

基于YOLOv11的水下中华鲟视频智能检测方法

贾永红*, 周温晖*, 范睿#, 陈曦

武汉大学遥感信息工程学院, 湖北 武汉

收稿日期: 2026年5月29日; 录用日期: 2026年9月8日; 发布日期: 2026年9月20日

摘要

为克服水下中华鲟检测存在的光照不均、水下水体浑暗、目标遮挡影响问题, 本文对YOLOv11检测算法, 引入全局注意力机制增强模型对关键特征区域的表达能力, 在特征融合阶段采用混洗卷积结构以提升多尺度特征融合效果, 从而提高中华鲟检测精度。改进的YOLOv11算法用于水下中华鲟视频检测的试验结果表明: 改进的YOLOv11模型精度指标mAP@0.5相比YOLOv11模型提高了3.2%, 达到了91.9%。因此, 改进YOLOv11对实际应用场景中华鲟视频检测具有更高的精度与更强的鲁棒性。

关键词

YOLOv11, 目标检测, 全局注意力机制, 混洗卷积

Intelligent Video Detection Method of Underwater Chinese Sturgeon Based on YOLOv11

Yonghong Jia*, Wenhui Zhou*, Rui Fan#, Xi Chen

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan Hubei

Received: May 29, 2026; accepted: September 8, 2026; published: September 20, 2026

Abstract

In order to overcome the problems of uneven illumination, dark underwater water and target occlusion in underwater Chinese sturgeon video image detection, the global attention mechanism was introduced to YOLOv11 detection algorithm to enhance the expression ability of the model for key feature regions, and the group shuffle convolution structure was used in the feature fusion stage to improve the multi-scale feature fusion effect, so as to improve the detection accuracy of Chinese sturgeon. The experimental results of the improved YOLOv11 algorithm for underwater Chinese sturgeon

*共同第一作者。

#通讯作者。

文章引用: 贾永红, 周温晖, 范睿, 陈曦. 基于YOLOv11的水下中华鲟视频智能检测方法[J]. 人工智能与机器人研究, 2026, 15(5): 1259-1266. DOI: 10.12677/airr.2026.155115

video detection show that the accuracy index of the improved YOLOv11 model, mAP@0.5, compared with YOLOv11 model, increased by 3.2% and reached 91.9%. Therefore, the improved YOLO11 has higher accuracy and stronger robustness for the actual application scene of Chinese sturgeon video detection.

Keywords

YOLOv11, Object Detection, Global Attention Mechanism, Group Shuffle Convolution

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

中华鲟是中国长江中十分古老且具有珍稀价值的鱼类,有“长江鱼王”的称号。其主要栖息地位于长江中下游干流。由于过度捕捞、水质污染等因素导致中华鲟天然种群规模不断缩小,自然繁殖的中华鲟幼鱼更是少之又少,野生的中华鲟种群延续状况岌岌可危。因此,采用智能检测水下环境中的中华鲟技术成为保护和养殖中华鲟的关键支撑技术之一。

随着深度学习技术的发展,YOLO 系列模型不断演进,从早期版本逐步发展至最新的 YOLO11。在这一演进过程中,模型在网络结构设计、特征表达能力以及检测机制等方面持续优化,为复杂环境下目标检测任务提供了更强的技术支撑。

