DOI:10.19651/j.cnki.emt.2416325

基于改进 RT-DETR 的浅水海洋生物识别方法*

蒋智臣[1] 胡俐蕊[2]
(1.广西大学计算机与电子信息学院 南宁 530004;2.北部湾大学电子与信息工程学院 钦州 535011)

摘 要:针对现有浅水海洋生物识别方法在水下环境中对浅水海洋生物识别效果不佳的问题,提出了一种以 RT-DETR 为基准模型的改进浅水海洋生物识别方法。首先,使用重参数化网络 RepViT 作为模型的主干网络,提升模型的特征提取能力。然后,构建基于重参数化的并行膨胀卷积 RepPDC 并引入颈部网络中,使模型能够有效获取长距离上下文信息,有利于提升模型的识别精度。最后,基于注意力机制构建了双向特征融合模块 CAFM,提升模型在水下环境中对重点信息的关注能力。实验结果表明,改进后的方法,mAP50 提升至 87.5%,mAP75 提升至 70.9%,mAP50:95 提升至 64.9%,且参数量更少,有望应用到实际浅水海洋生物识别任务中。

关键词:海洋生物识别;目标检测;深度学习;RT-DETR
中图分类号:TP391.4;TN919.8 **文献标识码**:A **国家标准学科分类代码**:520.6040

Marine life identification method based on improved RT-DETR

Jiang Zhichen[1] Hu Lirui[2]
(1. School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China;
2. College of Electronics and Information Engineering, Beibu Gulf University, Qinzhou 535001, China)

Abstract: Addressing the issue of subpar performance in identifying shallow water marine life in underwater environments using existing methods, we propose an improved method based on the RT-DETR benchmark model. Initially, the reparameterization network RepViT is utilized as the backbone of the model, enhancing its feature extraction capabilities. Subsequently, a reparameterized parallel dilated convolution (RepPDC) is constructed and incorporated into the neck network, enabling the model to effectively capture long-range contextual information, thereby improving the model's recognition accuracy. Lastly, a bidirectional feature fusion module (CAFM) is constructed based on the attention mechanism, enhancing the model's ability to focus on key information in underwater environments. Experimental results demonstrate that the improved method significantly boosts the mAP50 to 87.5%, mAP75 to 70.9%, and mAP50:95 to 64.9%, with fewer parameters, making it a promising candidate for practical applications in the identification of shallow water marine life.

Keywords: marine life identification;object detection;deep learning;RT-DETR

0 引 言

海洋作为生态圈的重要组成部分,其中蕴藏着丰富的海洋生物资源。目前各临海国家致力于开发海洋牧场,科研人员通过人工潜水以及拍照的方式对牧场中的浅水海洋生物的数量以及位置分布进行研究。然而,水下图像往往质量较低,不利于潜水员在水下寻找海洋生物。因此,目前需要一种能够在水下环境中准确、高效工作的海洋生物识别方法去代替肉眼进行工作。

目前基于深度学习的浅水海洋生物识别方法主要可以分为两类,分别是单阶段检测方法和双阶段检测方法。经典的双阶段检测方法主要有 Fast R-CNN(fast region convolutional neural network)[1]以及 Faster R-CNN[2]等,其使用多个不同的网络分别进行定位和分类。单阶段检测方法主要有 SSD(single shot multibox detector)[3]以及 YOLO(you only look once)[4]系列的方法,其使用同一个网络进行定位和分类。Carion 等[5]提出将 Transformer[6]应用于目标检测领域并提出了 DETR(detection transformer);Zhao 等[7]在 DETR 的基础上进一步提出了用于实时检测任务的 RT-DETR(real time detection transformer),其分析了 DETR 的检测速度较低的原因,并

收稿日期:2024-06-27
*基金项目:广西区科技计划项目(桂科 AC17195057)、广西钦州市科技计划项目(202116602)资助

提出了相应的解决方案，使得 RT-DETR 在速度明显提升的情况下没有明显影响检测精度。

不同于传统目标检测，浅水海洋生物识别往往受到水下复杂环境的影响，如水下图像由于光强和散射的因素导致其质量较低等，都会影响模型对海洋生物的识别，降低检测精度。史朋飞等[8]针对海洋生物检测提出了一种改进的 YOLOv4，其在主干网络中引入注意力机制，并改进了特征融合网络，同时提出了一种针对海洋生物检测的数据增强方法；Dai 等[9]针对海洋生物检测提出了一种门控跨域协作网络，引入了一种跨域信息交互模块以增强原始图像与特征提取后的图像的信息交互，并通过一个控制门对信息交互进行控制。Liu 等[10]提出了一种基于 YOLOv7 的水下目标检测方法，其在模型中引入了全局注意力机制，并引入了一种改进的残差连接块，在不影响模型检测速度的同时提升模型的检测能力。周新等[11]提出一种基于 YOLOv5 的轻量级水下生物识别方法，其使用经过改进的轻量化网络作为主干网络，并对损失函数进行了改进，在提高模型速度的同时提升了模型的精度。

