}$ 和 $K' \in R^{hwC/S^2}$ 矩阵。

本文为四个尺度特征图设置的 S 的值分别为8,4,2,1。最后根据多头自注意力求解公式计算不同特征块 Q, K' 矩阵之间的关联,并经过 Softmax 函数即可获得全局注意力权重。将全局注意力权重与 V' 矩阵进行加权求和后可以得到

最终的全局信息增强后的特征描述 $X_{att}^{q[13]}$ 。上述计算过程用公式表示为:

$$X_{att}^q = \sum_{l=1}^N \text{softmax} \left(\frac{Q_q \cdot K'_l}{\sqrt{d_{he}}} \right) V'_l, \quad (1)$$

其中: d_{he} 表示每个注意力头的维度。 Q_q 表示第 q 个特征块对应的 Q 矩阵, K'_l 和 V'_l 表示第 l 个特征块对应的 K' 和 V' 矩阵, N 表示特征块总数。对每个特征块求取增强特征描述后,将所有特征块进行整合得到最终的全局信息增强特征图 X_{att} 。

为了进一步增强网络对局部信息的提取能力,本文将全局注意力增强后的特征进一步送入构建的卷积注意力增强模块。首先经过全连接,将特征映射到更高维度的空间,并将特征尺度重塑为 $X'_{att} \in R^{h \times w \times 4C}$ 。接着利用残差连接引入 3×3 的深度卷积^[17]和 1×1 的普通卷积对 X'_{att} 在不同感受野范围内提取更多的局部细节信息。随后将 X'_{att} 与卷积后的特征进行相加,并经过激活层和全连接进行降维。上述操作可以用公式描述为:

$$\begin{cases} X''_{att} = X'_{att} + DWConv_{3 \times 3}(X'_{att}) + Conv_{1 \times 1}(X'_{att}) \\ X_{feat} = FC(GELU(X''_{att})) \end{cases}, \quad (2)$$

其中: $DWConv_{3 \times 3}$, $Conv_{1 \times 1}$ 分别表示卷积核大小为 3×3 的深度卷积和 1×1 的卷积, FC 表示全连接。

为了进一步融入通道维度所包含的信息,将不同感受野卷积操作增强后的特征 X_{feat} 经过全局平均池化操作,在空间维度进行特征压缩得到 $X_{sa} \in R^{1 \times 1 \times C}$ 。然后对压缩后的特征图通过两个线性层和激活层得到通道注意力增强特征图 $X_{chan} \in R^{1 \times 1 \times C}$ 。最后将学习到的通道注意力增强特征图 X_{chan} 与原始输入特征图 X_{feat} 逐通道相

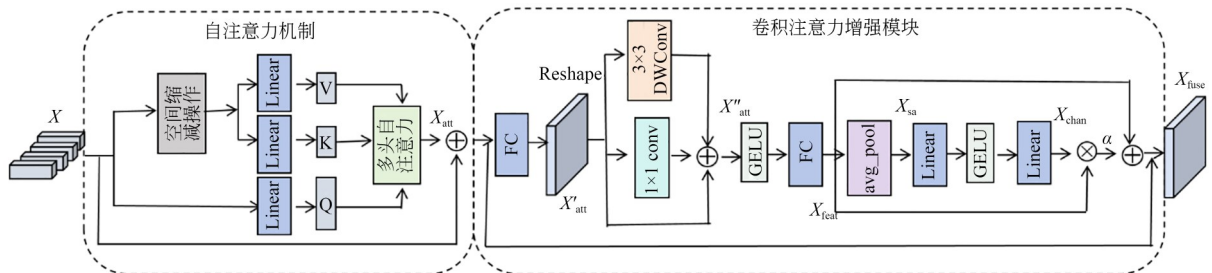


图2 全局-局部混合特征增强网络

Fig. 2 Global-local Hybrid Feature Enhancement Network

乘,输出最终的全局-局部混合增强特征图 X_{fuse} :

$$\begin{cases} X_{\text{sa}} = \text{avg_pool}(X_{\text{feat}}) \\ X_{\text{chan}} = \text{Linear}\left(\text{GELU}\left(\text{Linear}(X_{\text{spa}})\right)\right), (3) \\ X_{\text{fuse}} = \alpha(X_{\text{chan}} \times X_{\text{feat}}) + X_{\text{feat}} \end{cases}$$

