

DOI: 10.19650/j.cnki.cjsi.J2413033

BM3D-YOLOv8-s: 前视声呐图像目标检测算法*

陈美龙, 赵新华, 叶秀芬

(哈尔滨工程大学智能科学与工程学院 哈尔滨 150001)

摘要: 前视声呐作为海洋探测的重要传感器之一,能够远距离探测水下目标,被广泛应用于目标检测和跟踪领域中。然而,声呐数据采集时受海洋环境噪声影响,噪声分布不均匀,使得声呐图像的目标探测精度低。采用传统卷积神经网络对前视声呐目标进行跟踪时,因为声呐图像序列帧率较低、目标特征不清晰,容易出现目标丢失问题。针对前视声呐图像噪声污染严重的问题,结合前视声呐图像的特点,提出了一种改进的BM3D算法,减少3D转换处理的计算量,在基础估计的相似块匹配距离计算过程中,采用曼哈顿距离替代欧氏距离,更好地处理声呐图像中不同类型和强度的噪声;针对目标丢失问题,提出了基于YOLOv8-s改进网络的前视声呐图像目标检测算法,包括基于ConvNeXt的C2N改进算法、添加浅特征检测头和归一化Wasserstein距离(NWD)损失函数的改进。进行了声呐图像数据采集,并进行了实验验证。实验结果表明,改进后模型的准确率为87.2%,mAP_{0.5}为85.4%。与改进前的YOLOv8-s模型相比,虽然模型大小只增加了4.6 MB,但是精度增加了5.1个百分点,mAP@0.5增加了4个百分点,对比其他检测模型实验结果,改进后的YOLOv8-s能够有效提升声呐图像的目标检测精度。

关键词: 前视声呐;噪声;BM3D;目标检测;YOLOv8-s

中图分类号: TH741 TP391.41 文献标识码: A 国家标准学科分类代码: 510.40

BM3D-YOLOv8-s: Forward-looking sonar image target detection algorithm

Chen Meilong, Zhao Xinhua, Ye Xiufen

(College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: Forward-looking sonar is a crucial sensor in ocean exploration, widely used for target detection and tracking over long distances. However, sonar data acquisition is often compromised by environmental noise in the ocean, which is unevenly distributed and reduces the accuracy of target detection in sonar images. Traditional convolutional neural networks (CNNs) for tracking forward-looking sonar targets often fail due to the low frame rates of sonar image sequences and unclear target features. To address the issue of noise pollution in forward-looking sonar images, this paper proposes an enhanced BM3D (Block Matching and 3D Filtering) algorithm tailored to the specific characteristics of sonar images. The Manhattan distance is utilized in place of the Euclidean distance to compute similar block-matching distances, improving noise handling across different types and intensities. Additionally, to mitigate target loss, we introduce a forward-looking sonar image target detection algorithm based on an improved YOLOv8-s network. This enhancement includes modifications to the ConvNeXt-based C2N algorithm, the addition of a shallow feature detection head, and improvements to the normalized Wasserstein distance (NWD) loss function. Experimental results from sonar image data acquisition show that the accuracy of the improved model is 87.2%, with an mAP@0.5 of 85.4%. Compared to the original YOLOv8-s model, the modified model's size increased by only 4.6 MB, while precision improved by 5.1 percentage points, and mAP@0.5 rose by 4 percentage points. The improved YOLOv8-s outperforms other detection models, significantly enhancing target detection accuracy in sonar images.

Keywords: forward-looking sonar; noise; BM3D; target detection; YOLOv8-s

0 引言

前视声呐是海洋探测的重要传感器之一,它通过计算声波回声和返回时间来检测水下障碍物的大小和位置,并转换为声呐图像,用于水下环境观察和检测水下目标。前视声呐可应用于水产养殖、海洋资源探测开发、海底地形绘制、水下搜索与救援等工作^[1]。

前视声呐图像是由声波强度生成,最初得到的为灰度图。前视声呐图像和光学图像相比,存在低信噪比、噪声分布不均衡、目标特征不明显等问题^[2],传统图像处理方法往往无法很好地处理这些问题。因此,水下的目标检测与跟踪,需要结合声呐图像的特点,有针对性的改进,以取得对应算法的最优效果。

随着计算机性能和算力的增强,深度学习方法在前视声呐图像处理中应用逐渐深入。相对于传统卷积神经网络方法,深度学习的预测方法可以大幅改善目标检测和跟踪的效果,提高识别精度^[3]。与传统光学图像不同,在处理前视声呐图像时,检测算法和跟踪算法需要更加专注于处理噪声、低分辨率以及目标特征不明显等问题,目前主流的检测跟踪算法并不能取得很好的效果。因此,研究人员正在致力于设计适用于前视声呐图像特点的目标检测和跟踪算法。

Ma 等^[4]为了提高模型对噪声的鲁棒性,提出一种对抗性训练方法,在对抗训练阶段加入声呐模拟噪声,从而将光学检测算法适配到声呐图像中。Sledge 等^[5]提出了双重解码器,为声呐图像设计专门的解码器,使每个解码器关注目标的不同特征。霍冠英等^[6]提出了一种基于空间约束的快速模糊 C 均值聚类与马尔可夫随机场相结合的分割方法。

李鹏等^[7]提出基于双树双密度复小波子带模型匹配的方法自动判别水下目标有无。马珊^[8]提出了一种模糊图像增强算法,利用改进粒子滤波算法,提升了目标跟踪的准确性。崔杰等^[9]基于均值漂移算法,提出了一种自适应目标框,能够随着物体运动自动地调整大小。张伟正^[10]为提升模型的泛化能力,使用生成对抗网络做扩增,并改进了孪生网络的注意力机制,增强了分类回归网络,实现更准确的目标跟踪,但对多目标跟踪效果不理想。

综上所述,国内外关于前视声呐多目标检测与跟踪算法的研究现状表明,声呐图像在多目标检测与跟踪方面仍存在以下问题:

1) 由于恶劣的水下环境,声呐数据采集时受海洋环境噪声影响,且噪声分布不均匀。现有的降噪算法,大多是针对传统光学图像设计的,并不能解决噪声分布不均匀的问题,因此,需要结合声呐图像的特点进行降噪算法

改进。

2) 目前的主流的目标检测方法大多是基于光学图像设计的,声呐图像有别于传统的光学图像,表现为灰度值偏低、分辨率不高、细节少、噪声干扰严重和目标偏小,这些问题均会导致目标识别困难。因此,需要结合声呐图像的自身特征,改进现有的目标检测算法,实现最优的检测效果。

针对前视声呐图像受噪声干扰严重导致噪声分布不均匀问题,改进三维块匹配滤波(block-matching and 3D filter, BM3D)算法的块匹配距离计算方式,更好地处理声呐图像中不同类型的噪声。同时,根据前视声呐成像特征,提出了一种针对前视声呐图像的 BM3D 改进算法,大幅提升降噪性能,通过实验证明改进算法降噪后的数据具有更高的信噪比;针对前视声呐图像目标偏小、容易丢失的问题,提出了一种基于 YOLOv8-s (you only look once v8-s) 的声呐小目标检测改进算法,并通过实验,对比了多种已有方法,证明了改进算法的有效性。

1 前视声呐图像的获取及降噪处理

1.1 前视声呐的成像原理

前视声呐在工作时以特定的扇形面和角度向海底发射高频声波信号,声波遇到水中的物体或障碍物时,被反射回来,前视声呐系统会接收到这些回波信号。声波从发射到接收会有时间间隔,经过不同材质的物体反射,形成不同幅度的声波,最终呈现为声呐图像^[11]。成像原理如图 1 所示, a 表示目标近端与声呐的距离, b 表示目标远端与声呐的距离, c 表示目标阴影近端与声呐的距离, d 表示目标阴影远端与声呐的距离, e 表示海面与声呐的距离, f 表示声呐与海底的间距。声呐的每次扫描都可获得不同的 a 、 b 、 c 、 d 值,记录下来形成原始数据。

