

## 基于YOLOv8X和DIDSON的水域鱼类资源监测分析

沈蔚<sup>1,2</sup>, 董世泓<sup>1,2</sup>, 刘梦骐<sup>1,2</sup>, 殷兆伟<sup>1,2</sup>, 钱恩泽<sup>1,2</sup>, 龚小玲<sup>3</sup>

(1. 上海海洋大学 海洋科学与生态环境学院, 上海 201306; 2. 上海河口海洋测绘工程技术研究中心, 上海 201306; 3. 上海海洋大学 水产与生命学院, 上海 201306)

**摘要:** 为提高水域鱼类资源监测的自动化程度和实时分析能力, 结合YOLOv8X(You only look once version 8 - extra large)目标检测模型、ByteTrack(ByteTrack: a strong baseline for multi-object tracking)算法与双频识别声呐(Dual-frequency identification sonar, DIDSON)数据, 开发了1种快速、准确的鱼类目标识别与计数方法。实验结果表明, YOLOv8X与ByteTrack联合方法与传统的Echoview软件识别精度接近(偏差率仅为1.36%), 但处理时间显著减少(单条测线从约30 min减少至约3 min), 表现出较强的实时处理能力和泛化性能。同时, 通过重复实验验证了该方法的稳定性, 确认其不同场景中的可靠性。本研究方法与成果为水域鱼类资源的自动化监测提供了可靠的技术支持, 可广泛地应用于大范围高频次的渔业资源监测与管理工作中。

**关键词:** 鱼类资源监测; 深度学习模型; YOLOv8X模型; DIDSON声呐; ByteTrack算法

**中图分类号:** S 932.4 **文献标志码:** A

鱼类资源监测是渔业资源管理和可持续发展中的重要组成部分, 快速准确的鱼类资源调查为渔业资源的可持续开发和保护提供科学的数据支持<sup>[1]</sup>。然而, 传统的鱼类资源调查方法, 包括标记重捕和捕捞统计等, 因水下环境的复杂性, 技术的局限性和鱼类行为的多样性等面临着诸多挑战<sup>[2]</sup>。现代鱼类资源评估方法多借助声学仪器进行调查, 声学方法克服了光线在水下难以传播的问题, 并且对鱼类和生态环境影响较小, 被认为是非侵入性的技术<sup>[3]</sup>。

双频识别声呐(Dual-frequency identification sonar, DIDSON)也被称为“声学相机”, 能够在模糊、昏暗的水下条件下提供清晰的声学图像, 已被广泛应用于渔业管理、水下检查和环境监测<sup>[4]</sup>。研究<sup>[5-6]</sup>表明, DIDSON可以有效地取代浑浊水域中的光学系统, 为各种成像任务提供清晰接近摄影级的图像。它已被用于对养殖鱼的计数和测量<sup>[7]</sup>、估算鱼的丰度<sup>[8]</sup>、测量养殖的中国鲟鱼的游泳模式和体长<sup>[9]</sup>等方面。

Echoview (Echoview Software Pty Ltd, Hobart, Australia)数据分析软件目前广泛应用于水声学研究、渔业科学、海洋环境监测等领域<sup>[10]</sup>, 在声学数据处理和分析方面发挥了重要作用, 成为事实上的行业标准。但Echoview人机交互处理模式在处理声学大数据时存在耗时长、人工成本高等问题<sup>[11]</sup>。在此背景下, 利用深度学习技术, 特别是目标检测模型, 提高鱼类目标识别的效率, 成为了研究热点。

YOLO(You Only Look Once)系列算法作为当前领先的目标检测框架之一, 以其快速和准确的目标检测能力在多个领域得到广泛应用<sup>[12]</sup>。YOLOv8X由Ultralytics公司开发, 是YOLO系列中的一种先进目标检测模型, 相比之前版本在网络结构与检测效率方面有进一步优化, 其通过进一步优化网络结构和算法流程, 实现了更高的检测精度和实时性, 广泛应用于各种场景, 并在渔业声学领域中起到重要作用<sup>[13]</sup>。研究<sup>[14-15]</sup>表明, 与早期版本相比, 它在有挑战性的水下条件下更

收稿日期: 2024-11-27 修回日期: 2025-04-18

基金项目: 上海城投原水项目(D-8006-24-0509)

作者简介: 沈蔚(1977—), 男, 博士, 教授, 研究方向为海洋测绘和水深探测。E-mail: wshen@shou.edu.cn

通信作者: 龚小玲, E-mail: xlgong@shou.edu.cn

版权所有 ©《上海海洋大学学报》编辑部(CC BY-NC-ND 4.0)

Copyright © Editorial Office of Journal of Shanghai Ocean University (CC BY-NC-ND 4.0)

<http://www.shhydx.com>

具有卓越的性能,其在使用声呐图像检测鱼类方面表现出了更高的精度,使其成为鱼类计数和监测的有价值的工具。

在鱼类检测精度提升方面,ZHANG等<sup>[16]</sup>通过引入动态稀疏注意力机制与小目标检测模块,优化YOLOv8n模型,使其在复杂水域条件下对小型鱼类具有更高的识别能力。此外,CUI等<sup>[17]</sup>则综述了当前数字化养殖环境下鱼类检测与行为分析的发展趋势,强调了多模态传感器融合和深度神经网络在提高系统鲁棒性方面的重要性,同时指出目前仍存在缺乏标准化数据集与模型泛化能力不足等问题。

本研究将YOLOv8X与DIDSON数据相结合,以解决Echoview等传统方法需要大量人工参与的局限性,显著减少了处理声呐数据所需的时间,大大提高了自动化检测程度,避免了由于人工疲劳和主观偏见导致的识别偏差,确保了较高的识别准确性。

## 1 材料与方法

### 1.1 YOLOv8X目标检测模型

YOLOv8X是Ultralytics公司开发的YOLO系列目标检测模型之一,在YOLOv1到YOLOv7的基础上提升了精度和速度,广泛应用于目标检

测、图像分割和分类<sup>[18]</sup>。YOLOv8X模型由4个部分组成:输入层、主干网络(Backbone)、颈部(Neck)和检测层(Head)。输入层对图像数据进行预处理,主干网络通过卷积、归一化等操作提取特征,颈部整合了特征金字塔网络(Feature pyramid network, FPN)和路径聚合网络(Path aggregation network, PAN),增强了多尺度检测的能力,检测层负责边界框、类别和置信度的预测。YOLOv8X还引入了改进的空间金字塔池化(Spatial pyramid pooling, SPP)和焦点提取模块(Focus module),加强了对小物体和复杂场景的检测能力,使用复合损失函数优化分类和定位平衡性,并在推断阶段采用非最大抑制(Non-maximum suppression, NMS)以保证预测精度<sup>[19]</sup>。

### 1.2 ByteTrack多目标跟踪算法

使用ByteTrack多目标跟踪算法对识别到的目标进行跟踪计数,将检测结果分为高置信度和低置信度2个集合。高置信度检测结果通过卡尔曼滤波和匈牙利算法与上一帧的目标进行匹配,基于交并比(IOUS)进行匹配,未匹配的高置信度目标会与低置信度结果进行二次匹配,以捕捉漏检目标。对未匹配的检测目标,算法通过轨迹管理判断其是否为新目标或现有目标的延续。整个方法的流程如图1所示。

