

引用本文

展杰,胡斌,孟冠辰.基于改进YOLO11n的水下前视声呐人体目标检测算法[J].消防科学与技术,2026,45(8):57-63.

Zhan Jie, Hu Bin, Meng Guanchen. Forward-looking sonar human target detection based on improved YOLO11n[J]. Fire Science and Technology, 2026, 45(8): 57-63.

基于改进YOLO11n的水下前视声呐人体目标检测算法

展杰¹, 胡斌¹, 孟冠辰²

(1. 应急管理部上海消防研究所, 上海 200030;

2. 大连理工大学软件学院, 辽宁 大连 116620)

摘要:在水域应急救援任务中,受水体浑浊度、光照衰减等影响,光学成像设备可视距离极短甚至失效,前视声呐(FLS)成为探测水下遇难人员的关键设备。然而,声呐图像存在严重的散斑噪声及多径效应,且人体声学数据极度稀缺,导致现有通用目标检测算法在人体声呐图像小数据量条件下漏检率高、泛化能力差。针对人体声呐成像特征,本文面向小样本场景提出一种基于物理引导的改进YOLO11n水下前视声呐人体目标检测算法。首先,在主干网络引入基于小波的下采样模块(WDB),利用离散小波变换实现频域噪声抑制与轮廓保留;其次,提出声学阴影感知注意力模块(SHAA),联合建模高亮回波与声学阴影,在深层特征上显式注入声呐成像物理先验;最后,设计动态声学上下文模块(DACM),通过双分支几何建模自适应捕捉人体目标形态。在自制水下人体数据集的小样本训练设置下,与YOLO11n基线模型相比,改进算法的mAP₅₀最高提升了35.6%,同时模型参数量仅增加了3.1%,在保持实时性的前提下显著降低了漏检和误检,为水域应急救援中的水下人体目标快速探测提供了一种有效方案。

关键词:水下前视声呐;目标检测;声呐图像;轻量化网络;水域救援

中图分类号:TP181;TP391.4 **文献标志码:**A **DOI:**10.20168/j.1009-0029.2026.08.0057.07

Forward-looking sonar human target detection based on improved YOLO11n

Zhan Jie¹, Hu Bin¹, Meng Guanchen²

(1. Shanghai Fire Science and Technology Research Institute of MEM, Shanghai 200030, China; 2. School of Software, Dalian University of Technology, Dalian Liaoning 116620, China)

Abstract: In water emergency rescue tasks, the effective range of optical imaging devices is often severely limited or may even fail due to water turbidity and light attenuation. Therefore, forward-looking sonar (FLS) has become a key sensor for detecting underwater human targets. However, FLS images are affected by severe speckle noise and multipath effects, while annotated human acoustic data are extremely scarce, resulting in high missed-detection rates and poor generalization under small-sample conditions. Targeting human body sonar imaging characteristic, a physics-guided human target detection algorithm based on an improved YOLO11n network is proposed for underwater forward-looking sonar images in small-sample scenarios. A Wavelet-based Downs-sample Block (WDB) is introduced into backbone network to suppress frequency-domain noise and preserve target contours. A Shadow-aware Holistic Acoustic Attention (SHAA) module is designed to jointly model high-light echoes and acoustic shadows, so that sonar imaging priors can be injected into deep features. A Dynamic

基金项目:国家重点研发计划项目(2023YFC3010805-4)

作者简介:展杰,应急管理部上海消防研究所,助理研究员,主要从事水下搜寻、感知探测技术研究,上海市徐汇区中山南二路601号,200030。

通信作者:胡斌,应急管理部上海消防研究所,副研究员,1842376513@qq.com。

收稿日期:2025-12-10; **修回日期:**2026-06-08

Acoustic-Context Module (DACM) is further used to capture diverse human target shapes through dual-branch geometric modeling. Experiments on a self-collected underwater human dataset show that mAP₅₀ is improved by up to 35.6% compared with YOLO11n, while the number of parameters is increased by only 3.1%. Missed and false detections are reduced while real-time performance is maintained, providing an effective approach for rapid underwater human target detection in water emergency rescue tasks.

Key words: underwater forward-looking sonar; object detection; sonar image; lightweight network; water rescue

我国水域面积广阔,水上作业活动频繁,安全事故及洪涝灾害时有发生,水域应急救援任务呈现常态化趋势。受水体浑浊度、光照衰减等因素影响,传统的人工目视搜索和潜水员水下搜索效率较低且风险较高。前视声呐(FLS)具备探测距离远、不依赖自然光照的优势,已成为水下救援的重要技术手段。然而,目前对声呐设备的使用仍处于人工识别阶段,在应急救援场景下受操作员经验及作业环境限制,往往不具备人工精确识别的条件。因此,利用深度学习算法进行人体声呐目标的主动检测,是提升水域应急救援效率的关键技术路径。

近年来,以YOLO系列为代表的目标检测算法在应急消防领域得到广泛应用。Zhao L等^[1]提出面向火灾巡检场景的轻量化网络Fire-YOLO;Xue Z Y等^[2]及张术琳等^[3]分别提出基于改进YOLOv5的森林火灾与锂电池火灾检测模型;Wang H L等^[4]和Wang H J等^[5]分别提出引入轻量化与注意力机制的YOLO-LFD和DSS-YOLO模型;于春雨等^[6]提出改进坐标注意力的YOLOv8火焰检测模型。这些改进模型通过引入注意力机制、轻量化主干网络和多尺度检测头,有效提升了小尺度火焰和烟雾目标的检测精度与实时性。

