

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

多级特征筛选和任务动态对齐的声呐图像小目标检测

王燕, 王宏辉, 刘树东, 张艳, 郝泽玉

引用本文:

王燕, 王宏辉, 刘树东, 等. 多级特征筛选和任务动态对齐的声呐图像小目标检测[J]. *光电工程*, 2024, **51**(10): 240196.

Wang Y, Wang H H, Liu S D, et al. Small target detection in sonar images with multilevel feature screening and task dynamic alignment[J]. *Opto-Electron Eng*, 2024, **51**(10): 240196.

<https://doi.org/10.12086/oe.2024.240196>

收稿日期: 2024-08-20; 修改日期: 2024-09-21; 录用日期: 2024-09-21

相关论文

基于ZYNQ的轻量化YOLOv5声呐图像目标检测算法及实现

赵冬冬, 谢墩翰, 陈朋, 梁荣华, 沈伊, 郭新新

光电工程 2024, **51**(1): 230284 doi: [10.12086/oe.2024.230284](https://doi.org/10.12086/oe.2024.230284)

基于残差和注意力网络的声呐图像去噪方法

赵冬冬, 叶逸飞, 陈朋, 梁荣华, 蔡天诚, 郭新新

光电工程 2023, **50**(6): 230017 doi: [10.12086/oe.2023.230017](https://doi.org/10.12086/oe.2023.230017)

更多相关论文见光电期刊集群网站 



<http://cn.ojournal.org/oe>

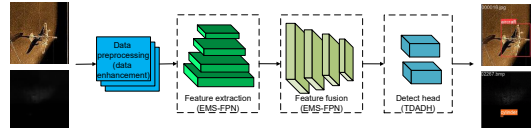


 OE_Journal



Website

多级特征筛选和任务动态对齐的声呐图像小目标检测



王燕, 王宏辉, 刘树东*, 张艳, 郝泽玉

天津城建大学计算机与信息工程学院, 天津, 300000

摘要: 针对声呐图像中小目标检测难度大、精度低、容易出现错检漏检的问题, 本文提出一种基于 YOLOv8s 的声呐图像小目标检测改进算法。首先, 考虑到声呐图像中的小目标通常具有低对比度且易被噪声淹没, 提出了高效多级筛选特征金字塔网络 (EMS-FPN)。其次, 由于解耦头的分类分支和定位分支是独立的, 会增加模型的参数量, 同时难以有效地适应不同尺度目标的检测需求, 导致对于小目标的检测效果不佳, 设计了任务动态对齐检测头模块 (TDADH)。最后为了验证本文模型的有效性, 在 URPC2021 和 SCTD 扩充声呐数据集上进行了相应的验证, mAP0.5 较 YOLOv8s 分别提高了 0.3% 和 1.8%, 参数量降低了 22.5%。结果表明, 本文提出的方法在声呐图像目标检测任务中不仅提高了精度, 还显著降低了模型参数量。

关键词: 水下目标检测; 声呐图像; 小目标检测; 高效多级筛选; 任务动态对齐; 轻量化

中图分类号: TP391.4

文献标志码: A

王燕, 王宏辉, 刘树东, 等. 多级特征筛选和任务动态对齐的声呐图像小目标检测 [J]. 光电工程, 2024, 51(10): 240196

Wang Y, Wang H H, Liu S D, et al. Small target detection in sonar images with multilevel feature screening and task dynamic alignment[J]. *Opto-Electron Eng*, 2024, 51(10): 240196

Small target detection in sonar images with multilevel feature screening and task dynamic alignment

Wang Yan, Wang Honghui, Liu Shudong*, Zhang Yan, Hao Zeyu

School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300000, China

Abstract: To solve the problem of small target detection in sonar images, which is difficult, low precision, and prone to misdetection and omission detection, this paper proposes an improved algorithm for small target detection in sonar images based on YOLOv8s. Firstly, considering that small targets in sonar images usually have low contrast and are easily overwhelmed by noise, an efficient multi-level screening feature pyramid network (EMS-FPN) is proposed. Secondly, since the classification branch and localization branch of the decoupled head are independent, which will increase the number of parameters of the model, and at the same time, it is difficult to effectively adapt to the detection needs of targets of different scales, resulting in poor detection of small targets, the task dynamic alignment detection head module (TDADH) is designed. Finally, to verify the effectiveness of the model in this

收稿日期: 2024-08-20; 修回日期: 2024-09-21; 录用日期: 2024-09-21

基金项目: 天津市哲学社会科学规划项目 (TJGL19XSX-045)

*通信作者: 刘树东, liushudong@tcu.edu.cn。

版权所有©2024 中国科学院光电技术研究所

paper, the corresponding validation was carried out on URPC2021 and SCTD expanded sonar dataset, mAP0.5 improved by 0.3% and 1.8% compared with YOLOv8s, respectively, and the number of parameters was reduced by 22.5%. The results show that the method proposed in this paper not only improves the accuracy but also significantly reduces the number of model parameters in the task of target detection in sonar images.

Keywords: underwater target detection; sonar image; small target detection; efficient multilevel screening; task dynamic alignment; lightweighting

1 引言

近年来, 利用目标检测技术对声呐图像中的目标进行定位和识别是水声领域的一个热点问题, 涉及到对回波成像后的声纳数据进行处理和分析。它被广泛应用于水下地形识别^[1]、海洋生物特征识别^[2-3]、鱼雷探测^[4]以及一些水下目标检测任务^[5-7]等民用和军事领域。随着深度学习技术的发展和普及, 基于深度学习的小目标检测方法逐渐成为研究的热点。深度学习模型, 尤其是卷积神经网络 (CNN), 具有强大的特征提取和分类能力, 能够有效应对声呐图像中的噪声、复杂背景等问题, 为小目标的准确检测提供了新的思路和方法。

水下目标检测技术在准确定位和识别水下目标方面发挥着至关重要的作用^[8]。水下环境复杂多变, 给数据传输、数据可用性和目标检测^[9]带来了巨大挑战。提高检测精度和效率, 最大限度地降低计算成本以及减少模型参数量是亟待解决的问题。水下目标检测采用了多种成像技术, 声呐成像技术^[10-11]因其探测距离远、穿透能力强以及在恶劣水下条件下的稳定性而脱颖而出, 然而, 水声信道固有的复杂性, 加上声波传播损耗和散射等因素, 往往会导致图像分辨率低、噪声大、目标边缘不清晰^[12-13], 从而使声呐图像的进一步处理变得复杂^[14], 并且声呐成像受到海洋环境、水下目标特性等多种因素的影响, 小目标在声呐图像中的检测一直以来都是一个具有挑战性的问题。小目标, 如圆笼、球等, 在声呐图像中往往面临着信号弱、背景复杂、噪声干扰^[15]等困难, 对它们进行目标定位任务时, 当前主流算法的检测精度往往不足^[16], 传统 YOLOv8 模型对于低分辨率和高噪声的声呐图像, 其提取的特征可能不足以准确区分目标和背景, 导致误检率或漏检率增高, 且声呐图像中的目标通常与背景的对比如较低, 尤其是在复杂水下环境中, 声学回波可能导致目标边缘模糊或与背景融合, 在处理低对比度、弱边界的声呐图像目标时, 表现欠佳, 其有效

检测对于保障水下航行安全和海洋资源开发至关重要。

针对上述声呐图像的小目标检测存在的问题, 本文提出一种基于改进 YOLOv8s 的轻量化声呐小目标检测算法, 通过对颈部网络的改进, 设计了一种高效多级筛选特征融合金字塔 (EMS-FPN), 该模块可以通过筛选机制突出重要特征, 抑制不相关的背景噪声, 实现多级融合, 从而提升对小目标检测能力的同时降低了网络模型的参数量; 并对网络的检测头进行改进, 设计了一种任务动态对齐检测头 (task dynamic align detection head, TDADH), 解决了解耦头使用独立的分类和定位分支导致两个任务之间缺乏交互的问题, 以有效适应不同尺度目标的检测需求, 能够更好地适应实时检测的要求。