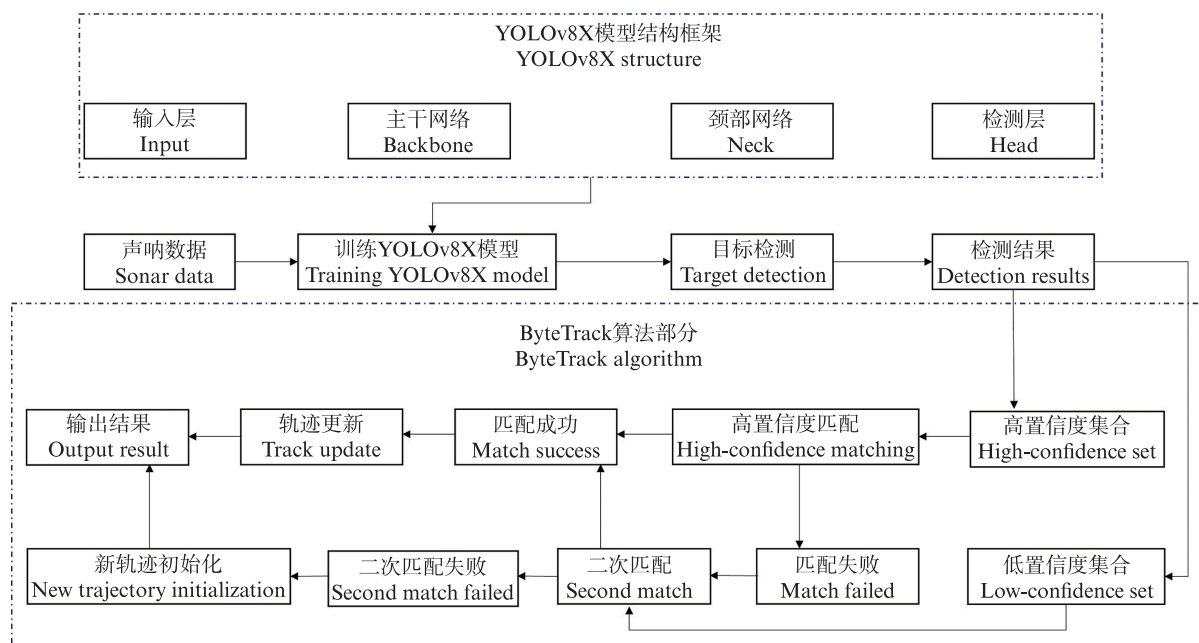


图1 算法流程图

Fig. 1 Algorithm flowchart

## 2 结果

### 2.1 数据获取

#### 2.1.1 DIDSON 双频识别声呐

DIDSON 是由 Sound Metrics 公司制造的高分辨率声呐,通过声透镜聚焦形成狭窄波束,实现低能见度水下的近乎光学级成像,原理如图 2 所示,声波通过 L1、L2、L3 三种透镜改变传播路径聚焦于传感器阵列。声透镜聚焦波束使得消耗功率低,易于发射和接收,提升了效率并缩小了设备体积<sup>[20]</sup>。DIDSON 水平视角为 29°,垂直视角为 14°,在 1.8 MHz 高频模式下,分辨率达 0.3°,适合 12 m 内高精度观察;在 1.1 MHz 低频模

式下,分辨率为 0.6°,探测距离可达 40 m,并具有自动聚焦功能,保证 1~40 m 内的清晰成像。具体参数见表 1。

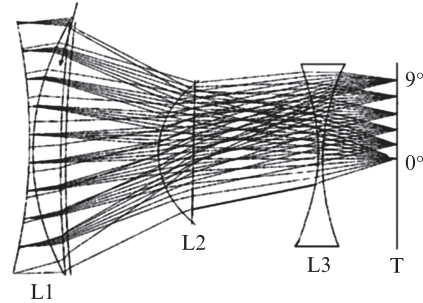


图2 DIDSON 成像原理图

Fig. 2 DIDSON imaging schematic

表1 DIDSON 声呐相关参数  
Tab. 1 DIDSON sonar related parameters

声呐性能参数 Sonar performance parameters	低频 Low frequency	高频 High frequency
工作频率 Operating frequency/MHz	1.1	1.8
波束宽度 Beam width	水平 0.6°, 垂直 12°	水平 0.3°, 垂直 12°
波束数量 Number of beams	48	96
声源级 Source level	202 dB 在 1 m 处, 参考值 1 $\mu$ Pa	206 dB 在 1 m 处, 参考值 1 $\mu$ Pa
工作范围 Start range/m	0.75 ~ 40.00	0.38 ~ 11.63
最大帧速率 Maximum frame rate/(帧/s)	4~21	
视场 Field of view	29°	
聚焦 Remote focusing	1 m 到最大范围	
空气中质量 Mass in air/kg	7.9	
水中质量 Mass in water/kg	1.0	
尺寸 Dimensions	30.7 cm×20.6 cm×17.1 cm	

#### 2.1.2 数据采集

1 年内对上海市金泽、陈行水库每季度进行 1 次数据采集,以春季上海陈行水库调查为例,利用机动船搭载 DIDSON 进行走航式鱼类资源调查,DIDSON 固定在船的一侧,水下 0.5 m 处,镜头向下 60°,采集方向与船前进方向一致,通过定制安装支架以减少航行过程中仪器的抖动,确保获得稳定清晰的高质量图像,调查航线如图 3 所示。考虑到陈行水库的具体情况,全程采用了高频模式,窗口起点设置为 0.83 m,窗长为 11.63 m。采样频率为 8 帧/s,接收增益设置为 25,阈值为 15 左右,根据实际情况进行调整。

### 2.2 声呐数据预处理

DIDSON 采集到的数据通常以矩形坐标系

形式储存,纵轴表示采样的深度,横轴表示声束的方向,为了进一步处理并增强数据的可视化效果,将矩形坐标系数据转换为极坐标视图,如图 4 所示,还原采集数据时的场景,其中转换的公式为

$$\rho = L_s + L_1 \times y/512 \quad (1)$$

$$x' = L_1 \times \sin(\alpha/2) - \rho \times \cos(90^\circ - \alpha/2 + B_w \times x) \quad (2)$$

$$y' = \rho \times \sin(90^\circ - \alpha/2 + B_w \times x) \quad (3)$$

式中: $y$ 为数组数据的 $y$ 坐标值; $L_s$ 为数据采集的起始距离; $L_1$ 为声呐探测距离; $\alpha$ 为声呐采集角度 29°; $B_w$ 为波束宽度,低频 0.6°,高频 0.3°。图像转换的过程如图 5 所示,直接读取 DIDSON 的原始图像,经过还原后生成扇形图像。