类似地,在水下声呐成像领域,研究者围绕低分辨率、强散斑噪声、目标形变大等难题,提出了多种基于YOLO或新型检测框架的声呐目标检测方法。例如,Sun Y等^[7]提出CCW-YOLOv5网络,在多波束前视声呐图像上引入坐标卷积和改进边界框损失函数,显著提升了小目标检测性能;Wang Z等^[8]提出YOLO-SONAR网络,通过语义-空间特征增强和竞争坐标注意机制,缓解了海底杂波和低分辨率带来的干扰;Wang K Q等^[9]提出ProNet网络,利用渐进敏感区域捕获策略,逐步聚焦于可能存在有效目标的区域,改善了复杂背景下的声呐目标检测效果;Lu Y Y等^[10]提出AquaYOLO网络,结合残差块、特征金字塔和注意力机制改进YOLO结构,能够在声呐图像小目标多、先验弱的条件下取得更优的检测精度。然而,现有方法多聚焦于水下结构物、船体残骸、海洋垃圾等一般目标,对人体这一与应急救援高度相关但声学反射较弱、成像形态多变的目标,效果相对有限。

与一般目标相比,水下人体目标具有突出的多样

性和不确定性。一方面,人体在不同姿态下回声形态各不相同,尺度变化大;另一方面,受声呐成像特点影响,目标有时会因较强的反射效应,在目标距离呈现为环形条带状阴影,干扰探测。现有方法多数未利用声呐成像的物理特征,对人体这类弱纹理、形态变化大的目标仍存在检测精度不足、小样本场景下鲁棒性较弱等问题。此外,受数据采集成本限制,用于应急救援的声呐数据稀缺,现有数据集规模有限,难以满足小样本训练下的高精度、实时性检测需求^[11-12]。

为此,本文在YOLO11n基础上提出一种改进的声呐人体目标检测模型YOLO11-FLSH,并构建水下前视声呐人体数据集。主要创新包括:设计基于小波的下采样模块(WDB),实现频域噪声抑制与轮廓保留;引入动态声学上下文模块(DACM),通过双分支几何建模自适应聚焦紧凑亮斑和细长条带;提出声学阴影感知注意力模块(SHAA),在深层特征上显式注入声呐成像先验。

1 YOLO11n模型与改进

1.1 网络整体结构

YOLO11^[13]是Ultralytics团队在YOLOv8基础上推出的新一代实时目标检测模型,适合部署在算力受限的水域救援终端设备上。本文以YOLO11系列中参数量最小的轻量级模型YOLO11n(深度因子0.5、宽度因子0.25、最大通道数1 024)作为基础模型,构建水下前视声呐人体检测网络YOLO11-FLSH。在保持YOLO11n网络主体结构不变的前提下,本文提出3个即插即用的轻量化模块(WDB、SHAA与DACM),从频域下采样、声学特征注意力和几何上下文建模3个层面增强网络对前视声呐图像的表征能力:在主干网络前两次下采样处使用WDB替代Conv卷积,从频域抑制水体扰动并保留人体轮廓细节;在主干网络最深层C3+SPPF之后接入SHAA,对高层语义特征进行声学阴影感知重加权;在Neck高层分支中用DACM替代部分C3k2模块,用于强化人体空间结构。在特征融合与预测阶段,网络沿用YOLO11原有的增强PANet架构进行多尺度特征高效融合,并通过解耦检测头结合分布焦点损失完成目标框与置信度的解码。YOLO11-FLSH网络结构如图1所示。

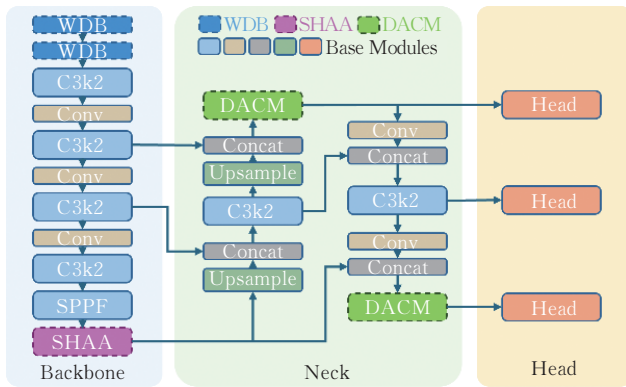


图1 YOLO11-FLSH网络结构

Fig. 1 Structure of YOLO11-FLSH

1.2 基于小波的下采样模块 WDB

在目标检测网络中,下采样操作通过逐步降低特征图空间分辨率、扩大感受野并聚合高层语义信息,是特征提取阶段的关键环节。传统检测网络通常采用步长为2的卷积或最大池化完成下采样,该类纯空间域算子在自然光学图像上表现良好,但在散斑噪声显著的前视声呐图像中存在明显局限。一方面,高频噪声在下采样过程中易通过混叠效应传播至深层特征,导致噪声放大;另一方面,低对比度的人体轮廓和阴影等关键信息在简单的空间采样与平滑过程中容易被削弱或抹除,从而影响后续检测精度。

为缓解上述问题,本文设计基于二维 Haar 小波的下采样模块 WDB。小波变换是一种典型的时频局部化分析工具,能够将信号分解为不同频带的子信号,低频子带主要刻画全局结构,高频子带则包含细节与局部纹理信息,在实现空间尺度压缩的同时引入频域分解与重构机制,实现对散斑噪声的频域抑制与目标轮廓细节的保留^[14]。Haar 小波具有结构简单、计算代价低等特点,适合实时检测场景下的网络模块设计。