2 YOLOv8 目标检测算法

YOLOv8 目标检测算法是在继承了前代算法优点的基础上, 引入了多项关键创新以提高其性能, 主要改进包括: 1) 提出了一个新的骨干网络; 2) 采用了 Anchor-Free 检测头; 3) 加入了新的损失函数。骨干网络从 YOLOv7 的 ELAN 设计中汲取灵感, 用 C2f 结构取代了 YOLOv5 的 C3 结构, 增强了梯度流。YOLOv7 针对不同的模型调整通道, 进一步提高了性能。Head 部分也发生了重大变化, 采用了解耦头结构, 将分类头和检测头分离, 并将 Anchor-Base 替换为 Anchor-Free。此外, 损失函数采用了任务对齐分配器的动态正样本分配策略, 并加入了分布焦点损失函数。YOLOv8 还采用了 Mosaic 数据扩增^[17]训练技术, 与 YOLOX^[18]类似, 在最后 10 次迭代中关闭了 Mosaic 增强操作, 进一步提高了准确率。得益于这些改进, YOLOv8 比 YOLOv5 获得了更高的精确度, 并提供了五种不同比例的模型: N/S/M/L/X。由于声呐图像的检测需要适应水下航行器的能耗要求, 确保系统运行的实时性, 须同时考虑模型参数量和检测精度, 采用 YOLOv8s 网络模型, 其结构如图 1 所示。

3 改进 YOLOv8s 目标检测算法

3.1 高效多级筛选特征金字塔 (EMS-FPN)

由于声呐图像中的水下目标相对较小, 声呐图像中的小目标通常具有低对比度且易被噪声淹没, 为提高声呐图像中小目标检测的精度, 在网络的 Neck 部分加入了高层次筛选特征融合金字塔 (high-level screening-feature fusion pyramid networks, HS-FPN)^[19] 模块, 该模块可以通过筛选机制突出重要特征, 抑制不相关的背景噪声, 实现多级融合, 从而提升对小目标的检测能力。这个模型使用高级特征作为权重以经由信道关注模块过滤低级特征信息, 然后将筛选后的信息与高级特征合并, 从而增强模型的特征表达能力。它的基本原理包括两个关键部分: 特征选择模块和特

征融合模块。首先, 不同尺度的特征图在特征选择模块中经历筛选过程。然后, 通过选择性特征融合 (selective feature fusion, SFF) 机制, 这些特征图中的高级和低级信息协同地整合在一起。这种融合产生了具有丰富语义内容的特点, 这对于检测声呐图像目标的细微特征非常有用, 从而增强了模型的检测能力。在原始的 HS-FPN 模块中坐标注意力机制 (coordinate attention, CA)^[20] 模块起着关键作用, 然而, CA 模块也表现出明显的局限性, 源于其不足的泛化能力和对通道维度降低的负面影响。CA 模块的设计过程相对复杂, 涉及两个方向上特征图的多次分离与合并。尽管两个二维卷积增强了坐标信息, 但它们也减少了通道维度, 对生成的注意力产生了负面影响。此外, 将 BN (batch normalization)^[21] 层整合到 CA 中引入了