沈军宇等基于 YOLO 算法,提出了端到端鱼群检测算法,不仅能在光线充足的环境下对鱼群进行更精确的检测,而且对远景目标实现更精确的检测定位,在光线非常暗的情况下,对单条鱼与鱼群的检测精度也非常高。但由于图像质量差,仍存在查全率较低的问题[1]。为了实现不同海底背景和光照条件下对鱼类目标的检测和计数,Labao 等为了在不同的海底背景和光照条件下实现对鱼类目标的检测和计数,提出了由一组卷积神经网络通过长、短期记忆网络串联而成的具有良好鲁棒性的鱼类检测系统[2]。朱世伟等提出了一种基于类加权 YOLO 网络的水下目标检测方法,引入了目标框自适应维度聚类算法来提升召回率,并用类加权损失来权衡不同样本的难易程度,在平均准确率上取得了比较理想的提升[3]。郑宗生等基于 YOLOv4 网络模型,在 CIoU 损失函数的基础上引入新的损失项,改进 YOLOv4 模型的平均精度均值比原模型有较大提升,但存在对模糊图像的识别结果较差的问题[4]。夏菲菲等在 YOLOv8 检测算法的基础上引入 EMA 注意力机制,增强模型对中华鲟关键特征的捕捉能力,改进后的 YOLOv8 模型在 mAP@0.5 上相比原始模型提升了 3.7%,验证了其在水下中华鲟检测中的有效性,但对复杂水下环境中的准确率有待进一步提高[5]。Chen 等提出的动态 YOLO 方法针对水下小目标检测性能下降问题,通过优化特征提取与多尺度融合机制,有效提升了检测精度[6]。Lin 等基于 YOLOv8 构建了轻量化水下检测模型,通过改进特征融合结构与损失函数,在保证实时性的同时提升了检测性能[7]。此外,Cheng 等引入注意力机制增强网络对关键区域的表达能力,有效改善了复杂水下环境中目标检测精度不足的问题[8]。然而,由于光照不均、水下浑暗及目标形态变化等因素影响,YOLOv8 模型在复杂水下环境中的检测鲁棒性有待改善[9][10]。

