

文章编号: 1674-5566(2025)01-0188-13

DOI: 10.12024/jsou.20240804619

基于改进YOLOv8n的水下鱼类目标识别轻量化模型

曹 宇, 李佳阳, 王 芳

(上海海洋大学 工程学院, 上海 201306)

摘 要: 针对复杂水下环境中鱼类目标识别精度不足、传统目标识别模型复杂度高、识别速度慢等问题, 出现了一种基于改进YOLOv8n的水下鱼类目标识别模型YOLOv8n-fish。本研究提出轻量级双卷积模块C2f-DualConv, 以改善YOLOv8n中C2f模块的特征学习能力; 基于高效的结构重参数化思想设计全新的颈部网络EffQAFPN, 以平衡目标模型的识别精度和速度; 采用二阶段微调方法, 提升水下弱光及干扰物环境下鱼类目标识别模型的识别精度。实验结果表明, YOLOv8n-fish模型在测试集的平均精度为97.47%, 较传统YOLOv8n模型提升了1.07%; 而改进后模型的参数量、浮点运算量和模型内存占用量仅为原始模型的56.1%、82%和66.7%。YOLOv8n-fish模型的识别速度仅次于YOLOv5n-P6, 可达到121 帧/s。结果表明, YOLOv8n-fish模型在保持高识别精度的同时显著降低计算成本, 为水产养殖的智能监测提供有效的技术支持。

关键词: 水产养殖; YOLOv8n; 双卷积模块; 轻量化; 深度学习

中图分类号: S 951.2; TP 391.41

文献标志码: A

目前, 世界水产品消费量增长显著, 2020年人均水产食品的消费量达到20.2 kg, 渔业将在提供食物和营养方面发挥着重要作用^[1]。智能化装备可有效地降低鱼类养殖的劳动强度, 采用计算机视觉和深度学习等人工智能技术识别水下养殖目标对于实现水产养殖健康智能化监测具有重要意义^[2-3]。

针对复杂海水环境下光线衰减导致颜色失真和鱼类传统识别方法识别精度低的问题, 矫萌璐等^[4]提出了一种基于Faster-RCNN的海洋牧场鱼类识别与分类方法。该方法通过串行数字接口(Serial digital interface, SDI)信号色彩补偿系统来改善图像质量, 优化特征提取网络与区域建议网络来实现海洋牧场鱼类识别与分类, 在测试集上的平均精度达到81.63%。熊海涛等^[5]提出基于Faster R-CNN的水下海参目标识别算法, 采用Swin Transformer作为特征提取网络, 同时融入多尺度特征提取层和实例分割功能, 提高算法的自适应特征融合能力; 针对商业渔船电子监控系统中鱼类识别依赖于人工完成等问题, 梅海彬

等^[6]提出一种轻量级鱼类实时识别模型YOLOv7-MRN。该模型首先将YOLOv7的骨干网络替换为MobileNetv3骨干网络, 以降低运算量, 并添加了感受野模块(Receptive field block, RFB)来增强网络的特征提取能力, 其次, 通过引入基于归一化的注意力机制模块(Normalization-based attention module, NAM)重新设计颈部特征融合网络, 以抑制无关紧要的权重。实验结果表明: 在HNY768远洋渔船电子监控视频渔业数据集上, YOLOv7-MRN模型的平均精度(Mean average precision, mAP₅₀)为86.5%, 运算量仅为原模型YOLOv7的9.8%, 模型在图形处理器(Graphics processing unit, GPU)和中央处理器(Central processing unit, CPU)上的运算速度分别提高了121.69%和219.09%; 针对识别海洋生物和生态的多样性研究, LIU等^[7]提出了一种复杂水下环境鱼类识别模型DP-FishNet。通过组合双路径特征提取网络, 分别通过视觉转换器路径和卷积神经网络路径来提取特征图中的全局特征和局部特征, 旨在提升模型对复杂环境中小目

收稿日期: 2024-08-18 修回日期: 2024-11-30

基金项目: 国家自然科学基金(52371282)