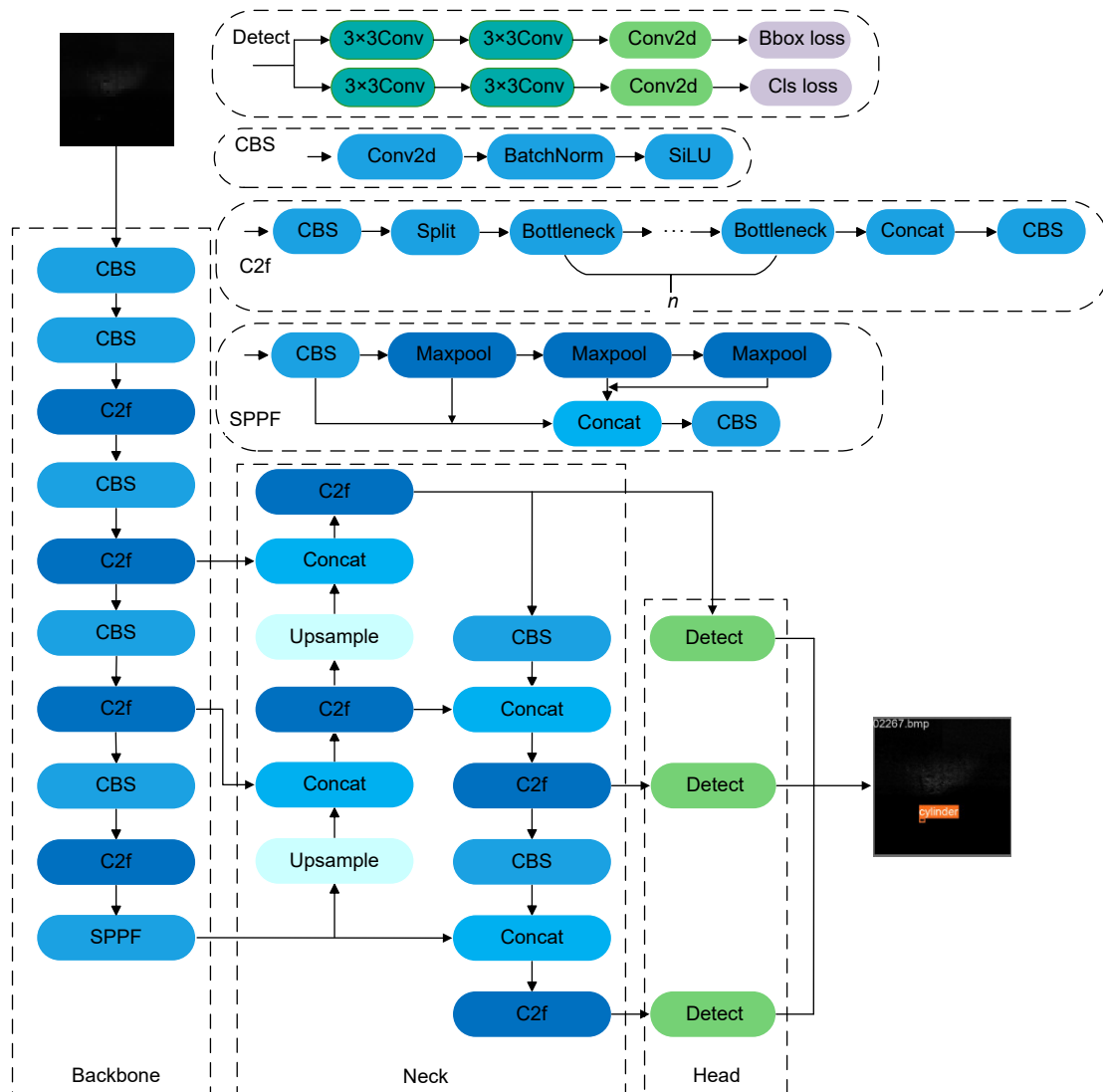


图 1 YOLOv8s 网络模型
Fig. 1 YOLOv8s network model

显著的弱点。例如，过小的 batch size 批次大小可能会对模型产生不利影响并阻碍其泛化能力。为此，利用 ELA (efficient local attention) 模块^[22] 替换 HS-FPN 中的 CA 模块，提出了高效多级筛选特征金字塔网络 (efficient multi-level screening feature pyramid network, EMS-FPN)，结构如图 2 所示，在特征选择模块中，它准确地捕获了感兴趣区域的位置，保持了输入特征图通道的维度及其轻量级特性。与 CA 相似，ELA 注意力机制在空间维度上采用 strip pooling^[23]，结构如图 3 所示，以获取水平和垂直方向的特征向量，保持狭窄的核形状以捕获长距离依赖关系，并防止不相关区域影响标签预测，从而在各自的方向上产生丰富的目标位置特征。ELA 注意力机制独立地处理上述每个方向上的特征向量以获得注意力预测，然后使用乘积运算将它们组合起来，确保感兴趣区域的准确位置信息。具体来说，在第二步中，应用一维卷积与两个特征向量局部交互，可以选择调整核大小来表示局部交互的覆盖范围。产生的特征向量经过分组归一化 (group normalization, GN)^[24] 和非线性激活函数处理，以生成两个方向的位置注意力预测。最终的位置注意力通过将两个方向的位置注意力相乘得到。与二维卷积相比，一维卷积更适合处理序列信号，并且更轻量、更快。GN 与 BN 相比，展现出更高的性能和更强的泛化能力。在特征融合之前，跨各种尺度的特征图的维度匹配至关重要，因为它们具有不同的通道数。为了实现这一点，使用一个卷积将每个尺度特征图的通道数减少到 256。

在特征融合模块中，Backbone 网络生成的多尺度特征图包含丰富的语义信息，但目标定位相对较差。相反，低尺度特征提供精确的目标位置，但包含有限

语义信息。解决这个困境的常见方法是直接将上采样的高级特征和低尺度特征像素值求和，为每个层增加语义信息。然而，这种技术没有进行特征选择，只是将多个特征层的像素值简单相加。为了解决这一局限性，开发了 ELA-SFF 模块，结构如图 4 所示，该模块通过使用高级特征作为权重来过滤低尺度特征中包含的必要语义信息。

如图 4 所示，输入一个给定的高级特征 $f_{high} \in R^{C \times H \times W}$ 和一个低尺度特征 $f_{low} \in R^{C \times H_1 \times W_1}$ ，高级特征首先使用一个大小为 2、核大小为 3×3 的转置卷积 (T-Conv) 进行扩展，得到特征大小 $\hat{f}_{high} \in R^{C \times 2H \times 2W}$ 。然后，为了统一高级特征和低尺度特征的维度，使用双线性插值来向上或向下采样高级特征，得到特征 $f_{att} \in R^{C \times H_1 \times W_1}$ 。接下来，使用 ELA 模块将高级特征转换为相应的注意力权重，以过滤低尺度特征，在获得具有相同维度的特征后。最后，将过滤后的低尺度特征与高级特征融合，以增强模型的特征表示，并得到 $f_{out} \in R^{C \times H_1 \times W_1}$ 。式 (1) 和 (2) 说明了特征选择的融合过程：

$$f_{att} = BL(T - Conv(f_{high})), \tag{1}$$

$$f_{out} = f_{low} * CA(f_{att}) + f_{att} . \tag{2}$$

3.2 任务动态对齐检测头

由于 YOLOv8 中使用的目标检测器解耦头^[18] 通常使用独立的分类和定位分支，这会导致两个任务之间缺乏交互，分类分支负责预测物体的类别，而定位分支负责预测物体的位置，然而这两个分支之间往往缺乏有效的信息交流和协同工作机制。这导致了在一些复杂场景下，目标的定位准确性不高，甚至出现了

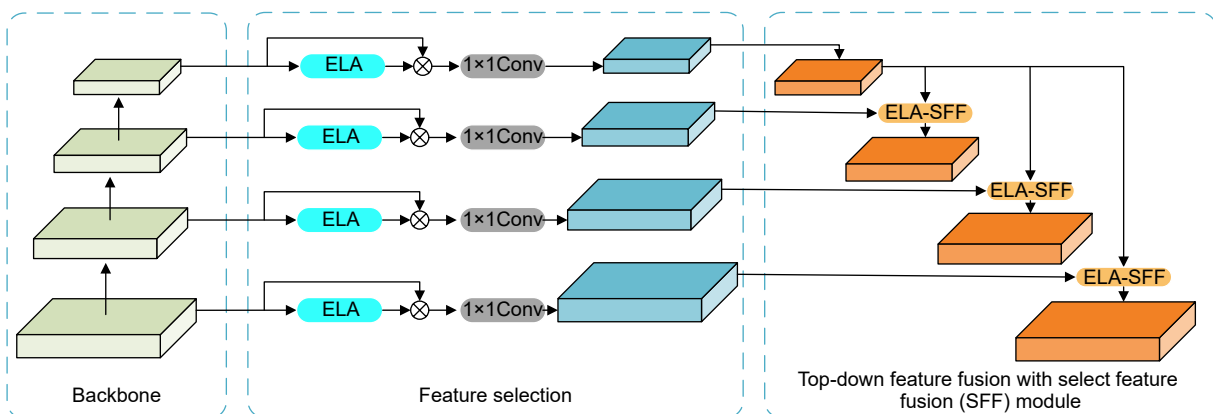


图 2 EMS-FPN 模块
Fig. 2 EMS-FPN module

误检和漏检的情况。其次，解耦头在处理目标尺度变化时存在一定的困难。由于分类分支和定位分支是独立的，往往难以有效地适应不同尺度目标的检测需求，导致对于小目标的检测效果不佳，容易出现漏检或误

检的情况。另外，解耦头往往需要额外的设计和调优工作。相比传统的一阶段目标检测器，解耦头通常需要额外设计分类分支和定位分支，并进行分开的训练和优化，这增加了模型的复杂度和训练的难度，同时