尽管目前研究者们针对海洋生物识别提出了较多改进方法，但还是存在不足之处。现有的海洋生物识别方法大多是基于 YOLO 系列方法改进得到，其参数量普遍较低，在特征提取能力上存在劣势，导致其识别精度相对较低，且这些劣势难以通过模型改进的方式消除；同时，由于非极大值抑制操作的影响，YOLO 系列方法识别速度相对更慢，不利于应用在实际海洋生物识别任务中。此外，由于前文所述的海洋生物识别在实际应用中遇到的困难，一般 RT-DETR 模型的识别精度相对较低，其仍然具有改进空间。因此，本文提出了一种基于 RT-DETR 的水下目标检测方法 RDC-RT-DETR。具体来说，本文的主要工作有：1）针对水下图像质量较低，导致一般模型检测精度较低的问题，使用基于结构重参数化的多分支网络 RepViT[12]作为本文方法的主干网络，以提升主干网络的特征提取能力；2）为充分利用图像背景信息以帮助模型对海洋生物进行识别，构建了重参数化并行膨胀卷积[13]，并引入模型的颈部网络中，扩大了模型的感受野；3）为降低水下图像中的冗余特征对模型精度的影响，基于通道注意力构建了双向特征融合模块，提升模型对重要特征的关注能力。

1　方法改进

1.1　基准模型

本文所作改进基于 RT-DETR-ResNet18 进行，其由主干网路、编码器、颈部网络、交并比感知查询模块以及解码器组成。主干网络用于对输入图像进行特征提取；编码器采用 Transformer 中标准的编码器，使用多头自注意力机制从不同角度捕捉图像特征信息；颈部网络采用基于结构重参数化的特征融合网络，用以聚合多尺度特征；交并比感知查询模块在训练期间对模型进行约束，使用于解码器中目标查询的特征同时拥有高分类分数和高交并比分数；解码器基于 DINO[14]解码器设计得到。RT-DETR 会直接预测出一定数量的预测框，其在模型训练阶段将预测框与真实框进行一一对比，对每个真实框按一定的规则寻找最适配的预测框，再计算损失函数。相较于 YOLO 系列方法，RT-DETR 不会在所有位置上都生成预测框，因此 RT-DETR 不需要在模型给出预测结果后进行非极大值抑制以消除大量冗余的预测框，显著减少了模型进行后处理的时间。

本文提出的 RDC-RT-DETR 主要包含 3 个网络结构改进点，分别是使用基于结构重参数化的特征提取网络 RepViT 替换了原主干网络，以及基于重参数化并行膨胀卷积模块改进特征融合网络，并基于通道注意力机制对特征融合网络中的特征拼接操作进行改进，如图 1 所示。

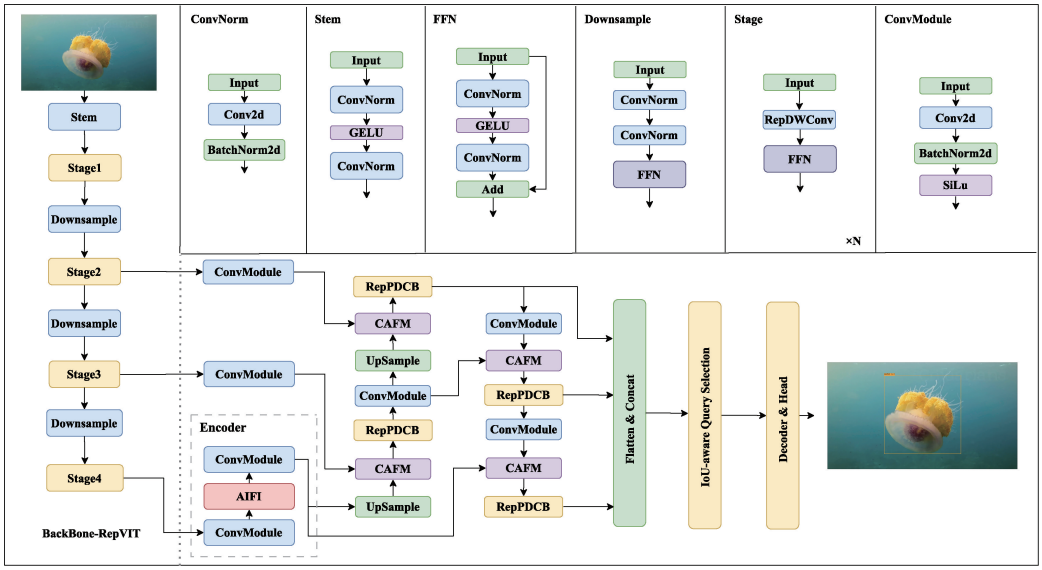


图 1　RDC-RT-DETR 结构图

Fig. 1　The structure of RDC-RT-DETR

1.2 基于重参数化的特征提取网络

在实际浅水海洋生物识别任务中,可能会出现由于光的散射以及水底光强较低等现象,导致图像较为模糊的情况,从而导致一般主干网络难以提取特征。采用多分支主干网络可以增强模型的特征提取能力,在模糊的图像中更好地获取海洋生物的特征,从而提升模型在水下环境中的对海洋生物的检测能力,但多分支网络也会带来额外的显存开销,一定程度上影响了模型的训练以及推理速度。结构重参数化方法[15]能够使模型在拥有多分支网络高性能的同时,推理速度与单分支网络相当,其在模型训练时使用多分支网络,并在推理时将多余的分支去除。在模型训练阶段,结构重参数化卷积使用 3 分支网络,分别为 3×3 卷积分支、1×1 卷积分支以及批归一化(batch normalization,BN)分支,并将这三个分支的输出进行相加。记 $\boldsymbol{x}$ 表示输入特征图,对于单个普通卷积核的输出结果,其可表示为:

$$Conv(\boldsymbol{x}) = \boldsymbol{w} * \boldsymbol{x} + \boldsymbol{b} \tag{1}$$