作者简介: 曹 宇(1982—), 男, 副教授, 研究方向为图像识别技术和海洋渔业装备。E-mail: y_cao@shou.edu.cn

通信作者: 王 芳, E-mail: wangfang@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>

标识别的准确性。ZHOU 等^[8]提出一种基于 YOLO 算法的海洋生物识别方法,该方法提出具有双端注意力机制的残差构建单元模块,将改进通道注意力机制融入残差结构中,结果表明,所提出的方法极大地提高了复杂海洋环境中生物识别精度。上述鱼类识别方法未考虑水下环境存在鱼类特征相似性等问题,另外,其模型参数量和模型内存占用量大,难以部署到资源受限的边缘设备。

为提高水下复杂环境鱼类目标识别的精确度,以及解决模型较大导致部署受限的问题,本研究提出一种适用于水下弱光及干扰物环境的轻量级鱼类识别模型 YOLOv8n-fish。在 YOLOv8n 的基础上,根据双卷积 DualConv 和 C2f 模块提出 C2f-DualConv 模块,旨在实现模型轻量化的同时改进模型特征提取网络的学习能力;针对水下环境中存在类似鱼类的背景干扰物,提出全新的高效重参数化颈部网络 (Efficient QARepNeXt feature pyramid network, EffQAFPN) 来增强模型学习重要鱼类特征的位置与信息能力,并提高对鱼类的识别准确度和速度。此外,提出一种高效的二阶段微调方法,在不增加模型复杂度的基础上,采用辅助数据集训练模型生成权重来微调目标数据集的训练过程,进一步优化模型识别能力,旨在便于后续的智能化水产养殖设备的部署。本研究在模型部署上为水下鱼类目标识别提供了新思路,利用机器视觉技术实现水产养殖健康智能化监测。

1 轻量化鱼类识别模型

1.1 改进 YOLOv8 网络结构

YOLO 是一种实时目标识别模型,通过全局推理来获得信息,可实现识别速度和识别精度的良好平衡^[8-13]。其中,YOLOv8 是一种端到端的深度学习模型,可直接从原始图像中识别和定位目标,并通过设置深度因子 d 、宽度因子 w 和比例

因子 r 等 3 个参数调节主干网络的参数量,现已划分出 5 个量级的模型,分别是 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x,其模型尺寸和参数依次增加。为分析目标识别模型的复杂度和精确度,使用原始鱼类图像数据集在 YOLOv8 的 5 个版本上进行了训练,训练结果如表 1 所示,YOLOv8n 表现出最轻量化的模型复杂度,模型参数量为 4.85×10^6 ,模型内存占有量大小为 8.70 MB,模型浮点数计算量为 8.9 B,同时 YOLOv8n 保持最快的识别速度,达到 83 帧/s,模型识别精度接近于 YOLOv8 其他模型。

因此本研究选用 YOLOv8n 作为基础模型进行改进:设计基于 DualConv 和 C2f 模块的轻量级卷积神经网络 C2f-DualConv,提高模型的特征信息提取能力;为提升模型在水下环境中提取有效特征的能力,优化了 EffQAFPN 卷积神经网络架构。改进后的 YOLOv8n-fish 网络架构如图 1 所示。

1.2 C2f-DualConv 模块

水产养殖装备由于受高温、高湿、高盐环境的限制,大多采用计算资源受限的嵌入式设备,将目标识别算法直接部署到嵌入式设备中,可能会导致识别速度缓慢和精度降低等问题^[14]。针对普通卷积存在参数量和计算量大的问题,通过设计轻量级的卷积模块可以大幅度降低卷积神经网络的模型复杂度。如图 2 所示, DualConv 并行采用 3×3 和 1×1 卷积,可以同时提取输入特征图中的通道信息,并结合组卷积和异构卷积,通过降低模型复杂度和共享特征图中的全局信息来提升目标模型的识别精度^[15]。在输入特征图上采用 1×1 卷积操作可以保留特征图中的原始信息,便于深层卷积更好地提取特征并在卷积操作之间共享。 M 为输入通道数,即输入特征图的深度; N 为卷积滤波器数,也是输出通道数,即输出特征图的深度; G 为组卷积和双卷积中的组数。