2. 研究方法

2.1. YOLOv11 网络

在 YOLOv8 的基础上,YOLO11 创新架构和优化参数,实现了卓越检测性能。YOLO11 由骨干网络、

颈部和检测头三个主要部分组成。其具体网络结构如图 1 所示。

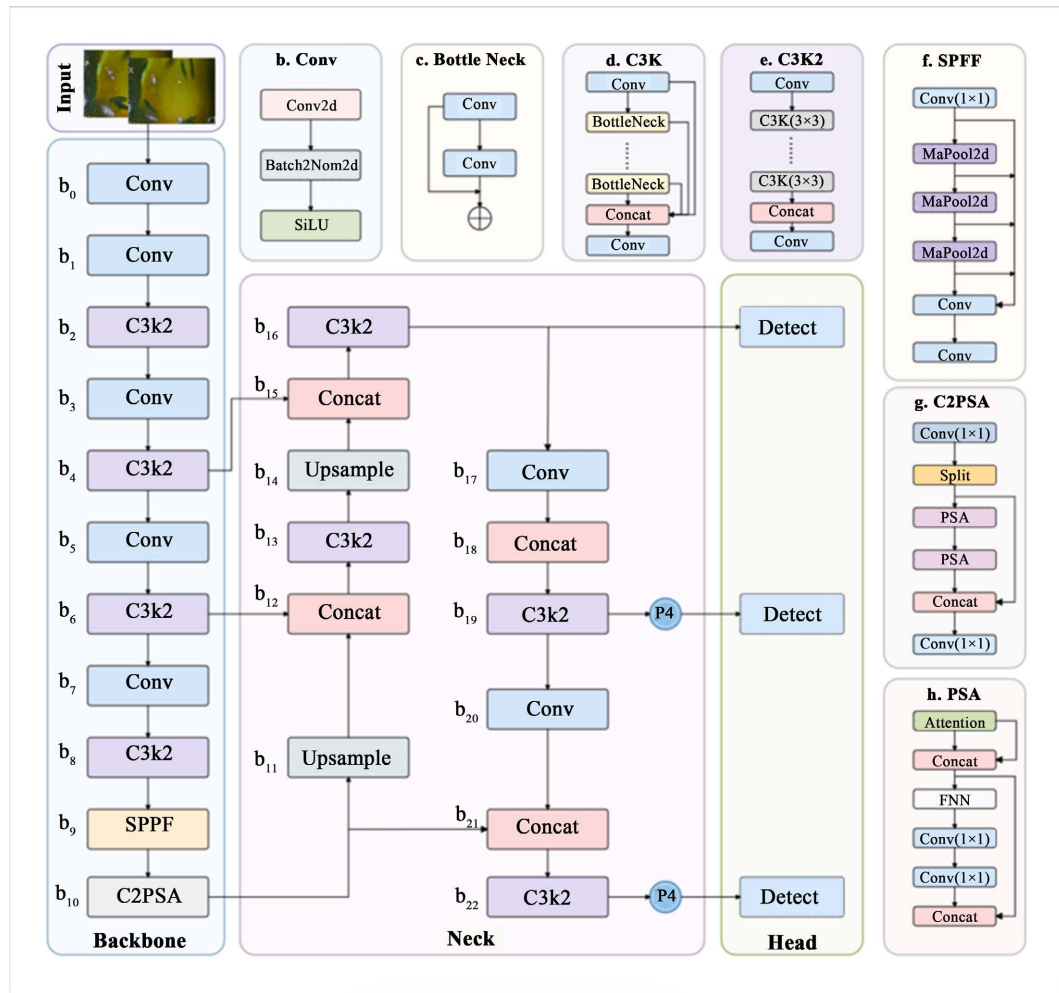


Figure 1. YOLO11 network structure

图 1. YOLO11 网络结构

1) 骨干网络

YOLO11 的骨干网络采用了全新的 C3k2 模块(Cross-Stage Partial Bottleneck with 2 Convolutions)。该模块在 YOLOv8 的基础上进行了重构,通过调整残差块的连接方式和内核大小,在保持梯度流丰富性的同时,降低了浮点运算量(FLOPs)。此外,Backbone 的末端采用了改进后的 SPPF 模块,通过多尺度池化层并行处理,增强了模型对多尺度特征的感受野。

2) 颈部组件

颈部采用增强型的 PAN-FPN 结构,通过双向路径融合机制,实现浅层空间信息与深层语义信息的深度交互。YOLO11 在此阶段对下采样和特征融合(Concat)操作进行精细化调整,有效减缓了特征传递过程中的信息丢失。

3) 检测头

YOLO11 采用了高效的解耦头(Decoupled Head)架构。该设计将分类任务与回归任务完全分离,通过独立的分支并行计算。同时,YOLO11 坚持 Anchor-free (无锚点)检测范式,直接预测目标中心点及其边

界框偏置, 极大地简化了后处理过程并提升了模型对长宽比剧烈变化目标的鲁棒性。

2.2. YOLOv11 算法的改进

为提升 YOLO11 模型在复杂水下环境中的检测性能, 从特征表达能力与特征融合能力两个方面出发, 对模型结构进行针对性改进。具体包括全局注意力机制(Global Attention Mechanism, GAM)和 GSConv (Group Shuffle Convolution)优化特征融合结构。

1) 基于 GAM 的全局注意力机制

GAM 核心思想是在不显著增加计算复杂度的前提下, 同时强化通道注意力与空间注意力, 即在通道维度与空间维度上有效地捕捉不同通道之间的相关性, 进而更好地区分不同的目标。其工作原理如图 2 所示。

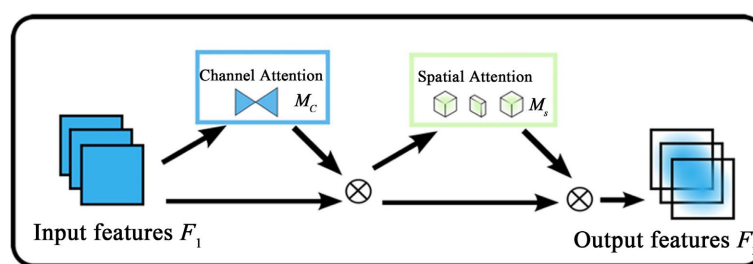


Figure 2. Working principle diagram of GAM attention module

图 2. GAM 注意力模块工作原理图

具体而言, GAM 主要包括以下三个部分:

① 通道注意力子模块(Channel Attention Submodule)