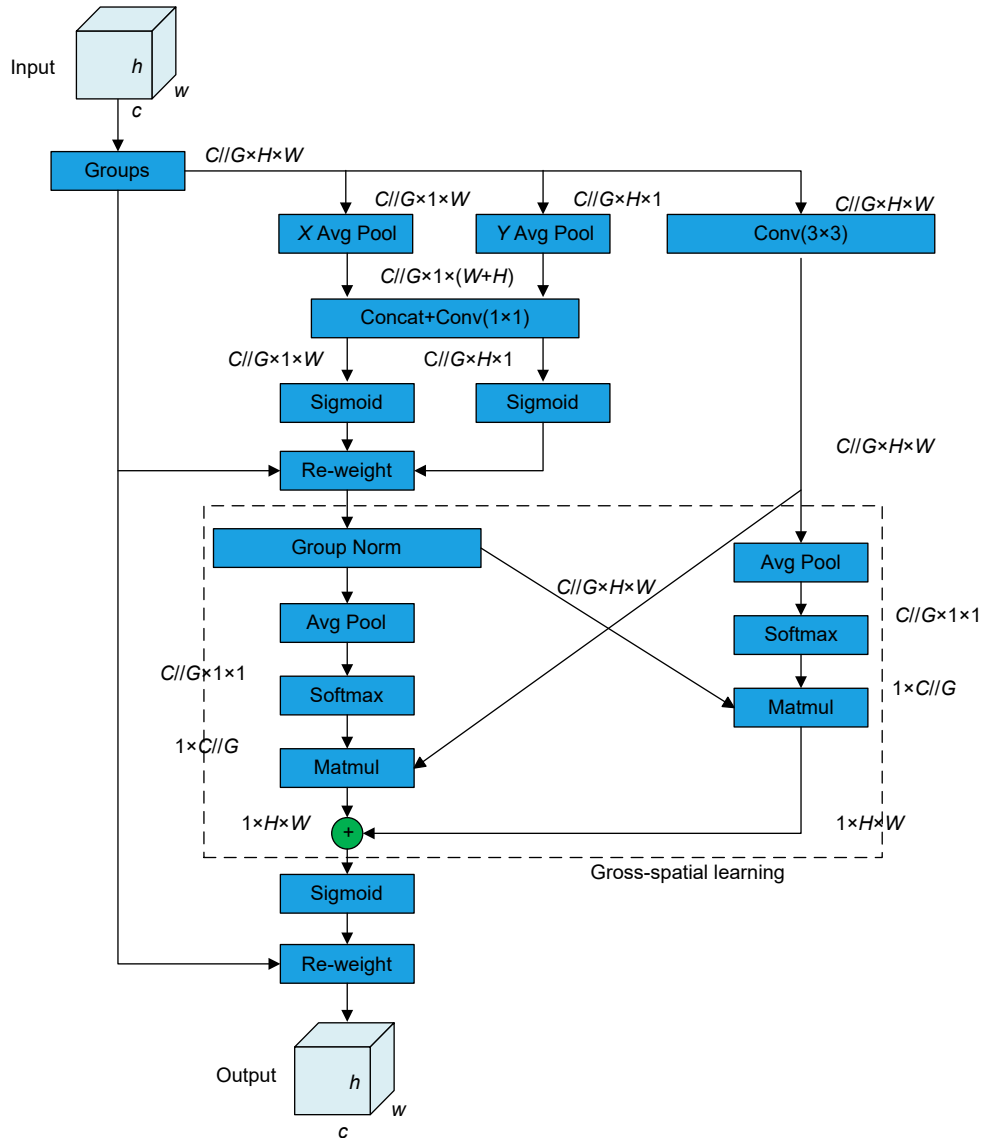


图 3 ELA 注意力机制

Fig. 3 ELA attention mechanisms

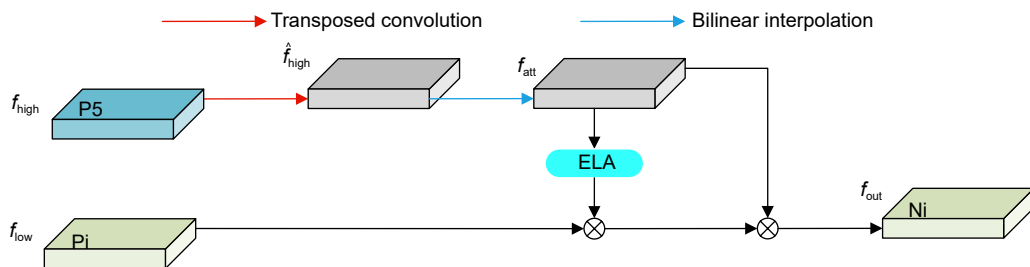


图 4 ELA-SFF 模块

Fig. 4 ELA-SFF module

也增加了调参的工作量。此外, 解耦头可能会增加模型的存储和计算成本。由于需要额外的分类分支和定位分支, 解耦头往往会增加模型的参数量和计算复杂度, 导致模型的存储和推理成本增加, 不利于模型的部署和实际应用。针对上述出现的问题对检测头进行了改进优化, 设计了任务动态对齐检测头 (task dynamic align detection head, TDADH) 模块, 提出了一个 GN-Conv 模块, 该模块将普通卷积的原始 BN^[21] 替换成 GN^[24], BN 是深度学习中常使用的归一化方法, 在提升训练以及收敛速度上发挥了重要的作用, 但是其仍然存在一些问题, 而 GN 解决了 BN 归一化对 batch size 依赖的影响, GN 与 BN 相比, 展现出更高的性能和更好的泛化能力。GN 在 FCOS^[25] 论文中已经证实可以提升检测头定位和分类的性能。为了降低检测头的参数量, 通过使用共享卷积, 可以大幅减少参数量, 这使得模型更轻量化, 特别是在资源受限的设备上, 并且在使用共享卷积的同时, 为了应对每个检测头所检测的目标尺度不一致的问题, 使用 Scale 层对特征进行缩放。TDADH 通过特征提取器从多个卷积层中学习任务交互特征, 得到联合特征, 定位分支使用 DCNv2 和交互特征生成 DCNv2 的 offset 和 mask, 分类分支使用交互特征进行动态特征选择, 动态特征选择由 1×1 Conv 和 3×3 Conv 组成, 结构如图 5 所示。

由于定位与分类的目标不同, 两类任务共用一种特征不可避免会带来特征冲突, 各自关注的特征也不相同。因此, 本文将任务交互特征通过任务分解 (task decomposition) 模块计算特征权重, 结构如图 6 所示, 动态计算任务交互特征促进定位和分类任务分

解。 w 是由跨层任务交互特征计算而来, 捕获层与层之间的依赖关系。

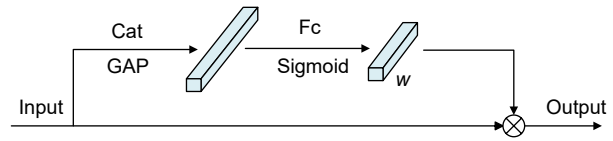


图 6 任务分解模块

Fig. 6 Task decomposition module

基于上述 YOLOv8s 模型的改进, 针对声呐图像中小目标的检测, 提出全新的 YOLOv8s 模型, 模型结构如图 7 所示。

4 实验与分析

4.1 声呐图像数据集

本文选取了 URPC2021 公共数据集进行实验, URPC2021 是一个水下声呐数据集^[26], 该数据集包括人体、球、圆笼、方笼、轮胎、金属桶、立方体与圆柱体八种类别, 共选取 4000 张声呐图像, 其中 3200 张图像作为训练集, 800 张作为测试集进行实验。为验证本文算法的有效性, 选取了声呐通用目标检测数据集 (SCTD)^[27] 对本文算法进行验证, 该数据集有 357 张声呐图像, 由于数据量较少, 采用随机去除像素点、锐化处理、仿射变换、改变亮度、色调随机、图像翻转、加噪等方式随机增强将数据集扩充到 2142 张, 分别由飞机、人体和船 3 类目标组成, 图像数据增强结果如图 8 所示。