设 WDB 模块的输入为归一化后的声呐图像,见式(1)。

$$X \in R^{C \times H \times W} \quad (1)$$

式中: C 为通道数; H 、 W 为空间尺寸。

二维 Haar 小波由一对一维低通滤波器 h 和高通滤波器 g 构造,其卷积核见式(2)。

$$h = \frac{1}{\sqrt{2}}[1, 1], g = \frac{1}{\sqrt{2}}[-1, 1] \quad (2)$$

低通滤波器 h 用于提取信号的低频结构,高通滤波器 g 用于提取高频细节。沿水平与垂直方向分别进行 h 和 g 滤波并下采样,可得到1个低频子带 X_{LL} 和3个高频子带 X_{LH} 、 X_{HL} 、 X_{HH} ,见式(3)。

$$X_{LL}, X_{LH}, X_{HL}, X_{HH} = \text{DWT}(X) \quad (3)$$

式中: $\text{DWT}(\cdot)$ 表示二维离散小波变换; X_{LL} 为低频-

低频子带,主要保留人体轮廓、阴影等宏观结构; X_{LH} 与 X_{HL} 分别对应水平方向与垂直方向的边缘与细节; X_{HH} 则包含对角方向的细节成分及大部分高频噪声。

首先将4个子带在通道维度上进行拼接,并经由 1×1 卷积处理后得到频域融合特征,见式(4)。

$$F_f = \phi_{1 \times 1}([X_{LL}, X_{LH}, X_{HL}, X_{HH}]) \quad (4)$$

本文进一步引入一条空间域特征分支 F_s ,用于保留原始卷积下采样中的基础空间结构,采用 3×3 卷积操作将输入 X 映射到与 F_f 相同的尺度,见式(5)。

$$F_s = \phi_{3 \times 3}(X) \quad (5)$$

最终,WDB模块的输出由频域分支 F_f 与空间域分支 F_s 特征相加所得,见式(6)。

$$F_{\text{out}}^{\text{WDB}} = F_f + F_s \quad (6)$$

与传统空间域下采样相比,WDB模块利用小波分解的频域分离特性,将人体结构等低频信息集中于 X_{LL} ,有助于减弱高频散斑噪声在下采样过程中的累积影响。同时,由于固定 Haar 小波系数,模块内仅门控分支与通道映射层参与训练。这使得网络在完成特征降维并保留声学阴影等边缘细节的同时,保证了模型的轻量化特性。此外,通过同时保留并融合4个频域子带,WDB在完成尺度压缩的同时保留了人体目标边缘和声学阴影等关键信息,可增强网络对弱反射目标的辨识度。

1.3 声学阴影感知注意力模块 SHAA

前视声呐成像基于主动声波探测原理:发射阵列向水体中发射声波,当声波遇到目标时发生反射和散射,回波信号被接收并经时域-空间域重建,形成二维图像。对于任意强散射体而言,若其在某一量程位置上对入射声束产生较强反射与散射,将消耗该距离处波束的大部分能量,使得同一量程上的后续路径几乎无有效回波返回接收阵列,从而在成像中形成显著的声学阴影区域。对于前视声呐,这种效应会在反射距离形成环状的暗带,并在部分照射角度下较为明显。

在水下人体检测任务中,人体目标通常兼具局部强散射表面和显著遮挡效应,因而在声呐图像上往往表现为前景高亮回波与其后方声学阴影的组合模式:人体躯干或肢体表面对声束产生较强反射,形成局部高亮区域;同时,由于人体对声束的遮挡,其后方一定范围内回波能量显著衰减,形成阴影区。相应的,孤立的高亮斑点、局部暗区或环状阴影既可能来源于散斑噪声、气泡,也可能对应固定的水下结构或地形起伏。因此,当高亮回波与后方阴影在量程、方位上呈现合理的遮挡布局时,更有可能对应真实的目标。

基于上述物理先验,本文提出声学阴影感知注意力模块 SHAA,通过双分支结构分别提取与回波强度相关的显著性特征以及与阴影区域相关的结构特

征,并采用乘积式融合构建强回波阴影协同约束,从特征空间显式编码声呐成像的遮挡机理。SHAA的整体结构如图2所示,可直接插入YOLO11骨干的深层语义特征提取阶段,对疑似目标区域进行物理先验指导下的重加权。

设SHAA模块D输入为深层特征张量,见式(7)。

$$F_{in}^{SHAA} \in R^{B \times C \times H \times W} \quad (7)$$

式中: B 为批次大小。

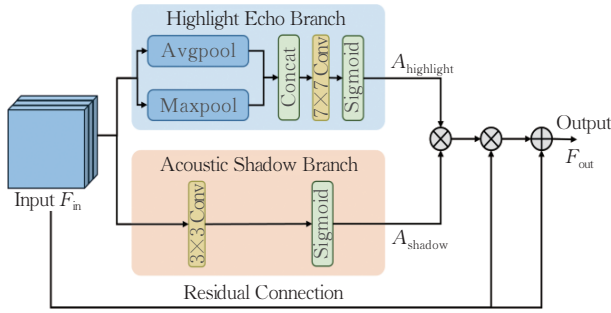


图2 SHAA模块结构

Fig. 2 Structure of SHAA

回波强度分支:为突出与强回波相关的显著性区域,首先在通道维度上对输入特征分别进行平均池化和最大池化,将二者在通道维度上拼接得到回波强度特征,见式(8)。

$$F_c = \text{Concat}[\text{Mean}(F_{in}^{SHAA}), \text{Max}(F_{in}^{SHAA})] \quad (8)$$