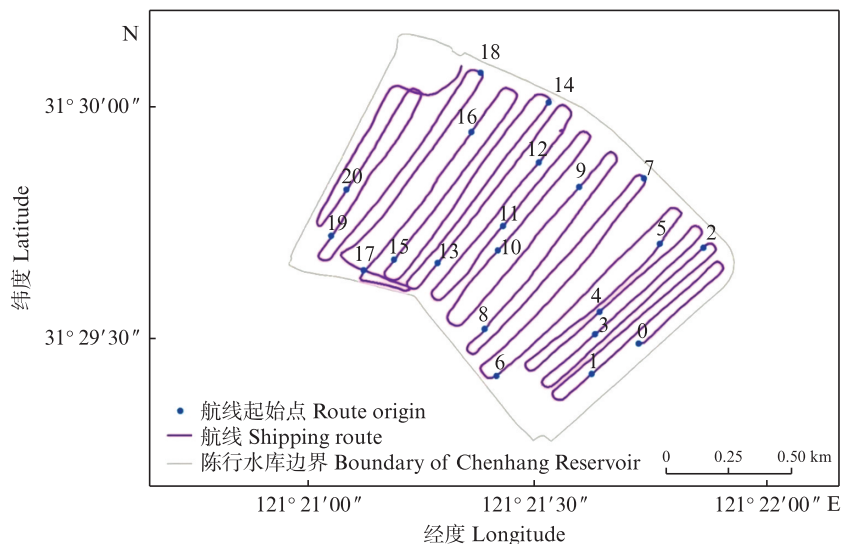


图3 陈行水库春季声呐调查航线  
Fig. 3 Spring sonar survey route of Chenhang Reservoir

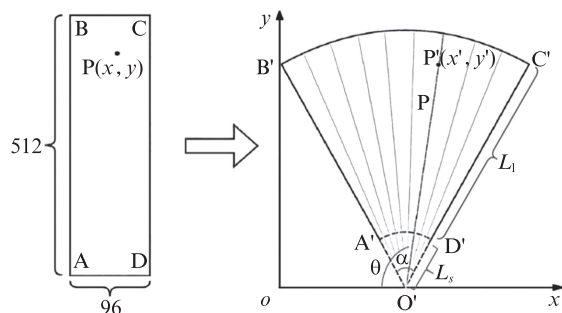


图4 坐标转换  
Fig. 4 Coordinate transformation

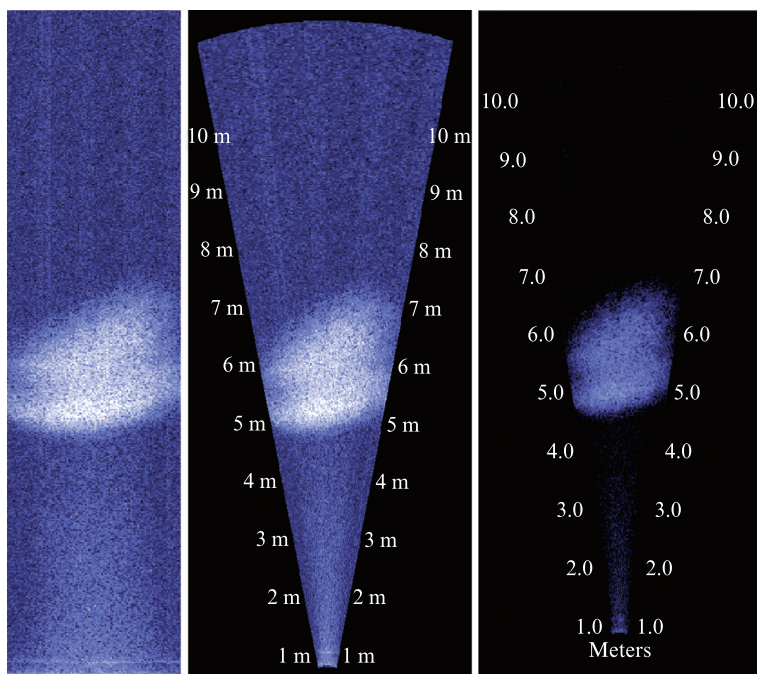


图5 预处理过程  
Fig. 5 Preprocessing process



在声呐图像预处理中,为提高鱼类目标的识别准确性,需要对原始图像中的噪声和弱反射信号进行有效滤除。由于声呐图像中每个像素的强度值代表了回波信号的反射强度,反映了水下物体的声波反射特性,因此在预处理阶段设置合理的强度阈值至关重要。

本研究中,根据实际采集水域的声呐图像特征,设置滤波强度阈值为70。即将像素强度值小于70的区域视为低回波区域,主要对应背景水体、悬浮颗粒或系统噪声等非目标信息,通过滤波将其从图像中去除。该阈值的设定基于对大量样本图像的统计分析 with 实验验证,在保留目标特征信息的同时,有效抑制了背景干扰和图像噪声,从而增强了鱼类目标的边缘轮廓与回波亮

度,提高了后续检测与识别的稳定性与精度。

### 2.3 YOLOv8X 模型训练

#### 2.3.1 模型训练

本研究模型训练在 Windows 11 操作系统上进行,硬件配置包括 NVIDIA GeForce RTX 4060 GPU 和 AMD Ryzen 9 7945HX CPU,运行环境为 CUDA 12.6,深度学习框架使用 PyTorch 2.6.0。所有原始图像和标注文件放置在 YOLOv8X 模型训练所需的相应目录中。经过多次实验并对不同训练参数的影响进行比较,最终确定使用最终版本的 YOLOv8X。这一模型既能保证较高的训练效率,又能充分利用硬件资源进行加速。模型主要参数设置见表2。

表2 模型参数  
Tab. 2 Model parameter

名称 Name	参数 Parameter	名称 Name	参数 Parameter
迭代轮次 Epochs	300	动量 Momentum	0.937
批量大小 Batch	8	权重衰减 Weight_decay	0.000 5
图像尺寸 Imgsz	640	预热轮次 Warmup_epochs	3.0
数据加载线程数 Workers	4	目标框损失权重 Box	7.5
初始学习率 lr0	0.002	类别损失权重 Cls	0.5
学习因子 lrf	0.01		

本研究从总共 88 763 帧 DIDSON 图像数据中随机抽取 8 876 帧进行初步筛选,再进一步挑选出含有明显鱼类目标的图像共 1 000 张,以确保训练数据质量和效率。为降低模型泛化风险,数据的筛选覆盖了不同水下光照条件(不同水体浑浊度)以及不同噪声场景(如船舶运动产生的噪声、气泡干扰、水底反射杂波)下的多种典型场景,力求涵盖实际应用过程中可能遇到的典型声呐图像特征,增强模型泛化能力和实际应用鲁棒性。

具体筛选标准:(1)光照条件,考虑不同时间段水下光照差异,增加训练数据多样性。(2)噪声条件,涵盖典型噪声干扰场景,包括气泡、水中悬浮杂质、底部反射、船舶振动噪声等不同噪声条件下的声呐图像。(3)鱼类行为与密度,所选数据涵盖不同鱼类种类、大小、游动姿态以及个体和群体分布的情况。

通过以上严格的数据筛选标准,训练集能够更好地反映实际监测环境下可能出现的多样场景,有效降低了训练模型的样本偏差风险,提高了模型在实际复杂环境中的泛化能力。

使用 Labelimg 图像标注工具,依次对图像数据

中的鱼类目标进行人工标注,如图6所示,用于后续模型训练。其中,1~4 m 的区域显示了由船舶噪声、气泡和水中杂质等因素引起的斑点噪声,框选的部分显示为鱼类目标,从7~10 m 的大反射面积是由水底的反射引起的。将所有的原始图像和标注文件放到 YOLOv8X 训练模型的对应目录之下进行训练。

#### 2.3.2 训练效果评估

训练结束后,依据精确度(Precision, P)、召回率(Recall, R)、平均精度(Mean average precision, mAP)以及  $mAP_{50-95}$  (在不同 IoU 阈值下以 5 为步长计算的平均精度)等指标,对模型性能进行了评估,并绘制了如图7所示的折线图。

其中,各评价指标的计算公式:

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$mAP = \int_0^1 P(R) dR \quad (6)$$

式中:TP 为正确识别鱼类目标的数量;FP 为错误识别成鱼类目标的数量;FN 为没有识别成鱼类的数量。

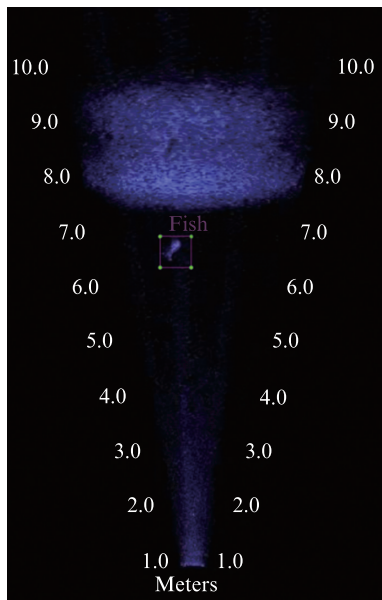


图6 Labeling标注示意

Fig. 6 Labeling annotation schematic

由图7可发现,随着迭代次数的增加,图7a的精度稳定在0.95左右,在300次迭代时达到0.953 42的峰值。图7b的召回率稳定在0.85左右,在240个迭代达到了最大值0.901 63。图7c在IoU阈值为0.5的情况下,显示的 $mAP_{50}$ 值在第257轮训练时达到0.953 03,且整体性能曲线保持平稳,说明在较低IoU阈值下,模型检测性能较优。图7d的 $mAP_{50-95}$ (IOU阈值从0.5到0.95每0.05一个计算平均精度取均值)稳定在0.43左右,在210个迭代达到的峰值为0.453 21,效果较为理想,可以用来对图像数据中的鱼类目标进行识别。