4.2 实验环境和参数配置

实验环境采用的硬件平台是 GPU 为 3080 显卡, 软件平台是 Ubuntu 操作系统, 深度学习框架为 Pytorch。

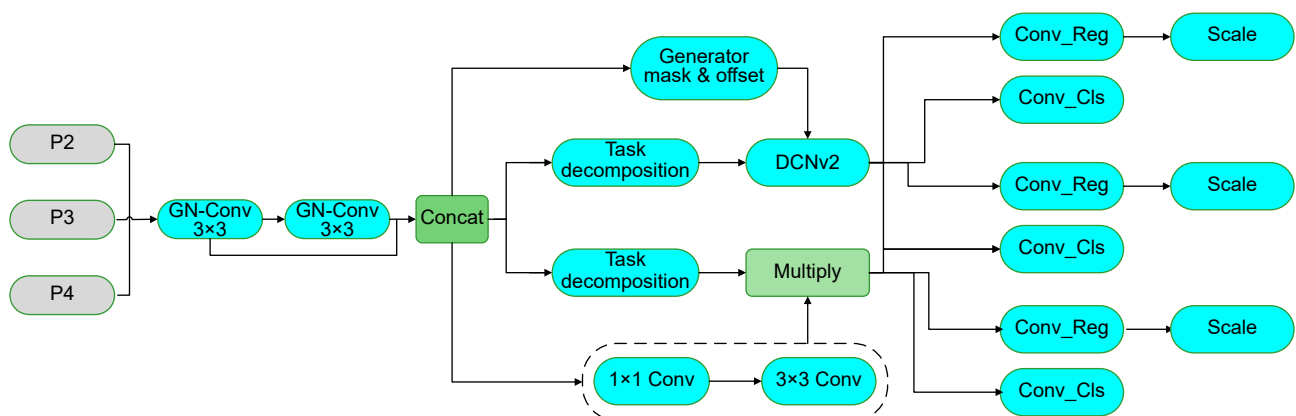


图 5 TDADH 模块

Fig. 5 TDADH module

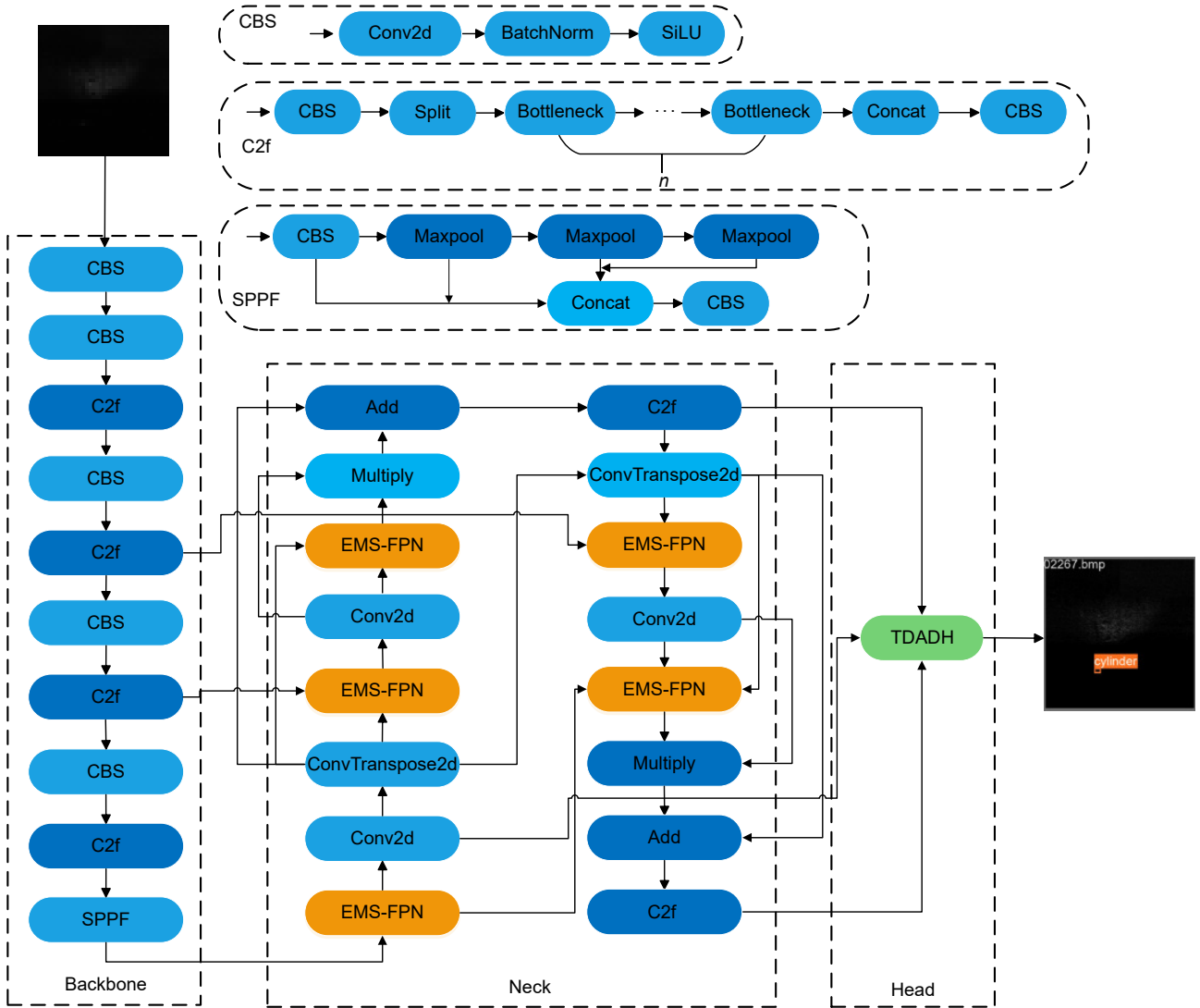


图 7 改进的 YOLOv8s 目标检测网络模型
Fig. 7 Improved YOLOv8s target detection network model

4.3 评价指标

在本文进行的实验中, 利用平均精度 (mAP)、精确度 (Precision)、召回率 (Recall)、参数量 (Params) 来评估各种模型的性能。mAP 衡量每个类别数据的平均精度, 在评估模型定位性能, 目标检测模型性能和分割模型性能方面是非常重要的。Precision 和 Recall 的计算如式 (3) 和 (4) 所示:

$$Precision = \frac{TP}{TP + FP}, \quad (3)$$

$$Recall = \frac{TP}{TP + FN}, \quad (4)$$

其中: TP 表示正确检测到的声呐目标数量; FP 表示已检测到声呐目标, 但检测不正确的数量; FN 表示未检测到声呐目标的数量。

4.4 实验结果与分析

4.4.1 消融实验

消融实验经常用来探索某些网络改进或者训练策略对网络模型性能的影响。针对本文提出的 EMS-FPN 和 TDADH 模块两部分改进方案, 为了验证其有效性, 设置了多组消融实验, 实验结果见表 1 和表 2。表 1 中“√”代表实验算法中使用了对应的模块, “×”代表算法中没有使用对应的模块。

从表 1 可以看出, 融合的改进模型 mAP0.5 提高了 0.3%, 参数量下降了 22.5%, 消融实验清晰地论证了本文模型在准确性和轻量化方面的优越性和有效性, 证明了设计方案应用于声呐图像目标检测的可行性。

从表 2 可以看出, 当加入 EMS-FPN 模块和

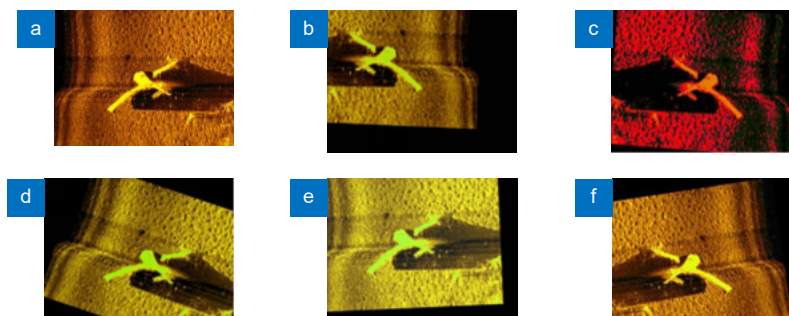


图 8 SCTD 声呐图像数据增强。(a) 原图; (b) 随机去除像素点; (c) 锐化处理; (d) 调整亮度; (e) 调节色调; (f) 图像翻转
Fig. 8 SCTD sonar image data enhancement. (a) Original image; (b) Random removal of pixel points; (c) Sharpening process; (d) Adjustment of brightness; (e) Adjustment of hue; (f) Image flipping

表 1 URPC2021 数据集消融实验结果

Table 1 Results of ablation experiments on URPC2021 dataset

Algorithm	HSFPN	EMS-FPN	TDADH	Precision/%	Recall/%	Params/M	mAP0.5/%
YOLOv8s	×	×	×	97.3	96.4	11.1	97.9
YOLOv8s	√	×	×	97.4	96.1	7.1	97.8
YOLOv8s	×	√	×	96.9	96.6	7.3	98.1
YOLOv8s	×	×	√	96.9	96.2	8.8	97.8
YOLOv8s	×	√	√	97.4	96.8	8.6	98.2