随后通过一个 7×7 的卷积对其进行空间建模,得到高亮注意力图,见式(9)。

$$A_{\text{highlight}} = \text{Sigmoid}(\phi_{7 \times 7}^{\text{high}}(F_c)) \quad (9)$$

式中: $\phi_{7 \times 7}^{\text{high}}(\cdot)$ 表示卷积操作。

较大的卷积核能够提供更大的感受野,有助于捕获目标区域跨像素的回波强度上下文信息。

声学阴影分支:为刻画由目标遮挡引起的声学阴影及其边界形态,该分支采用一个 3×3 卷积层对输入深层特征 F_{in}^{SHAA} 进行局部空间编码,以有效捕获暗区的边缘梯度与上下文结构。随后,经Sigmoid函数激活,生成阴影注意力图,见式(10)。

$$A_{\text{shadow}} = \text{Sigmoid}(\phi_{3 \times 3}^{\text{shadow}}(F_{in}^{SHAA})) \quad (10)$$

式中: $\phi_{3 \times 3}^{\text{shadow}}(\cdot)$ 表示卷积操作。

该分支在保留量程方向亮度变化的同时,更关注暗区的局部梯度,使模型能够对人体后方的连续阴影区域形成稳定响应。

为了显式编码目标高亮回波及后方对应阴影的物理先验,SHAA采用逐元素乘积将两支注意力进行融合,见式(11)。

$$A = A_{\text{highlight}} \odot A_{\text{shadow}} \quad (11)$$

式中: $\odot$ 表示Hadamard乘积。

仅当局部高亮回波与其后方声学阴影同时显著时,模块才予以高响应,从而抑制了声呐图像上孤立

亮斑或背景杂波的干扰。

最终,SHAA以残差方式对输入特征进行重加权,见式(12)。

$$F_{out}^{SHAA} = F_{in}^{SHAA} + A \odot F_{in}^{SHAA} \quad (12)$$

残差连接保证了在注意力退化或物理先验不明显的场景下,网络可以平滑地退回到原始特征表示,提高了整体训练稳定性与泛化鲁棒性。同时,在存在显著高亮加阴影的区域,SHAA能够自适应强化与声学目标相关的响应,实现对声呐人体等目标的物理感知注意增强。

1.4 动态声学上下文模块DACM

前视声呐图像中的人体目标受姿态、入射角和距离等因素影响,在几何外观上呈现出显著多样性。结合大量前视声呐图像样本可以观察到,两类成像模式尤为典型:一类为由躯干或头部正向入射产生的局部高反射区域,通常表现为尺度较小、能量集中的紧凑亮斑;另一类为由侧向姿态、肢体伸展等因素引起的细长结构,沿量程方向或方位方向呈条带状延伸。

传统卷积网络多依赖固定尺寸卷积核、进行特征提取,这种单尺度局部算子难以同时兼顾不同的几何模式:小卷积核对紧凑结构敏感,却难以覆盖细长型目标的完整形态;较大卷积核虽能扩大感受野,却易引入背景噪声,削弱对小尺度高亮区域的定位精度。在复杂声呐图像背景噪声影响下,这种尺度与形状不匹配容易导致人体结构与背景杂波的混淆。

为此,本文提出动态声学上下文模块DACM,通过双分支几何建模与形状选择注意力机制,自适应融合不同尺度与形状声学上下文,从而增强对多样化人体目标形态的刻画能力。本模块受选择性核网络^[15]中自适应选择卷积核尺度思想的启发,一条分支侧重局部紧凑结构建模,另一条分支侧重沿特定方向延伸的条带结构建模,最终通过形状选择注意力完成动态融合。DACM的整体结构如图3所示。

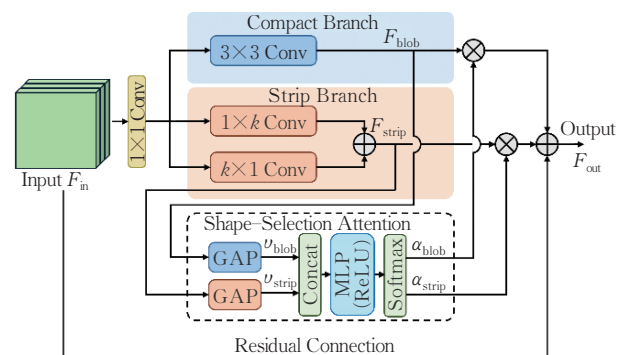


图3 DACM模块结构

Fig. 3 Structure of DACM

设DACM的输入特征见式(13)。

$$F_{in}^{DACM} \in R^{B \times C \times H \times W} \quad (13)$$

式中: B 为批次大小; C 为输入通道数; H 、 W 为空间尺寸。

首先通过 1×1 卷积进行通道压缩, 见式(14)。

$$F_{\text{mid}} = \phi_{1 \times 1}(F_{\text{in}}^{\text{DACM}}) \in R^{B \times C_m \times H \times W} \quad (14)$$

式中: $\phi_{1 \times 1}(\cdot)$ 表示 1×1 卷积算子; $C_m = \alpha C$, $\alpha \in (0, 1)$ 为压缩系数, 本文取 $\alpha = 0.5$ 。