其中: avg_pool 表示平均池化。 α 为通道信息增强权重系数,本文根据前期实验结果,设置该参数取值为 0.1。整个网络利用自注意力求解,有效建立了图像中每个位置与其他所有位置特征之间的关联,因此可以更好地捕获特征之间的长程全局信息。通过引入不同大小的局部卷积操作,在不同感受野范围下提取了更多的局部细节信息。进一步利用通道注意力增强处理,对不同通道的局部特征进行选择增强,因此提高了模型对水下图像的特征描述力。

2.3 目标感知增强双阶段检测头

经过全局-局部混合特征增强网络,对不同尺度特征进行了增强后,通过特征金字塔网络^[14]采用自顶向下的结构,对不同尺度特征进行融合。然后,将输出的多尺度特征送入检测头网络求取目标检测结果。由于水下图像通常具有成像模糊失真、对比度低等特点,很难将目标与背景区分开。为了提高双阶段目标分类与定位精度,本文构建了目标感知增强双阶段检测头。在一阶段区域提议生成网络中加入了具有组归一化的卷积层,通过增强网络深度提高网络对水下复杂成像环境中的目标语义特征的提取能力。同时,加入了额外的目标检测交并比分支,求取更多的目标先验信息。

如图1所示,通过特征金字塔网络对多尺度特征进行融合后,首先会将其送入一阶段区域提议生成网络,输出包含目标区域的高质量候选框。本文一阶段区域提议生成网络在每个空间位置生成多个具有不同尺度和纵横比的锚框,经过 3 个 3×3 组归一化卷积和并行的 $3 \times 1, 1 \times 3$ 和 3×3 的组归一化卷积后,利用分类子网络、回归子网络和交并比子网络对每个锚框生成三个输出:分类子网络输出第 i 个锚框内有目标的概率 $\hat{P}_i$ 。回归子网络输出该锚框与真实边界框的偏移量,进而可以通过调整该锚框的位置和尺寸得到目标的初始定位结果。交并比子网络输出第 i 个锚框和目标真值之间的区域交并比 $\hat{g}_i$ 。利用非极大抑制后,可以从一系列锚框中筛选出包

含有目标区域的高质量候选框。

通过额外增加的交并比分支可以为水下目标提供更多的先验信息,进而提高后续二阶段目标定位和分类精度。具体而言,目标先验信息 P_{ri} 通过将交并比分支得到的预测结果与分类子网络所得到的分类分数相乘后开平方根进行求取^[15]:

$$P_{ri} = \sqrt{\hat{g}_i * \hat{P}_i}. \quad (4)$$

随后,进入二阶段求解。二阶段首先将一阶段输出的候选框所包含区域对应的特征图经过 ROI Align 操作^[15],得到固定大小的特征图描述。然后经过两个全连接层减少特征通道数,得到输出特征 X_{fc} 。为了提高目标兴趣区域的特征描述力,抑制复杂背景对目标检测的干扰,进一步引入特征自注意力增强运算^[16]。将特征 X_{fc} 经过线性投影分别得到 D_k, D_v 特征矩阵,再将 X_{fc} 与 D_k 进行点积运算,并对得到的特征行和列分别进行归一化,减少梯度消失或爆炸的风险,提高模型训练过程的稳定性,进而得到 X_v :

$$X_v = D \text{Norm}(X_{fc} D_k^T). \quad (5)$$

接着利用 X_v 与 D_v 进行计算得到增强后的特征描述 X_o :

$$X_o = X_v D_v + X_{fc}. \quad (6)$$

最后,利用 X_o 进行二阶段的分类和回归求解。通过在二阶段中引入自注意力机制计算不同区域之间的相似度,可以将与目标相关性较大的特征进行增强,并将与目标相关性小的信息进行过滤,达到抑制背景或次要信息干扰的目的。