式中:$\boldsymbol{w}$ 表示卷积核权重,$\boldsymbol{b}$ 表示卷积核训练得到的偏置项。对于 BN 层的输出结果,可表示为:

$$BN(\boldsymbol{x}) = \frac{\boldsymbol{x} - \mu}{\sqrt{\sigma^2 + \varepsilon}}\gamma + \beta \tag{2}$$

式中:μ,σ,γ,β 分别表示用于批归一化的卷积核权重的方差、标准差、缩放因子和偏置项;ε 为较小值,以避免式中分母为 0。由此可以得到卷积层的输出结果:

$$BN(Conv(\boldsymbol{x})) = \frac{\gamma}{\sqrt{\sigma^2 + \varepsilon}}\boldsymbol{w} * \boldsymbol{x} + \frac{\boldsymbol{b} - \mu}{\sqrt{\sigma^2 + \varepsilon}}\gamma + \beta \tag{3}$$

根据此式修改卷积核的权重与偏置项可将 BN 操作融入卷积中。此外,结构重参数化还需要将 1×1 卷积核融入 3×3 卷积中,首先使用零填充的方法将 1×1 卷积核填充至 3×3 的大小,BN 层可以视为一个特殊的 1×1 的卷积核,同样使用零填充将其填充为 3×3 的大小,并将 3 个填充后的卷积核的权重及其偏置进行直接相加,将其作为新的 3×3 卷积核的权重与偏置,最后将 1×1 分支与 BN 分支删除,如图 2 所示。

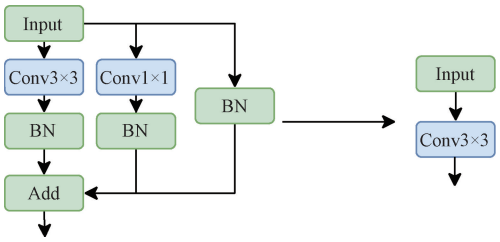


图 2 结构重参数化

Fig. 2 Structural re-parameterization

RepViT 基于结构重参数化思想设计得到,其结构展示在图 1 中。RepViT 包含较多的 Stage 模块以增加整个网络的深度,每个 Stage 模块均含有一个基于结构重参数化的深度可分离卷积(reparameterization depthwise convolution, RepDWConv),其将卷积的所有通道都分到不同的组内,每个组的输入以及输出通道数均为 1,通过结构重参数化卷积进行单独的卷积操作。结合上述多分支网络以及结构重参数化的特点,本文采用了基于结构重参数化的特征提取网络 RepViT-M1.1 作为本文模型的主干网络,在减少主干网络的参数量与计算量的同时提升模型的特征提取能力。考虑到参数量和推理速度等因素,本文并未使用 RepViT 中包含的通道注意力机制。

1.3 基于重参数化并行膨胀卷积的特征提取模块

膨胀卷积相比一般的卷积拥有更大的感受野,对于卷积核大小为 k,膨胀率为 d 的膨胀卷积,其感受野大小 s 可表示为:

$$s = k + (d - 1) \times (k - 1) \tag{4}$$

根据此式可知,当膨胀卷积的膨胀率为 1 时,其感受野与普通卷积相同;当膨胀率为 2 时,其感受野与核大小为 5 的普通卷积相当;当膨胀率为 3 时,其感受野进一步增大,与核大小为 7 的普通卷积相当,如图 3 所示。膨胀卷积在保持模型参数量不变的同时有效增大了模型的感受野。

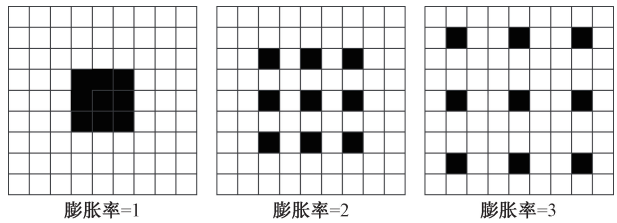


图 3 膨胀卷积感受野

Fig. 3 The receptive field of dilated convolution

考虑到部分海洋生物(如海参、扇贝等)与背景区分度较低,并且可能出现相互遮挡的情况,使用感受野更大的膨胀卷积将有助于模型提取长距离上下文信息,有利于模型通过目标周围的信息对目标进行识别。为使模型的感受野更大且更加灵活,本文提出了重参数化并行膨胀卷积(reparameterization parallel dilated convolution, RepPDC)。并构建了基于 RepPDC 的特征提取模块(RepPDC block, RepPDCB)。RepPDC 使用 3 个并行且膨胀率不同的卷积有效捕捉不同长距离的上下文信息,然后在通道维度上将这些卷积的输出进行拼接,使输出特征图中包含更加丰富的多尺度特征。同时,为进一步提升输出特征图的质量,有效利用输入特征图中的包含的信息,本文使用重参数化卷积替代了膨胀率为 1 的普通卷积。最后,使用一个逐点卷积将拼接后的特征图的通道数减半,以匹配输入通道数。RepPDC 的结构如图 4 所示。