通过全局平均池化(Global Average Pooling)对特征图进行压缩, 并结合全连接层生成通道权重系数, 对不同通道的重要性进行自适应调整, 从而突出具有判别能力的特征通道。

② 空间注意力子模块(Spatial Attention Submodule)

在通道压缩后的特征基础上, 通过卷积操作提取空间信息, 并生成空间权重分布, 对图像中关键区域进行增强, 同时抑制背景噪声干扰。

③ 跨维度融合模块

对通道注意力与空间注意力结果进行融合, 实现特征在通道与空间维度上的协同优化, 可避免单一维度处理带来的信息损失。

2) 基于 GSConv 的特征融合优化

YOLO11 采用深度可分离卷积(Depthwise Separable Convolution, DSC)在一定程度上降低了计算量, 但其通道间信息交互能力有限, 可能导致特征表达能力不足, 尤其在复杂背景下易产生信息损失。

特征融合阶段引入 GSConv 结构对原有卷积操作进行替换, 则可在保持模型轻量化的同时增强特征表达能力。其结构如图 3 所示。

GSConv 通过将分组卷积(Group Convolution)与通道重排(Channel Shuffle)相结合, 在减少参数数量的同时实现不同通道之间的信息交互。其基本过程为: 首先对输入特征进行分组卷积操作, 以降低计算复杂度; 随后通过通道重排操作打破各组之间的独立性, 使不同通道特征充分融合, 从而提升特征表达的多样性与完整性。

GSConv 结构在提升模型效率与特征表达能力之间取得了良好平衡, 有助于 YOLO11 提升多尺度特

征的融合效果, 使 YOLO11 模型对尺度变化显著及密集分布的目标场景检测, 仍能够保持较高的检测性能。

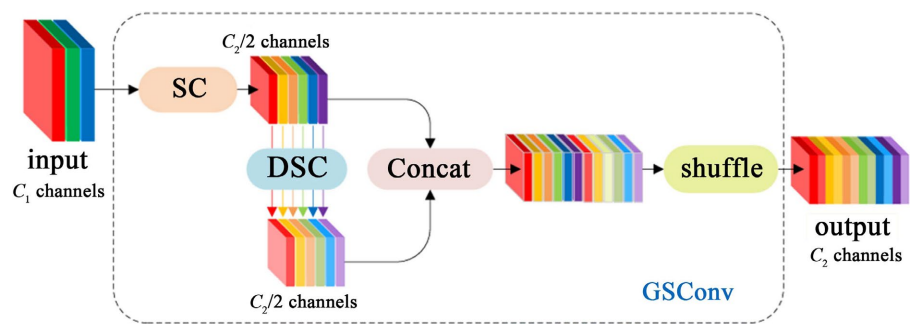


Figure 3. GSConv structure
图 3. GSConv 结构

水下中华鲟视频检测流程如图 4 所示。流程主要包括数据准备、模型训练、模型测试以及结果分析四个阶段。

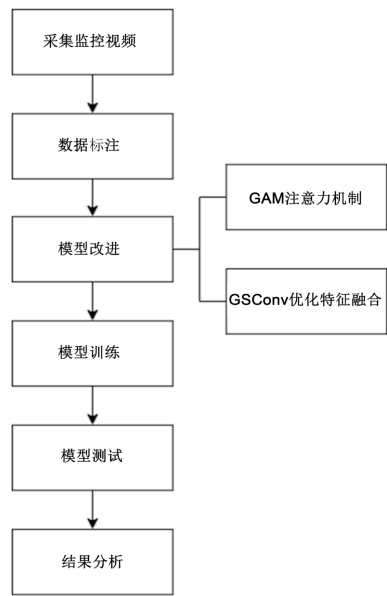


Figure 4. Testing process
图 4. 检测试验流程

① 数据准备阶段

对采集的中华鲟养殖池视频数据进行抽帧处理, 筛选出包含目标的图像样本, 并利用 LabelImg 工具进行人工标注, 生成符合 YOLO 格式的数据集。随后按照训练集、验证集与测试集进行划分, 用于模型训练与性能评估。