在二阶段分类求解中,本文将一阶段得到的目标先验信息融入目标类别标签求解过程中,进一步提高二阶段检测器的鲁棒性,从而获得最终的分类结果 S_{final} :

$$S_{\text{final}} = \sqrt{P_{ri} * cls_k}, \quad (7)$$

其中, cls_k 表示第二阶段的分类结果。

2.4 损失函数

目标感知增强双阶段检测头包含三个子网络:分类子网络、回归子网络和交并比子网络。在网络模型训练过程中分别采用不同的损失函数,使网络达到更好的检测性能。

对分类子网络,利用焦点损失函数计算分类损失^[18],公式如下:

$$L_{\text{cls}}(\hat{p}_i) = \begin{cases} -\beta(1-\hat{p}_i)^\gamma \log(\hat{p}_i), & y_i = 1 \\ -(1-\beta)\hat{p}_i^\gamma \log(1-\hat{p}_i), & y_i = 0 \end{cases}, \quad (8)$$

$$L_{\text{cls}} = \frac{1}{n} \sum_{i=1}^n L_{\text{cls}}(\hat{p}_i), \quad (9)$$

其中: $\hat{p}_i$ 表示第*i*个锚框中目标存在的预测概率, β 表示平衡参数, γ 表示聚焦参数,用于调整权重减小的速度。 n 表示锚框的数量。 $y_i \in \{0, 1\}$ 表示第*i*个锚框的标签,如果是正样本则取值为1,否则为0。实际计算时,本文将锚框与真实边界框的交并比大于0.5的样本作为正样本,其余作为负样本。

对回归子网络,为了提高损失求解收敛速度,并克服奇异值对损失求解的影响,选用FIOU损失^[19]进行模型训练,计算公式如下:

$$L_{\text{FIOU}}(\hat{b}_i) = g_i^\eta \left(1 - g_i + \sum_{j \in \{x, y, w, h\}} \|\hat{t}_{i,j} - t_{i,j}^*\|_2^2 \right), \quad (10)$$

$$L_{\text{reg}} = \frac{1}{m} \sum_{i=1}^m L_{\text{FIOU}}(\hat{b}_i), \quad (11)$$

其中: η 来控制异常值的抑制程度, m 是正样本的数量, g_i^η 能够有效缓解异常值的问题。

对于交并比子网络,选用交叉熵损失^[20],具体计算公式如下:

$$L_{\text{iou}} = \frac{1}{m} \sum_{i=1}^m -[g_i \log(\hat{g}_i) + (1 - g_i) \log(1 - \hat{g}_i)], \quad (12)$$

其中: $\hat{g}_i$ 表示第*i*个锚框所预测的交并比, m 是正样本的数量。

3 实验结果与分析

本节首先介绍实验所选用的数据集、实验机器配置和重要的实验参数设置。然后,将本文所提算法与水下目标检测算法和通用目标检测算法在公测数据集上进行实验对比,通过定量和定性的实验证明本文方法的有效性。最后,进行了消融实验,证明在网络模型中设计的子模块均对检测性能的提升有所贡献。

3.1 实验数据集与参数设定

本文使用两大公开水下数据集 TrashCan^[8]和 WPBB^[21]验证所提方法的有效性。TrashCan是一个具有像素级标注的水下目标检测数据集,

共计包括水下垃圾、水下机器人和各种海底动植物等在内的16个目标类别。训练集和测试集分别包含了6 008张图像和1 204张图像。WPBB数据集是一个包含水下2种常见垃圾:塑料袋和塑料瓶的检测数据集。该数据集由900张带注释的图像组成,训练集和测试集分别包含了720张图像和180张图像。两个数据集存在环境背景复杂、成像对比度低和目标遮挡等难点。