RepPDCB 基于泛化高效层聚合网络(generalized efficient layer aggregation network, GELAN)[16]架构进行构建,其结构如图 5 所示。RepPDCB 包含两个 RepPDC 模块,其中第 2 个 RepPDC 模块的输入基于第 1 个 RepPDC 模块的输出得到,从而获取更长距离的图像上下

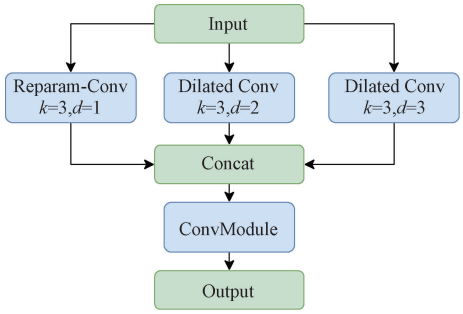


图 4　重参数化并行膨胀卷积

Fig. 4　Reparam-parallel dilated convolution

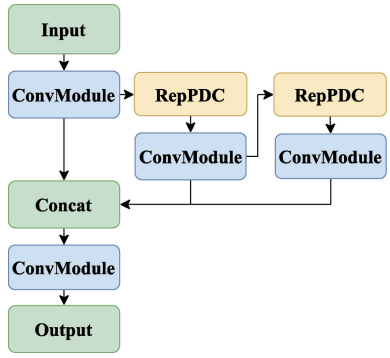


图 5　RepPDCB 结构

Fig. 5　The structure of RepPDCB

文关系，最后使用 1 个卷积模块对拼接后的特征图的通道数进行压缩，以减少模型体积。记 $\boldsymbol{x}$ 为输入特征图，$\boldsymbol{y}$ 为输出特征图，这一过程可以描述为：

$$\begin{cases} \boldsymbol{x}_1 = Conv_1(\boldsymbol{x}) \\ \boldsymbol{x}_2 = RepPDC_1(\boldsymbol{x}_1) \\ \boldsymbol{x}_3 = RepPDC_2(\boldsymbol{x}_2) \\ \boldsymbol{y} = Conv_2(Concat(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)) \end{cases} \tag{5}$$

本文使用 RepPDCB 作为颈部网络中的特征提取模块，使颈部网络向检测头输入的特征图包含更丰富的长距离上下文信息。

1.4　基于通道注意力的特征双向融合模块

海洋生物识别任务中可能会出现由于水下图像背景较为复杂，其中可能包含冗余特征，导致模型提取的特征图的质量受到影响，进而影响模型的识别精度的情况。为使模型更加关注有效的特征，抑制冗余特征对模型造成的干扰，本文构建了基于通道注意力的双向特征融合模块（channel attention fusion module，CAFM），并替代颈部网络中的拼接操作，其结构如图 6 所示。CAFM 主要作用在于通过通道注意力机制从 RepPDCB 模块和主干网络输出的特征图中捕捉并利用重要的上下文信息，引导模型学习目标的有效信息，并通过将注意力特征图与另一分支的输入特征图进行相加，使特征之间的信息传递更加充分，从而进一步增强融合后的特征图的表达能力。

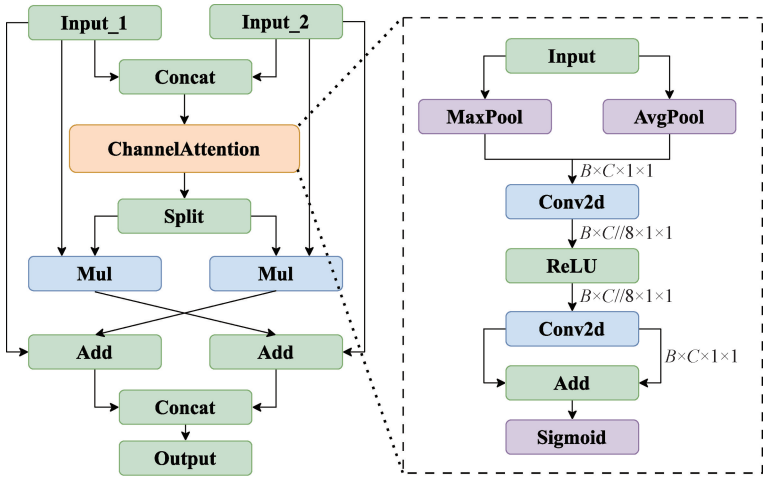


图 6　CAFM 结构

Fig. 6　The structure of CAFM

CAFM 首先对将两个输入特征图拼接，并通过通道注意力模块获取各个通道的权重，分别记两个输入特征图为 $\boldsymbol{F}_1$ 和 $\boldsymbol{F}_2$，则获得的通道权重 $\boldsymbol{W}$ 为：

$$\boldsymbol{W} = ChannelAttention(Concat(\boldsymbol{F}_1, \boldsymbol{F}_2)) \tag{6}$$