② 模型训练阶段

分别构建 YOLO11 以及改进 YOLO11 模型, 在统一训练参数条件下进行训练。训练过程中通过验证集对模型性能进行实时评估, 并保存最优权重参数。

③ 模型测试阶段

将训练完成的模型应用于测试集及典型水上、水下中华鲟视频数据, 对中华鲟进行检测, 并记录模型在不同场景下的检测结果。

④ 结果分析阶段

对检测结果进行定性分析, 依据查准率、查全率及 mAP 等评价指标, 进行定量分析, 从而全面评估模型的实际应用效果。

3. 试验与分析

中华鲟检测试验软硬件环境配置如下: 处理器为 AMD Ryzen 9 5900HX with Radeon Graphics 八核, 内存 32 GB, GPU 为 NVIDIA GeForce RTX 3080 Laptop。操作系统为 64 位 Windows 11, Python 版本 3.9, 深度学习框架为 Pytorch 1.13.1 + cu116、CUDA 11.6 和 CUDNN 8.3.0。

试验数据集是采用 LabelImg 工具进行人工标注水上养殖池中华鲟与水下养殖池中华鲟视频样本各 5000 张, 共 10,000 张。标注数据集按照 8:1:1 的比例划分为训练集、验证集与测试集, 用于模型训练与评估。

在模型训练过程中, 设置训练轮次为 150, 批次大小为 16, 初始学习率为 0.01, 采用 SGD 优化器完成参数更新。所有对比试验均在相同训练参数与数据集条件下进行, 以保证试验结果的可比性。

从平衡检测精度与效率角度出发, 选用了 YOLO11n.pt 作为预训练权重, 采用标注的中华鲟数据集对模型进行了训练。

训练完成后, 在测试集及实际视频数据上进行检测试验, 通过 Precision、Recall、mAP@0.5、Params、FPS 和 FLOPs 指标对模型进行定量评价。

为分析改进模块对模型性能的贡献, 设计了消融试验, 对不同模块组合进行对比分析。试验结果定量评价如表 1 所示。

Table 1. Evaluation of ablation test results

表 1. 消融试验结果评价

Model	Precision	Recall	mAP@0.5	Params (M)	FPS	GFLOPs
原始 YOLO11	91.7%	74.8%	88.7%	2.6	2.64	6.5
YOLO11 + GAM	95.5%	75.6%	91.2%	3.4	2.36	10.8
YOLO11 + GSConv	92.7%	79.8%	89.5%	2.6	2.74	6.5
YOLO11 + GAM + GSConv	96.4%	83.7%	91.9%	3.4	2.59	10.8

从表 1 消融试验结果评价可以看出, 仅采用 GAM 注意力机制后, 模型的 Precision 由 91.7%提升至 95.5%, mAP@0.5 由 88.7%提升至 91.2%, 但计算量增加 66.2%。这表明模型误检率显著降低但计算量增加。其主要原因在于 GAM 通过在通道与空间维度联合注意力建模机制, 使网络能够更加聚焦于中华鲟目标区域, 同时抑制水体、背景噪声带来的干扰, 从而提升目标与背景之间的区分能力。仅采用 GSConv 结构, 模型的 Recall 由 74.8%提升至 79.8%, 表现出较为明显的提升, 计算量没有明显变化。这说明模型对目标的检出能力增强, 漏检情况有所减少。主要源于 GSConv 通过分组卷积与通道重排机制增强了特征融合过程中的跨通道信息交互能力, 使多尺度特征表达更加充分, 从而提高了对弱特征目标的检测能力。

当同时引入 GAM 与 GSConv 时, 模型在精度指标上均达到最优水平, 其中 Precision 达到 96.4%, Recall 递增至最大值 83.7%, mAP@0.5 提升至 91.9%, 但计算量增加 66.2%。这表明两种改进策略具有良好的性能: GAM 侧重于增强特征表达能力, 而 GSConv 侧重于优化特征融合过程, 二者协同作用显著提升了模型整体检测性能。