表 1 预实验训练结果
Tab. 1 Pre-experimental training results

模型 Models	平均精度 $mAP_{50}/\%$	参数量 Params	内存占用量 Model size/MB	浮点运算量 FLOPs/GB	识别速度 FPS/(帧/s)
YOLOv8n	95.10	4.85×10^6	8.70	8.9	83
YOLOv8s	95.80	9.95×10^6	21.35	25.8	67
YOLOv8m	96.40	21.90×10^6	50.14	78.7	54
YOLOv8l	96.14	35.90×10^6	83.58	134.9	38
YOLOv8x	96.21	56.10×10^6	130.53	210.3	29

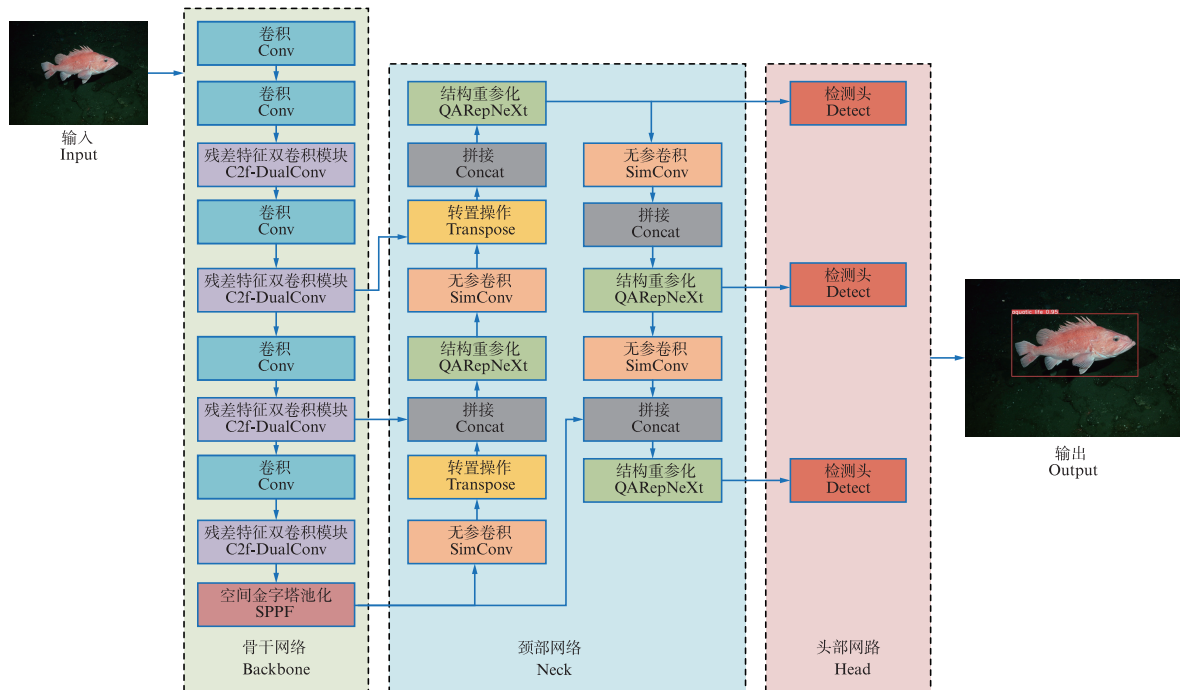


图1 改进后的YOLOv8n-fish网络架构图
 Fig. 1 Improved YOLOv8n-fish network frame structure

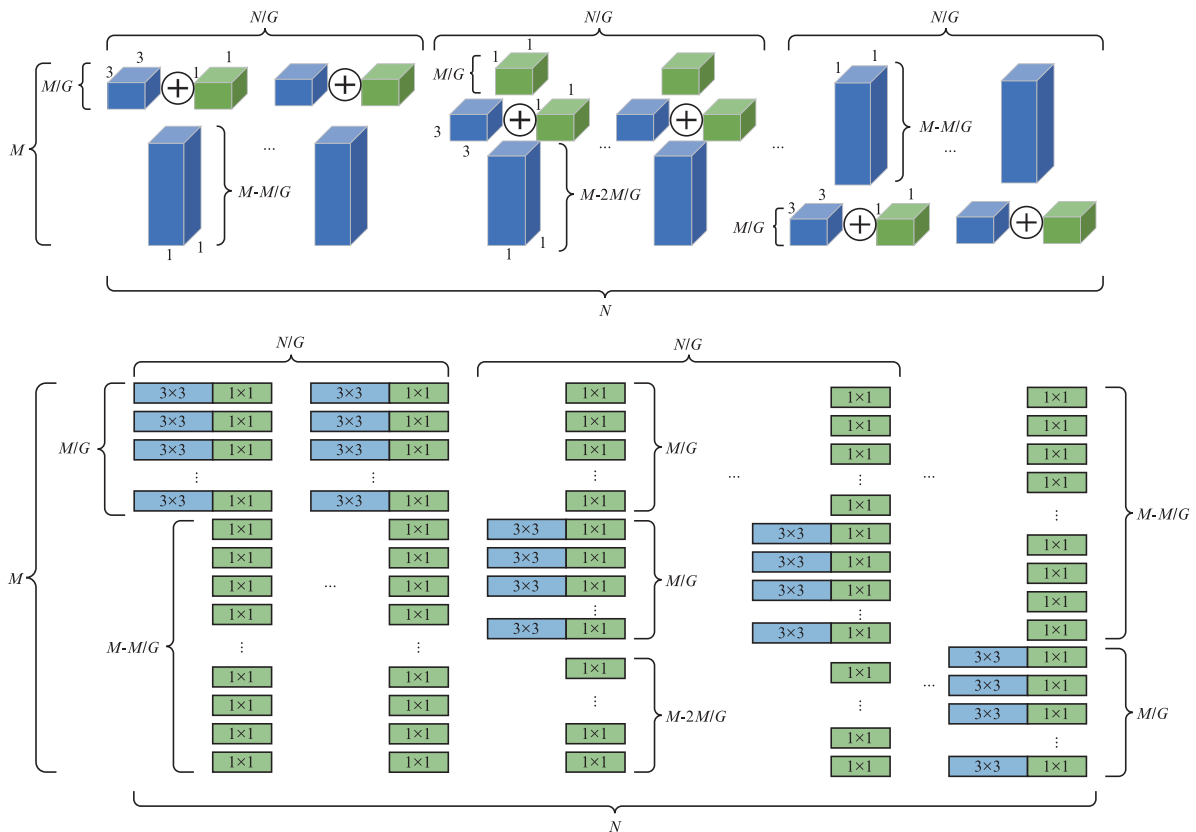


图2 DualConv架构图
 Fig. 2 DualConv frame structure

为降低模型复杂度,便于模型部署从而达到实时监测的目的,本研究基于C2f模块和

DualConv设计了C2f-DualConv模块,其架构如图3所示。

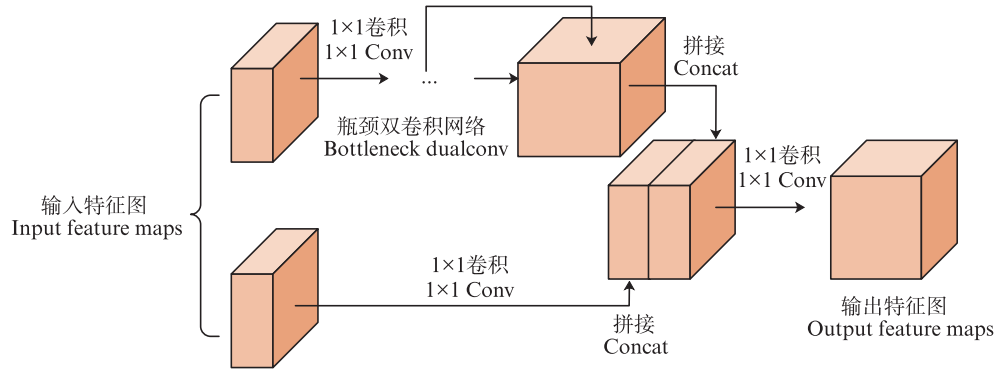


图3 C2f-DualConv 架构图

Fig. 3 C2f-DualConv frame structure