然后，将获得的权重按 $\boldsymbol{F}_1$ 和 $\boldsymbol{F}_2$ 的通道数进行拆分，得到两个部分的通道权重 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$，并将两个输入特征与各自的通道权重进行相乘，再与另一方的输入特征图进行相加，得到增强后的包含注意力的特征图 $\boldsymbol{F}'_1$ 和 $\boldsymbol{F}'_2$：

$$\boldsymbol{F}'_1 = \boldsymbol{F}_1 \oplus (\boldsymbol{F}_2 \otimes \boldsymbol{W}_2) \tag{7}$$

$$\boldsymbol{F}'_2 = \boldsymbol{F}_2 \oplus (\boldsymbol{F}_1 \otimes \boldsymbol{W}_1) \tag{8}$$

最后将 $\boldsymbol{F}'_1$ 和 $\boldsymbol{F}'_2$ 进行拼接，得到最终的输出结果 $\boldsymbol{O}$：

$$\boldsymbol{O} = Concat(\boldsymbol{F}'_1, \boldsymbol{F}'_2) \tag{9}$$

通道注意力模块基于通道优先卷积注意力机制[17]进行构建。在通道注意力模块中，输入特征图将分别经过最大池化层以及平均池化层进行压缩，以尽可能保留原输入特征图的信息，然后通过两个逐点卷积以及一个 ReLU 激

活函数在压缩通道数的同时获取两个特征图的权重,将其相加后进行归一化,得到各个通道的权重。记输入特征图为 $\boldsymbol{x}$,输出通道权重为 $\boldsymbol{y}$,两个二维卷积分别为 C_1 和 C_2,则这一过程可以描述为:

$$\begin{cases}\boldsymbol{x}_1 = C_1(ReLU(C_2(AvgPool(\boldsymbol{x}))))\\ \boldsymbol{x}_2 = C_1(ReLU(C_2(MaxPool(\boldsymbol{x}))))\\ \boldsymbol{y} = Sigmoid(\boldsymbol{x}_1 \oplus \boldsymbol{x}_2)\end{cases} \tag{10}$$

2 实验与结果分析

2.1 所用数据集

本文使用的数据集为 RUOD 数据集[18]。RUOD 数据集由 Fu 等整理并发布,总共包含 14 000 张图像,其中 9 800 张为训练图像,4 200 张为测试图像。RUOD 数据集包含了九种常见的浅水海洋生物,分别为海胆、海参、海星、扇贝、普通鱼类、珊瑚、墨鱼、海龟以及水母,且包含了潜水员。本文将训练集按照 8∶2 的方式随机划分为训练集和验证集,最终训练集中含有 7 920 张图像,验证集中含有 1 980 张图像。

2.2 实验环境与训练相关设置

本文所作改进基于 RT-DETR 官方仓库实现,训练所使用机器操作系统为 ubuntu20.04,使用显卡为 RTX4090,处理器为 AMD EPYC 9654 96-Core Processor,Pytorch 版本为 2.0.1,CUDA 版本为 11.8。训练设置基本采用 RT-DETR 的默认训练设置,初始学习率设置为 0.000 1,Batch Size 设置为 8,优化器使用 AdamW,训练轮次设置为 150,输入图像大小调整为 640×640。由于 DETR 系列模型较为依赖主干网络的预训练权重,在不使用预训练权重的情况下训练效果较差,且长时间不能收敛,因此本文为所有 DETR 系列模型的主干网络载入了在 ImageNet-1K 上进行训练得到的预训练权重,以保证模型正常训练。

2.3 评价指标

本文采用所有类别目标的平均准确率的平均值(mean average precision, mAP)、模型参数量以及单幅图像的平均推理时间作为模型综合性能的评价指标,并使用了 TIDE 错误分析工具[19]对模型性能进行分析,其中 mAP 的定义如下所示:

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{TP}{TP + FN} \tag{12}$$

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{13}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{14}$$

式中:P 表示查准率,R 表示查全率,其针对每一类目标分别计算。对每一类目标而言,真阳性(true positive,TP),表示该类目标中,与真实框交并比大于给定阈值的预测框数;假阳性(false positive,FP)表示该类目标中,与真实框交并比低于给定阈值,或多余的预测框数;假阴性(false negative,FN)表示该类目标中,没有被检测到的目标数量。在式(13)中,$P(R)$ 即为 $P-R$ 曲线,单一类别目标的平均准确率(average precision, AP)定义为该类目标的 $P-R$ 曲线下方的面积。在式(14)中,N 表示数据集中的目标类别数,AP_i 表示第 i 类目标的平均准确率。

本文计算 mAP 时交并比阈值分别取为 0.5 以及 0.75,记为 mAP50 和 mAP75;并对所有大于等于 0.5,小于等于 0.95,且间隔为 0.05 的交并比阈值分别计算 mAP,然后计算这些 mAP 的平均值,记为 mAP50:95。本文实验得出的 mAP 均使用 Pycocotools(目标检测的相关工具)计算得到。

TIDE 包含多种指标,其分别计算单独将错误分类(CLS)、错误定位(LOC)、错误定位及错误分类(BOTH)、预测框重叠(DUP)、误检(BKGD)、漏检(MISS),以及假阳性和假阴性的错误修正(即认为是正确的)之后,模型的 mAP50 的增加量。若某类错误的增加量更小,则说明该类错误对模型的精度影响更小。

