doi:10.6041/j.issn.1000-1298.2026.13.030

基于双流注意力机制引导和多尺度特征表达的轻量化水下目标检测算法

李　莹　王耀政　王应楠　何自芬
（昆明理工大学机电工程学院，昆明 650500）

摘要：在复杂的自然环境中，提升水下生物资源探测效率对中国海洋经济发展具有重要意义。为了应对水下目标检测中计算资源受限、水下环境复杂性所导致的检测效果差的问题，提出了一种基于改进 YOLO v8s 的水下目标检测算法 CBM－YOLO。首先，设计了轻量化特征提取模块 CSP－DLN，该模块融合了 3×3 和 1×1 卷积核的优势，能够高效捕捉细节空间特征，减少冗余计算。其次，为解决水下生物目标特征信息丢失的问题，引入新的特征融合网络 BGL－FPN，通过跨尺度连接结合全局和局部空间注意力机制，有效提高检测精度。最后提出最大池化下采样（MPD），通过最大池化和卷积分支的并行处理，更好地捕捉小目标的边缘和细节，从而增强小目标的检测能力。实验结果表明，该算法在 URPC2020 数据集和 UDD 数据集上的 mAP@0.5 分别达到 78.2% 和 69.1%，相较基准模型 mAP@0.5 各自提高 1.5、2.6 个百分点，同时参数量和计算量分别下降 47.3% 和 25.5%，相比于其他主流目标检测算法表现性能也最优。该模型被部署到 Jetson TX2 嵌入式端并经过 TensorRT 加速推理后，mAP@0.5 为 77.4%，检测速度达到 37.6 f/s，能够在保证高检测精度的同时实现水下生物的实时检测。

关键词：水下生物检测；YOLO v8s；轻量化；特征融合；Jetson TX2
中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2026)13-0359-10　　**OSID**：

Lightweight Underwater Target Detection Algorithm Driven by Dual-stream Attention Mechanism and Multi-scale Feature Representation

Li Ying　Wang Yaozheng　Wang Yingnan　He Zifen
(*Faculty of Mechanical and Electrical Engineering*, *Kunming University of Science and Technology*, *Kunming* 650500, *China*)

Abstract: In complex natural environments, enhancing the detection efficiency of underwater biological resources is vital for China's marine economic development. To address the issues of limited computational resources and poor detection due to underwater complexity, CBM－YOLO, an improved YOLO v8s-based underwater target detection algorithm was proposed. Firstly, CSP－DLN, a lightweight feature extraction module, was designed, combining the advantages of 3×3 and 1×1 convolutional kernels to efficiently capture detailed spatial features and reduce redundant calculations. Secondly, to address the loss of feature information in underwater biological targets, a new feature fusion network, BGL－FPN, was introduced, which effectively improved detection accuracy through cross-scale connections combined with global and local spatial attention mechanisms. Lastly, max pooling downsampling (MPD) was proposed, leveraging parallel processing of max pooling and convolutional branches to better capture edges and details of small targets, thereby enhancing their detection capability. Experimental results indicated that the algorithm attained mAP@0.5 of 78.2% and 69.1% on the URPC2020 and UDD datasets, respectively, with improvements of 1.5 and 2.6 percentage points over the baseline model, while reducing parameters and computations by 47.3% and 25.5%. It outperformed other mainstream target detection algorithms. Deployed on the Jetson TX2 embedded device and accelerated by TensorRT, the model achieved mAP@0.5 of 77.4% and a detection speed of 37.6 f/s, enabling real-time underwater detection with high accuracy.

Key words: underwater biological detection; YOLO v8s; lightweight; feature fusion; Jetson TX2

收稿日期：2025－03－02　修回日期：2025－04－01
基金项目：国家自然科学基金项目（62171206）
作者简介：李莹（1977—），女，副教授，博士，主要从事智能制造和目标检测研究，E-mail：7114091@qq.com

0 引言

随着探测技术和资源勘探领域的飞速发展,海洋以其巨大的资源储备,已经成为全球各国经济发展的关键战略领域。现代海洋经济正逐步成为推动国家经济增长的核心引擎,而人工智能技术与先进的目标检测算法,已成为推动这一领域技术突破的关键力量。然而,当前的水下资源勘探和海洋生态研究主要依赖于大型勘探设备和人力资源,这些方法的机动性有限,容易受到人为失误和其他外部因素的影响,导致探测精度降低。水下环境的复杂性,包括光线衰减、悬浮颗粒干扰以及低对比度[1],都增加了目标检测的难度,尤其是目标物体遮挡问题,进一步加剧了探测的漏检率。面对这些挑战,传统的水下目标检测方法已无法满足现代工业对高标准、高精度的需求,特别是在复杂水下环境中的应用表现出明显不足。因此,研发轻量化、高性能的目标检测算法尤为重要。而将这些算法集成至小型嵌入式设备,并广泛应用于水下机器人技术,将推动水下资源勘探迈向智能化、自动化的新阶段,为现代海洋经济的可持续发展注入强劲动力。

近年来,深度学习技术的蓬勃发展极大地推动了目标检测领域的进步。目标检测算法主要被划分为两大类:双阶段算法与单阶段算法。双阶段算法通过先生成候选框再进行分类的方式来实现目标检测,其中 R-CNN 系列算法例如(Faster R-CNN[2]和 Mask R-CNN[3])在物体检测和实例分割方面表现尤为突出。这些方法在复杂场景中具有较高的检测精度,但由于流程复杂和计算开销大,难以满足实时性的要求。相比之下,单阶段算法通过将目标检测任务简化为单一回归问题,实现了端到端的高效检测,以 YOLO(You Only Look Once)系列[4-13]、SSD[14]、DETR[15-16]系列为代表。其中 YOLO 系列不仅在实时性方面表现突出,同时也能在多种场景中达到较高的检测精度,因而在水下目标检测中得到了广泛应用。

目前,国内外众多学者针对水下目标检测领域展开了大量研究[17-24],从引入多尺度特征提取与跨尺度融合模块、嵌入注意力机制与轻量化结构、优化相似性度量与损失函数,以及融合可变形卷积和频域变换等多个角度进行改进优化,以提升对多尺度目标尤其是小目标的检测能力,并将这些改进分别应用于 Faster R-CNN、SSD、RT-DETR 以及 YOLO 系列等不同的模型,有效增强了复杂水下环境下对不同尺寸目标的自适应特征融合能力、识别精度与定位准确性。