双分支几何建模: 将中间特征 F_{mid} 分别送入紧凑分支和条带分支, 针对两类典型几何形态进行建模。

紧凑分支通过 3×3 卷积提取局部结构, 见式(15)。

$$F_{\text{blob}} = \phi_{3 \times 3}(F_{\text{mid}}) \quad (15)$$

式中: $\phi_{3 \times 3}(\cdot)$ 表示 3×3 卷积算子。

较小的感受野有利于精确刻画局部高亮区域及其边界, 适合描述正向入射角下人体躯干等紧凑亮斑型目标。

条带分支旨在增强对沿量程或方位方向延伸的细长结构的敏感性, 引入一对条形卷积核, 见式(16)。

$$F_{\text{strip}} = \phi_{1 \times k}(F_{\text{mid}}) + \phi_{k \times 1}(F_{\text{mid}}) \quad (16)$$

式中: $\phi_{1 \times k}(\cdot)$ 与 $\phi_{k \times 1}(\cdot)$ 分别为水平方向与垂直方向的条形卷积, 本文取 $k = 7$ 。

两个条形卷积对不同方向的细长结构具有各向异性的响应, 将其输出相加可同时覆盖沿方位与量程方向延伸的条带目标, 在保证计算代价可控的同时突出条带几何特征。

形状选择注意力: 为在两种几何模式之间实现内容自适应选择, DACM 模块引入形状选择注意力机制。首先对两分支输出进行全局平均池化, 得到通道描述向量, 见式(17)。

$$v_{\text{blob}} = \text{GAP}(F_{\text{blob}}), v_{\text{strip}} = \text{GAP}(F_{\text{strip}}) \quad (17)$$

式中: $\text{GAP}(\cdot)$ 表示全局平均池化操作, 得到长度为 C_m 的通道特征向量, 用于表征当前特征中紧凑结构与条带结构的响应强度。

随后将两向量拼接, 经由共享权重的两层感知机 (MLP) 映射并归一化, 生成形状选择权重, 见式(18)。

$$[w_b, w_s] = \text{Softmax}\left(w_2 \text{ReLU}\left(w_1[v_{\text{blob}}, v_{\text{strip}}]\right)\right) \quad (18)$$

式中: w_1 与 w_2 为 MLP 的权重矩阵; $\text{ReLU}(\cdot)$ 为激活函数; $\text{Softmax}(\cdot)$ 用于对两路权重进行归一化。

最后, 将两分支特征根据动态权重进行融合, 并通过可学习的缩放因子与输入特征进行残差叠加, 见式(19)~式(20)。

$$F_{\text{dyn}} = w_b \cdot F_{\text{blob}} + w_s \cdot F_{\text{strip}} \quad (19)$$

$$F_{\text{out}}^{\text{DACM}} = \lambda(F_{\text{dyn}} - F_{\text{in}}^{\text{DACM}}) + F_{\text{in}}^{\text{DACM}} \quad (20)$$

其中, λ 为可学习的残差缩放因子 (本文中设置

初始值为 0.2), 用于控制上下文特征的注入比例, 使得网络能够平滑地学习目标的多尺度几何表征。

2 试验准备与评价指标

2.1 人体声呐图像数据集

为验证所提出算法在真实救援场景中的有效性, 本文基于 Oculus M750D 多波束二维前视声呐和 BlueView M900 多波束二维前视声呐采集多批次水域人体目标数据。采集场景为训练水池, 潜水员在水中以站立、漂浮、俯卧、仰游等多种姿态运动, 包含 M750D 声呐采集的数据 700 张, M900 声呐采集的数据 3 000 张。数据集中包含前文所述的各类人体声呐成像样本, 如图 4 所示。

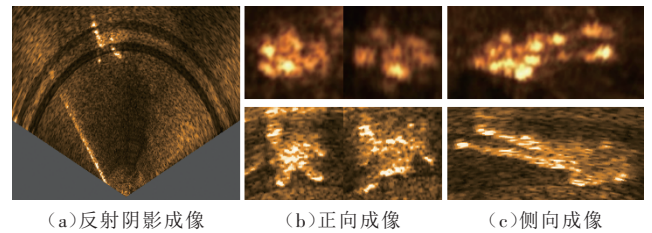


图 4 人体目标声呐成像

Fig. 4 Samples of human target sonar imaging

2.2 FSOD 数据集

为验证所提出算法的多目标检测能力, 本文使用 Li C 等^[16]提出的 FSOD 数据集进行测试, 该数据集涵盖了钢板、混凝土柱、铁罐、块体、钢架、轮胎、圆形笼、方形笼、机器人以及潜水员 10 个不同的水下目标类别。

2.3 人体声呐试验设计

在数据量充足的情况下, 现有通用算法通常能够取得较好的检测精度。然而, 真实水域救援面临着水下目标样本极度稀缺的问题。为验证本文算法在数据受限场景下的检测精度与鲁棒性, 本文专门针对小样本情况设计试验: M900 数据集选取训练集的 50% 进行训练, M750 数据集选取训练集的 25% 进行训练, 验证集和测试集保持不变。对于多目标 FSOD 数据集, 分别使用 25% 与 50% 训练数据进行训练。

考虑到声呐与光学图像存在显著的模态差异, 所有模型均采用从头训练策略, 未加载预训练权重, 全网络参数均统一参与端到端更新。

2.4 试验环境与训练策略

所有试验均在 Windows 11 操作系统下完成。硬件平台为搭载 AMD R9 7900X CPU 与 NVIDIA RTX 3090 GPU 的工作站。配置环境为 PyTorch 2.7.1, CUDA 11.8, Python 3.10。网络输入分辨率统一为 512×512 , 批次大小为 32, 共训练 50 轮。WDB 中的频带门控缩放因子初始化为 0.5, SHAA 和 DACM 中的残差插值缩放因子初始化为 0.2。