2.4 实验结果及分析

表 1 展示了本文方法与其他方法的对比。表中推理速度由 TensorRT(英伟达官方推理库)测量得到,测速时使用的所有 YOLO 系列方法均插入了高效非极大值抑制模块。由表可见,本文方法相比改进前的方法,在 mAP50 上提升了 1.0 个百分点,在 mAP75 上提升了 1.9 个百分点,在 mAP50:95 上提升了 1.6 个百分点,且参数量减少了约 1.1 M。相较于 YOLO 系列目前最常见的方法 YOLOv8m,本文方法在 mAP50 上提升了 2.2 个百分点,在 AP75 上提升了 3.3 个百分点,在 mAP50:95 上提升了 2.8 个百分点,且参数量减少了约 7.0 M;相较于 YOLOv6m[20],本文方法在 mAP50 上提升了 1.0 个百分点,在 mAP75 上提升了 1.2 个百分点,在 AP50:95 上提升了 1.4 个百分点。相较于针对海洋生物识别进行优化的方法 UW-YOLOv8[21]以及 DyFish-DETR[22],本文方法在精度方面同样拥有较大优势。此外,本文使用改进的方法在 COCO2017 数据集[23]上进行了 36 个轮次的训练,并在 RUOD 数据集上进行了微调,结果表明,经过预训练后,本文方法仅用 50 个轮次的微调就达到了原方法的精度,大幅节省了训练需要的时间,如表 1 中的最后一行所示。

在推理速度方面,当 Batch Size 为 1 时,本文方法对单幅图像的平均推理时间为 1.511 ms,相比基准模型增加了约 0.5 ms,速度降低的主要原因来源于主干网络 RepViT 较大的深度,其中使用的深度可分离卷积虽然参数量较少,但对速度的提升作用并不明显;此外,RepPDCB 模块由于采用了两个重参数化膨胀卷积,使得模型的参数量和

表 1　不同方法比较

Table 1　Comparison of different methods

方法	训练轮次	mAP50/%	mAP75/%	mAP50:95/%	参数量(10^6)	单张图像平均推理时间(Batch Size=1)/ms	单张图像平均推理时间(Batch Size=4)/ms
YOLOv3t[24]	300	80.7	56.8	52.4	12.1	1.755	0.449
YOLOv5m	300	85.3	65.9	60.9	20.9	2.598	1.113
YOLOv6m	400	86.5	69.7	63.5	34.8	2.401	0.914
YOLOv8m	500	85.3	67.6	62.1	25.9	2.439	0.969
DINO-R50	50	85.2	65.0	60.2	47.6	—	—
Deformable-DETR-R50[25]	50	83.7	61.7	56.2	40.1	—	—
UW-YOLOv8	—	86.8	—	—	—	—	—
DyFish-DETR	—	83.2	—	57.9	—	—	—
RT-DETR-R18	150	86.5	68.9	63.3	20.1	1.012	0.613
本文	150	87.5	70.9	64.9	18.9	1.511	0.825
本文改进方法	50	87.5	70.7	64.9	18.9	1.511	0.825

计算量有所增加，同样对推理速度造成了影响。相较于需要进行非极大值抑制操作的 YOLO 系列方法，本文方法对单幅图像的推理时间相比 YOLOv8m 减少了约 1 ms，仅为 YOLOv8m 的约百分之六十，相比 YOLOv6m 减少了约 0.9 ms，可见本文方法在检测精度明显高于 YOLO 系列方法的同时，推理速度也明显更快，相比 YOLO 系列方法更适合部署于海洋生物探测设备中。对于其他 DETR 系列模型，由于其并非实时检测模型，推理速度较慢，因此本文没有对其进行测速。

表 2 展示了本文方法对每一类目标的识别精度(取交并比阈值为 0.5～0.95，且间隔为 0.05 的平均准确率的平均值)。相比改进前的方法，本文方法对所有类别的海洋生物以及潜水员的识别精度均有所提升。在鱼类和珊瑚这两种常见海洋生物的识别上，本文方法提升最为明显，相比改进前的方法分别提升了 2.3 以及 2.0 个百分点。相比 YOLOv8m，本文方法在鱼类、墨鱼、海龟以及水母的识别上具有较大优势，分别提升了 6.0、3.9、3.8 以及 5.3 个百分点。

表 2　各类目标识别精度

Table 2　Detection accuracy for all types of targets in RUOD　　%

方法	海参	海胆	扇贝	海星	鱼类	珊瑚	潜水员	墨鱼	海龟	水母
YOLOv5m	49.5	52.2	53.1	55.6	49.7	52.4	71.2	80.0	78.8	57.4
YOLOv6m	52.6	54.6	53.9	57.6	53.9	56.5	74.5	84.6	84.8	61.0
YOLOv8m	50.6	52.5	53.7	55.5	52.2	55.2	73.9	83.1	83.4	61.2
RT-DETR-R18	51.1	52.4	53.7	56.0	55.9	55.2	73.9	85.9	85.3	64.6
本文	52.0	54.2	54.3	56.8	58.2	57.2	75.4	87.0	87.2	66.59