表 2 URPC2021 数据集中小目标类别消融实验结果

Table 2 Results of small target category ablation experiments in the URPC2021 dataset

Algorithm	HSFPN	EMS-FPN	TDADH	Params/M	Precision/%	Recall/%	Ball/%	Cylinder/%	Tyre/%
YOLOv8s	×	×	×	11.1	97.3	96.4	98.6	96.8	96.7
YOLOv8s	√	×	×	7.1	97.4	96.1	98.8	96.8	97
YOLOv8s	×	√	×	7.3	96.9	96.6	98.9	97.4	97.5
YOLOv8s	×	×	√	8.8	96.9	96.2	98.3	96.6	96.5
YOLOv8s	×	√	√	8.6	97.4	96.8	99.3	98.2	97.5

TDADH 检测头时, 数据集的小目标类别检测精度均有所提高, ball、cylinder、tyre 小目标类别的 mAP 分别提高了 0.7%、1.4%、0.8%, 消融实验清晰地论证了本文模型在小目标识别准确性和轻量化方面的优越作用效果。

4.4.2 不同算法对比实验分析

为了评估本文的模型在声呐目标检测性能上的性能, 在 URPC2021 数据集上进行了实验, 以评估本文模型与其他 SOTA 目标检测算法 (如 RetinaNet^[28]、PAA^[29]、CenterNet^[30]、Sparse-RCNN^[31]、YOLOF^[32]、TOOD^[33]、VarifocalNet^[34]、DSA-Net^[35]、YOLOv5、YOLOv6、YOLOv7、YOLOv8s) 的优越性。如表 3 所示, 很明显, 在所有对比方法中, 本文提出的方法表现出最好的声呐图像目标检测性能。

为了比较本文模型与其他算法在不同类别上的检测效果, 对 URPC2021 数据集上的测试结果进行了可视化实验, 比较结果如图 9、图 10 所示。从图中可以看出, 本文模型的检测效果最佳, 与其他算法相比, 本文模型在加入 EMS-FPN 模块和 TDADH 检测头的作用下, 可以通过筛选机制突出重要特征, 抑制不相关的背景噪声, 实现多级融合, 改进检测头部分的分类分支和定位分支的信息交流和协同工作机制得到加强, 有效地适应不同尺度目标的检测需求, 从而提升对小目标的检测能力, 对于检测小目标类别 (特别是 ball 和 cylinder) 方面表现出更优越的性能, 有效地提高了声呐图像中对小目标特征的关注。

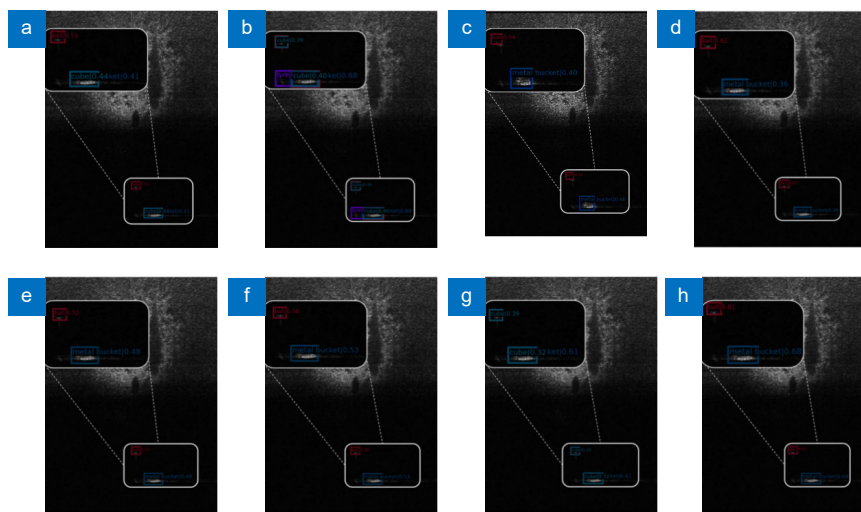
4.4.3 多数据集对比验证

为了验证本文算法的有效性, 在其他声呐图像数

表 3 URPC2021 数据集上不同水下物体精度与 mAP@50 与其他方法的比较

Table 3 Comparison of accuracy and mAP@50 for different underwater objects on URPC2021 dataset with other methods

Algorithm	Cube/%	Ball/%	Cylinder/%	Human body/%	Trye/%	Circle cage/%	Square cage/%	Metal bucket/%	mAP0.5/%
RetinaNet ^[28]	92.6	94.0	83.0	86.5	72.7	83.3	92.0	58.5	82.8
PAA ^[29]	95.3	93.6	88.2	94.1	73.7	91.5	95.1	81.4	89.1
CenterNet ^[30]	91.9	96.8	84.2	96.4	89.5	85.2	97.7	88.1	91.2
SparseRCNN ^[31]	97.4	96.8	89.0	97.6	89.2	93.1	98.5	88.0	93.7
YOLOF ^[32]	95.5	91.9	87.1	92.0	66.7	72.9	88.8	62.0	82.1
TOOD ^[33]	96.4	95.1	88.6	93.4	76.9	91.2	96.9	85.8	90.5
VarifocalNet ^[34]	96.4	94.6	90.1	95.0	79.5	91.5	96.7	85.4	91.2
DSA-Net ^[35]	97.9	98.0	94.2	98.4	94.9	93.2	99.2	94.0	96.2
YOLOv5	97.9	98.3	95.9	98.8	97.5	96.0	99.4	95.5	97.4
YOLOv6	98.0	98.7	96.2	99.4	96.9	96.1	99.3	97.4	97.7
YOLOv7	96.2	97.5	92.9	94.9	91.3	91.6	98.5	98.0	95.1
YOLOv8s	97.7	98.6	96.8	99.5	96.7	97.5	99.2	97.2	97.9
本文模型	97.7	99.3	98.2	99.5	97.5	97.3	98.7	97.8	98.2

图 9 URPC 2021 中声呐图像不同方法检测结果的比较。(a) RetinaNet^[28]; (b) PAA^[29]; (c) CenterNet^[30]; (d) SparseRCNN^[31]; (e) YOLOF^[32]; (f) TOOD^[33]; (g) VarifocalNet^[34]; (h) DSA-Net^[35]Fig. 9 Comparison of detection results of different methods for sonar images in URPC 2021. (a) RetinaNet^[28]; (b) PAA^[29]; (c) CenterNet^[30]; (d) SparseRCNN^[31]; (e) YOLOF^[32]; (f) TOOD^[33]; (g) VarifocalNet^[34]; (h) DSA-Net^[35]

数据集上也进行了相应的验证, 选择声呐通用目标检测数据集 (SCTD) 进行验证, 实验结果如表 4 所示。

5 结论

本文针对声呐图像分辨率低、存在噪声强度大和小目标检测精度不佳等问题, 提出了一种基于 YOLOv8s 的高效多级特征融合的轻量化声呐图像小目标检测算法, 首先为了解决声呐图像中小目标检测

结果不佳的问题, 设计了 EMS-FPN 模块, 该模块通过高层次筛选机制来筛选和融合特征, 从而增强特征表示能力。声呐图像中的小目标通常具有低对比度且易被噪声淹没, 该模块可以通过筛选机制突出重要特征, 抑制不相关的背景噪声, 从而提升对小目标的检测能力。其次由于解耦头的分类分支和定位分支是独立的, 往往难以有效地适应不同尺度目标的检测需求, 导致对于小目标的检测效果不佳, 容易出现漏检或误