本文所有模型均基于PyTorch1.7.0深度学习框架MMDetection实现。硬件设备为内存128 GB,显卡 GeForce RTX 2080Ti的工作站。训练过程中,优化器选用SGD,批量处理大小设置为1,初始学习率为0.001 25,动量设置为0.9,衰减率设置为0.000 1。实验中所有方法均使用对应数据集给定的训练和验证样本进行模型训练和参数微调,并在统一的测试样本上使用 $mAP_{0.5:0.95}$ 和 AP_{50} 评测指标对不同算法的检测结果进行评估。其中, $mAP_{0.5:0.95}$ 计算了交并比阈值范围从0.5到0.95(步长为0.0.5)的平均精度:

$$mAP_{0.5:0.95} = \frac{1}{M} \sum_{j=1}^M AP_j, \quad (13)$$

其中: M 表示交并比阈值数量。 AP 表示精度-召回曲线下的面积:

$$AP = \int_0^1 Pre(r) dr, \quad (14)$$

$$Pre = \frac{TP}{TP + FP}, \quad (15)$$

$$r = \frac{TP}{TP + FN}, \quad (16)$$

其中: Pre 表示计算精度, r 表示召回率, TP 表示正样本判别为正类, FP 表示负样本判别为正类, FN 表示负样本判别为负类。 AP_{50} 则计算了交并比阈值为0.5时求取获得的 AP 值。

3.2 定量分析

本文选用Faster-RCNN^[22],AutoAssign^[23],Deformable DETR^[24],PVT^[13],ERL-Net^[6],Boosting R-CNN^[15],RoIAttn^[10]在内的七种经典算法进行实验对比。

表1和表2分别给出了不同方法在TrashCan和WPBB数据集上的检测精度。本文方法在 $mAP_{0.5:0.95}$ 和 AP_{50} 上分别获得了37.8%,61.8%和82.0%,98.9个百分点%的检测精度,在两个

评测指标下都能得到优于其他算法的检测性能。

表 1 不同算法在 TrashCan 数据集上的定量对比
Tab.1 Quantitative comparison of different algorithms on the TrashCan dataset

方法	$mAP_{0.5:0.95}/\%$	$AP_{50}/\%$
Faster R-CNN	31.2	55.3
AutoAssign	30.7	51.1
Deformable DETR	24.1	43.4
PVT	31.4	52.7
ERL-Net	37.0	58.9
Boosting R-CNN	36.8	57.6
RoIAttn	32.6	57.2
本文	37.8	61.8

表 2 不同算法在 WPBB 数据集上的定量对比
Tab.2 Quantitative comparison of different algorithms on the WPBB dataset

方法	$mAP_{0.5:0.95}/\%$	$AP_{50}/\%$
Faster R-CNN	75.7	98.7
AutoAssign	75.9	98.8
Deformable DETR	73.2	98.3
PVT	79.5	98.5
ERL-Net	78.5	97.1
Boosting R-CNN	70.2	88.1
RoIAttn	79.7	98.5
本文	82.0	98.9

具体来说,在 TrashCan 数据集上本文方法通过构建全局-局部混合特征增强网络,在 $mAP_{0.5:0.95}$ 和 AP_{50} 评价指标上,相比于经典的双阶段 Faster R-CNN 算法分别高出 6.6% 和 6.5%;相比于自适应单阶段检测器 AutoAssign 算法分别高出 7.1% 和 10.7%。近年来 Transformer 模型在目标检测领域获得了广泛的应用,并取得了良好的检测性能。本文算法相较于基于可变形自注意力模块的 Deformable DETR 高出 13.7% 和 18.4%,相较于基于多头自注意力机制的 PVT 算法高出 6.4% 和 9.1%。同样,在 WPBB 数据集上,相比于经典的 Faster R-CNN, AutoAssign, Deformable DETR 和 PVT,本文方法在 $mAP_{0.5:0.95}$ 和 AP_{50} 评价指标上均有所提升。实验结果证明,本文算法相比经典和最新的通用目标检测算法,可以获得更好的水下目标检测