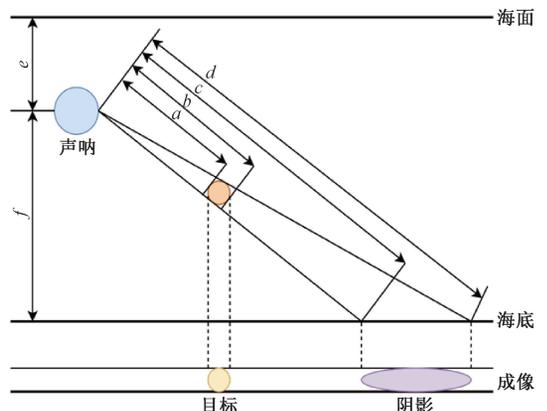


图1 声呐回波成像原理

Fig. 1 Schematic diagram of sonar echo imaging

1) 原始数据解析

声呐每发射一次声波脉冲,保存1帧数据,包括数据头和图像数据区两部分。在每帧数据头中定义数据长度、序号、声速、俯仰角、横滚角、航向角等基础信息,图像数据区按波束次序排列。

不同前视声呐设备具有不同的文件存储格式,如 Oclous、Son、Dat 等,所以解析二进制声呐数据是数据预处理不可缺少的环节。通过对 Oclous 声呐协议分析出声呐数据结构体,使用 Python 的结构体模块解析原始声呐文件。如表 1 所示,首先校验声呐头文件表示是否为 0×1FFF1FFF,校验成功后继续解析声呐 PING (packet internet groper) 头文件,再依次解析出波束 1~500,循环往复,直到解析完毕。

表 1 原始声呐文件结构

Table 1 Original sonar file structure

文件头标志	头文件	波束 1	...	波束 500
1FFF1FFF/字节	512	4 000	...	4 000

2) 前视声呐成像的几何原理

原始前视声呐图像是极坐标图像。在极坐标系下,声呐图像矩阵如图 2(a) 所示,图像上方对应声呐的远端,图像下方表示声呐的近端,图像的宽度则代表波束的数量,图像的长度表示波束的长度。极坐标图像中目标出现严重的扭曲,需要坐标变换还原目标物体,真实的形状如图 2(b) 所示。

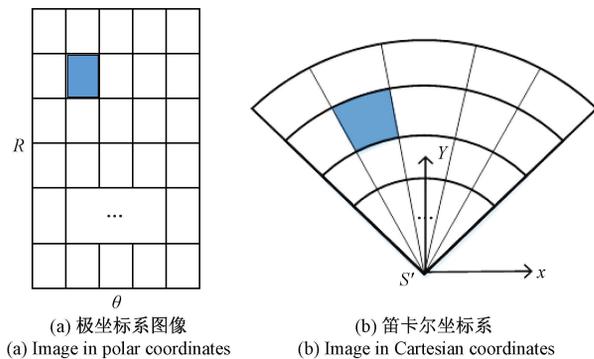


图 2 坐标转换示意图

Fig. 2 Schematic diagram of coordinate transformation

使用坐标变换公式,由极坐标下的坐标算出该像素在笛卡尔坐标系下的坐标并填充。转换完成后需要进行插值,图 3(a) 为极坐标系下的前视声呐图像,图 3(b) 为使用双三次插值算法转换后的笛卡尔坐标系下的前视声呐图像。

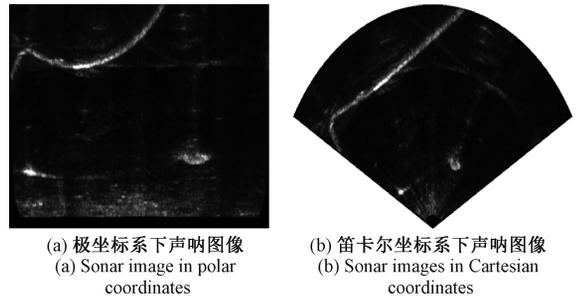


图 3 极坐标系转笛卡尔坐标系的前视声呐图像
Fig. 3 Forward-looking sonar image transformation from polar coordinates to Cartesian coordinates

1.2 前视声呐图像的采集系统

为了实现前视声呐图像的采集,在水下机器人上搭建前视声呐设备,前视声呐采用型号 blueprint subsea 的声呐,性能为:750 kHz/1.2 MHz 双频水平波束角 130°,垂直波束角 20°,512 波束,距离分辨率 4 mm。前视声呐采集的数据经过 NVIDIA Jetson Xavier 系统存储,采用离线的方式在上位机软件上显示图像,如图 4 所示。前视声呐采集的原始图像存在噪声并且对目标的检测精度较低,因此,基于采集的原始图像,对降噪方法以及检测方法进行改进。

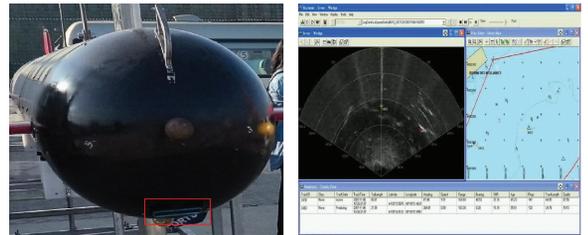


图 4 前视声呐图像检测装置以及检测效果
Fig. 4 Forward-looking sonar image detection device and detection effect

1.3 改进的 BM3D 算法

BM3D 算法结合了非局部去噪和转换算法的优点,将图像去噪领域提升到了一个新的水平。然而,BM3D 算法存在时间复杂度过高,降噪速度偏慢的问题,且难以解决声呐图像噪声分布不均匀问题。针对 BM3D 算法偏慢的问题,结合前视声呐图像特征区域少的特点,在相似块匹配的过程中添加预过滤器,减少 3D 转换处理的计算量。针对声呐图像噪声分布不均匀的问题,在基础估计的相似块匹配距离计算过程中,采用曼哈顿距离替代原来的欧氏距离,从而更好地处理声呐图像中的不同类型和不同强度的噪声。

1) 块匹配预过滤器算法改进

在声呐图像中,噪声定义为:

$$z(t) = y(t) + n(t), x \in I \quad (1)$$

式中: $z(t)$ 是声呐图像; $z(t)$ 表示 x 位置的噪声 $n(t)$ 与特

征 $y(t)$ 之和。

理想情况下,欧氏距离衡量参考块和 x 处相似程度为:

$$d^{ideal}(Z, Z_x) = \frac{\|Z - Z_x\|_2^2}{(N^{ht})^2} \quad (2)$$

式中: Z 和 Z_x 分别表示目标块和 x 处的图像块。

如果 Z 和 Z_x 不重叠,则期望为非中心基数随机分布:

$$\min_x \left\{ \sum_p (S_p - I_p)^2 + \lambda \times c(S) \right\} \quad (3)$$

$$\text{var} \{ d^{noisy}(Z, Z_x) \} = \frac{8\sigma^4 + 8\sigma^2 d^{noisy}(Z, Z_x)}{(N^{ht})^2} \quad (4)$$

前视声呐图像噪声严重,会导致方差接近于 $\sigma(\sigma^4)$, 由于 σ 较大, N^{ht} 较小,使得 $d^{noisy}(Z, Z_x)$ 方差较大,进而导致不正确的分组。为了避免这个问题,使用预过滤器来计算距离。根据前视声呐图像特征区域少的特点,对匹配的图块设置预过滤器,预过滤器是一个均值滤波函数,计算当前图像块是否包含特征,若没有特征则直接使用最大距离,从而提高了去噪效率,降低了计算复杂度。改进后的噪声方差计算公式为:

$$d^{noisy}(Z, Z_x) = \frac{\|\gamma'(T_{2D}^{ht}(Z) - T_{2D}^{ht}\gamma''(Z_x))\|_2^2}{(N^{ht})^2} \quad (5)$$

式中: T_{2D}^{ht} 表示二维线性变换; γ' 表示硬阈值滤波; γ'' 表示预滤波器。

$$N_{\max}^{ht} = \{x \in I; d(Z, Z_x) \leq \tau_{match}^{ht}\} \quad (6)$$

式中: τ_{match}^{ht} 表示相似块之间的最大距离; N_{\max}^{ht} 表示相似块的数量; Z_N 表示 N_{\max}^{ht} 相似组的数目; Z_N 通过 3D 变换后,再使用硬阈值滤波,之后,完成 3D 线性变换,可表示为:

$$Y_{N_{\max}^{ht}}^{ht} = T_{3D}^{ht-1}(\gamma(T_{3D}^{ht}(Z_{N_{\max}^{ht}}))) \quad (7)$$