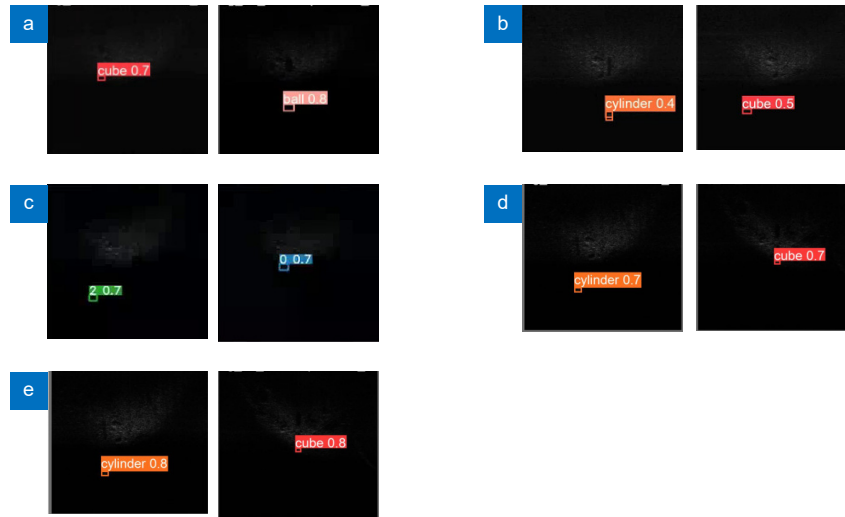


图 10 URPC 2021 中声呐图像不同方法检测结果的比较。(a) YOLOv5; (b) YOLOv6; (c) YOLOv7; (d) YOLOv8s; (e) 本文模型

Fig. 10 Comparison of detection results of different methods for sonar images in URPC 2021. (a) YOLOv5; (b) YOLOv6; (c) YOLOv7; (d) YOLOv8s; (e) This paper

表 4 SCTD 数据集对比实验结果

Table 4 Results of comparison experiments on SCTD dataset

Algorithm	HSFPN	EMS-FPN	TDADH	Params/M	Precision/%	Recall/%	Ship(small)/%	mAP0.5/%
YOLOv8s	×	×	×	11.1	94.0	92.4	93.4	96.3
YOLOv8s	√	×	×	7.1	95.2	91.5	95.4	97.0
YOLOv8s	×	√	×	7.3	96.1	92.1	95.6	97.0
YOLOv8s	×	×	√	8.8	97.4	94.1	97.3	97.4
YOLOv8s	×	√	√	8.6	97.1	94.4	97.8	98.1

检的情况。设计了任务动态对齐检测头模块，在降低模型参数数量的同时更好地提取小目标特征，提高了模型的检测精度。最后，为了验证本文算法的有效性，进行了多数据集验证，实验结果表明，本文提出的模型在降低模型参数数量的前提下提高了检测精度，以适用于水下声呐图像实时检测任务中。下一步的研究工作，将重点放在模型的适应性上，旨在针对不同类型的声呐系统进行优化，确保其能够在多种硬件配置下保持高效表现。此外，研究还将探索如何提升模型的实时处理能力，以满足实时监控和快速响应等实际应用场的需求，从而推动声呐技术在更多复杂环境中的广泛应用。

参考文献

- [1] Li H S, Xu C, Zhou T. High-resolution integrated detection of underwater topography and geomorphology based on multibeam interferometric echo sounder[J]. *Appl Mech Mater*, 2012, **212-213**: 345-350
- [2] Wang L, Ye X F, Wang S L, et al. ULO: an underwater light-weight object detector for edge computing[J]. *Machines*, 2022, **10**(8): 629.
- [3] Wang Z Y, Ye X F, Han Y T, et al. Improved real-time target detection algorithm for similar multiple targets in complex underwater environment based on YOLOv3[C]//*Global Oceans 2020: Singapore – U. S. Gulf Coast, Biloxi*, 2020: 1-6. <https://doi.org/10.1109/IEEECONF38699.2020.9389108>.
- [4] Lange H, Vincent L M. Advanced gray-scale morphological filters for the detection of sea mines in side-scan sonar imagery[J]. *Proc SPIE*, 2000, **4038**: 362-372.
- [5] Zhang W Y, Zhou T, Li J H, et al. An efficient method for detection and quantitation of underwater gas leakage based on a 300-kHz multibeam sonar[J]. *Remote Sens*, 2022, **14**(17): 4301.
- [6] Li J W, An W, Xu C, et al. Sunken oil detection and classification using MBES backscatter data[J]. *Mar Pollut Bull*, 2022, **180**: 113795.
- [7] Zhou T, Si J K, Wang L Y, et al. Automatic detection of underwater small targets using forward-looking sonar images[J].

- IEEE Trans Geosci Remote Sens*, 2022, **60**: 4207912.
- [8] Park C, Kim Y, Lee H, et al. Development of a 2 MHz sonar sensor for inspection of bridge substructures[J]. *Sensors*, 2018, **18**(4): 1222.
- [9] Zhao D D, Xie D H, Chen P, et al. Lightweight YOLOv5 sonar image object detection algorithm and implementation based on ZYNQ[J]. *Opto-Electron Eng*, 2024, **51**(1): 230284.
赵冬冬, 谢墩翰, 陈朋, 等. 基于 ZYNQ 的轻量化 YOLOv5 声呐图像目标检测算法及实现[J]. *光电工程*, 2024, **51**(1): 230284.
- [10] Abu A, Diamant R. A statistically-based method for the detection of underwater objects in sonar imagery[J]. *IEEE Sensors J*, 2019, **19**(16): 6858–6871.
- [11] Negahdaripour S. Application of forward-scan sonar stereo for 3-D scene reconstruction[J]. *IEEE J Oceanic Eng*, 2020, **45**(2): 547–562.
- [12] Shang Z G, Zhao C H, Wan J. Application of multi-resolution analysis in sonar image denoising[J]. *J Syst Eng Electron*, 2008, **19**(6): 1082–1089.
- [13] Jin Y, Ku B, Ahn J, et al. Nonhomogeneous noise removal from side-scan sonar images using structural sparsity[J]. *IEEE Geosci Remote Sens Lett*, 2019, **16**(8): 1215–1219.
- [14] Wang Z, Zhang S W, Huang W Z, et al. Sonar image target detection based on adaptive global feature enhancement network[J]. *IEEE Sensors J*, 2022, **22**(2): 1509–1530.
- [15] Zhao D D, Ye Y F, Chen P, et al. Sonar image denoising method based on residual and attention network[J]. *Opto-Electron Eng*, 2023, **50**(6): 230017.
赵冬冬, 叶逸飞, 陈朋, 等. 基于残差和注意力网络的声呐图像去噪方法[J]. *光电工程*, 2023, **50**(6): 230017.
- [16] Ge X Y, Wei N Y, Zhou H K, et al. Research on small underwater target detection technology based on side-scan sonar[J]. *Digit Ocean Underwater Warf*, 2023, **6**(2): 155–161.
葛锡云, 魏柠阳, 周宏坤, 等. 基于侧扫声呐的水下小目标检测技术研究[J]. *数字海洋与水下攻防*, 2023, **6**(2): 155–161.
- [17] Wang C Y, Bochkovskiy A, Liao H Y M. Scaled-yolov4: scaling cross stage partial network[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 13024–13033.
<https://doi.org/10.1109/CVPR46437.2021.01283>.
- [18] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding yolo series in 2021[Z]. arXiv: 2107.08430, 2021.
<https://doi.org/10.48550/arXiv.2107.08430>.
- [19] Chen Y F, Zhang C Y, Chen B, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases[J]. *Comput Biol Med*, 2024, **170**: 107917.
- [20] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 13708–13717.
<https://doi.org/10.1109/CVPR46437.2021.01350>.
- [21] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//*Proceedings of the 32nd International Conference on Machine Learning*, Lille, 2015: 448–456.
- [22] Xu W, Wan Y. ELA: efficient local attention for deep convolutional neural networks[Z]. arXiv: 2403.01123, 2024.
<https://doi.org/10.48550/arXiv.2403.01123>.
- [23] Hou Q B, Zhang L, Cheng M M, et al. Strip pooling: rethinking spatial pooling for scene parsing[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2020: 4002–4011.
<https://doi.org/10.1109/CVPR42600.2020.00406>.
- [24] Wu Y X, He K M. Group normalization[C]//*Proceedings of the 15th European Conference on Computer Vision*, Munich, 2018: 3–19. https://doi.org/10.1007/978-3-030-01261-8_1.
- [25] Tian Z, Shen C H, Chen H, et al. FCOS: a simple and strong anchor-free object detector[J]. *IEEE Trans Pattern Anal Mach Intell*, 2022, **44**(4): 1922–1933.
- [26] Zhou Y, Chen S C, Wu K, et al. SCTD 1.0: sonar common target detection dataset[J]. *Comput Sci*, 2021, **48**(11A): 334–339.
周彦, 陈少昌, 吴可, 等. SCTD1.0: 声呐常见目标检测数据集[J]. *计算机科学*, 2021, **48**(11A): 334–339.
- [27] Xie K B, Yang J, Qiu K. A dataset with multibeam forward-looking sonar for underwater object detection[J]. *Sci Data*, 2022, **9**(1): 739.
- [28] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, Venice, 2017: 2999–3007.
<https://doi.org/10.1109/ICCV.2017.324>.
- [29] Kim K, Lee H S. Probabilistic anchor assignment with IoU prediction for object detection[C]//*Proceedings of the 16th European Conference on Computer Vision*, Glasgow, 2020: 355–371. https://doi.org/10.1007/978-3-030-58595-2_22.
- [30] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points[Z]. arXiv: 1904.07850, 2019. <https://doi.org/10.48550/arXiv.1904.07850>.
- [31] Sun P Z, Zhang R F, Jiang Y, et al. Sparse R-CNN: end-to-end object detection with learnable proposals[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 14449–14458.
<https://doi.org/10.1109/CVPR46437.2021.01422>.
- [32] Chen Q, Wang Y M, Yang T, et al. You only look one-level feature[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 13034–13043.
<https://doi.org/10.1109/CVPR46437.2021.01284>.
- [33] Feng C J, Zhong Y J, Gao Y, et al. TOOD: task-aligned one-stage object detection[C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*, Montreal, 2021: 3490–3499. <https://doi.org/10.1109/ICCV48922.2021.00349>.
- [34] Zhang H Y, Wang Y, Dayoub F, et al. VarifocalNet: an IoU-aware dense object detector[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 2021: 8510–8519.
<https://doi.org/10.1109/CVPR46437.2021.00841>.
- [35] Li Z K, Xie Z J, Duan P H, et al. Dual spatial attention network for underwater object detection with sonar imagery[J]. *IEEE Sens J*, 2024, **24**(5): 6998–7008.