尽管这些改进方法取得了良好的效果,但在水下复杂环境中,如何在保证检测精度的基础上进一步提升模型的轻量化和部署效率是当前面临的重要难题。因此,本文以 YOLO v8s 为基础,首先设计一种轻量化的特征提取模块,在降低计算复杂度的同时保持特征信息的完整性。其次对颈部网络进行优化重构,通过跨尺度连接和自注意力机制,加强特征信息的融合。最后,引入新的下采样模块,通过多分支的并行处理,提高捕捉目标边缘和细节的能力。

1 CBM-YOLO 网络

YOLO v8 是当前较为领先的一阶段实时目标检测模型,由主干网络(Backbone)、特征融合网络(Neck)和检测头(Head)3 部分组成,协同完成高效的目标检测任务。主干网络通过改进的 CSPNet[25]架构结合动态感受野,实现了轻量化的同时提升了特征提取能力。在特征融合网络中,YOLO v8 融合了特征金字塔网络[26](Feature pyramid network, FPN)和路径聚合网络[27](Path aggregation network, PAN)两种结构,从而实现多尺度特征的有效融合。YOLO v8 的检测头则采用 Anchor-free 设计,直接回归目标中心点和尺寸,减小了锚框生成的复杂性,同时引入分布焦点损失(DFL Loss)[28]和解耦头结构(Decoupled Head),提升了检测精度和推理效率。YOLO v8 的优势不仅在于其强大的检测性能,还体现在其灵活的部署支持上,兼容多种优化工具(如 TensorRT 和 ONNX),并具备良好的迁移学习能力。

本文基于 YOLO v8s 模型提出 CBM-YOLO,通过引入轻量化特征提取模块,提出一种全新的特征融合网络,采用新的下采样机制,在显著降低计算复杂度的同时提高其在复杂水下环境中的检测性能,其整体结构如图 1 所示。

2 改进方法

2.1 轻量化 CSP-DLN 模块

YOLO v8 中的 C2f 模块结构依赖于标准卷积以及通道间的特征信息融合,但由于其卷积核设计较为单一,在处理复杂的水下场景如光线不均匀、背景噪声等时,可能无法捕捉到目标细节的高效特征,同时标准卷积操作占据了较大的计算资源,特别是在高分辨率输入图像或嵌入式平台部署时,这种问题显得尤为明显。

为了解决上述问题,本文提出一种新的模块 CSP-DLN(CSP dual lightweight network),其结构如图 2 所示。CSP-DLN 借鉴了 CSPNet 结构理念,通过 DLN 进行堆叠,从而实现特征提取和融合。DLN

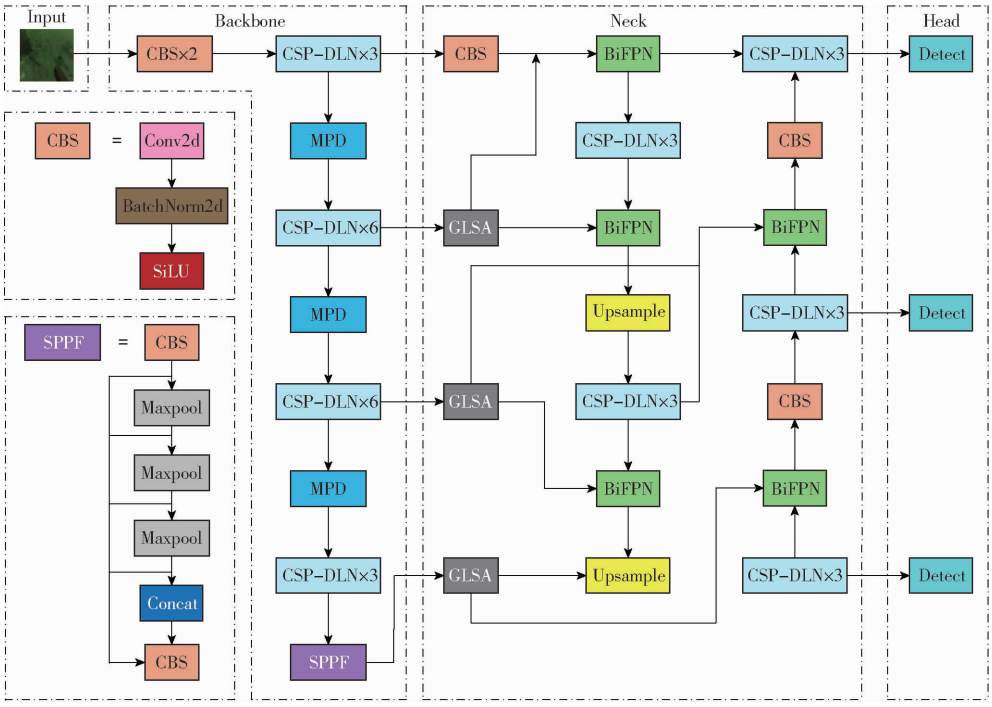


图 1　CBM－YOLO 网络框图

Fig. 1　CBM－YOLO network

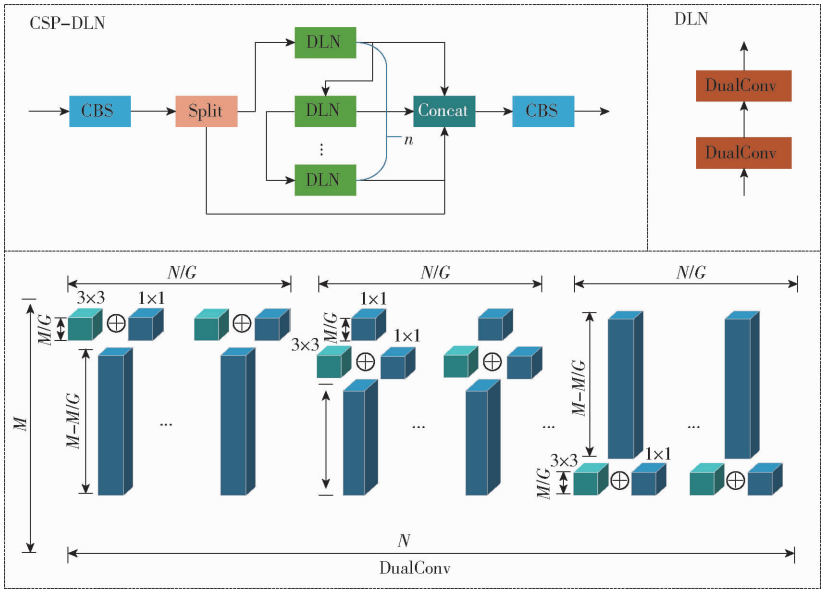


图 2　CSP－DLN 模块

Fig. 2　CSP－DLN module

模块由两个单独的 DualConv[29] 组成，结合 3×3 和 1×1 卷积核以及组卷积技术，实现了局部细节和全局上下文的高效结合。较大的 3×3 卷积核在识别水下生物特征所需的丰富空间细节方面表现卓越，而 1×1 卷积核则专注于降低特征通道维度，从而加速处理过程。这种设计不仅提升了捕捉水下生物复杂特征的能力，还增强了对目标细节的敏感度。