为了进一步评估YOLOv8X在水下鱼类目标检测任务中的性能,本研究采用混淆矩阵(Confusion matrix)进行分析,以量化模型在不同置信度阈值下的检测能力和分类精度。

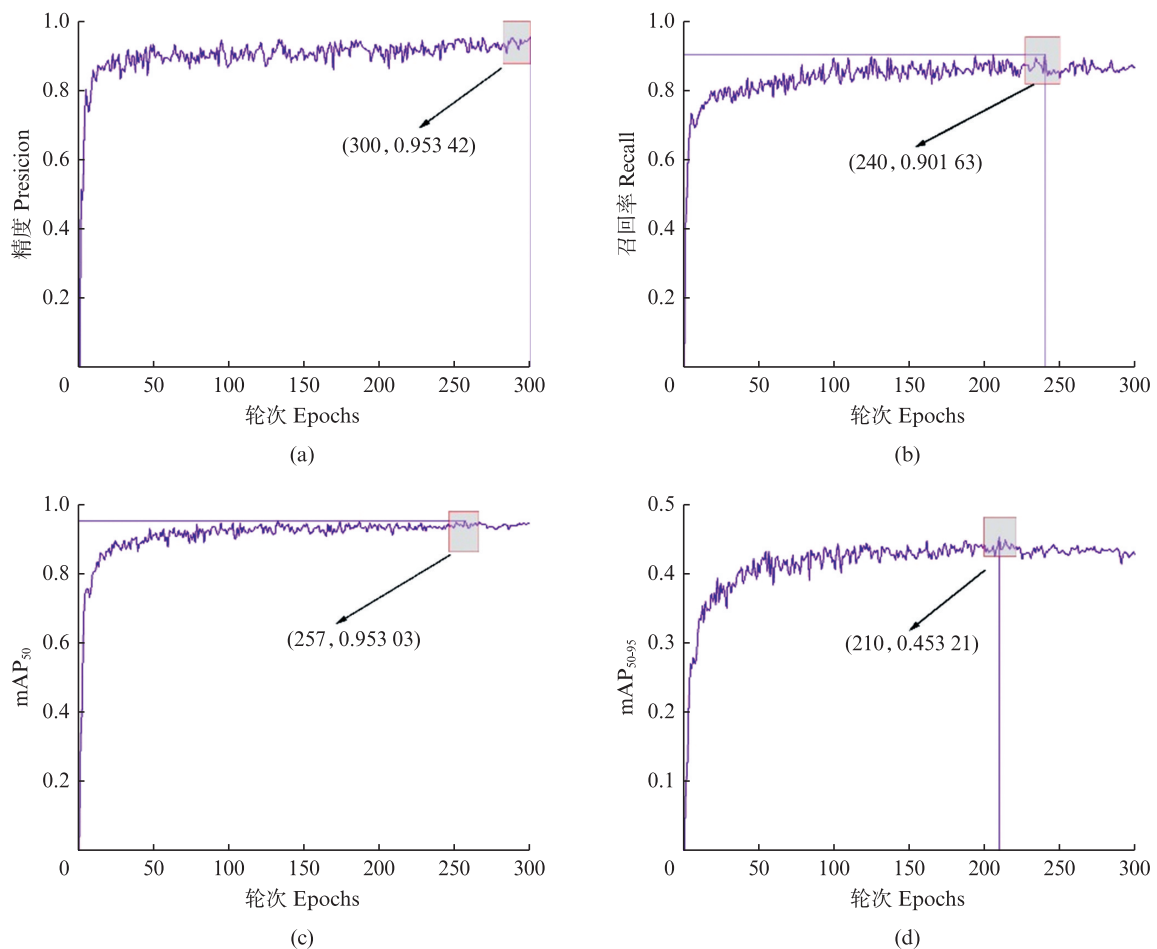


图7 模型评估参数

Fig. 7 Model evaluation parameters

如图8所示,混淆矩阵展示了YOLOv8X模型在测试数据集上的分类效果。矩阵的纵轴表示真实类别,横轴表示模型预测的类别,颜色深度反映了分类准确率。从混淆矩阵可以看出:(1)YOLOv8X在鱼类目标检测任务中表现良好,识别正确率较高;(2)绝大部分鱼类目标被正确检测,误判较少,表明模型的分类能力较强;(3)少部分目标被错误分类,通常是由于水中噪声干扰、回声重叠、目标遮挡等因素造成。

YOLOv8X在测试集上的混淆矩阵表明,其检测准确度较高,能够较好地适应水下声呐数据的复杂性。

#### 2.4 目标识别与计数

利用训练好的YOLOv8m模型,对预处理后的声呐数据进行鱼类识别,将检测到的鱼类目标传递给ByteTrack进行跟踪,并为每一个检测到的鱼类目标分配唯一ID跟踪其运动,实现精确快速的鱼类目标识别计数,如图9所示。本次调

查测线共21条,处理后得到结果与传统Echoview处理结果进行对比,如表3所示,人工重复实验为另一个有长期处理经验者使用Echoview的处理结果,三者结果比较接近。

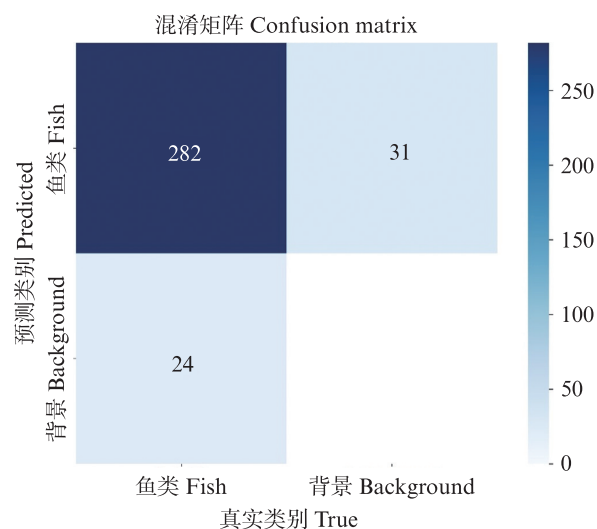


图8 YOLOv8X混淆矩阵  
Fig. 8 YOLOv8X confusion matrix

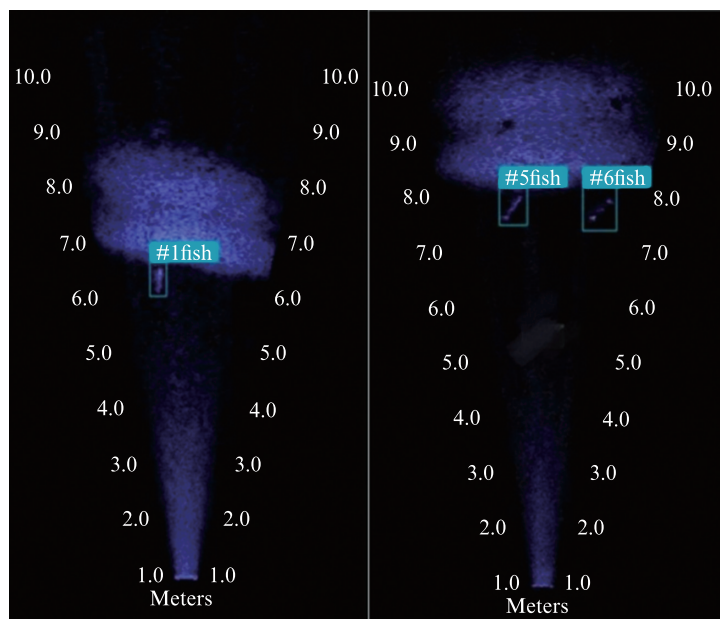


图9 鱼类目标识别  
Fig. 9 Fish target identification

表3 自动识别计数  
Tab. 3 Automatic identification count

测线 Survey	000	001	002	003	004	005	006	007	008	009	010
YOLOv8m	189	242	147	124	157	94	66	52	62	99	59
Echoview	202	251	152	130	149	83	68	66	76	90	63
人工重复实验 Manual repeated experiment	200	246	147	122	152	87	73	57	71	98	59
测线 Survey	011	012	013	014	015	016	017	018	019	020	总数
YOLOv8m	64	75	10	41	37	43	125	35	18	13	1 752
Echoview	77	73	12	41	42	44	151	36	25	22	1 853
人工重复实验 Manual repeated experiment	70	67	10	50	39	48	149	42	25	12	1 824