2.5　消融实验

表 3 展示了本文方法消融实验的结果。由表可见，在使用 RepViT-M1.1 作为本文方法的特征提取网络后，模型的特征提取能力得到明显提升，mAP75 提高了 1.0 个百分点，mAP50:95 提高了 0.7 个百分点，且参数量降低了约 4 M；在此基础上将 RepPDCB 引入颈部网络后，模型的感受野得到提升，mAP75 进一步提升了 0.4 个百分点，mAP50:95 进一步提升了 0.7 个百分点；此外，在单独将 RepPDCB 模块引入颈部网络后，模型的 mAP75 以及 mAP50:95 均提升了 0.6 个百分点，验证了较大的感受野对于浅水海洋生物识别的有效性；最后，在前两种改进的基础上引入基于通道注意力的特征拼接模块，提升模型对重点信息的关注能力后，mAP75 进一步提升了 0.6 个百分点，mAP50:95 进一步提升了 0.2 个百分点。实验结果表明本文所做的改进均有效提升了模型对海洋生物的检测精度，且没有明显影响模型推理的实时性。

2.6　TIDE 错误分析

表 4 展示了本文方法与改进前的方法在使用 TIDE 错误分析工具进行分析后得到的结果。由表可见，改进前的方法在分别修正分类错误、定位错误、重叠框错误、误检以及漏检后，其 mAP50 的增加量均多于本文方法；表明本文方法受单一错误影响更小，拥有更强的鲁棒性。

表 3　消融实验

Table 3　Results of ablation experiment

基准模型	RepViT	RepPDCB	CAFM	mAP50/%	mAP75/%	mAP50:95/%	参数量(10^6)	单张图像平均推理时间(Batch Size=1)/ms
RT-DETR	×	×	×	86.5	68.9	63.3	20.1	1.012
RT-DETR	√	×	×	87.2	69.9	64.0	16.3	1.207
RT-DETR	×	√	×	87.0	69.5	63.9	22.5	1.220
RT-DETR	×	×	√	86.6	69.0	63.4	20.4	1.103
RT-DETR	√	√	×	87.6	70.3	64.7	18.7	1.402
RT-DETR	√	√	√	87.5	70.9	64.9	18.9	1.511

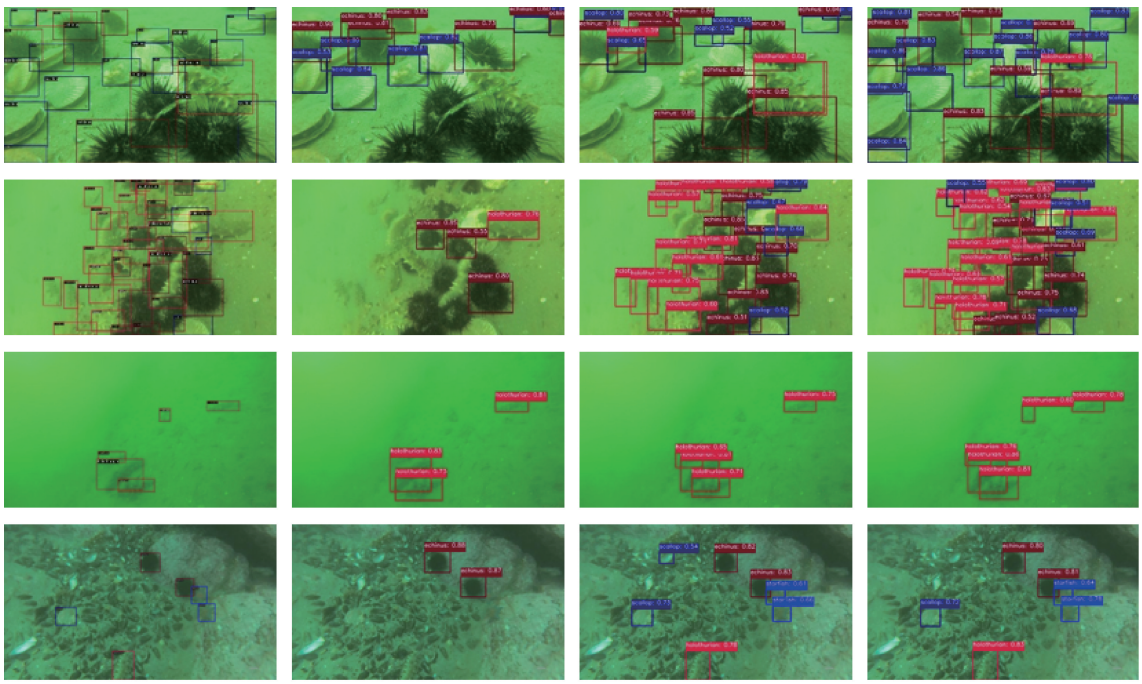
表 4　TIDE 指标

Table 4　Results of TIDE　　%

类别	RT-DETR-ResNet18	本文
CLS	0.66	0.60
LOC	2.17	2.07
BOTH	0.26	0.27
DUP	0.30	0.25
BKGD	4.55	4.30
MISS	1.56	1.34
FP	9.12	8.55
FN	3.37	3.07