假设输出特征图的尺寸为 $D_0 \times D_0 \times N$，当使用标准卷积时，输入特征图通过 N 个大小为 $K \times K \times M$ 的卷积核进行卷积，因此，标准卷积模块的浮点运算量 L_S 计算公式为[29]

$$L_S = D_0^2 K^2 MN \tag{1}$$

在 DualConv 中，对于给定的 G，所有通道中 $(K \times K + 1 \times 1)$ 的组卷积的比例为 $1/G$，剩余的 1×1 卷积核的比例为 $(1 - 1/G)$。因此，组卷积的浮点运算量 L_C 计算公式为

$$L_C = D_0^2 K^2 M \frac{N}{G} + D_0^2 M \frac{N}{G} \tag{2}$$

剩余的 1×1 卷积的浮点运算量 L_P 计算公式为

$$L_P = D_0^2 MN(1 - G) \tag{3}$$

故 DualConv 模块的总浮点运算量 L 为

$$L = D_0^2 K^2 M \frac{N}{G} + D_0^2 MN \tag{4}$$

则 DualConv 模块与标准卷积模块的浮点运算量之比 $R_{D/C}$ 为

$$R_{D/C} = \frac{L}{L_S} = \frac{1}{G} + \frac{1}{K^2} \tag{5}$$

尽管 CSP－DLN 模块降低了计算复杂度，但其设计仍然保留了足够的网络表征能力。通过高效的特征提取与融合，CSP－DLN 能够捕获丰富的特征信息，同时在复杂水下检测任务中保持良好的性能。

2.2 BGL－FPN 模块

水下生物图像由于其特有的光照不均匀、模糊和亮斑等特点，导致图像特征信息丢失，对目标检测任务提出了更高的挑战。因此，高效的特征融合方法是解决这一问题的关键。

在 YOLO v8 中，采用了 FPN－PAN 架构，如图 3a 所示，通过构建自上而下和自下而上的多尺度特征融合路径，增强了特征表达能力。然而，FPN－PAN 对不同特征层级采用无差别的融合方式，忽略了各特征层对检测任务贡献的差异性，可能导致算法性能下降。

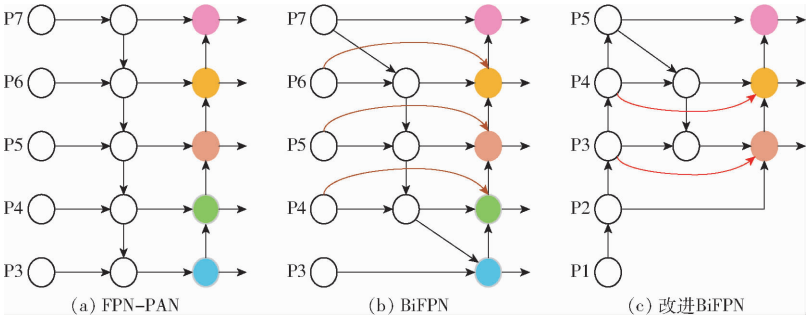


图 3 不同特征融合网络对比

Fig. 3 Comparison of different feature fusion networks

为了解决这一问题，本文提出了一种全新的特征融合网络——BGL－FPN，其结构如图 4 所示。BGL－FPN 从加权双向特征金字塔网络（Bidirectional feature pyramid network，BiFPN）的思想出发，并在此基础上进一步对其结构进行优化。BiFPN[30] 结构如图 3b 所示，通过动态加权并融合不同尺度特征图，使模型能够更准确地学习特定任务中各尺度特征的重要性。然而，BiFPN 未能充分结合水下生物检测的特点。因此，本文在 BiFPN 的基础上，结合 YOLO v8 自身特性，进行了针对性的改进，如图 3c 所示。

首先，通过引入大尺寸 P2 特征层，加强了对小目标特征信息的捕捉能力；其次，删除了 P5 的中间节点，精简了网络结构以避免冗余；同时新增了 P3 和 P4 层的跳跃连接，加强了特征信息间的语义联系。这些调整在保持参数成本的同时，显著提升了网络对水下生物目标的检测能力。

为了进一步提升特征的表达能力，BGL－FPN 还融合了全局与局部特征增强模块 GLSA[31]，从局部和全局两个维度对特征进行强化，GLSA 模块通过一种双流设计，分别提取局部空间注意力（LSA）和全局空间注意力（GSA），从而同时保留细粒度特征和长距离依赖信息。输入的特征图首先通过 Split 模块被均匀分割为两份，分别送入 LSA 和 GSA 模块进行处理，随后通过 Concat 融合两者的输出，