## 2.5 方法对比

使用不同版本的YOLO模型对此次采集的DIDSON数据进行处理,并以Echoview处理结果为基础做偏差对比(表4),其中YOLOv8X模型表现最优,偏差率仅为1.26%。Echoview在处理过程中由于人工长期介入的方式准确率得以保证,但是处理速度较慢,并且由于主观偏差和工

作疲劳,不同人处理甚至单人两次处理的结果可能会出现不同。当DIDSON在较深的水域采集数据时,声呐数据将收到更多的噪音和干扰,人工工作所需的时间和精力还会加倍增加。因此YOLOv8X和DIDSON相结合的方法具有较强的实用性,节省了人工处理所需的大量时间和精力,保证了识别的一致性和高精度。

表4 不同处理方法对比  
Tab. 4 Comparison of different methods for processing

方法 Methods	Echoview	YOLOv5m	YOLOv6m	YOLOv8m	YOLOv8X	YOLOv8L	YOLOv9c
计数 Counting/尾	1 853	1 574	1 674	1 752	1 814	1 805	1 734
处理时间(单测线) Processing time (single transect)/min	30	3	3	3	3	3	3
统计偏差 Statistical bias/%	0	14.41	8.97	4.73	1.36	1.85	5.71
均方根误差 Root mean square error	0	18.29	14.75	8.32	7.75	8.68	9.51

在表4中,不同版本的YOLO模型在DIDSON数据处理中的偏差率存在显著差异。其主要原因可以归结为YOLOv8版本模型在网络结构设计和训练策略上的改进。

相比之下,YOLOv5、YOLOv6等早期模型,在上述特征提取细节能力和泛化性能方面相对较弱,导致在复杂的声呐数据处理时产生较高的误差。

此外,在YOLOv8系列模型中,YOLOv8X作为最强版本,相较于YOLOv8m、YOLOv8L具备更高的特征提取能力和更深的网络结构,使其在高复杂度任务中表现更优。

## 2.6 鱼类资源量与时空分布分析

为全面评估水库中鱼类资源的数量与分布特征,本研究采用平面密度法对鱼类资源总量进行估算,并结合空间可视化方法分析鱼类在不同区域与水层中的时空分布。

首先根据每条航线探测到的鱼类的数量和

每条航线的仪器探测面积得出每条航线的鱼类密度。其次,取所有航线鱼类密度平均值作为整个水库的鱼类密度,由于每条航线测量长度基本相同,所以视作所有航线权重相同。最后根据整个水库的面积和鱼类密度计算结果和相关参数计算得出鱼类总数量。计算公式:

$$S_i = 2h_i \times \tan \frac{\alpha}{2} \times L_i \quad (7)$$

$$\rho_i = \frac{N_i}{S_i} \times 1.058 \quad (8)$$

$$\rho = \frac{1}{n} \sum_{i=1}^n \rho_i \quad (9)$$

式中: $\alpha$ 为仪器开角; $N_i$ 为第*i*条测线的鱼类数量,尾; $S_i$ 为探测面积, $\text{m}^2$ ; $\rho_i$ 为鱼类的密度,尾/ $\text{m}^2$ ; $L_i$ 为航线的长度,m; $h_i$ 为平均水深,m; $\rho$ 为水库内鱼的平均密度,尾/ $\text{m}^2$ ; $n$ 为总的航线数量。

经过上式得出本次调查航线总长度为22.78 km,共21条航线,探测面积为41 618  $\text{m}^2$ ,根据模型和Echoview计数的差距得出校正系数为1.058,得



出探测所得水库鱼类平均密度为  $0.044\ 5$  尾/ $\text{m}^2$ 。在时空分布分析方面,本研究选取金泽和陈行两座水库在 2023 年 9 月至 2024 年 6 月的 4 个典型时间点数据,分别开展水平与垂直分布分析。水平分布:利用 ArcMap 的空间点密度分析方法构建热力图(图 10 和图 11),结果显示:金泽水库的鱼类资源主要集中在西北角和中心岛周边水域,呈现季节性波动;而陈行水库的鱼类资源则主要分布于东南与东北区域,2023 年 12 月数量较少,2024 年 6 月探测密度最高。

垂直分布:依据鱼类出现在不同深度层的数

据,统计各水层内的鱼类数量分布比例。结果表明,金泽水库地形西高东低,西侧鱼类目标略高,整体集中于水体中下层(图 12);陈行水库底部较平坦,鱼类主要集中于 2~6 m 水深,且趋于靠近水底(图 13)。两水库在冬季(12 月)探测到的鱼类数量普遍低于其他季节,反映出一定的季节性分布规律。