2.5 评价指标

本文采用精确率(Precision)、召回率(Recall)、平均精度均值(mAP)、参数量(Params)、帧率(FPS)指标对模型进行评价。主要采用 mAP_{50} 以及 $mAP_{50:95}$ 作为核心指标。

3 试验结果与分析

3.1 整体性能分析

为全面评估 YOLO11-FLSH 模型的检测性能,本文在 M750 和 M900 两个数据集上,分别与 YOLO 系列基线模型以及改进的水下救援目标检测模型 ShuffleNetv2-Db-YOLOv5s^[17] 进行对比试验,检测效果如图 5 所示。

表 1 为不同模型在两个数据集上的性能对比结果。从模型复杂度来看,YOLO11-FLSH 的参数量相比 ShuffleNetv2-Db-YOLOv5s 模型降低 16.0%,相比 YOLOv5s 减少 70.8%,在保持较低参数量的同时取得更高检测精度。在检测精度方面,YOLO11-

FLSH 在 M900 数据集上的 mAP_{50} 相比 YOLO11n 基线模型相对提升 1.8%,相比 ShuffleNetv2-Db-YOLOv5s 模型相对提升 5.2%。在更具挑战性的 M750 小样本数据集上,YOLO11-FLSH 的 mAP_{50} 相比 YOLO11n 基线模型相对提升 35.6%,说明在标注样本有限的声学目标的场景中,本文所提出的网络能够显著增强模型对水下人体目标的检测能力。

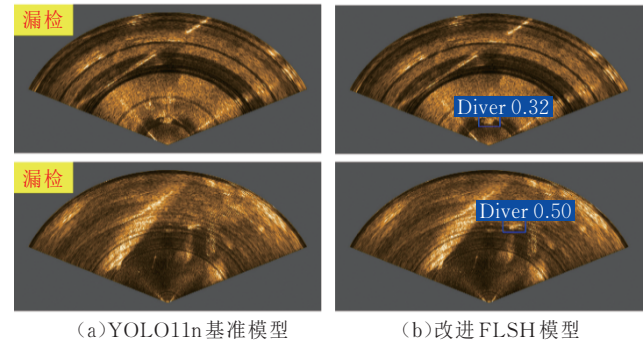


图 5 改进 FLSH 模型检测效果

Fig. 5 Detection results of proposed FLSH model

表 1 模型性能对比

Table 1 Model performance comparison

数据集	模型	参数量/M	帧率/f/s	精确率	召回率	mAP_{50}	$mAP_{50:95}$
M900	YOLOv5n	2.65	1 150.29	0.917	0.887	0.946	0.459
	YOLOv5s	9.15	1 152.40	0.927	0.909	0.954	0.477
	ShuffleNetv2-Db-YOLOv5s	3.18	—	0.880	0.867	0.910	0.395
	YOLO11n	2.59	1 046.57	0.908	0.889	0.940	0.454
	Ours	2.67	1 024.28	0.934	0.904	0.957	0.461
M750	YOLOv5n	2.65	1 268.39	0.496	0.570	0.525	0.183
	YOLOv5s	9.15	1 180.83	0.672	0.671	0.714	0.272
	ShuffleNetv2-Db-YOLOv5s	3.18	—	0.169	0.075	0.063	0.016
	YOLO11n	2.59	1 054.44	0.533	0.524	0.494	0.159
	Ours	2.67	908.78	0.619	0.632	0.670	0.224

表 2 展示了模型在 FSOD 数据集上的多目标检测效果。在 25% 训练数据条件下,YOLO11-FLSH 相较 YOLO11n 基线 mAP_{50} 提升 1.5%, $mAP_{50:95}$ 提升 2.4%;在 50% 训练数据条件下,YOLO11-FLSH 相较 YOLO11n 基线 mAP_{50} 提升 0.3%, $mAP_{50:95}$ 提升 0.5%,说明 YOLO11-FLSH 模型在小样本多类别声呐场景下仍具有适用性。

表 2 FSOD 数据集多类别目标检测结果

Table 2 Multi-class target detection results on FSOD

训练数据	模型	mAP_{50}	$mAP_{50:95}$
25%	YOLO11n	0.828	0.537
25%	本文	0.841	0.550
50%	YOLO11n	0.956	0.672
50%	本文	0.959	0.676

3.2 消融实验分析

消融实验结果如表 3 所示。与基线 YOLO11n 模

型配置相比,单独引入 3 种模块的配置在 M750 数据集上均取得了有益的效果。在 mAP_{50} 指标上,WDB、SHAA 和 DACM 分别相对提升 5.5%、4.5% 和 14.4%;在 $mAP_{50:95}$ 指标上,相对提升为 7.6%、13.8% 和 24.5%。其中,DACM 配置在 mAP_{50} 上取得了最高的相对增幅,表明动态声学上下文建模对于小样本、形态变化显著的声呐人体场景具有更强的适应与刻画能力;WDB 和 SHAA 在高 IoU 区间同样带来了相对积极的提升,验证了基于小波的下采样在抑制散斑噪声并保留目标轮廓方面的有效性,以及声学阴影感知注意在提升检测精度上的积极作用。相对而言,在 M900 数据集上,3 种单模块配置对 mAP_{50} 的提升幅度整体均低于 1%,且在 $mAP_{50:95}$ 上存在轻微波动,说明在样本更加充足、场景分布更为完备的条件下,单一结构改动的边际收益有限,更高性能的提升依赖于多模块之间的协同设计与联合优化。