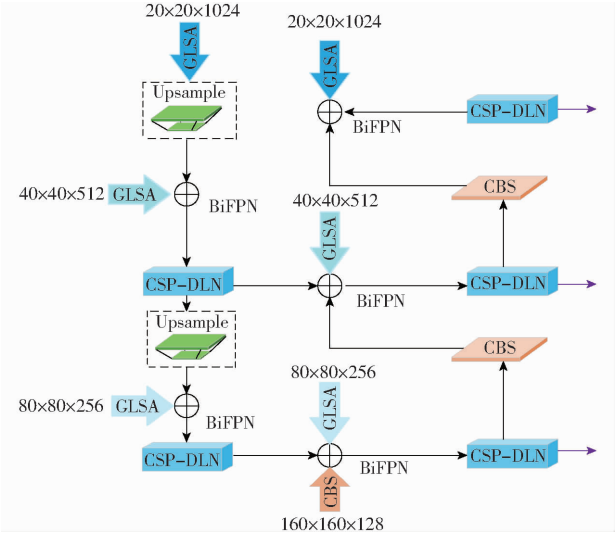


图 4 BGL－FPN 网络结构图

Fig. 4 BGL－FPN network

并通过 1×1 卷积进一步优化。这种设计平衡了模型的精确度和计算资源，同时增强了模型对复杂场景下目标特征的表达能力。GLSA 结构如图 5 所示。

通过结合 BiFPN 的优化设计与 GLSA 模块的增强策略，BGL－FPN 在特征层之间引入跨尺度连接，丰富了语义信息与底层信息，并通过自注意力机制提升了全局和局部特征的表达能力。BGL－FPN 在降低计算成本的同时，提升了对水下生物的检测效果，对于水下目标检测任务更具适应性。

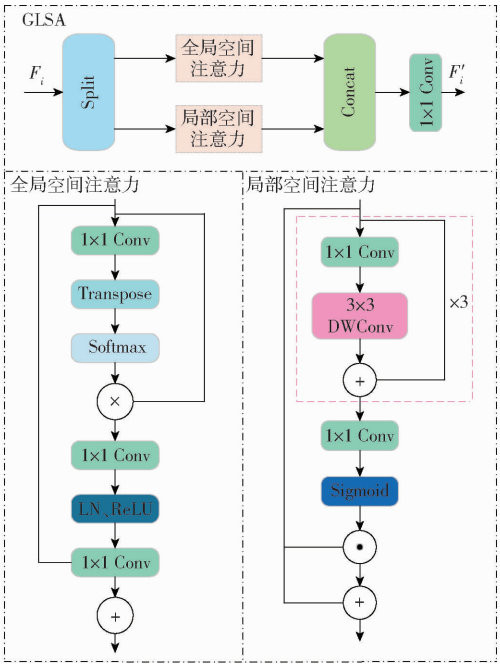


图 5　GLSA 模块

Fig. 5　GLSA module

2.3　下采样模块

YOLO v8 的下采样模块仅基于步长为 2 的卷积实现，采用单一路径特征提取方式。单一特征可能不足以同时捕获全局信息和细节特征，在复杂水下环境下容易导致小目标漏检或特征混淆。针对这一问题，提出了一种新的下采样模块（Max pooling downsampling，MPD），MPD 采用多路径聚合下采样机制，最大池化分支通过选择每个局部区域内的最大值作为输出，能够突出最显著的特征，从而增强模型对重要特征信息的捕捉能力，有助于识别水下场景中的目标区域，卷积分支通过提取局部细节特征，能够捕捉小目标的边缘和细节，两条分支通过 Concat 进行融合，得到更加丰富的特征信息，最后再通过 1×1 卷积减少通道冗余，同时进一步增强网络的特征表达能力。MPD 结构如图 6 所示。

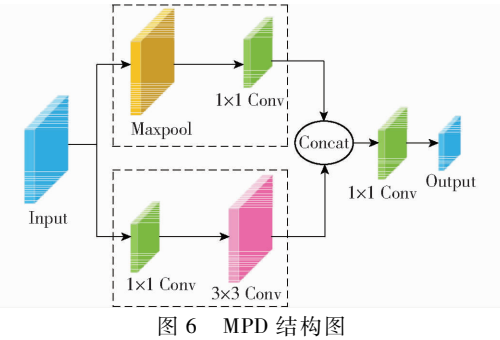


图 6　MPD 结构图

Fig. 6　MPD structure

3　实验结果与分析

3.1　实验环境和数据集

实验所用操作系统为 Windows 10，使用 Python 3.9、CUDA 12.1、Pytorch 2.1.0 深度学习框架开发环境，在 NVIDIA GeForce RTX 4060Ti 显卡、16 GB 显存中进行训练。训练时，使用 SGD 优化器。初始化学习率设置为 0.01，迭代批次为 4，共训练 500 个周期，早停设置为 50。嵌入式设备为 Jetson TX2，搭载双核 Denver2 与四核 Cortex－A57 CPU，均主频 2 GHz，8GB LPDDR4 内存，带宽 59.7 GB/s，32 GB eMMC 5.1 存储，256 核 Pascal 架构 GPU，AI 算力 1.33×10^{12} s^{-1}，功耗 7.5～15 W。配套的 JetPack SDK 拥有完整的开发环境，包含 TensorRT、CUDA、cuDNN、OpenCV 等，支持 PyTorch、TensorFlow、Caffe、MXNet 等框架，可以运行深度神经网络并且能够获得较高的准确性，功耗相较于 PC 和服务器很低。

本文使用了两个常用的水下生物数据集：URPC2020[32]和 UDD[33]。URPC2020 数据集来自水下机器人目标抓取大赛（URPC），包括海参、海胆、扇贝、海星 4 种类别，共有 7 543 幅图像，其中 5 543 幅用来训练，1 200 幅用来验证，800 幅用来测试。UDD 由 3 类类别（海参、海胆和扇贝）组成，这是在真实的公海农场收集的数据集，共有 2 227 幅图像，其中 1 827 幅用来训练，400 幅用来测试。样图如图 7 所示。

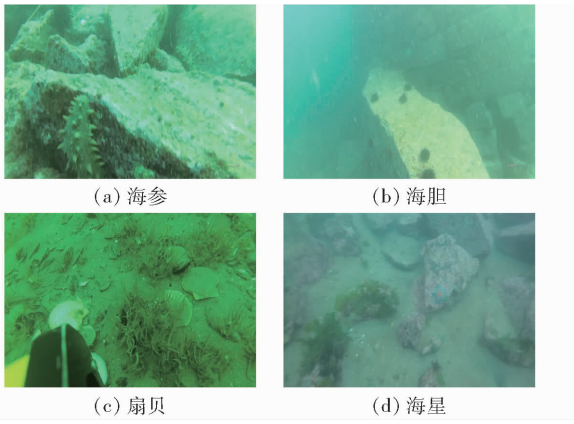


图 7　数据集样图

Fig. 7　Sample image of dataset

3.2　评价标准

采用平均精度均值（Mean average precision，mAP）指标来衡量多个类别的平均检测精度，采用浮点运算量来衡量模型复杂程度，采用参数量来衡量模型大小，采用帧率来衡量模型检测速度。