综上,通过密度估算与空间分布可视化的结合分析,本方法能够高效、直观地刻画水域内鱼类资源的数量特征与分布格局,为鱼类资源评估与动态管理提供了科学支撑。

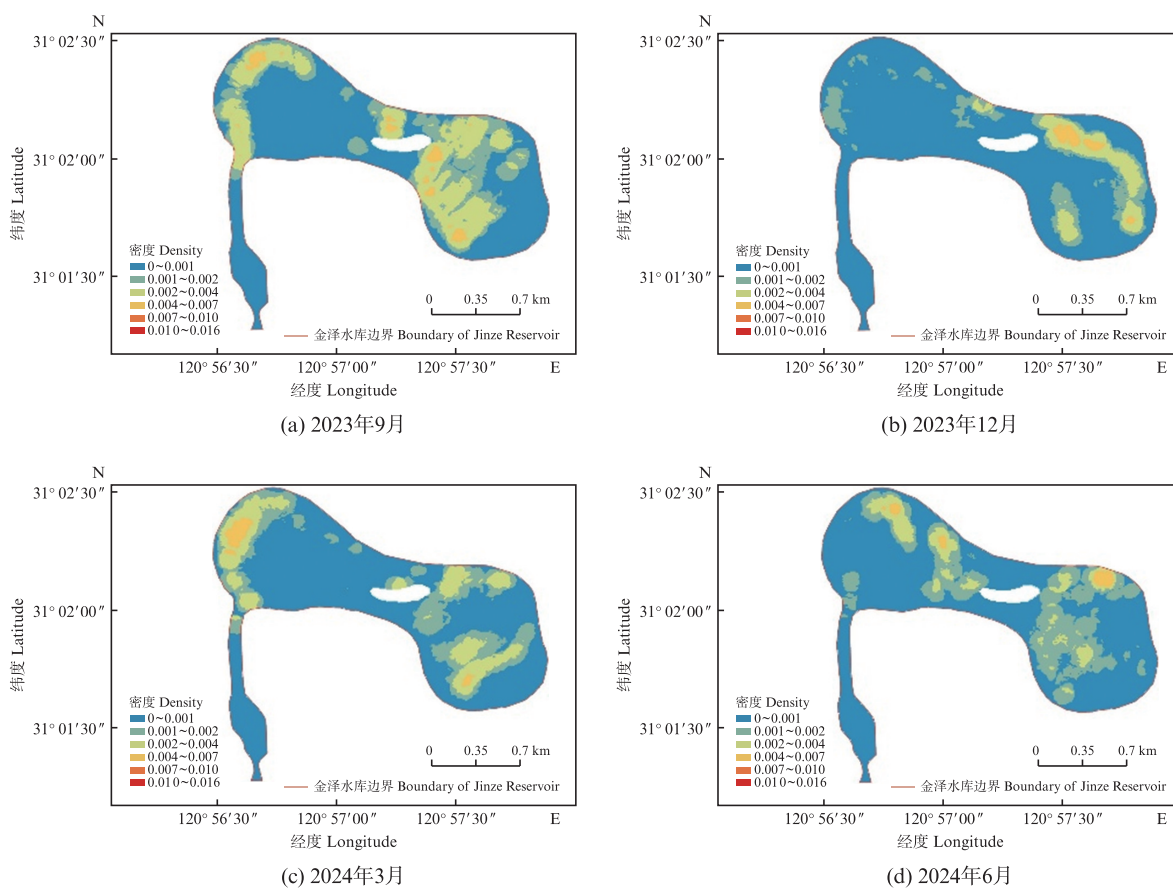


图 10 金泽水库鱼类资源四季水平分布图

Fig. 10 Seasonal horizontal distribution of fish resources in Jinze Reservoir

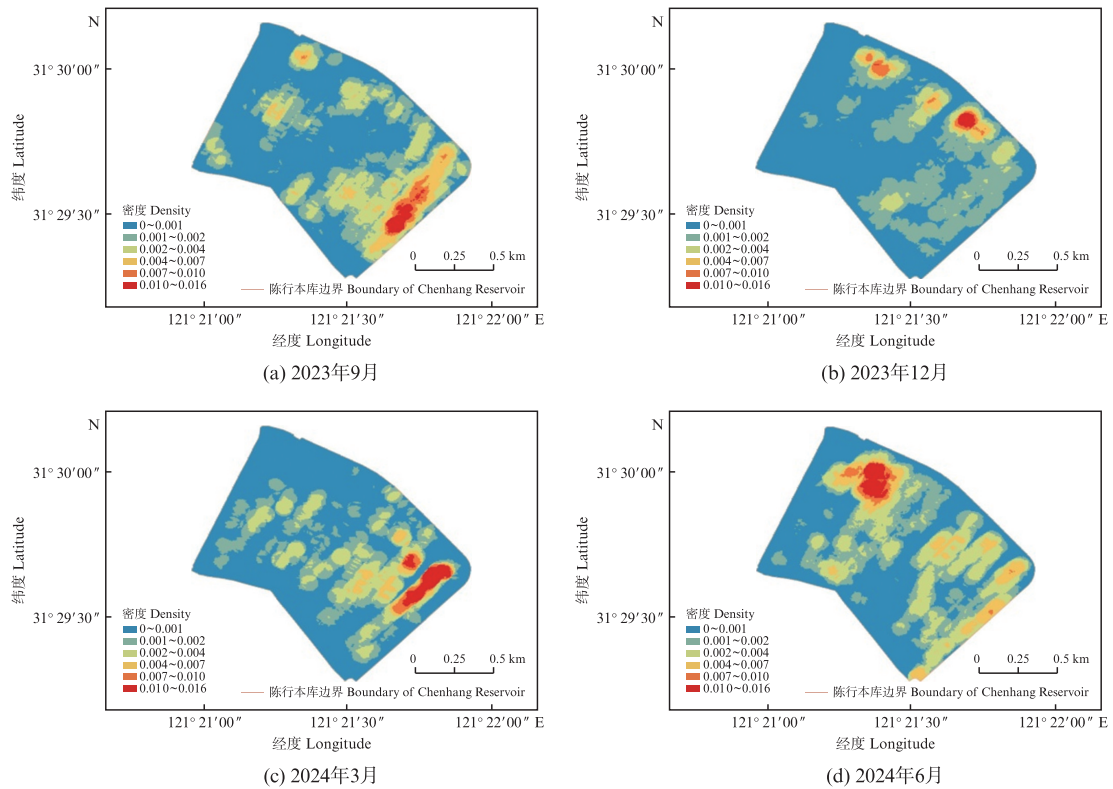


图 11 陈行水库鱼类资源四季水平分布图

Fig. 11 Seasonal horizontal distribution of fish resources in Chenhang Reservoir

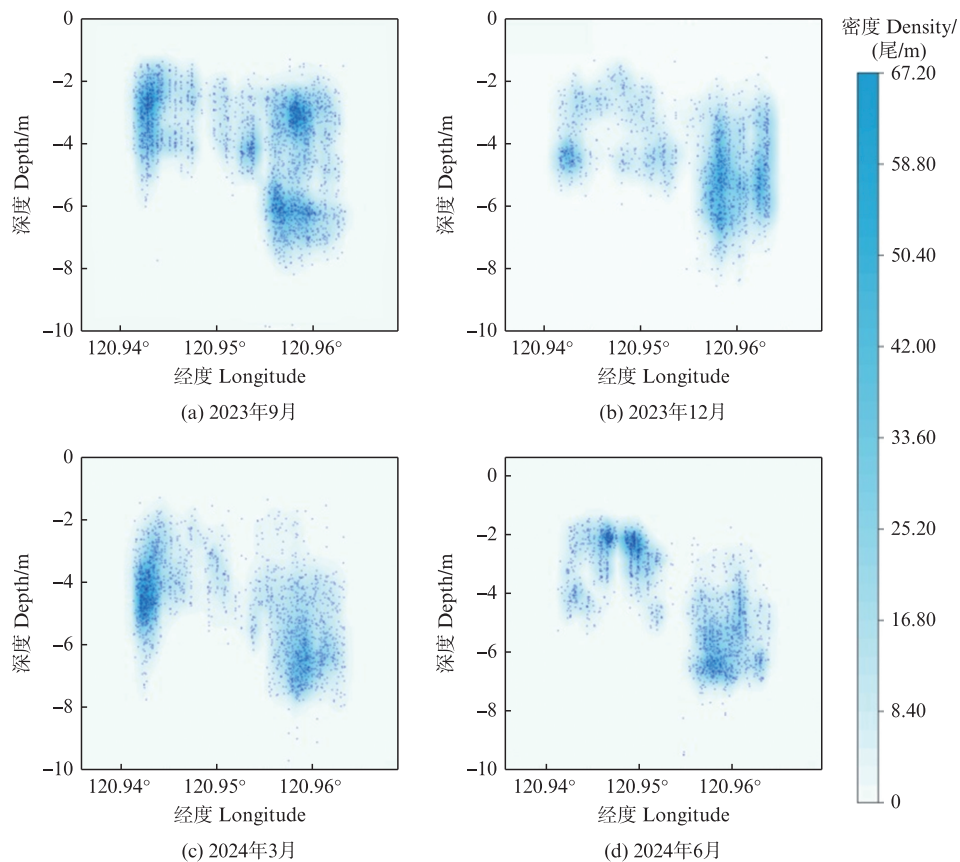


图 12 金泽水库鱼类垂直分布

Fig. 12 Vertical distribution of fish in Jinze Reservoir

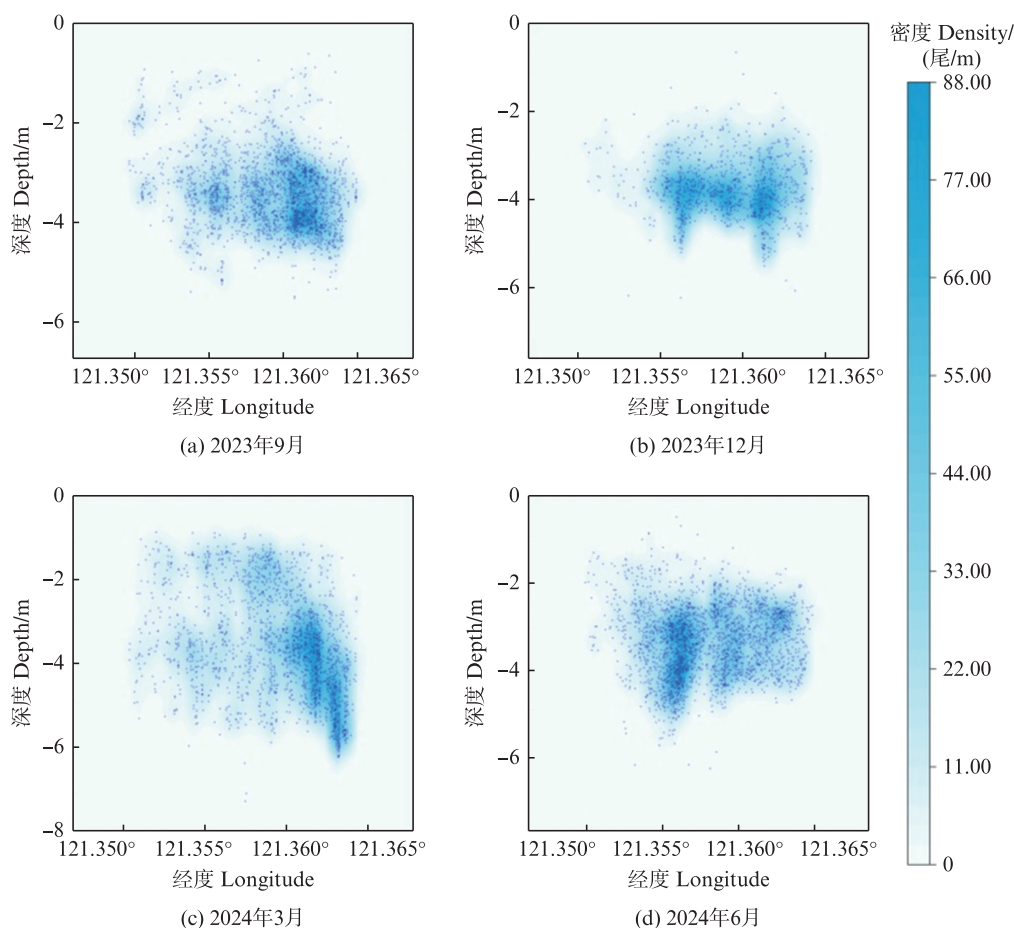


图 13 陈行水库鱼类垂直分布  
Fig. 13 Vertical distribution of fish in Chenhang Reservoir

### 3 讨论

#### 3.1 声呐图像

DIDSON作为一款高分辨率识别声呐,能够清晰的展现出鱼在水下的形态,与常规的鱼探仪相比,具有所见即所得的优势,但一些因素也会影响其探测精度。水体中存在的气泡、浮游生物、树枝杂物等反射回来的声呐回波会在声呐图像中产生噪声<sup>[21]</sup>,采集过程中复杂的环境噪声如船舶发动机噪声、波浪产生的噪声等也会影响声呐图像造成鱼类目标难以检测<sup>[22]</sup>。BALK等<sup>[23]</sup>在研究中指出,一旦船速较快,采集到的鱼体目标就会出现锯齿状轮廓,当速度降低时锯齿状特征就有明显改善。另外,鱼类目标的行为同样会产生影响,当鱼类集群时或靠近水底时,鱼类目标会产生重叠<sup>[24]</sup>或融合,造成数量的误差。因此,在采集过程中选择噪音小的船、适当的船速、确保安装牢固等可以减少噪音获取高质量的声

呐图像。

#### 3.2 识别计数

YOLOv8X加ByteTrack在进行鱼类目标识别和计数在实际应用时具有显著优势。YOLOv8X以其高精度、快速响应的特点,适用于实时目标识别和计数领域<sup>[25]</sup>。为了达到最佳的识别效果,需要大量的数据进行训练和验证,具备足量、多样化的特点,涵盖可能出现的各种情况,包括鱼的不同种类、姿态和密度,以及各种水下环境和光照条件以确保模型在各种应用场景下的准确度<sup>[26]</sup>。ByteTrack主要负责在视频帧中保持对目标的连续跟踪和计数,其核心在于利用卡尔曼滤波器结合线性运动模型预测目标位置,并通过匈牙利算法解决检测框与轨迹之间的数据关联问题,同时有效利用低置信度目标检测以提升关联准确性<sup>[27]</sup>。然而,在目标高度集群、快速移动或发生遮挡的情况下,ByteTrack仍可能出现跟踪错误或目标丢失的问题,尽管此类问题已被显著减