2.7　识别结果对比

图 7 (a)～(d)分别展示了原始图片,以及本文方法与其他方法在过滤掉置信度低于 0.5 的预测框后的识别结果,识别时使用的模型均通过 TensorRT 转换为 engine 格式(优化后的模型格式,常用于部署到边缘设备中)。由图可见,当海洋生物之间出现大量聚集且相互遮挡,以及图像能见度较低的情况时,改进前的方法相对更容易出现漏检的情况,而本文方法漏检的目标数量明显更少;此外,当图像背景较为复杂,含有较多冗余特征时,更容易出现误检的情况,以较高的置信度将背景的碎石块识别为海洋生物,而本文方法认为其为海洋生物的概率低于 0.5,更不容易出现误检的情况,体现了本文改进方法的有效性。



(a) 原图 (a) Original image　(b) YOLOv8m　(c) RT-DETR-ResNet18　(d) 本文 (d) Ours

图 7　识别效果对比

Fig. 7　Comparison of detection results

3　结　　论

针对目前浅水海洋生物识别方法精度较低的问题，本文以 RT-DETR-ResNet18 为基准模型，提出了一种改进的水下目标实时检测方法 RDC-RT-DETR，具体而言，本文使用基于结构重参数化的网络 RepViT 作为本文方法的特征提取网络，提出了重参数化并行膨胀卷积 RepPDC，并基于 RepPDC 构建了特征提取模块 RepPDCB，最后基于通道注意力机制构建了双向特征融合模块。实验结果表明，本文所提方法拥有更少的参数量，推理速度满足实时性需求，且识别精度达到先进水平，有望应用到实际海洋生物识别任务中。

未来将在继续在此方法的基础上进行优化。考虑到现有水下探测设备大多为小型设备，利用合适的剪枝方法进一步降低模型的参数量以及计算量，使其能够广泛运用于小型水下探测设备上，是下一步的主要研究方向。

参考文献

[1] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015: 1440-1448.

[2] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. ArXiv preprint arXiv: 1506.01497, 2015.

[3] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, 2016: 21-37.

[4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779-788.

[5] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.

[6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv preprint arXiv: 1706.03762, 2017.

[7] ZHAO Y, LYU W Y, XU SH L, et al. DETRs beat YOLOs on real-time object detection[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

[8] 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 P F, HAN S, NI J J, et al. Underwater object detection algorithm combining data enhancement and improved YOLOv4 [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(3): 113-121.

[9] 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.

[10] LIU K Y, SUN Q, SUN D M, et al. Underwater target detection based on improved YOLOv7 [J]. Journal of Marine Science and Engineering, 2023, 11(3): 677.

[11] 周新，张春堂，樊春玲. 基于 YOLOv5_PGS 的轻量级水下生物识别目标检测[J]. 电子测量技术，2024，46(21)：168-175.
ZHOU X, ZHANG CH T, FAN CH L. Lightweight YOLOv5_PGS based objective detection for underwater biological identification [J]. Electric Measurement Technology, 2024, 46(21): 168-175.

[12] WANG AO, CHEN H, LIN Z J, et al. RepViT: Revisiting mobile CNN from ViT perspective [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 15909-15920.

[13] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions [J]. ArXiv preprint arXiv: 1511.07122, 2015.

[14] ZHANG H, LI F, LIU SH L, et al. DINO: DETR with improved denoising anchor boxes for end-to-end object detection [J]. ArXiv preprint arXiv: 2203.03605, 2022.

[15] DING X H, ZHANG X Y, MA N N, et al. RepVGG: Making VGG-style ConvNets great again [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13733-13742.

[16] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information [J]. ArXiv preprint arXiv: 2402.13616, 2024.

[17] HUANG H J, CHEN Z G, ZOU Y, et al. Channel prior convolutional attention for medical image segmentation[J]. Computers in Biology and Medicine, 2024, 178: 108784.

[18] FU CH P, LIU R SH, FAN X, et al. Rethinking general underwater object detection: Datasets, challenges, and solutions[J]. Neurocomputing, 2023, 517: 243-256.

[19] BOLYA D, FOLEY S, HAYS J, et al. Tide: A general toolbox for identifying object detection errors [C]. Computer Vision-ECCV 2020: 16th European Conference, 2020: 558-573.

[20] LI CH Y, LI L L, JIANG H L, et al. YOLOv6: A

single-stage object detection framework for industrial applications [J]. ArXiv preprint arXiv: 2209.02976, 2022.

[21] GUO AN, SUN K Q, ZHANG Z Y. A lightweight YOLOv8 integrating FasterNet for real-time underwater object detection[J]. Journal of Real-Time Image Processing, 2024, 21(2): 49.

[22] WANG ZH W, RUAN ZH K, CHEN CH. DyFish-DETR: Underwater fish image recognition based on detection transformer[J]. Journal of Marine Science and Engineering, 2024, 12(6): 864.

[23] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]. Computer Vision-ECCV 2014: 13th European Conference, 2014: 740-755.

[24] REDMON J, FARHADI A. YOLOV3: An incremental improvement[J]. ArXiv preprint arXiv: 1804.02767, 2018.

[25] ZHU X ZH, SU W J, LU L W, et al. Deformable DETR: Deformable transformers for end-to-end object detection[J]. ArXiv preprint arXiv:2010.04159, 2020.

作者简介

蒋智臣,硕士研究生,主要研究方向为深度学习、目标检测。

E-mail:451560812@qq.com

胡俐蕊(通信作者),博士,教授,主要研究方向为嵌入式系统、图像识别、深度学习。

E-mail:hulr163@163.com