3.3　消融实验

为了对比验证改进模型各个模块的有效性，以 YOLO v8s 为基准模型，图像输入尺寸为 640 像素×640 像素，在 URPC2020 数据集上进行消融实验。通过相同的实验条件来评估不同的模块对检测效果的影响，实验结果见表 1（表中 mAP@0.5 表示 IoU

阈值为 0.5 时所有类别的平均检测精度)。其中“√”表示使用改进方法,“×”表示未使用改进方法。

表 1 消融实验结果

Tab.1 Results of ablation experiment

CSP－DLN	BGL－FPN	MPD	参数量	浮点运算量	mAP@0.5/%
×	×	×	1.11×10^{7}	2.86×10^{10}	76.7
√	×	×	8.71×10^{6}	2.24×10^{10}	76.9
×	√	×	7.94×10^{6}	2.71×10^{10}	77.8
×	×	√	1.05×10^{7}	2.80×10^{10}	77.7
√	√	×	6.46×10^{6}	2.17×10^{10}	77.9
√	√	√	5.87×10^{6}	2.13×10^{10}	78.2

从表 1 结果来看,在引入改进模块后,水下目标检测任务的各项性能指标均实现了提升。在应用轻量化特征提取模块 CSP－DLN 后,mAP@0.5 指标提升了 0.2 个百分点的同时,参数量和浮点运算量较基线模型分别减少 2.39×10^{6} 和 6.2×10^{9},说明 CSP－DLN 模块在实现轻量化的同时又能够保证模型性能。替换颈部网络为 BGL－FPN 后,mAP@0.5 指标提升了 1.1 个百分点的同时,参数量和浮点运算量较基线模型分别减少 3.16×10^{6} 和 1.5×10^{9},有效地提高了模型的特征融合能力。引入新的下采样模块 MPD 后,mAP@0.5 指标提升了 1.0 个百分点的同时,参数量和浮点运算量较基线模型分别减少 6×10^{5} 和 6×10^{8},验证了 MPD 模块对重要特征信息的捕捉能力。最后,将 CSP－DLN 模块、BGL－FPN 模块、MPD 模块同时作用于基线模型,在mAP@0.5 指标上较基线模型提升 1.5 个百分点,参数量降低 5.23×10^{6},浮点运算量降低 7.3×10^{9},验证了所提 CBM－YOLO 模型能够在实现轻量化的同时提高检测精度,对水下生物目标检测任务是有效的。

3.4 多模型对比实验

3.4.1 在 URPC2020 数据集上对比

为了验证改进算法的有效性,本文首先在 URPC2020 数据集上展开一系列实验研究。图 8(图中 AP@0.5 表示 IoU 阈值为 0.5 时单个类别 P－R 曲线下的面积)展示了在 URPC2020 数据集上分类下的精确率-召回率(P－R)曲线,从图中可以看到,其中检测结果最好的是海胆,AP@0.5 达到了 90.1%。然而,对于扇贝,鉴于其在数据集中的稀少性,目前难以取得较高的精确度。最后,CBM－YOLO 算法在 URPC2020 数据集上 mAP@0.5 达到了 78.2%,实现了较好的性能。

为了进一步验证 CBM－YOLO 算法的优势,本

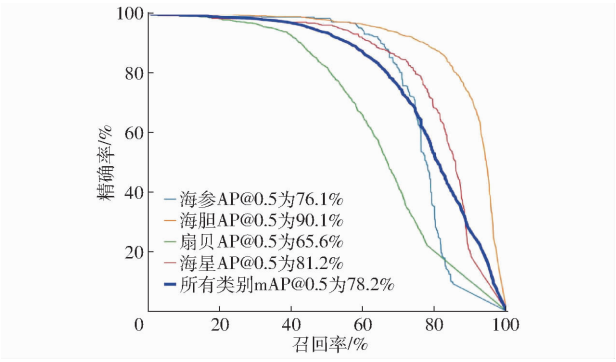


图 8 URPC2020 数据集的 P－R 曲线

Fig.8 P－R curves of URPC2020 dataset

文选取了一系列前沿且具有代表性的目标检测算法,如 SSD、Fast R－CNN、YOLO v5s、YOLO v7－tiny、YOLO v8s、YOLO v9s、YOLO v10s、YOLO 11s、RT－DETR 在 URPC2020 数据集上进行对比分析。实验结果如表 2 所示,mAP@0.5:0.95 表示在多个 IoU 阈值(0.5 到 0.95,步长 0.05)下平均精度的均值。相比于其他目标检测算法,本文提出的 CBM－YOLO 展现出了更优异的性能。mAP@0.5 和 mAP@0.5:0.95 分别达到了 78.2% 和 45.6%,均优于其他模型,同时模型浮点运算量 2.13×10^{10},参数量 5.87×10^{6} 最低。综合评估,在水下生物目标检测任务中,CBM－YOLO 以较低的浮点运算量和参数量实现了更高的检测精度。

表 2 URPC2020 数据集上各模型性能对比

Tab.2 Comparison of model performance on URPC2020

模型	参数量	浮点运算量	mAP@0.5/%	mAP@0.5:0.95/%
SSD	2.40×10^{7}	6.11×10^{10}	54.9	26.2
Faster R－CNN	1.37×10^{8}	3.70×10^{11}	50.8	24.3
YOLO v5s	7.02×10^{6}	1.58×10^{10}	74.8	42.5
YOLO v7－tiny	6.01×10^{6}	1.30×10^{10}	75.5	41.5
YOLO v8s	1.11×10^{7}	2.86×10^{10}	76.7	44.2
YOLO v9s	9.78×10^{6}	3.98×10^{10}	76.6	43.6
YOLO v10s	8.04×10^{6}	2.45×10^{10}	76.2	43.1
YOLO 11s	9.41×10^{6}	2.13×10^{10}	76.7	43.5
RT－DETR	1.99×10^{7}	5.70×10^{10}	77.6	44.0
本文	5.87×10^{6}	2.13×10^{10}	78.2	45.6

此外,为了验证所提特征融合网络的有效性,在保证其他条件一致的情况下,将 BGL－FPN 与FPN－PAN 和改进 BiFPN 进行了对比,如表 3 所示。通过特征融合网络的对比实验可知,在 YOLO v8 中引入改进 BiFPN 后,增强了特征融合效果及对多尺度目标的检测能力,mAP@0.5 和 mAP@0.5:0.95 分别提升了 0.7、0.4 个百分点,同时参数量和浮点运算量有明显下降;进一步采用 BGL－FPN 作为特征融