少,但仍不能完全避免<sup>[28]</sup>。总体而言,通过合理利用ByteTrack的优势,进一步提升鱼类目标识别和跟踪的整体性能。

### 3.3 周期性监测

水库鱼类资源的动态变化受季节、气候和水体环境等多重因素影响,传统方法难以满足高效、精准、非侵入性监测的需求。DIDSON以其高分辨率成像和非破坏性特点,可在浑浊水体中清晰捕捉鱼类分布与行为信息,同时结合深度学习算法,实现鱼类目标的自动识别与量化分析。通过周期性检测,跟踪水库鱼类数量变化,支持制定捕捞和保护策略,为生态系统健康评价和渔业资源可持续管理提供科学依据,为推动渔业资源管理与生态保护的智能化提供助力。

## 4 结论

水域鱼类资源的周期性监测对于渔业资源管理与可持续发展具有重要意义。本研究提出了一种基于DIDSON高清声呐数据,结合YOLOv8X目标检测算法与ByteTrack多目标跟踪算法的鱼类资源探测方法。相比传统的Echoview方法,该方法在识别精度基本一致(偏差率为1.36%)的基础上,将单条测线数据的处理时间从约30 min缩短至3 min,显著提升了处理效率与自动化水平,适用于水库、河流、湖泊等多类水域的高频次监测任务。

该方法的准确性在实际应用中受到多个因素影响,包括声呐图像质量、背景噪声干扰、鱼类的体型、密度及其接近水底的行为变化等。因此,未来研究需进一步优化模型结构与参数设计,丰富训练数据集,提升模型对不同水域环境中鱼类目标的识别与计数精度。

此外,为进一步拓展本方法在更复杂场景下的适用性,未来可从以下几个方向开展研究。

(1)复杂动态水域的适用性提升:目前方法在静态水库环境(如上海金泽、陈行)中表现良好,但在流速较快、湍流显著、泥沙较多的动态水域中,声呐图像的噪声水平可能显著增加,影响检测与跟踪效果。后续应在河流、河口等复杂环境中进行实地测试,并通过引入自适应滤波、深度学习去噪等图像预处理技术,提升图像质量与模型鲁棒性。

(2)鱼类遮挡与密集群体检测优化:当鱼类

出现高密度聚集或相互遮挡时,目标易发生重叠,导致跟踪失败或个体丢失。为提升在此类复杂场景下的检测性能,未来可探索引入时空注意力机制增强模型对遮挡目标的辨识能力;进一步优化ByteTrack算法以减少ID切换和跟踪中断现象;结合轨迹预测模型与滤波算法提升对鱼类运动连续性的判断能力;同时构建面向群聚行为的训练策略与标注体系,增强密集区域内个体区分能力。

(3)智能实时监测平台构建:为推动该方法在实际生产中的广泛应用,未来可进一步开发船载或水下嵌入式平台,实现YOLOv8X与ByteTrack的边缘部署和实时处理,从而显著降低数据延迟、提升鱼类监测的效率与响应速度。

综上所述,本研究方法兼具较高的识别精度与处理效率,可有效避免传统人工识别中存在的主观偏差,具备良好的实用性和扩展性,为未来水域鱼类资源的动态监测与科学管理提供了一种高效、智能的技术路径。