式中: γ 表示硬阈值操作; T_{3D}^{ht} 表示 3D 线性变换; ω^{ht} 表示聚合计算的权重。

将成功匹配的区域块变换后, $\gamma(T_{3D}^{ht}(Z_{N_{\max}^{ht}}))$ 表示为 N_{har} , 得到聚合权重计算公式,即:

$$\omega^{ht} = \begin{cases} \frac{1}{\sigma^2 N_{har}}, & N_{har} \geq 1 \\ 1, & \text{其他} \end{cases} \quad (8)$$

式中: σ^2 表示噪声方差。

每个目标块的基础估计表示为:

$$y^{basic}(x) = \frac{\sum_{x \in I} \sum_{\omega \in N_{\max}^{ht}} \omega^{ht} Y_{x_{\omega}}^{ht}(x)}{\sum_{x \in I} \sum_{\omega \in N_{\max}^{ht}} \omega^{ht} \chi_{x_{\omega}}(x)}, \quad \forall x \in X \quad (9)$$

式中: y^{basic} 表示基础估计; $\chi_{x_{\omega}}$ 是相似块的特征函数。

2) 块匹配距离计算的改进

针对声呐图像中的非均匀噪声,基础估计中,运用曼哈顿距离代替原始的欧氏距离。BM3D 算法首先对图像

进行块匹配,计算块之间的欧氏距离,并进行二次求和,扩大块的范围,而利用曼哈顿距离进行等权重求和,不会扩大块的范围。这种曼哈顿距离计算处理非均匀噪声,通过有效的块匹配,使其在图像去噪方面表现优于其他方法。

块匹配分组时,在原始的噪声图像选择 $n \times n$ 区域的目标匹配块,在目标匹配块周围选取 $k \times k$ 区域作为搜索区域,假设 P 为目标匹配块, P_{match} 为区域 $k \times k$ 作为搜索区域内搜索到的匹配块。在搜索的过程中,目标块 P 与匹配块 P_{match} 通过曼哈顿距离来度量,其计算公式为:

$$d(P, P_{match}) = \frac{1}{k^2} \sum_i^k \sum_j^k |P(i, j) - P_{match}(i, j)| \quad (10)$$

式中: $P(i, j)$ 表示目标块 P 在 (i, j) 位置的像素灰度值; $P_{match}(i, j)$ 表示匹配块的像素灰度值。根据式(10)可以得到以目标块 P 附近区域内所有 P_{match} ; 按照它们与 P 之间的距离 $d(P, P_{match})$ 排序,再取前 $MAXB$ 个相似块,组成一个三维矩阵集合,其计算公式为:

$$P_{set} = \{P_{match}; d(P, P_{match}) \leq MAXB\} \quad (11)$$

式中: d 表示两个图块之间的曼哈顿距离; $MAXB$ 为超参数,可以调整。

通过上述公式,将这些匹配块 P_{set} 与目标块 P 一起整合为新的三维矩阵,之后,进行协同滤波。

1.4 实验结果与分析

为了验证改进算法的可靠性,对真实的前视声呐图像进行去噪,设置了 2 组实验,分别为块匹配距离计算方法对比实验和多种降噪算法对比实验。降噪结果使用主观和客观的评价方式。主观上,通过观察目标噪声情况和边缘特征保持情况进行评价;客观上,通过公认的图像评价指标评价,其中包括均值误差(mean square error, MSE)、峰值信噪比(peak signal-to-noise ratio, $PSNR$)、结构相似度(structural similarity, $SSIM$)。

1) 评价指标

MSE 表示降噪前后的像素差异,值越小表示差异越少,降噪效果越好,计算方式为:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [Q(i, j) - K(i, j)]^2 \quad (12)$$

式中: m 为图像宽度; n 为图像高度; Q 表示清晰图像, K 表示噪声图像。

峰值信噪比 $PSNR$ 是一种基于 MSE 的图像质量评价指标,计算公式为:

$$PSNR = 10 \lg \left(\frac{P^2}{MSE} \right) \quad (13)$$

式中: P 表示图像中像素值中最大值,声呐图像中 P 取 255 即可; $PSNR$ 衡量了两幅图像之间的像素点误差, $PSNR$ 越大表示图像的插值效果越好^[12]。

结构相似度 *SSIM* 评估了降噪前后的相似度,包括结构相似度(式(14))、亮度相似度(式(15))和对比度的相似度(式(16))。

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (14)$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (15)$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (16)$$

SSIM 计算公式为:

$$SSIM = [l(x,y)]^\alpha [c(x,y)]^\beta [s(x,y)]^\gamma \quad (17)$$

式中: α, β, γ 表示 3 个模块的比例分配。一般取值为 1:1:1。将式(14)~(16)带入式(17),整理后可得 *SSIM* 的计算表达式:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)^\alpha (2\sigma_{xy} + C_2)^\beta (\sigma_{xy} + C_3)^\gamma}{(\mu_x^2 + \mu_y^2 + C_1)^\alpha (\sigma_x^2 + \sigma_y^2 + C_2)^\beta (\sigma_x\sigma_y + C_3)^\gamma} \quad (18)$$

2) 块匹配距离计算方法对比实验

对比基于欧氏距离 (Euclidean distance, ED) 匹配和本文提出的基于曼哈顿距离 (Manhattan distance, MD) 匹配方法,在各种参数一致的情况下,对比降噪结果。设置了 4 组实验,比较了在基础估计和最终估计采用 ED 和 MD 的降噪效果。本节随机抽取了来自不同环境、不同目标的 3 张图片数据,如图 5 所示,图 5(a) 为球体,图 5(b) 为蛙人,图 5(c) 为鱼类,将这 3 张图作为基准图像,原图呈现出或明或暗的且分布不均的噪声。

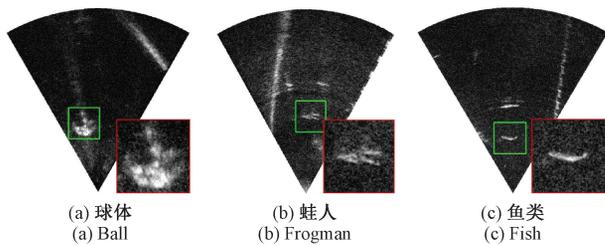


图 5 实验原图

Fig. 5 Original diagram of the experiment

实验结果如图 6 所示。左侧 ED+ED 表示在基础估计和最终估计的块匹配阶段距离计算都采用基于欧氏距离算法,即原始的 BM3D 算法;MD+MD 表示在基础估计和最终估计的块匹配阶段距离计算都采用基于曼哈顿距离算法;ED+MD 表示基础估计采用欧氏距离计算,最终估计采用曼哈顿距离计算;MD+ED 为本文提出算法,右下角为图像方框内容放大的结果。