综上所述, 消融试验验证了各改进模块的有效性, 并表明所提出的改进方法能够显著提升中华鲟目标检测性能。

为了检验改进方法的实际应用效果, 分别对养殖池水上中华鲟视频和养殖池水下中华鲟视频进行了检测试验。

对比如图 5 养殖池水上中华鲟视频的检测结果, 可以看出改进YOLO11 漏检情况要少于 YOLO11, 印证了模型的可靠性。

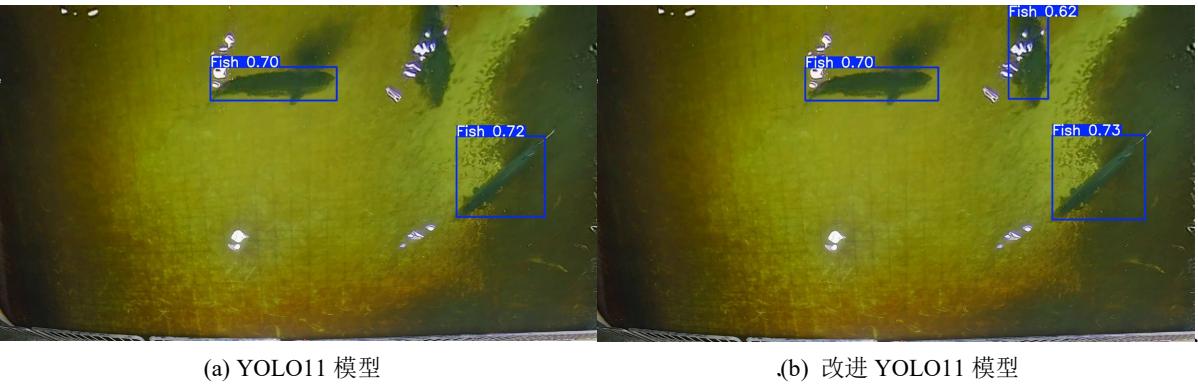


Figure 5. Comparison of detection results between YOLO11 and improved YOLO11 models (overwater)
图 5. YOLO11 与改进 YOLO11 模型检测结果对比(水上)

对比如图 6 养殖池水下中华鲟视频的检测结果, 可见改进YOLO11 模型在能见度较差环境下对中华鲟的检测效果仍好于 YOLO11。

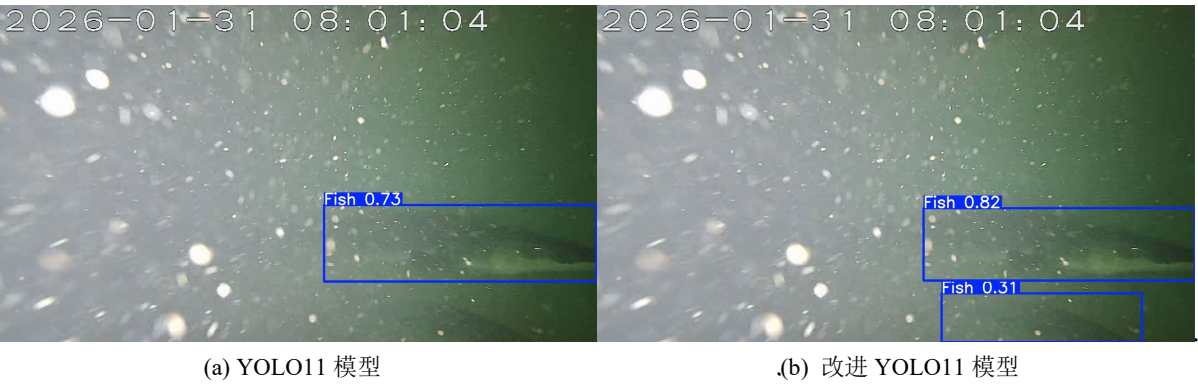


Figure 6. Comparison of detection results between YOLO11 and improved YOLO11 models (underwater)
图 6. YOLO11 与改进 YOLO11 模型检测结果对比(水下)