合网络后，该方法更有效地融合了局部空间细节信息和全局空间语义信息，mAP@0.5 和 mAP@0.5:0.95 分别提升了 1.1、1.3 个百分点，相较于改进 BiFPN，精度有明显提升且浮点运算量和参数量仅略微增长，表明 BGL－FPN 模块在针对水下生物检测任务中具有更好的适应性。

表 3　不同特征金字塔性能对比

Tab. 3　Performance comparison of different feature pyramids

模型	参数量	浮点运算量	mAP@0.5/%	mAP@0.5:0.95/%
FPN－PAN	1.11×10^{7}	2.86×10^{10}	76.7	44.2
改进 BiFPN	7.37×10^{6}	2.51×10^{10}	77.4	44.6
BGL－FPN	7.94×10^{6}	2.71×10^{10}	77.8	45.5

3.4.2　在 UDD 数据集上对比

为了验证 CBM－YOLO 算法在其他水下生物数据集的普适性，进一步在 UDD 数据集上与 SSD、Faster R－CNN、YOLO v5s、YOLO v7－tiny、YOLO v8s、YOLO v9s、YOLO v10s、YOLO 11s、RT－DETR 进行对比，结果如表 4 所示。CBM－YOLO 在检测精度方面，mAP@0.5 达到 69.1%，同时 mAP@0.5:0.95 达到 30.3%，均为表中最佳结果。同时 CBM－YOLO 的参数量为 5.87×10^{6}，均显著低于其他主流模型，浮点运算量仅为 2.13×10^{10}。实验结果表明，CBM－YOLO 即使在 UDD 数据集上依然展现出卓越性能，充分体现了其在水下生物检测任务中的广泛适用性。

表 4　UDD 数据集上各模型性能对比

Tab. 4　Comparison of model performance on UDD

模型	参数量	浮点运算量	mAP@0.5/%	mAP@0.5:0.95/%
SSD	2.40×10^{7}	6.11×10^{10}	55.7	24.4
Faster R－CNN	1.37×10^{8}	3.70×10^{11}	55.2	24.1
YOLO v5s	7.02×10^{6}	1.58×10^{10}	65.6	28.1
YOLO v7－tiny	6.01×10^{6}	1.30×10^{10}	64.3	26.2
YOLO v8s	1.11×10^{7}	2.86×10^{10}	66.5	29.2
YOLO v9s	9.78×10^{6}	3.98×10^{10}	65.7	28.2
YOLO v10s	8.04×10^{6}	2.45×10^{10}	66.8	28.5
YOLO 11s	9.41×10^{6}	2.13×10^{10}	67.3	28.9
RT－DETR	1.99×10^{7}	5.70×10^{10}	68.3	29.3
本文	5.87×10^{6}	2.13×10^{10}	69.1	30.3

3.5　可视化分析

通过可视化分析，可以更直观地展示模型的检测效果。如图 9、10 所示，分别展示了 CBM－YOLO 在 URPC2020 数据集和 UDD 数据集上的可视化检测结果，以及与 YOLO v8s 的对比结果，图中每个检测框上方标注的是该检测框的类别以及模型预测该目标属于该类别的置信度。从图中可以看出，CBM－YOLO 算法在目标多有遮挡、目标少无遮挡以及模糊 3 种场景下检测精度、漏检率和误检率方面均优于基准算法。相比基准模型，CBM－YOLO 能够准确识别并定位图像中的所有目标，且未出现任何错误的检测框。这些对比结果充分证明了本文提出模型在水下生物目标检测领域的有效性和优越性。

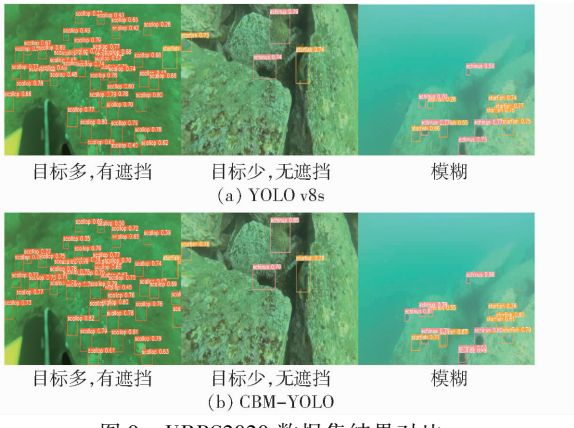


图 9　URPC2020 数据集结果对比

Fig. 9　Comparison of URPC2020 dataset results

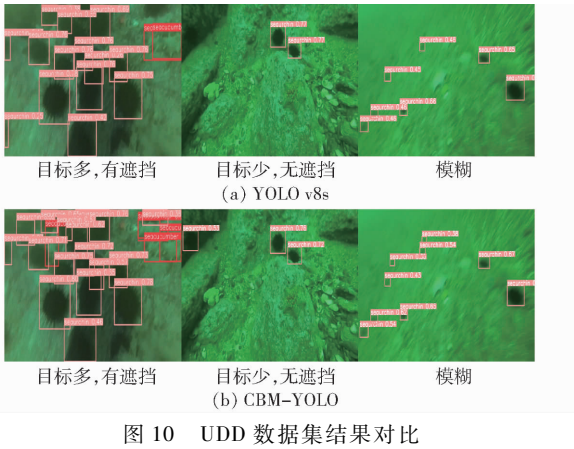


图 10　UDD 数据集结果对比

Fig. 10　Comparison of UDD dataset results

3.6　嵌入式部署

随着小型嵌入式平台计算性能的提升，它们已经能够满足部分基于深度学习技术的目标检测任务。本文选用了高性能且低功耗的 Jetson TX2 嵌入式设备作为推理平台。由于深度学习模型在推理阶段需要依赖庞大框架，且频繁地调用网络各层的函数可能会带来额外计算负担，因此对水下生物检测模型实施推理优化显得尤为关键。

本文利用 TensorRT 进行加速与优化，旨在提升网络处理速度并减少能耗。具体部署流程如下：首先在 PC 端完成对 CBM－YOLO 模型的训练，并将训练好的模型文件传输到 Jetson TX2 开发板。随后在 Jetson TX2 上对模型进行优化与编译，确保其能够充分利用硬件资源。在模型优化完成后，进行格式