块匹配距离计算改进算法对比如表 2 所示,可以看出,提出的距离计算方法取得了良好的去噪效果,*MSE* 越小表示降噪误差越小,*PSNR* 越大表示信号与噪声的

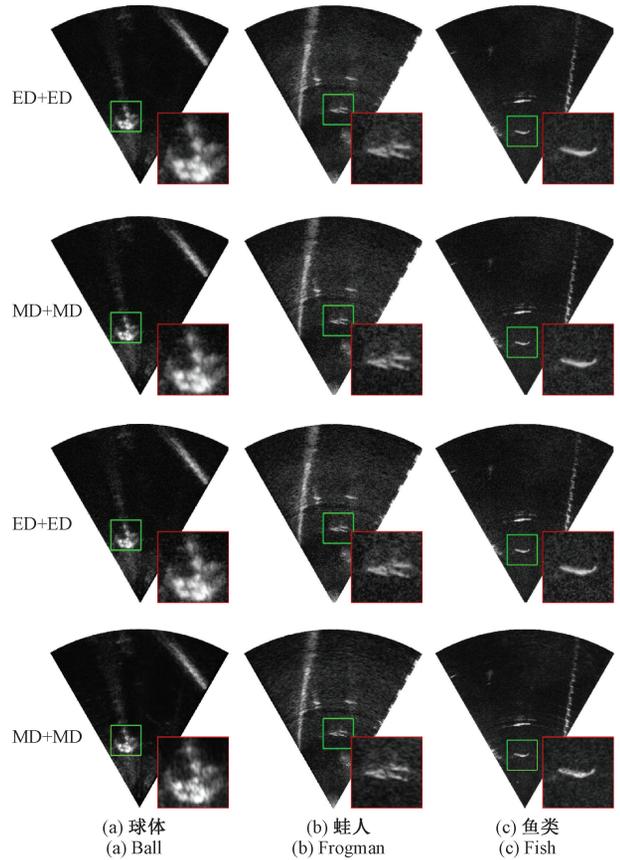


图 6 块匹配距离计算改进对比

Fig. 6 Comparison of improved block matching distance calculation

表 2 块匹配距离计算改进算法对比

Table 2 Comparison of improved block matching distance calculation algorithms

基础估计	最终估计	均方误差 ↓	峰值信噪比/dB ↑	结构相似度 ↑	时间/s ↓
ED	ED	139.72	26.68	0.92	5.33
MD	MD	122.81	27.24	0.92	5.42
ED	MD	128.60	27.04	0.93	5.37
MD	ED	114.88	27.53	0.93	5.36

比值越大,*SSIM* 越大表示图像效果越好,Time 越小表示单张图像降噪速度越快。比较 ED+ED 和 MD+MD 的距离计算方法可以看出,使用基于曼哈顿的距离计算方法比基于欧氏的距离计算方法 *MSE* 降低 16.91,*PSNR* 提升了 0.56 dB,只牺牲了 0.09 s 的速度。这是因为采用 ED 的块匹配,在二次求和中扩大了块的范围,而利用 MD 进行等权重求和,不会扩大块的范围,从而在有非均匀噪声的块匹配中找到更多的相似块,用于相似块的多帧平均。比较 MD+MD 和 MD+ED 的距离计算方法可以看出,并不是两次估计都采用 MD 的方式效果最优,最终估计使

用 ED 计算效果反而更好,这是因为在基本估计中,已经滤除大量噪声,继续采用 MD 难以找到真正相似的图像块,结合图 6 也可以看出,最终估计采用 MD 会损失更多的细节。

使用 MD+ED 改进后的方法相比于 ED+ED 方法, MSE 降低了 21.6%, PSNR 提高了 2.9%, SSIM 提高了 1.1%,时间增加了 0.6%。

3) 多种降噪算法对比实验

实验使用 Python3.8.10 中的 Opencv 和 Numpy 模块实现,将图像搜索半径设置为 16,图像块的尺寸设置为 4×4,搜索步长为 2,相似块的最大数量为 8。为了验证本研究提出的算法的有效性,将改进后的 BM3D 算法与其

他主流去噪算法对比测试。所有模型都使用相同的数据测试,遵循控制变量的原则。

对图 5 的 3 幅带噪声的前视声呐图像分别进行中值滤波、均值滤波、高斯滤波、维纳滤波、双边滤波、小波变换、深度卷积神经网络滤波 (deep convolutional neural network filtering, DCNN)、原始 BM3D 算法和本研究改进算法处理,效果图如图 7~9 所示。图 7(a)~9(a)为原图,图 7(b)~9(b)和图 7(c)~9(c)为中值滤波和均值滤波算法的降噪效果,通过对比原图可以发现去除了大量噪声也模糊了边缘,这不利于后续的目标识别。图 7(d)~9(d)、图 7(e)~9(e)和图 7(g)~9(g)分别为高斯滤波、维纳滤波和小波变换去噪算法的降噪结果,通