精度。

进一步,本文与最新提出的水下目标检测算法进行性能比较。相比于在检测头提出进行不确定性建模和困难样本挖掘的 Boosting R-CNN 算法,本文得益于目标感知增强双阶段检测头,得到了更加精确的检测结果,算法总体性能在两个数据集上分别提升了 1.0%, 4.2% 和 11.8%, 10.8%。相比于提出在检测头中引入自注意力模块,并采用分离头进行分类和回归的 RoIAttn 算法,本文方法通过构建全局-局部混合特征增强网络,可以得到更加精确的检测结果,总体检测性能分别提升了 5.2%, 4.6% 和 2.3%, 0.4%。相比于基于特征聚合和边缘特征提取的 ERL-Net 算法,本文得益于网络更有效的目标特征提取能力,算法总体性能提升了 0.8%, 2.9% 和 3.5%, 1.8%。

3.3 定性分析

图 3 给出了不同方法在 TrashCan 和 WPBB 数据集上检测结果对比图。第一行为 AutoAssign 算法,第二行为 Faster R-CNN 算法,第三行为 PVT 算法,第四行为 Boosting R-CNN 算法,第五行为 RoIAttn 算法,第六行为本文所提算法的实验结果图。

第 1 列图中包含尺度大小不一的多个目标,同时图像背景较为复杂。本文算法依靠全局-局部混合特征增强网络可以捕获更多的目标全局和局部信息,因此对目标的定位更加准确。AutoAssign 算法、Boosting R-CNN 和 PVT 算法在定位过程中出现了目标框定位范围过大等定位不准的现象。Faster R-CNN 算法出现了不同的目标框混叠情况,并对植物这一类别出现了漏检的现象。RoIAttn 算法将水下机器人误检为其他类别,并在检测过程中出现了目标框混叠情况。第 2 列图像出现了水下环境中光照不平衡现象, Faster R-CNN 算法存在不同目标框重叠现象; PVT 算法出现了漏检现象; Boosting R-CNN 算法将植物这一类别误检为垃圾; RoIAttn 算法将背景中的目标误检为贝壳,将植物误检为垃圾。

第 3 列与第 4 列展现了水下环境中存在的模糊和低对比度情况,并且图像中包含尺度较小的目标,容易造成误检和漏检。本文算法依靠目标感知增强双阶段检测头有效提高了最终检测结



图3 不同方法在TrashCan数据集和WPBB数据集上的检测结果对比

Fig. 3 Comparison results of different detection methods on TrashCan and WPBB dataset

果的准确性。在第3列中AutoAssign算法出现目标框重叠现象,并对鳗鱼这一类别出现漏检情况;FasterR-CNN和Boosting R-CNN算法均出现目标框重叠现象。在第4列中AutoAssign, Faster R-CNN, PVT, Boosting R-CNN和RoIAttn算法均将鱼误检为其他类别。在第5列中RoIAttn算法出现目标框混叠的情况。综上所述,本文方法通过加入全局-局部混合特征增强网络和目标感知增强双阶段检测头对水下目标的识别和定位更加准确。

3.4 消融实验

为了验证本文方法中每个模块的有效性,我们在TrashCan数据集上进行了消融实验。为了公平比较,所有方法均使用相同的实验设置,不同模型下的实验结果如表3所示。当同时去掉目标感知增强双阶段检测头和卷积注意力增强模块后,模型退化为经典PVT模型, $mAP_{0.5:0.95}$ 为31.4, AP_{50} 为52.7。当仅加入卷积注意力增强模块时,网络模型性能分别提升了2.3%和3.9%。由此可见,通过在主干网络中将自注意力机制和

卷积操作进行有效结合,能够获取更多的全局和局部信息,增加特征的判别性,提高网络模型的检测性能。当仅加入目标感知增强双阶段检测头后,模型检测性能分别提升了6%和7.5%。由此可见,通过构建目标感知增强双阶段检测头,可以有效提高目标兴趣区域的特征描述力,减少背景信息的干扰,提高水下模糊成像环境下的目标检测精度。当两个模块同时加入后,本文方法在 $mAP_{0.5:0.95}$ 和 AP_{50} 上取得了最高的检测精度。综上所述,本文方法中每个模块对