转换,以适配 TensorRT 的推理加速框架。格式转换的流程依次为:先将模型从 PyTorch 格式(.pt)转换为开放神经网络交换格式(ONNX),然后进一步转换为 TensorRT 格式(.TRT)。模型运行效果如图 11 所示。



图 11 Jetson TX2 运行效果

Fig. 11 Operation effect of Jetson TX2

在完成模型的优化与部署后,为验证其优化效果,本文对模型优化前后的性能进行了全面对比分析。尽管该优化方法在较大程度上维持了模型的训练精度,但由于推理阶段网络结构的融合,在加速过程中可能会导致检测精度一定程度下降。为此,通过在 URPC2020 数据集上的实验,对比优化前后模型的性能,从检测精度和检测速度两个方面进行评估,结果如表 5 所示。

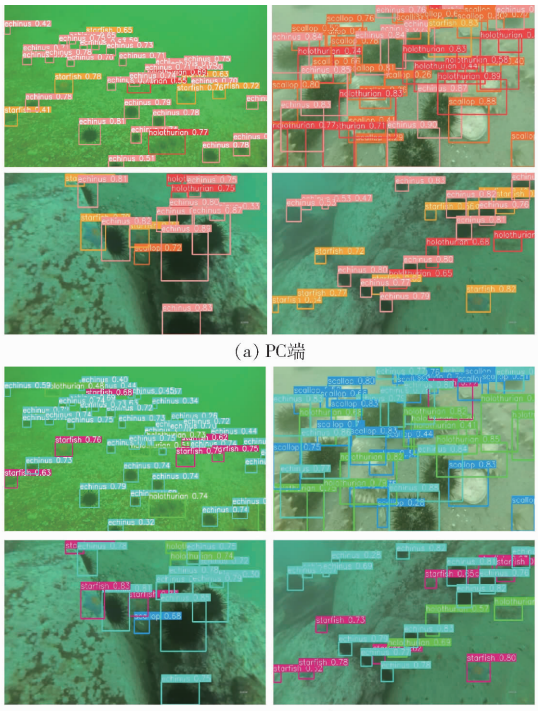
表 5 不同平台检测结果对比

Tab. 5 Comparison of test results on different platforms

处理平台	模型格式	检测速度/($f \cdot s^{-1}$)	mAP@0.5/%	mAP@0.5:0.95/%
RTX4060Ti	PT	98.0	78.2	45.6
Jetson TX2	PT	12.0	78.2	45.6
Jetson TX2	TRT	37.6	77.4	44.9

如表 5 所示,模型在转换前后其 mAP 并未出现明显下降,但处理速度却大幅度提高。其中,优化后的 CBM-YOLO 模型在 Jetson TX2 开发板上的处理速度提升至优化前的 3.1 倍(37.6/12.0),满足实时检测的需求。同时为了进一步验证优化模型在不同设备上的检测性能,本文在 PC 端和 Jetson TX2 端分别运行优化后的模型,并对其检测效果进行对比,结果如图 12 所示。

从图 12 可以看出,优化后的 CBM-YOLO 网络能够精准检测出水下生物个体目标。在利用 TensorRT 加速的条件下,模型不仅显著提高了水下目标检测速度,同时仍然能够保持较高的检测精度。这表明本文提出的优化方法在速度与精度之间达到了良好的平衡。



(a) PC端

(b) Jetson TX2端

图 12 PC 端和 Jetson TX2 端检测效果对比

Fig. 12 Comparison of detection effects between PC and Jetson TX2 terminals

4 结束语

水下目标检测任务面临着严峻的挑战。一方面,受限的计算资源使得高效处理和实时分析海量水下数据变得困难;另一方面,复杂多变的水下环境,包括光线衰减、悬浮颗粒干扰以及低对比度等因素,显著降低了检测的性能和准确性。本文提出了一种水下生物目标检测方法 CBM-YOLO。首先,在 YOLO v8s 模型中引入轻量化 CSP-DLN 特征提取模块,降低计算复杂度的同时保持了模型的特征提取能力。其次,替换颈部网络为 BGL-FPN,加强了特征信息之间的多尺度融合以及全局和局部特征的表达能力。最后引入下采样 MPD 模块,保留了更完整的特征信息,增强了对小目标的检测能力。CBM-YOLO 在 URPC2020 和 UDD 数据集上 mAP@0.5相较于原始 YOLO v8s 分别提高 1.5、2.6 个百分点同时参数量降低了 47.3%,浮点运算量降低了 25.5%。在多模型对比实验中,CBM-YOLO 在 mAP@0.5、mAP@0.5:0.95 和参数量指标上均优于主流的目标检测算法。该模型最终被部署到 Jetson TX2 嵌入式设备上,通过 TensorRT 加速推理后,mAP@0.5 为 77.4%,检测速度达到 37.6 f/s,能够满足水下实时检测的需求。

参 考 文 献

[1] Jian M, Liu X, Luo H, et al. Underwater image processing and analysis: a review[J]. Signal Processing: Image Communication, 2021, 91(5): 116088.

[2] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[3] He K, Gkioxari G, Doallar P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017:2961-2969.

[4] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 779-788.

[5] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 6517-6525.

[6] Redmon J, Farhadi A. YOLOv3: an incremental improvement[PP/OL]. V1. arXiv(2018-04-08)[2025-04-01]. https://arxiv.org/abs/1804.02767.

[7] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[PP/OL]. V1. arXiv(2020-04-23)[2025-04-01]. https://arxiv.org/abs/2004.10934.

[8] Ge Z, Liu S, Wang F, et al. YOLOX: exceeding YOLO series in 2021[PP/OL]. V2. arXiv(2021-08-06)[2025-04-01]. https://arxiv.org/abs/2107.08430.

[9] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[PP/OL]. V1. arXiv(2022-09-07)[2025-04-01]. https://arxiv.org/abs/2209.02976.

[10] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[PP/OL]. V1. arXiv(2022-07-06)[2025-04-01]. https://arxiv.org/abs/2207.02696.

[11] Wang C Y, Yeh I H, Liao H Y M. YOLO v9: learning what you want to learn using programmable gradient information[C]// European Conference on Computer Vision. Cham: Springer, 2025.