在YOLOv8中,C2f结构使用了分流操作,导致模型参数量大幅降低,相应地增加了小参数量的卷积次数^[16]。这种复杂的结构不利于硬件计算,其频繁的内存读取可能导致低浮点运算量(Floating point operations, FLOPs)但高延迟的情况。此外,随着网络深度的增加,C2f结构的首尾卷积缺少残差结构,会导致梯度消失,影响目标识别的精度。因此,引入跨阶段局部网络(Cross stage partial network, CSPNet)^[17]中的残差结构,提升模型的实时识别能力。C2f-DualConv模块首先沿通道维度将输入特征图划分为2个大小相同的特征图,并采用2个1×1卷积进行特征提取操作,以增加网络的梯度信息;其次,中间特征图输入到瓶颈网络结构中,将输入通道数降低50%,以降低50%的计算复杂度和参数量;然后在瓶颈网络结构中引入DualConv卷积降低模型复杂度和共享特征图中的全局信息;最后使用拼接操作对2个分支的特征图进行通道合并。

为了验证本研究提出的轻量化卷积神经网络模块的有效性,分别统计普通卷积和DualConv的计算量,根据公式(1)~(5),本研究设计的C2f-DualConv模块的 K 为3时,分组卷积的 G 值足够大,模型的浮点数计算量可下降为原模型的1/9。

$$F_s = D_o^2 \times K^2 \times M \times N \quad (1)$$

$$F_c = \frac{D_o^2 \times K^2 \times M \times N + D_o^2 \times M \times N}{G} \quad (2)$$