表 3 消融实验结果

Table 3 Ablation study results

模型	WDB	SHAA	DACM	M750 数据集		M900 数据集	
				mAP ₅₀	mAP _{50:95}	mAP ₅₀	mAP _{50:95}
YOLO11n	—	—	—	0.494	0.159	0.940	0.454
M1	✓	—	—	0.521	0.171	0.951	0.441
M2	—	✓	—	0.516	0.181	0.948	0.445
M3	—	—	✓	0.565	0.198	0.940	0.447
M4	—	✓	✓	0.579	0.203	0.956	0.459
M5	✓	—	✓	0.517	0.195	0.950	0.452
M6	✓	✓	—	0.537	0.192	0.952	0.456
FULL	✓	✓	✓	0.670	0.224	0.957	0.461

从多模块组合试验的结果来看,双模块组合中M4模型在两个数据集上都取得了较为显著且稳定的增益:相较基线YOLO11n模型,在M750数据集上的mAP₅₀相对提升17.2%,mAP_{50:95}相对提升27.7%;在M900数据集上,mAP₅₀与mAP_{50:95}也分别获得1.7%和1.1%的提升。该现象表明,DACM与SHAA在特征空间具有良好的互补性,能够提升模型在高IoU阈值下的目标定位精度。在此基础上,FULL模型同时引入WDB、SHAA与DACM,在两个数据集上均取得了最优表现,体现出显著的协同效应,使得模型在小样本场景与相对充足样本场景下都能稳定提升整体检测性能与高精度定位能力。

3.3 实战应用

为验证所提算法的实战表现,本文将其集成于自研的水下机器人感知系统,并参与了青海某大桥绳索断裂事故的遂行搜救保障任务。该算法搭载于集成有Oculus M750D前视声呐的水下机器人操作终端,配合无人船完成了水域内10处疑似目标点的发现与排查工作。应用结果表明,该算法能够在一定程度上抑制水下复杂环境中的声呐图像噪声干扰,鲁棒性较强,性能优于传统的检测方法。机器人系统与算法应用效果如图6所示。

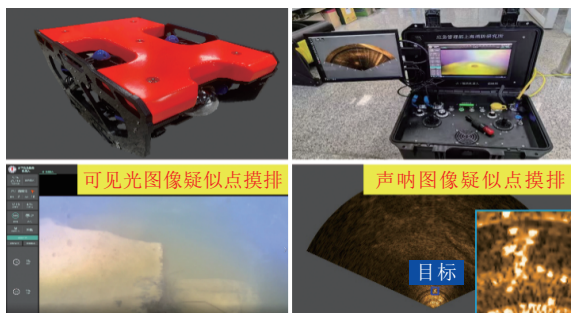


图6 机器人系统与现场应用

Fig. 6 Underwater robot system and field application

4 结论

针对水域应急搜救任务提出基于改进YOLO11n

的轻量化检测网络YOLO11-FLSH,并构建了相应的水下人体目标声呐图像数据集。通过在骨干网络引入WDB、SHAA和DACM这3个模块,改进网络能够从频域、几何空间和高层语义3个层面联合建模前视声呐图像的声学成像特征。试验结果表明,YOLO11-FLSH在保持参数量基本不变的前提下,实现了mAP₅₀和mAP_{50:95}指标检测精度的提升,尤其在M750小样本数据集上具有显著优势。此外,基于低参数量的轻量化设计,模型实现了检测精度与执行效率的良好平衡,能够切实满足水域搜救任务中水下机器人等边缘算力受限设备上的实时性需求。未来工作将扩充数据集场景类型,结合时序信息实现目标跟踪,结合可见光与侧扫声呐图像实现多模态融合检测,并探索半监督学习策略。

参考文献:

- [1] Zhao L, Zhi L Q, Zhao C, et al. Fire-YOLO: A small target object detection method for fire inspection[J]. Sustainability, 2022, 14(9):4930.
- [2] Xue Z Y, Lin H F, Wang F. A small target forest fire detection model based on YOLOv5 improvement[J]. Forests, 2022, 13(8): 1332.
- [3] 张术琳,王澜凝,文拙,等.基于改进YOLOv5s算法的锂电池火灾识别模型[J].消防科学与技术,2025,44(8):1023-1028.
Zhang S L, Wang L N, Wen Z, et al. Identification model of lithium battery fire based on improved YOLOv5s algorithm[J]. Fire Science and Technology, 2025, 44(8):1023-1028.
- [4] Wang H L, Zhang Y Y, Zhu C. YOLO-LFD: A lightweight and fast model for forest fire detection[J]. Computers, Materials & Continua, 2025, 82(2):3399-3417.
- [5] Wang H J, Fu X Y, Yu Z X, et al. DSS-YOLO: An improved lightweight real-time fire detection model based on YOLOv8[J]. Scientific Reports, 2025, 15:8963.
- [6] 于春雨,李晓旭,李泊宁,等.基于改进YOLOv8的小目标火焰图像检测算法研究[J].消防科学与技术,2025,44(9):1274-1280.
Yu C Y, Li X X, Li B N, et al. Research on small target flame image detection algorithm based on improved YOLOv8[J]. Fire Science and Technology, 2025, 44(9):1274-1280.
- [7] Sun Y, Yin B. CCW-YOLOv5: A forward-looking sonar target method based on coordinate convolution and modified boundary frame loss[J]. PLoS One, 2024, 19(6):e0300976.
- [8] Wang Z, Guo J X, Zhang S W, et al. Marine object detection in forward-looking sonar images via semantic-spatial feature enhancement[J]. Frontiers in Marine Science, 2025, 12: 1539210.
- [9] Wang K Q, Liu P, Zhang C. ProNet: Underwater forward-looking sonar images target detection network based on progressive sensitivity capture[J]. Computers, Materials & Continua, 2025, 82(3):4931-4948.