[12] Wang A, Chen H, Liu L, et al. YOLOv10: real-time end-to-end object detection[PP/OL]. V2. arXiv(2024-10-30)[2025-04-01]. https://arxiv.org/abs/2405.14458.

[13] Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements[PP/OL]. V1. arXiv(2024-10-23)[2025-04-01]. https://arxiv.org/abs/2410.17725.

[14] Wei L, Dragomir A, Dumitru E, et al. SSD: single shot multibox detector[M]. Cham: Springer, 2016.

[15] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[PP/OL]. V3. arXiv(2020-05-28)[2025-04-01]. https://arxiv.org/abs/2005.12872.

[16] Zhao Y, Lv W, Xu S, et al. DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle: IEEE, 2024: 16965-16974.

[17] 熊海涛, 林琪, 宣魁, 等. 基于改进 Faster R-CNN 的海参目标检测算法[J]. 农业机械学报, 2022, 53(增刊 2): 204-209.
Xiong Haitao, Lin Qi, Xuan Kui, et al. Sea cucumber object detection algorithm based on improved Faster R-CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(S2): 204-209. (in Chinese)

[18] 周华平, 宋明龙, 孙克雷. 一种轻量化的水下目标检测算法 SG-Det[J]. 光电子·激光, 2023, 34(2): 156-165.
Zhou Huaping, Song Minglong, Sun Kelei. SG-Det: a lightweight underwater image target detection method[J]. Journal of Optoelectronics·Laser, 2023, 34(2): 156-165. (in Chinese)

[19] Han K, Wang Y, Tian Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle: IEEE, 2020: 1577-1586.

[20] 张路, 魏本昌, 魏鸿奥, 等. 基于改进 RT-DETR 的水下目标检测[J]. 计算机系统应用, 2024, 33(12): 131-140.
Zhang Lu, Wei Benchang, Wei Hongao, et al. Underwater target detection based on improved RT-DETR[J]. Computer Systems & Applications, 2024, 33(12): 131-140. (in Chinese)

[21] 张明华, 黄基萍, 宋巍, 等. 基于 UMS-YOLO v7 的面向样本不均衡的水下生物多尺度目标检测方法[J]. 农业机械学报, 2025, 56(1): 388-396,409.
Zhang Minghua, Huang Jiping, Song Wei, et al. Multi-scale object detection method for underwater organisms under unbalanced samples based on UMS-YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(1): 388-396,409. (in Chinese)

[22] Tong Z, Chen Y, Xu Z, et al. Wise IoU: bounding box regression loss with dynamic focusing mechanism[PP/OL]. V3. arXiv(2023-04-08)[2025-04-01]. https://arxiv.org/abs/2301.10051.

[23] Wang J, Xu C, Yang W, et al. A Normalized Gaussian Wasserstein Distance for tiny object detection[PP/OL]. V2. arXiv(2022-06-14)[2025-04-01]. https://arxiv.org/abs/2110.13389.

[24] 袁红春, 李春桥. 基于改进可变形卷积的 FDC-YOLO v8 水下生物目标检测方法研究[J]. 农业机械学报, 2024,

55(11): 140 - 146.
Yuan Hongchun, Li Chunqiao. Research on FDC - YOLO v8 underwater biological object detection method improved by deformable convolution[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 140 - 146. (in Chinese)

[25] Wang C Y, Liao H Y M, Wu Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle: IEEE, 2020: 1571 - 1580.

[26] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos: IEEE Computer Society, 2017:2117 - 2125.

[27] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City: IEEE, 2018: 8759 - 8768.

[28] Li X, Wang W, Wu L, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[C]// Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: NeurIPS, 2020: 21002 - 21012.

[29] Zhong J, Chen J, Mian A. DualConv: dual convolutional kernels for lightweight deep neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(11): 9528 - 9535.

[30] Tan M, Pang R, Le Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle: IEEE, 2020: 10778 - 10787.

[31] Tang F, Huang Q, Wang J, et al. DuAT: dual-aggregation transformer network for medical image segmentation[C]//Pattern Recognition and Computer Vision. Cham: Springer, 2023: 343 - 356.

[32] Liu C, Li H, Wang S, et al. A dataset and benchmark of underwater object detection for robot picking[C]//2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). Shenzhen: IEEE, 2021: 1 - 6.

[33] Wang Z, Liu C, Wang S, et al. UDD: an underwater open-sea farm object detection dataset for underwater robot picking[PP/OL]. V2. arXiv(2021 - 07 - 28)[2025 - 04 - 01]. https://arxiv.org/abs/2003.01446.

(上接第 346 页)

[25] Ma Jiajia, Wang Keqiang. Research progress of optical nondestructive testing technology for fruit quality[J]. Science and Technology of Food Industry, 2021, 42(23):427 - 437.

[26] 邵志明, 王怀彬, 董志城, 等. 基于近红外相机成像和阈值分割的苹果早期损伤检测[J]. 农业机械学报, 2021, 52(增刊 1): 134 - 139.
Shao Zhiming, Wang Huaibin, Dong Zhicheng, et al. Early bruises detection method of apple surface based on near infrared camera imaging technology and image threshold segmentation method[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(S1): 134 - 139. (in Chinese)

[27] 徐超, 肖潇, 骆燕, 等. 基于距离变换的新型骨架提取方法[J]. 仪器仪表学报, 2012, 33(12): 2851 - 2856.
Xu Chao, Xiao Xiao, Luo Yan, et al. A new skeleton extraction method based on distance transformation[J]. Chinese Journal of Scientific Instrument, 2012, 33(12): 2851 - 2856. (in Chinese)

[28] 魏亚辉, 黄耿楠, 吴福培. 基于改进分水岭算法的苹果识别方法[J]. 包装工程, 2021, 42(8): 255 - 260.
Wei Yahui, Huang Gengnan, Wu Fupei. Apple recognition method based on improved watershed algorithm[J]. Packaging Engineering, 2021, 42(8): 255 - 260. (in Chinese)

[29] Tang Huapeng, Qin Danyang, Yan Mengying, et al. Research on color image matching method based on feature point compensation in dark light environment[J]. High Technology Letters, 2023, 29(1): 78 - 86.

[30] Gan Yong, Yu Jianghao, Xue Fei, et al. Analysis of effect of light source angle on image edge of metal surfaces[J]. Journal of Applied Optics, 2021, 42(4): 614 - 621.