$$F_p = (D_o^2 \times M \times N) \times \left(1 - \frac{1}{G}\right) \quad (3)$$

$$F_d = D_o^2 \times K^2 \times M \times \frac{N}{G} + D_o^2 \times M \times N \quad (4)$$

$$R_{\frac{D}{S}} = \frac{1}{G} + \frac{1}{K^2} \quad (5)$$

式中: M 为特征图输入通道数; N 为特征图输出通

道数; K 为卷积核大小; F_s 为普通卷积的计算量; F_c 为组卷积的计算量; F_p 为逐点卷积的计算量; F_d 为DualConv的计算量; D_o^2 为输出特征图的大小,其中 D_o 为输出特征图的宽度和高度尺寸; G 为组卷积和双卷积中的组数; $R_{\frac{D}{S}}$ 为双卷积与标准卷积计算成本的缩减比。

1.3 EffQAFPN 网络

YOLOv8n模型的颈部网络采用路径聚合网络(Path aggregation network, PANet)结构,对比传统的特征金字塔结构(Feature pyramid network, FPN),PANet增加了一条自下而上的路径,可实现双向特征融合。该方法使得特征图同时具备较高的语义信息和位置信息,突破了单向信息流的局限,有效解决了FPN中浅层特征信息丢失严重的问题^[18]。但由于PANet存在多尺度特征的融合和上采样等操作,直接导致模型所需的计算量和内存占用骤升;并且对于复杂水下环境中的小目标和姿态不规则的鱼类而言,PANet在捕捉特征图的细节特征上仍存在一定的局限性。