(下转第70页)

- [11] Yousefinezhad R, Jamchi N, Amrolahi M, et al. Impact of thermal-acoustic exposure on human performance of healthy adults: Simulation of traffic noise and thermal conditions in the urban environment[J]. Scientific Reports, 2026, 16: 10173.
- [12] 潘博, 薛茂梅, 王君霞. FLARESIM 软件在火炬设计中的应用[J]. 安全、健康和环境, 2014, 14(3): 33-35.
- [13] ISO 23251:2019, Petroleum, petrochemical and natural gas industries — Pressure-relieving and depressuring systems[S].
- [14] 章庆军. 放空火炬系统的计算与安全因素[J]. 江汉石油职工大学学报, 2007, 20(4): 79-80.
Zhang Q J. Computation and secure factors of flare stack system[J]. Journal of Jiangnan Petroleum University of Staff and Workers, 2007, 20(4): 79-80.
- [15] ISO 9613-2: 2024. Acoustics: Attenuation of sound during propagation outdoors—Part 2: Engineering method for the prediction of sound pressure levels outdoors[S].
- [16] Hannah L. Wind and temperature effects on sound propagation[J]. New Zealand Acoustics, 2007, 20: 22-29.
- [17] 魏东, 梁强. 消防水幕衰减火灾辐射热的理论研究[J]. 中国安全科学学报, 2008, 18(10): 75-81.
Wei D, Liang Q. Theoretical study on fire radiation attenuation with water curtains[J]. China Safety Science Journal, 2008, 18(10): 75-81.
- [18] 刘欣, 张龙, 夏振炎, 等. 海洋石油平台火炬辐射及水幕隔热特征分析[J]. 天津大学学报(自然科学与工程技术版), 2014, 47(9): 790-795.
Liu X, Zhang L, Xia Z Y, et al. Analysis of flare radiation characteristics of offshore oil platform and thermal shielding properties of water curtain system[J]. Journal of Tianjin University (Science and Technology), 2014, 47(9): 790-795.
- [19] 程贵海, 张勤彬, 于秋, 等. 消声降噪技术发展现状及展望[J]. 职业卫生与应急救援, 2017, 35(2): 113-115.
Cheng G H, Zhang Q B, Yu Q, et al. The current situation and perspective of noise reduction technology[J]. Occupational Health and Emergency Rescue, 2017, 35(2): 113-115.
- [20] 李孟东, 陶大闯, 吴子庚, 等. 建筑施工降噪研究现状与趋势[J]. 工程建设与设计, 2026(3): 138-140.
Li M D, Tao D C, Wu Z G, et al. Current status and trends of noise reduction research in construction[J]. Construction & Design for Engineering, 2026(3): 138-140.
- [21] 代昌明, 王戡, 徐明林, 等. 基于统计能量法泡沫铝三明治板隔声性能研究[J]. 舰船科学技术, 2023, 45(15): 65-70.
Dai C M, Wang Y, Xu M L, et al. Research sound insulation performance of aluminum foam sandwich panel based on statistical energy method[J]. Ship Science and Technology, 2023, 45(15): 65-70.
- [22] 徐金荣, 王影, 张弛, 等. 泡沫铝复合结构隔声性能的仿真和实验研究[J]. 佳木斯大学学报(自然科学版), 2024, 42(7): 94-96.
Xu J R, Wang Y, Zhang C, et al. Simulation and experimental investigation on sound insulation performance of aluminum foam composite structure[J]. Journal of Jiamusi University (Natural Science Edition), 2024, 42(7): 94-96.

(责任编辑:毛 星)

(上接第 63 页)

- [10] Lu Y Y, Zhang J J, Chen Q L, et al. AquaYOLO: Enhancing YOLOv8 for accurate underwater object detection for sonar images[J]. Journal of Marine Science and Engineering, 2025, 13(1):73.
- [11] Xie K B, Yang J, Qiu K. A dataset with multibeam forward-looking sonar for underwater object detection[J]. Scientific Data, 2022, 9:739.
- [12] Zeng W J, Li R Z, Zhou H, et al. Underwater target tracking method based on forward-looking sonar data[J]. Journal of Marine Science and Engineering, 2025, 13(3):430.
- [13] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[PP]. arXiv preprint arXiv: 2410.17725, 2024.
- [14] Liu B, Liu G Y, Feng W, et al. Undecimated dual-tree complex wavelet transform and fuzzy clustering-based sonar image denoising technique[J]. Journal of Shanghai Jiaotong University (Science), 2025, 30(5):998-1008.
- [15] Li X, Wang W H, Hu X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019:510-519.
- [16] Li C, Guo P, Meng G, et al. RSOD: Reliability-Guided Sonar Image Object Detection with Extremely Limited Labels[C]//In Proceedings of the AAAI Conference on Artificial Intelligence, 2026.
- [17] Hu S F, Liu T. Underwater rescue target detection based on acoustic images[J]. Sensors, 2024, 24(6):1780.

(责任编辑:邢玉军)