表3 在TrashCan数据集上的消融实验

Tab. 3 Ablation experiments on the TrashCan dataset

卷积注意力 增强模块	目标感知增强 双阶段检测头	$mAP_{0.5:0.95}/\%$	$AP_{50}/\%$
×	×	31.4	52.7
✓	×	33.7	56.6
×	✓	37.4	60.2
✓	✓	37.8	61.8

水下目标检测性能的提高均有不同程度的贡献。

图4进一步给出了从TrashCan数据集中选取的两幅图像在送入卷积注意力增强模块前后网络提取的特征图对比。从图中可以看出,经过所设计的卷积注意力增强模块可以有效地从目标中提取出更多的局部细节信息,目标的轮廓也更为清晰,直观地证明了该模块具有良好的特征增强效果。

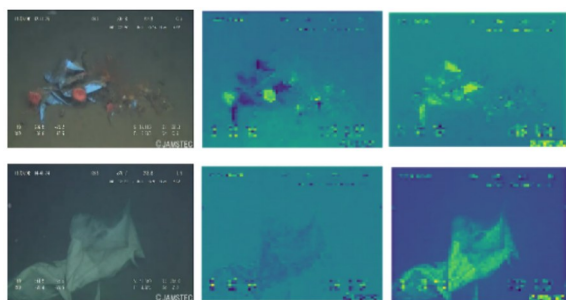


图4 经卷积注意力增强模块前后的特征图

Fig. 4 Feature maps before and after convolutional attention enhancement module

参考文献:

- [1] XU S 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.
- [2] JIANG L, WANG Y, JIA Q, *et al.* Underwater species detection using channel sharpening attention [C]. *Proceedings of the 29th ACM International Conference on Multimedia*, 2021: 4259-4267.
- [3] SHEN X, WANG H B, LI Y F, *et al.* Criss-cross global interaction-based selective attention in YOLO for underwater object detection [J]. *Multimedia Tools and Applications*, 2024, 83 (7): 20003-20032.
- [4] CHEN L, XIE Y Z, LI Y X, *et al.* CWSCNet: channel-weighted skip connection network for underwater object detection [J]. *IEEE Transactions on Image Processing*, 2024, 33: 5206-5218.
- [5] YU H F, LI X B, FENG Y K, *et al.* Multiple attentional path aggregation network for marine object detection [J]. *Applied Intelligence*, 2023, 53 (2): 2434-2451.

4 结 论

本文提出了基于感知增强混合网络的水下目标检测模型。构建了全局-局部混合特征增强网络,通过充分提取图像中的全局和局部信息,强化了特征描述的判别性。同时,构建了目标感知增强双阶段检测头,利用卷积和注意力操作进一步提高了水下模糊、低对比度环境下的目标检测性能。所提方法在TrashCan与WPBB数据集上的 $mAP_{0.5:0.95}$ 和 AP_{50} 分别达到了37.8%, 61.8%和82.0%,98.9%,证明了方法对复杂水下环境中不同大小的目标均能够获得更加准确的检测结果。未来工作重点是进一步优化网络模型,对检测网络进行轻量化设计,在保证网络检测精度的同时,达到更优的实际应用效果。

作者贡献声明:

姚婷婷:模型整体构思,创新性设计;

李宁:相关实验执行,数据测量;