作者简介



王燕 (1979-), 女, 讲师, 博士, 研究方向为无线感知、水下传感器网络、人工智能。

E-mail: ellawangyan2016@outlook.com



张艳 (1982-), 女, 副教授, 博士, 主要研究方向为图像处理、机器视觉。

E-mail: yanzhang0910@163.com



王宏辉 (1998-), 男, 硕士研究生, 主要研究方向为深度学习、目标检测。

E-mail: wanghh133@163.com



郝泽玉 (2000-), 男, 硕士研究生, 主要研究方向为深度学习、小样本学习。

E-mail: Haozeyu_tcu@163.com



【通信作者】刘树东 (1965-) 男, 教授, 博士, 主要研究方向为图像处理、无线传感网络。

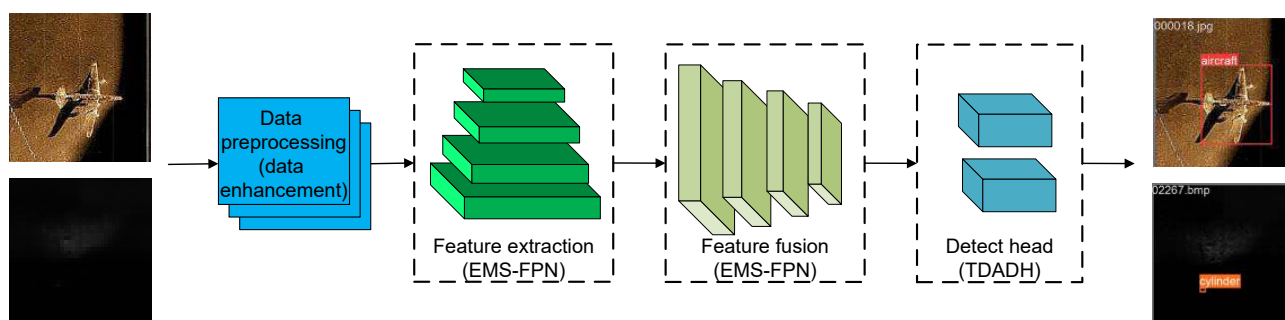
E-mail: liushudong@tcu.edu.cn



扫描二维码, 获取PDF全文

Small target detection in sonar images with multilevel feature screening and task dynamic alignment

Wang Yan, Wang Honghui, Liu Shudong*, Zhang Yan, Hao Zeyu



Improved YOLOv8s target detection network model

Overview: Sonar technology has an important application value in the marine field, and is widely used in seabed geological exploration, marine environmental pollution monitoring, underwater target detection, marine resources development, and other fields. However, the detection of small targets in sonar images has always been a challenging problem due to the fact that sonar imaging is affected by a variety of factors, such as the marine environment and underwater target characteristics. Small targets, such as round cages and balls, often face difficulties such as weak signals, complex backgrounds, low resolution, and noise interference in sonar images, and their effective detection is crucial to ensure the safety of underwater navigation and the development of marine resources. To solve the problem of small target detection in sonar images, which is difficult, low precision, and prone to wrong detection and leakage, a lightweight sonar image small target detection algorithm based on YOLOv8s with efficient multilevel feature fusion is proposed. Firstly, considering that small targets in sonar images usually have low contrast and are easily overwhelmed by noise, an efficient multilevel screening feature fusion pyramid EMS-FPN module is proposed. It can highlight the important features through the screening mechanism, suppress irrelevant background noise, and extracting features from different scales achieve multilevel fusion so as to improve the detection capability of small targets. Secondly, since the classification branch and the localization branch of the decoupling head are independent, it will increase the number of parameters of the model and lead to the problem of lack of interaction between the two tasks. It is difficult to effectively adapt to the detection needs of targets at different scales, resulting in poor detection of small targets, therefore, the task dynamic align detection head (TDADH) module is designed to learn the task interaction features from multiple convolutional layers through a feature extractor to obtain joint features to effectively adapt to the detection needs of targets at different scales, and finally, to validate the effectiveness of the model in this paper, corresponding validation is carried out on the URPC2021 and SCTD sonar datasets, and the detection accuracy mAP50 is improved compared with that of YOLOv8s respectively by 0.3% and 1.8%, and the number of parameters is reduced by 22.5%. The results show that the sonar image target detection algorithm proposed in this paper improves the accuracy and significantly reduces the number of model parameters.

Wang Y, Wang H H, Liu S D, et al. Small target detection in sonar images with multilevel feature screening and task dynamic alignment[J]. *Opto-Electron Eng*, 2024, 51(10): 240196; DOI: [10.12086/oe.2024.240196](https://doi.org/10.12086/oe.2024.240196)

Foundation item: Project supported by Tianjin Philosophy and Social Science Planning Project (TJGL19xSX-045)

School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300000, China

* E-mail: liushudong@tcu.edu.cn