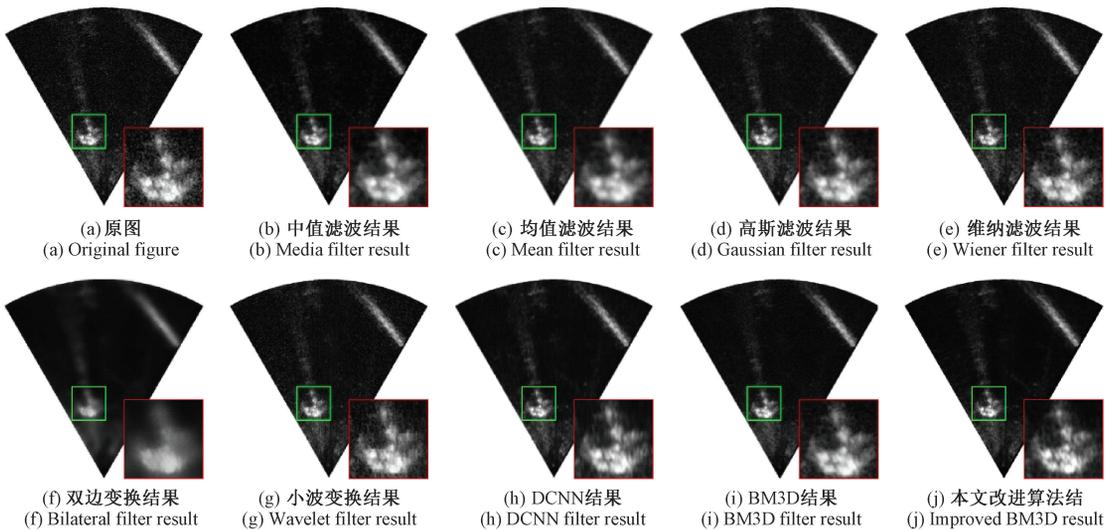


图 7 球类目标降噪算法效果对比

Fig. 7 Comparison of denoising algorithm effects for ball target

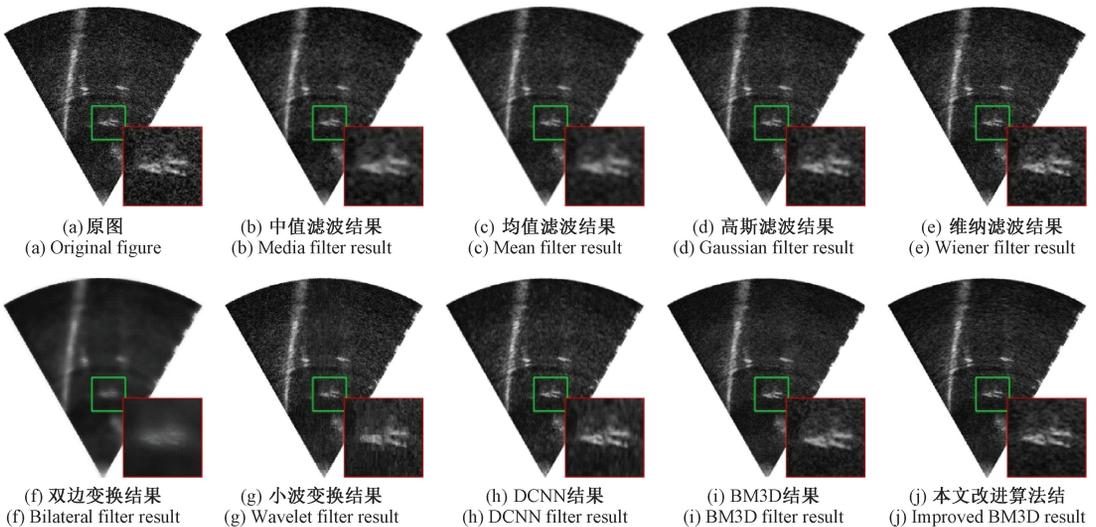


图 8 蛙人目标降噪算法效果对比

Fig. 8 Comparison of the denoising algorithm effect for frogman target

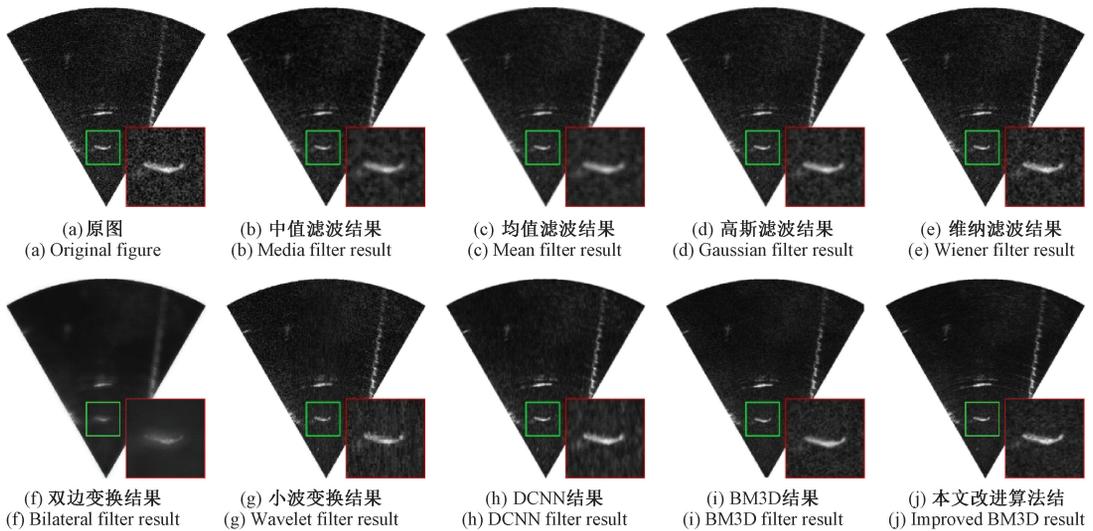


图9 鱼类目标降噪算法效果对比

Fig. 9 Comparison of denoising algorithms for fish image

过对比原图可以发现目标特征得到了保留,但噪声仍然存在,去噪效果并不好。图7(f)~9(f)双边滤波算法去噪效果最好,但目标大部分特征丢失了。图7(h)~9(h)和图7(i)~9(i)分别为DCNN和BM3D的去噪效果,可以看到对大部分区域的噪声有良好的剔除效果,并且保留了大部分边缘,比较可以发现BM3D去噪效果优于DCNN,但特征保留能力弱于DCNN,噪声最强烈的地方去噪效果不理想。图7(j)~9(j)为改进后的BM3D算法结果,可以看出,有良好的降噪效果,并很好地保留了边缘和细节,达到了抑制不均匀噪声的效果。

本研究改进算法与主流降噪算法的比较结果如表3所示。

表3 降噪算法效果对比

Table 3 Comparison of denoising algorithms

方法	均方误差↓	峰值信噪比/dB↑	结构相似度↑	时间/s↓
中值滤波	261.80	23.95	0.83	0.002
均值滤波	214.50	24.82	0.86	0.002
高斯滤波	210.63	24.90	0.83	0.003
维纳滤波	282.51	23.62	0.76	0.04
双边滤波	411.12	21.99	0.63	0.02
小波变换	386.31	22.26	0.65	0.31
DCNN	191.16	25.32	0.90	1.92
BM3D	139.72	26.68	0.92	5.33
改进 BM3D (本文)	116.75	27.46	0.93	3.63

对比目前最先进的基于深度学习的DCNN去噪算法,可以看出,本研究算法在降噪效果上均高于DCNN算法,但速度稍逊,其中, MSE 降低了74.41; $PSNR$ 提升了2.14 dB; $SSIM$ 提升了0.03;速度落后了1.71 s。所以相对于DCNN算法,本研究算法具有更好的去噪效果、不均匀噪声抑制和保存细节的能力。

对比改进前的BM3D算法,本研究算法不仅在降噪效果上均高于改进前的算法,而且速度也有很大提升,其中, MSE 提升了19.2%; $PSNR$ 提升了3.9%; $SSIM$ 提升了1.1%;速度提升了46.8%。

2 改进的YOLOv8-s目标检测算法

由于前视声呐图像目标特征不明显、目标偏小和亮度不均衡的特点,需要更高的特征提取能力和更好的信息保留能力的网络^[13]。然而,强大优秀的网络往往计算量也大,需要强力的算力设备支持^[14]。基于前视声呐图像的这些特点,本研究选用一种更深更复杂且更为轻量的网络作为YOLOv8-s的主干网络,进一步提升水下声呐场景的模型检测能力。

2.1 基于ConvNeXt的C2N算法改进

在YOLOv8-s中,主干网络是指C2f,其主要作用是特征提取,以提高目标检测的准确性。具体来说,主干网络通常采用特定的网络结构,可以对输入图像进行多层次特征提取,使得网络能够学习到不同层次的抽象特征。通过主干网络的引入,YOLOv8-s可以实现更加准确的目标检测。

早期特征提取采用不断的卷积以提取更深层次的特征,如图10(a)所示,这会导致梯度消失问题,后续的特

征提取网络都采用图 10(b)的残差结构,这种结构将卷积提取出的特征与卷积前的特征叠加,形成 2 倍于输入通道数的特征图。YOLO 系列的主干网络也采用这种结构,让梯度流更加丰富^[15]。

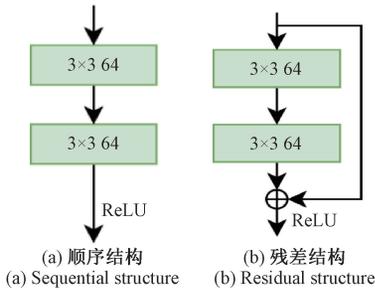


图 10 特征提取网络

Fig. 10 Feature extraction network

受到 Liu 等^[16-18]提出的 ConvNeXt 的启发,ConvNeXt 结合了 Transformer 的各种模块组件,包括训练策略、模型架构、反向瓶颈和大内核大小。这些增强功能与基本卷积神经网络堆栈和参数的优化相结合,可以显著提高推理速度。因此,在 ImageNet-22K 上进行评估时,该模型展现了优异的精度与性能。

原始 YOLOv8-s 的 C2f 模块如图 11(a)所示,其中的 Bottleneck 模块采用常见的卷积神经网络进行特征提取,这导致它无法捕获关键的局部特征和上下文信息。因此,在小目标检测中,导致检测精度的降低。相反,如图 11(b)传统的卷积运算在 ConvNeXt 模块中被丢弃,并由深度可分离卷积代替。与常见卷积相比,深度可分离卷积具有减少参数和计算量的优点。此外,在与 MobileNetv2^[19]进行比较之后,ConvNeXt 模块实现了一个反向瓶颈层结构,其特征在于一个宽的中心部分和一个窄的端点。这种设计使 ConvNeXt 模块能够增强捕获特

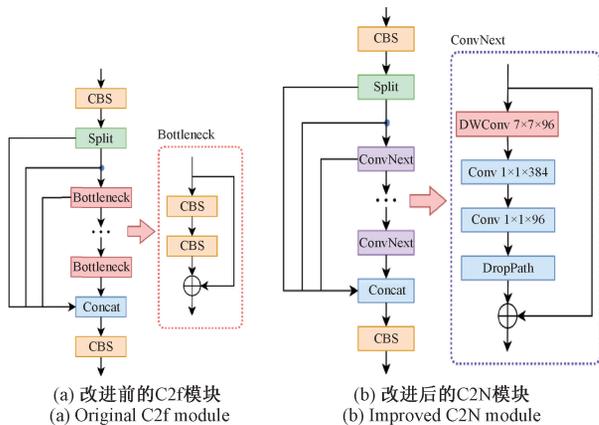


图 11 改进前的 C2f 模块与改进后的 C2N 模块结构对比
 Fig. 11 Comparison of the structure of the C2F module before improvement and the C2N module after improvement

征相关性的能力,同时有效地减少在特征空间内的变换期间由维度压缩引起的信息损失。

本研究将 ConvNeXt 模块集成到 YOLOv8-s 模型的颈部网络结构的 C2f 模块中,从而结合了两者的优势。得到的 ConvNeXt2(C2N)模块,由 2 个卷积层和 1 个包含多个 ConvNext 层的顺序模块组成。与 C2f 模块相比,C2N 模块具有更多的卷积层和全连接层,增强了网络深度和复杂性,并最终增强了模型的表达能力。值得注意的是,C2N 模块具有更高的特征提取能力和更好的信息保留能力,使其更适合处理声呐目标检测中固有的复杂性。