张煜:论文写作及编排。

- [6] DAI L, LIU H, SONG P, *et al.* Edge-guided representation learning for underwater object detection [J]. *CAAI Transactions on Intelligence Technology*, 2024:1078-1091.
- [7] 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.
- [8] DAI L H, LIU H, SONG P H, *et al.* A gated cross-domain collaborative network for underwater object detection [J]. *Pattern Recognition*, 2024, 149: 110222.
- [9] 陶洋,朱腾,钟邦乾,等. RepViTS-YOLOX: 水下模糊及遮挡目标检测方法[J/OL]. *计算机工程与应用*, 2024: 1-12. (2024-03-10). <http://kns.cnki.net/KCMS/detail/detail.aspx?filename=JSGG20240307008&dbname=CJFD&dbcode=CJFQ>.
- TAO Y, ZHU T, ZHONG B Q, *et al.* RepViTS-YOLOX: Underwater Blurred and Occluded Object Detection Method [J/OL]. *China Industrial Economics*, 2024: 1-12. (2024-03-10). <http://kns.cnki.net/KCMS/detail/detail.aspx?filename=JSGG20240307008&dbname=CJFD&dbcode=CJFQ>.

- CJFQ. (in Chinese)
- [10] LIANG X T, SONG P H. Excavating RoI attention for underwater object detection [C]. 2022 *IEEE International Conference on Image Processing (ICIP)*. October 16-19, 2022, Bordeaux, France. IEEE, 2022: 2651-2655.
 - [11] LIU Y, WANG S N. A quantitative detection algorithm based on improved faster R-CNN for marine benthos [J]. *Ecological Informatics*, 2021, 61: 101228.
 - [12] ZHANG J Y, WANG Y P, XU X C, *et al.* YoloXT: a object detection algorithm for marine benthos [J]. *Ecological Informatics*, 2022, 72: 101923.
 - [13] WANG W H, XIE E Z, LI X, *et al.* Pyramid vision transformer: a versatile backbone for dense prediction without convolutions [C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2021: 548-558.
 - [14] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 936-944.
 - [15] GUO M H, LIU Z N, MU T J, *et al.* Beyond self-attention: external attention using two linear layers for visual tasks [J]. *IEEE Trans Pattern Anal Mach Intell*, 2023, 45(5): 5436-5447.
 - [16] SONG P H, LI P T, DAI L H, *et al.* Boosting R-CNN: Reweighting R-CNN samples by RPN's error for underwater object detection [J]. *Neurocomputing*, 2023, 530: 150-164.
 - [17] SHI D. TransNeXt: robust foveal visual perception for vision transformers [C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 17773-17783.
 - [18] LIN T Y, GOYAL P, GIRSHICK R, *et al.* Focal loss for dense object detection [C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017, Venice, Italy. IEEE, 2017: 2999-3007.
 - [19] ZHANG Y F, REN W Q, ZHANG Z, *et al.* Focal and Efficient IOU Loss for Accurate Bounding Box Regression [EB/OL]. 2021: 2101.08158. <https://arxiv.org/abs/2101.08158v2>.
 - [20] MAO A Q, MOHRI M, ZHONG Y T. Cross-entropy Loss Functions: Theoretical Analysis and Applications [EB/OL]. 2023: 2304.07288. <https://arxiv.org/abs/2304.07288v2>.
 - [21] ZOCCO F, LIN T C, HUANG C I, *et al.* Towards more efficient EfficientDets and real-time marine debris detection [J]. *IEEE Robotics and Automation Letters*, 2023, 8(4): 2134-2141.
 - [22] 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.
 - [23] ZHU B J, WANG J F, JIANG Z K, *et al.* AutoAssign: Differentiable Label Assignment for Dense Object Detection [EB/OL]. 2020: 2007.03496. <https://arxiv.org/abs/2007.03496v3>.
 - [24] ZHU X Z, SU W J, LU L W, *et al.* Deformable DETR: Deformable Transformers for End-to-End Object Detection [EB/OL]. 2020: 2010.04159. <https://arxiv.org/abs/2010.04159v4>.

作者简介:



姚婷婷(1988—),女,辽宁大连人,博士,副教授,硕士生导师。2016年于合肥工业大学取得博士学位。现就职于大连海事大学信息科学技术学院,主要从事计算机视觉,目标检测,图像增强等方面的研究。

Email: ytt1030@dlmu.edu.cn