综上所述, 本文提出的改进 YOLO11 模型在中华鲟养殖池视频场景中具有更高的检测精度与更强的鲁棒性。

4. 结论

改进 YOLO11 模型能有效满足水上与水下中华鲟养殖池环境下的视频检测需求, 验证了方法的有效性与实用价值。从定量结果来看, 改进模型在 Precision、mAP 等指标上均实现提升, 表明其在减少误检与漏检方面均具有较好效果, 但计算量增加。

从方法机制分析来看, 性能提升主要来源于两个方面: 一是 GAM 全局注意力机制增强了模型对关键特征区域的表达能力, 使模型能够在低对比度条件下有效区分目标与背景; 二是 GSConv 结构优化了特征融合过程, 提高了多尺度信息交互能力, 从而增强模型对不同尺度目标的适应能力。

从应用场景角度来看, 改进模型在低能见度、水体浑浊及目标密集分布的养殖池环境中表现出较好的鲁棒性。未来将进一步补充样本集数据, 降低计算量, 以提高检测模型的适用性和效率, 为后续多目标跟踪与三维表型测量提供更可靠的检测结果。

基金项目

该成果获得三峡集团“中华鲟产卵场建设方案设计 & 监控设备研发(合同编号: 202203404)”项目、武汉大学大学生创新训练计划项目资助。

参考文献

- [1] 沈军宇, 李林燕, 夏振平, 等. 一种基于 YOLO 算法的鱼群检测方法[J]. 中国体视学与图像分析, 2018, 23(2): 174-180.
- [2] Labao, A.B. and Naval, P.C. (2019) Cascaded Deep Network Systems with Linked Ensemble Components for Underwater Fish Detection in the Wild. *Ecological Informatics*, **52**, 103-121. <https://doi.org/10.1016/j.ecoinf.2019.05.004>
- [3] 朱世伟, 杭仁龙, 刘青山. 基于类加权 YOLO 网络的水下目标检测[J]. 南京师大学报 (自然科学版), 2020, 43(1): 129-135.
- [4] 郑宗生, 李云飞, 卢鹏, 等. 改进 YOLOv4 模型在鱼类目标检测上的应用研究[J]. 渔业现代化, 2022, 49(1): 82-88, 96.
- [5] 夏菲菲, 贾永红, 周温晖, 程欣悦. 基于 YOLOv8 的中华鲟智能检测方法[J]. 计算机科学与应用, 2024, 14(4): 213-218.
- [6] Chen, J. and Er, M.J. (2024) Dynamic YOLO for Small Underwater Object Detection. *Artificial Intelligence Review*, **57**, Article No. 165. <https://doi.org/10.1007/s10462-024-10788-1>
- [7] Lin, Y., Zhang, Y., Yang, D., Zhou, Y. and Wang, L. (2025) DLFE-YOLO: An Enhanced Framework for Underwater Object Detection Based on YOLOv8. *Intelligent Marine Technology and Systems*, **3**, Article No. 37. <https://doi.org/10.1007/s44295-025-00088-x>
- [8] Cheng, S., Wang, Z., Liu, S., Han, Y., Sun, P. and Li, J. (2024) Attention-Based Lightweight YOLOv8 Underwater Target Recognition Algorithm. *Sensors*, **24**, Article No. 7640. <https://doi.org/10.3390/s24237640>
- [9] Wang, X., Li, K., Fan, F., Wu, Y. and Zhang, Y. (2025) Optimized YOLOv8s Framework with Deformable Convolution for Underwater Object Detection. *Scientific Reports*, **15**, Article No. 45446. <https://doi.org/10.1038/s41598-025-28989-1>
- [10] Zhang, C. and Jiao, P. (2023) YOLO Series Target Detection Algorithms for Underwater Environments. arXiv: 2309.03539.