2.2 检测头模块算法改进

在 YOLOv8-s 中,检测头是指网络结构的最后几层,其主要作用是将主干网络提取到的目标特征映射转化为目标的位置和类别,以实现目标检测的功能。检测头通常采用卷积层和全连接层的组合,它可以将抽象的特征映射到目标检测的空间坐标和类别信息,使得网络能够准确地定位和分类目标物体。

YOLOv8-s 模型的骨干网络通过 3 轮下采样获取 3 个尺度的目标检测头,表示为 P3、P4 和 P5,如图 12(a)所示。这些层对应于颈部中不同尺寸的特征图,其中 P_i表示对应于原始图像的分辨率。模型分别从三级特征图导出的检测头上进行目标检测。当输入图像尺寸为 640 pixels×640 pixels 像素时,检测特征图的相应尺寸为 80×80、40×40 和 20×20,分别被设计用于尺寸为 8×8、16×16 和 32×32 的目标检测。基于特征图的大小来划分网格,并且使用预测框来预测每个图像元素并将其分配给目标。最后,该模型将目标位置和分类信息存储在输出特征图中。然而,在声呐目标检测中,由于声呐的检测范围较大,经常遇到较小的鱼类目标。来自这些物体的大部分特征信息在多个下采样阶段期间丢失,尽管用更高分辨率的 P3 检测头处理它们,要检测这些小目标仍具有挑战性。

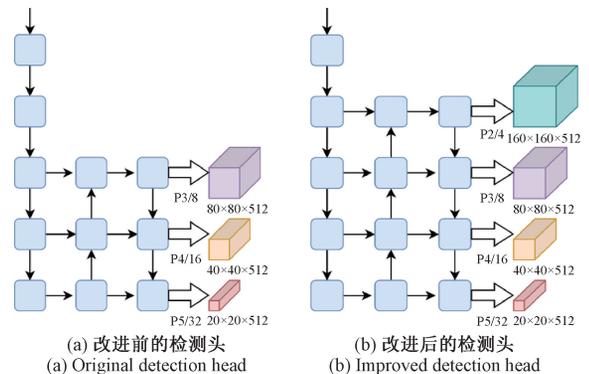


图 12 改进前后检测头对比
 Fig. 12 Comparison of the detection head before and after improvement

为了提高检测较小目标的准确性,如图 12(b)所示,引入了一个额外的浅检测头,表示为 P2,使用 160×160 检测特征图,旨在检测大于 $4 \text{ pixels} \times 4 \text{ pixels}$ 的目标。在改进后,尽管增加了计算量和模型存储量需求,但明显提高了小目标的检测精度。

1) NWD 损失函数的改进

考虑到交并比(intersection over union, IoU)度量在不同大小的目标之间显示的灵敏度的显著变化以及声呐数据集中小目标的普遍存在性,仅仅依赖于传统的 IoU 计算方法,很难评估较小目标的检测能力。引入了一种新方法来提高模型在识别此类微小目标时的精度,该方法利用归一化 Wasserstein 距离(normalized distance, NWD)来评估边界框之间的相似性,该方法对尺度差异不太敏感,特别适合小物体检测任务。除此之外,由于 NWD 评估边界框之间的相似性,即使面对广泛重叠的目标,仍能保持较高精度。因此,它更适合于前视声呐的小目标检测。边界框 A 和 B 需要被建模为高斯分布 N_a 和 N_b 。

$$A = (cx_a, cy_a, \omega_a, h_a) \quad (19)$$

$$B = (cx_b, cy_b, \omega_b, h_b) \quad (20)$$

以指数形式归一化距离度量以获得 NWD, NWD 通过计算两个高斯分布之间的距离来量化它们之间的相似性。

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right) \quad (21)$$

式中: C 是与数据集相关的常数。

在小目标检测中, NWD 在尺度、定位和遮挡目标相似性评估方面优于传统 IoU 模型。然而,采用 NWD 方法需要将边界框建模为二维高斯分布并评估相似性,导致网络收敛速度减慢,延长了训练过程。因此,直接使用 NWD 取代 IoU 并非最佳做法。将 IoU 和 NWD 合并为一个,更全面的评估边界框相似性,同时减少计算时间。通过微调 IoU 和 NWD 之间的权重,得到改进的损失函数。

$$LOSS_{Nl} = r_N \times LOSS_N + (1 - r_N) \times LOSS_I \quad (22)$$

式中: r_N 是 NWD 的权重; $1 - r_N$ 是 IoU 的权重。

2) 改进后的网络结构

首先,改进框架中的 C2N 模块,并将其应用于主干网络和颈部组件中,其中的深度可分离卷积和反向瓶颈层结构可以减少在下采样和前向传播期间的特征信息损失;其次,在网络预测层中引入了浅特征层,以增强大于 4×4 的像素的特征提取;最后,采用高斯分布的归一化加权距离 NWD 在梯度下降期间与交集 IoU 相结合。改进优化后的 YOLOv8-s 整体网络结构如图 13 所示。

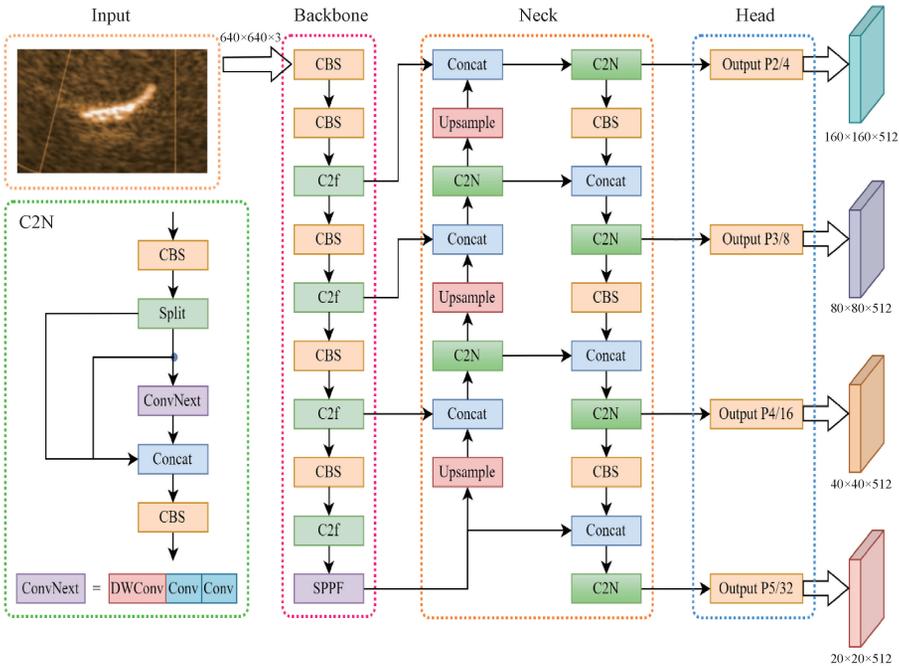


图 13 改进的 YOLOv8-s 模型结构

Fig. 13 Improved YOLOV8-S model structure

3 实验结果与分析

3.1 评价指标

在目标检测领域最常见的评价指标为平均精度均值

(mean average precision, mAP), 它由各个类别的精确率 (Precision) 和召回率 (Recall) 曲线面积平均值求得, 相关公式为:

$$Precision = \frac{TP}{TP + FP} \quad (23)$$

$$Recall = \frac{TP}{TP + FN} \quad (24)$$

预测结果为 TP 还是 FP 由预测框和真实框的交并比 IoU 和设置的阈值决定,计算示意图如图 14 所示。当分类正确且 IoU 大于设定阈值,则记为 TP ,相应的,计算出 FP 、 FN 、 TN ,就可以获得相应阈值的准确率与召回率。由它们围成的曲线面积求得 AP ,再算出所有类别的 AP 平均值,即为 mAP 。 $mAP@0.5$ 即 IoU 阈值为 0.5 时,计算的平均精度。