作者声明本文无利益冲突。

### 参考文献:

- [1] GARCIA G F, CORPETTI T, NEVOUX M, et al. AcousticIA, a deep neural network for multi-species fish detection using multiple models of acoustic cameras[J]. Aquatic Ecology, 2023, 57(4): 881-893.
- [2] SARR J M A, BROCHIER T, BREHMER P, et al. Complex data labeling with deep learning methods: lessons from fisheries acoustics[J]. ISA Transactions, 2021, 109: 113-125.
- [3] TALLON B, ROUX P, MATTE G, et al. Acoustic density estimation of dense fish shoals[J]. The Journal of the Acoustical Society of America, 2020, 148(3): EL234-EL239.
- [4] MOURSUND R A, CARLSON T J, PETERS R D. A fisheries application of a dual-frequency identification sonar acoustic camera[J]. ICES Journal of Marine Science, 2003, 60(3): 678-683.
- [5] BELCHER E, HANOT W, BURCH J. Dual-frequency identification sonar (DIDSON)[C]//Proceedings of the 2002 International Symposium on Underwater Technology. Tokyo: IEEE, 2002: 187-192.
- [6] MAXWELL S L, GOVE N E. The Feasibility of estimating migrating salmon passage rates in turbid rivers using a dual frequency identification sonar (DIDSON)[R]. Alaska: Alaska Department of Fish and Game



- Division of Commercial Fisheries Central Region, 2004.
- [7] HAN J, HONDA N, ASADA A, et al. Automated acoustic method for counting and sizing farmed fish during transfer using DIDSON [J]. *Fisheries Science*, 2009, 75(6): 1359-1367.
  - [8] JING D X, HAN J, WANG X D, et al. A method to estimate the abundance of fish based on dual-frequency identification sonar (DIDSON) imaging [J]. *Fisheries Science*, 2017, 83(5): 685-697.
  - [9] ZHANG H, WEI Q W, KANG M. Measurement of swimming pattern and body length of cultured Chinese sturgeon by use of imaging sonar [J]. *Aquaculture*, 2014, 434: 184-187.
  - [10] LADROIT Y, ESCOBAR-FLORES P C, SCHIMEL A C G, et al. ESP3: an open-source software for the quantitative processing of hydro-acoustic data [J]. *SoftwareX*, 2020, 12: 100581.
  - [11] HARRISON L M K, COX M J, SKARET G, et al. The R package *EchoviewR* for automated processing of active acoustic data using Echoview [J]. *Frontiers in Marine Science*, 2015, 2: 15.
  - [12] 李江川, 韩彦岭, 董传胜, 等. YOLO-U: 基于结构重参数化和双重注意力机制的水下目标检测算法 [J]. *上海海洋大学学报*, 2025, 34(3): 696-706.
  - LI J C, HAN Y L, DONG C S, et al. YOLO-U: An underwater object detection algorithm based on structural reparameterization and dual attention mechanism [J]. *Journal of Shanghai Ocean University*, 2025, 34(3): 696-706.
  - [13] VARGHESE R, SAMBATH M. YOLOv8: a novel object detection algorithm with enhanced performance and robustness [C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Chennai: IEEE, 2024: 1-6.
  - [14] 曹宇, 李佳阳, 王芳. 基于改进YOLOv8n的水下鱼类目标识别轻量化模型 [J]. *上海海洋大学学报*, 2025, 34(1): 188-200.
  - CAO Y, LI J Y, WANG F. An improved YOLOv8n based on lightweight model for underwater fish target detection [J]. *Journal of Shanghai Ocean University*, 2025, 34(1): 188-200.
  - [15] GUPTA C, GILL N S, GULIA P, et al. A novel finetuned YOLOv8 model for real-time underwater trash detection [J]. *Journal of Real-Time Image Processing*, 2024, 21(2): 1-16.
  - [16] ZHANG Y, LIU C, WANG X, et al. BSSFISH-YOLOv8: A novel fish detection model for natural waters based on improved YOLOv8n [J]. *Animals*, 2024, 14(14): 2022.
  - [17] CUI M, LIU X B, LIU H H, et al. Fish tracking, counting, and behaviour analysis in digital aquaculture: a comprehensive survey [J]. *Reviews in Aquaculture*, 2025, 17(1): e13001.
  - [18] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of yolo architectures in computer vision: from yolov1 to yolov8 and yolo-nas [J]. *Machine Learning and Knowledge Extraction*, 2023, 5(4): 1680-1716.
  - [19] HUSSAIN M. YOLOv5, YOLOv8 and YOLOv10: the go-to detectors for real-time vision [J]. *arXiv preprint arXiv:2407.02988*, 2024.
  - [20] BELCHER E O, LYNN D C, DINH H Q, et al. Beamforming and imaging with acoustic lenses in small, high-frequency sonars [C]//Oceans '99. MTS/IEEE. Riding the Crest into the 21st Century. Conference and Exhibition. Conference Proceedings (IEEE Cat. No. 99CH37008). Seattle: IEEE, 1999: 1495-1499.
  - [21] 沈蔚, 卢泉水, 彭战飞, 等. 基于主动识别声呐的养殖塘南美白对虾探测与初步分析 [J]. *渔业现代化*, 2024, 51(1): 64-70.
  - SHENW, LU Q S, PENG Z F, et al. Detection and preliminary analysis of penaeus vannamei in aquaculture ponds based on active identification sonar [J]. *Fisheries Information*, 2024, 51(1): 64-70.
  - [22] LUO Y Z, LU H Z, ZHOU X X, et al. Lightweight model for fish recognition based on YOLOV5-mobilenetV3 and sonar images [J]. *Guangdong Agricultural Sciences*, 2023, 50(7): 37-46.
  - [23] BALK H, LINDEM T. Improved fish detection in data from split-beam sonar [J]. *Aquatic Living Resources*, 2000, 13(5): 297-303.
  - [24] HELMINEN J, LINNANSAARI T. Object and behavior differentiation for improved automated counts of migrating river fish using imaging sonar data [J]. *Fisheries Research*, 2021, 237: 105883.
  - [25] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. *arXiv preprint arXiv: 1804.02767*, 2018.
  - [26] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//FLEET D, PAJDLA T, SCHIELE B, et al. *Computer Vision - ECCV 2014*. Cham: Springer International Publishing, 2014: 740-755.
  - [27] ZHANG Y F, SUN P Z, JIANG Y, et al. bytetrack: Multi-object tracking by associating every detection box [C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 1-21.
  - [28] QIAN Z M, CHEN X, JIANG H. Fish tracking based on YOLO and ByteTrack [C]//2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). Taizhou: IEEE, 2023: 1-5.

## Monitoring and analysis of fish resources in waters based on YOLOv8X and DIDSON

SHEN Wei<sup>1,2</sup>, DONG Shihong<sup>1,2</sup>, LIU Mengqi<sup>1,2</sup>, YIN Zhaowei<sup>1,2</sup>, QIAN Enze<sup>1,2</sup>, GONG Xiaoling<sup>3</sup>

(1. College of Oceanography and Ecological Science, Shanghai Ocean University, Shanghai 201306, China; 2. Shanghai Engineering Research Center of Estuarine and Oceanographic Mapping, Shanghai 201306, China; 3. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China)

**Abstract:** In order to improve the automation and real-time analysis ability of fish resources monitoring in waters, this paper combines YOLOv8X (You only look once version 8 - Extra Large) target detection algorithm, ByteTrack (ByteTrack: a strong baseline for multi-object tracking) algorithm and DIDSON sonar (Dual-frequency identification sonar) data to develop a fast and accurate fish target recognition and counting method. The experimental results show that the combined method of YOLOv8X and ByteTrack is close to the traditional Echoview software recognition accuracy (the deviation rate is only 1.36%), but the processing time is significantly reduced (the single line is reduced from approximate 30 minutes to approximate 3 minutes), showing strong real-time processing ability and generalization performance. Meanwhile, the stability of the method is verified by repeated experiments, and its reliability in different scenarios is confirmed. The research methods and results of this paper provide reliable technical support for the automatic monitoring of fish resources in waters, and can be widely used in the monitoring and management of fishery resources in a wide range of high frequency.

**Key words:** fish stock monitoring; deep learning model; YOLOv8X model; DIDSON Sonar; ByteTrack algorithm