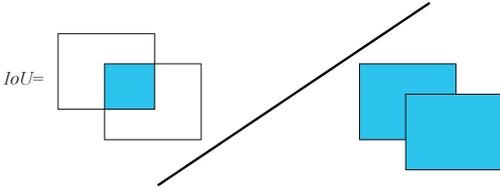


图 14 IoU 计算示意图

Fig. 14 Diagram of IoU calculation

3.2 多种检测算法对比实验

本次实验使用 500 张图像,主要包括鱼、潜水员。按照 8:1:1 的比例分为训练集、验证集、测试集。实验环境如表 4 所示,使用 s 模型权重作为预训练模型。训练的总数量为 200 轮,批次样本的数量为 8,初始学习率为 0.01,训练方法为随机梯度下降。

表 4 目标检测实验环境参数

Table 4 Experimental environment parameters of target detection

环境名称	配置参数
中央处理器	Intel i9-10900k CPU @ 3.70 GHz
显卡	NVIDIA GeForce RTX 3090 GPU
内存	SAMSUNG DDR4 3200MHz 64 GB
操作系统	Windows 11
编译语言	Python 3.8.18
集成开发环境	PyCharm 2021.3
CUDA	11.1
CUDNN	8.0.5
Torch	1.9.0
Torchvision	0.10.0
Numpy	1.24.4

为了验证提出的模型 DP-YOLOv8-s 的有效性,DP-YOLOv8-s 模型与其他主流检测模型进行了对比测试,遵循控制变量的原则。表 5 显示了本研究改进算法与其他先进检测算法之间的 mAP 和模型大小(model size, MS) 的比较。

表 5 检测算法性能对比

Table 5 Performance comparison of detection algorithms

模型	精度/ % ↑	$mAP@0.5$ / % ↑	模型大 小/MB ↓	帧率 ↑
Faster R-CNN ^[20]	71.2	67.3	331	20
YOLOv5-s ^[21]	76.0	73.4	7.2	156
YOLOX-s ^[22]	78.4	75.7	9.0	58
YOLOv6-s ^[23]	80.7	77.1	18.5	225
YOLOv7-s ^[24]	80.0	76.4	72.0	192
YOLOv8-n ^[25]	71.8	63.7	3.2	252
YOLOv8-s ^[25]	82.1	81.4	11.2	207
YOLOv8-m ^[25]	82.8	84.2	25.9	135
DP-YOLOv8-s(本研究)	87.2	85.4	15.8	174

分析表 5 可以发现,改进后的模型的准确率为 87.2%, $mAP@0.5$ 为 85.4%。与改进前的 YOLOv8-s 模型相比,本研究的模型大小只增加了 4.6 MB,但是精度增加了 5.1 个百分点, $mAP@0.5$ 增加了 4 个百分点,说明针对声呐小目标改进后的 YOLOv8-s 能有效的提升目标检测精度。对比 YOLOv8-m 模型,模型大小增加了 14.7 MB,但是 $mAP@0.5$ 的检测精度只提升了 2.8 个百分点,说明增加模型的深度并不能有效的提高声呐图像的检测效果。对比 Faster R-CNN 可以发现,双阶段目标检测模型与目前先进的单阶段目标检测模型相比各方面都处于劣势。

为了进一步验证改进算法的有效性,并排除误检率,从待检测数据集中随机选取了 3 张图片。如图 15 所示,最上侧是真实目标的标注情况,这 3 张图分别包含有 4 个鱼目标、3 个鱼目标和 2 个蛙人目标。通过对比不同的算法检测结果,单阶段的检测效果目前优于双阶段 Faster R-CNN。通过对比改进前后的 YOLOv8-s 的检测结果,可以看出,改进前的 YOLOv8-s 在图 15(a) 中遗漏了一个鱼目标。在改进后的 DP-YOLOv8-s 得到了解决,这是因为该算法添加了浅特征检测头,提高了对更小目标检测的敏感度。图 15(b) 改进前,将图中噪声误判为了鱼目标,在改进后得到了解决,这是因为使用了改进的 C2N 模块作为特征提取网络,增加了网络深度,提高了模型表达能力,从而模型具有更强的抗干扰能力。改进前的算法图 15(c) 中,将较小蛙人目标识别为了鱼目标。通过对比改进前后的目标检测置信度可以发现,改进后的 YOLOv8-s 可以有效提高小目标的置信度,在提高检测精度和 mAP 的同时,降低了漏检率和误检率。

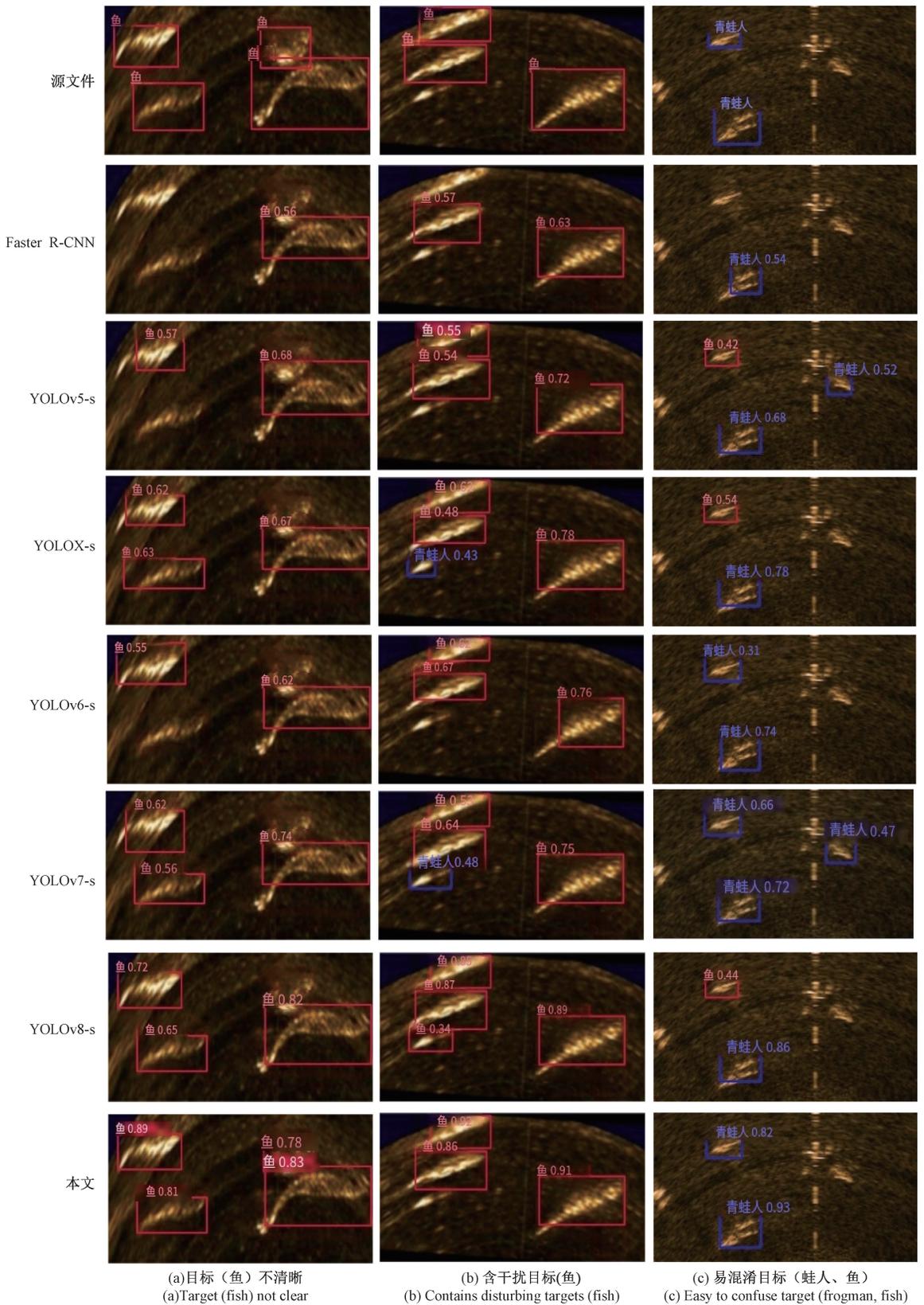


图 15 不同算法检测结果对比

Fig. 15 Comparison of detection results of different algorithms

4 结 论

在声呐图像降噪部分,对BM3D算法做了两点改进,针对BM3D算法偏慢的问题,根据前视声呐图像特征区域少的特点,在相似块匹配的过程中添加预过滤器,减少相似块的个数,从而减少3D转换的处理计算量;针对声呐图像噪声分布不均匀问题,在基础估计的相似块匹配距离计算过程中采用曼哈顿距离替代原来的欧氏距离,从而能够更好地处理声呐图像中的不同类型和强度的噪声;通过块匹配距离计算方法对比实验证明了改进后的曼哈顿距离算法计算相似块距离的优越性;与多种降噪算法对比,改进后的BM3D算法在PSNR、SSIM评价指标下都取得了最优,证明了改进方法的有效性。在前视声呐降噪处理之后,对目标识别方法进行了改进。

结合前视声呐图像的特点对YOLOv8-s算法网络进行改进,解决声呐图像中较小目标识别效果差的问题。改进包括3个关键方面:特征提取、检测头和损失函数。针对目标的特征丢失问题,在主干网络中引入改进的C2N模块,进一步减少下采样和前向传播过程中的特征信息损失;为了提高检测较小目标的准确性,在原始的YOLOv8-s网络通过添加P2浅层特征层进行扩展,以实现4种不同尺度的目标检测;针对现有损失函数对小物体不敏感问题,将NWD与IoU相结合,以帮助模型捕获更多的细节和空间信息。最后,进行多组不同算法的对比实验,结果表明,改进算法确保了最高的mAP,提高了检测精度和速度。

参考文献

- [1] NEVES G, RUIZ M, FONTINELE J, et al. Rotated object detection with forward-looking sonar in underwater applications[J]. *Expert Systems with Applications*, 2020, 140: 112870.
- [2] GASPAR A R, MATOS A. Feature-based place recognition using forward-looking sonar[J]. *Journal of Marine Science and Engineering*, 2023, 11(11): 2198.
- [3] NEUPANE D, SEOK J. A review on deep learning-based approaches for automatic sonar target recognition[J]. *Electronics*, 2020, 9(11): 1972.
- [4] MA Q X, JIANG L Y, YU W X, et al. Training with noise adversarial network: A generalization method for object detection on sonar image[C]. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020: 718-727.
- [5] SLEDGE I J, EMIGH M S, KING J L, et al. Target detection and segmentation in circular-scan synthetic aperture sonar images using semisupervised convolutional

encoder-decoders[J]. *IEEE Journal of Oceanic Engineering*, 2022, 47(4): 1099-1128.

- [6] 霍冠英,刘静,李庆武,等. 空间约束FCM与MRF结合的侧扫声呐图像分割算法[J]. *仪器仪表学报*, 2017, 38(1): 226-235.
- HUO G Y, LIU J, LI Q W, et al. Side scan sonar image segmentation algorithm based on space constrained FCM and MRF[J]. *Chinese Journal of Scientific Instrument*, 2017, 38(1): 226-235.
- [7] 李鹏,叶方跃,李剑乔,等. 基于小波子带模型匹配的同水域下目标探测[J]. *电子测量与仪器学报*, 2017, 31(11): 1860-1868.
- LI P, YE F Y, LI J Q, et al. Target detection in the same water area based on wavelet sub band model matching[J]. *Journal of Electronic Measurement and Instrumentation*, 2017, 31(11): 1860-1868.
- [8] 马珊. 水下机器人前视声呐多目标跟踪技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2016: 44-57.
- MA SH. Multi-target tracking of AUV based on forward looking sonar[D]. Harbin: Harbin Engineering University, 2016: 44-57.
- [9] 崔杰,胡长青,徐海东,等. 改进的Mean Shift算法在前视声呐运动目标跟踪中的应用[J]. *声学技术*, 2020, 39(3): 279-283.
- CUI J, HU CH Q, XU H D, et al. Application of improved Mean Shift algorithm in moving target tracking of forward-looking sonar[J]. *Technical Acoustics*, 2020, 39(3): 279-283.
- [10] 张伟正. 基于深度学习的前视声呐图像目标检测与跟踪算法研究[D]. 哈尔滨: 哈尔滨工程大学, 2022: 35-39.
- ZHANG W ZH. Research on target detection and tracking algorithm in forward-looking sonar image based on deep learning [D]. Harbin: Harbin Engineering University, 2022: 35-39.
- [11] FRANCHI M, RIDOLFI A, PAGLIAI M. A forward-looking sonar and dynamic model-based AUV navigation strategy: Preliminary validation with FeelHippo AUV[J]. *Ocean Engineering*, 2020, 196: 106770.
- [12] HUYNH-THU Q, GHANBARI M. Scope of validity of PSNR in image/video quality assessment[J]. *Electronics Letters*, 2008, 44(13): 800-801.
- [13] ZHANG H T, TIAN M, SHAO G P, et al. Target detection of forward-looking sonar image based on improved YOLOv5[J]. *IEEE Access*, 2022, 10: 18023-18034.
- [14] 高晗,刘卫东,高立娥. 基于前视声呐的水下目标检测算法研究[J]. *计算机测量与控制*, 2022, 30(11):

- 17-24.
GAO H, LIU W D, GAO L E. Research on underwater target detection algorithm based on forward looking sonar[J]. Computer Measurement & Control, 2022, 30(11): 17-24.
- [15] 刘熹,陈晨,双丰. 基于改进 YOLOv7-tiny 的多种类绝缘子检测算法[J]. 仪器仪表学报, 2024, 45(9): 101-110.
LIU X, CHEN CH, SHUANG F. Multi-type insulator detection algorithm based on improved YOLOv7-tiny[J]. Chinese Journal of Scientific Instrument, 2024, 45(9): 101-110.
- [16] LIU ZH, MAO H Z, WU CH Y, et al. A convnet for the 2020s[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 11966-11976.
- [17] WU Z F, SHEN CH H, VAN DEN HENGEL A. Wider or deeper: Revisiting the resnet model for visual recognition[J]. Pattern Recognition, 2019, 90: 119-133.
- [18] 刘明杰,何峥言,陈俊生,等. 基于循环跨视图转换和多状态特征融合的鸟瞰图生成方法[J]. 仪器仪表学报, 2024, 45(10): 133-142.
LIU M J, HE ZH Y, CHEN J SH, et al. Bird's eye view generation based on recurrent cross-view transformation and multi-state feature fusion[J]. Chinese Journal of Scientific Instrument, 2024, 45(10): 133-142.
- [19] WANG J W, XU CH, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[J]. ArXiv preprint arXiv:2110.13389, 2021.
- [20] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]. Advances in Neural Information Processing Systems, 2017, 39(6): 1137-1149.
- [21] YAO J, QI J M, ZHANG J, et al. A real-time detection algorithm for Kiwifruit defects based on YOLOv5[J]. Electronics, 2021, 10(14): 1711.
- [22] GE ZH, LIU S T, WANG F, et al. Yolox: Exceeding YOLO series in 2021[J]. ArXiv preprint arXiv: 2107.08430, 2021.
- [23] 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.
- [24] SUN K X, CONG CH. Research on chest abnormality detection based on improved YOLOv7 algorithm[C]. 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2022: 3884-3886.
- [25] WANG X Q, GAO H B, JIA Z M, et al. BL-YOLOv8: An improved road defect detection model based on YOLOv8[J]. Sensors, 2023, 23(20): 8361.

作者简介



陈美龙, 2021 年于哈尔滨学院获得学士学位, 2024 年于哈尔滨工程大学获得硕士学位, 现就职于美团, 任前端研发工程师, 主要研究方向为图像识别和 AIGC。

E-mail: mileschen666@163.com

Chen Meilong received his B. Sc. degree from Harbin college in 2021, received his M. Sc. degree from Harbin Engineering University in 2024. Now he is working at Meituan as front-end R&D engineer. His main research interests include image recognition and AIGC.



赵新华(通信作者), 2005 年、2008 年于哈尔滨工程大学获得硕士、博士学位, 现为哈尔滨工程大学副教授, 主要研究方向为人工智能、声呐图像识别与检测。

E-mail: zhaoxinhua@hrbeu.edu.com

Zhao Xinhua (Corresponding author) received her M. Sc. degree and Ph. D. degree from Harbin Engineering University in 2005 and 2008. Now she is an associate professor at Harbin Engineering University. Her main research interests include artificial intelligence and sonar image recognition and detection.