因此,基于无参卷积(Simple conv, SimConv)、转置函数和结构重参数化(Quantization-aware repNeXt, QARepNeXt)功能设计出高效快速增量神经网络(Efficient fast and progressive neural network, EffQAFPN)^[19]并替换原模型的颈部网络,EffQAFPN的网络结构如图4所示。YOLOv8n的颈部网络在进行多次的上采样和下采样操作后会改变特征图的大小,进而出现特征信息丢失并需要较高的计算复杂度,导致模型在计算资源受限的水产养殖智能化装备上运行效率低。EffQAFPN网络在上采样过程中,采用SimConv处理输入特征图,通过批量归一化层(Batch normalization, BN)规范化输出中间特

征图,最后通过线性整流函数(Rectified linear unit, ReLU)激活函数引入非线性,利用 SimConv 可以有效地增加特征图中的空间信息和表达能力,并降低计算量。在 YOLOv8n 中,通过转置卷积操作,恢复特征图的尺寸大小,从而使得该模型能够处理更高分辨率的特征图,并提高识别精度和计算效率。EffQAFPN 网络较 PANet 结构可

以在训练与推理模式之间切换,结构重参数化 QARepNeXt 在训练过程中采用多个重参数卷积(Re-parameterizable convolution, RepConv)以及恒等变换操作进行特征提取和融合;在推理时,多分支被融合成一个卷积层,便于简化计算和提高识别速度。

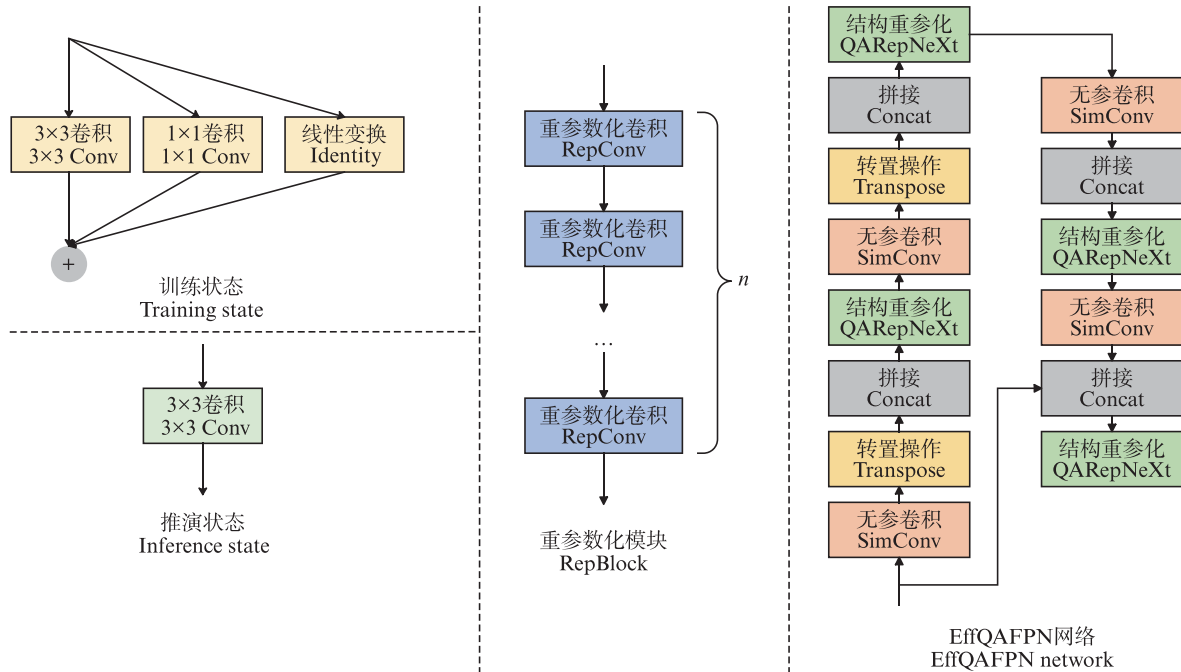


图4 EffQAFPN网络架构图
Fig. 4 EffQAFPN network frame structure

1.4 二阶段微调方法

模型预训练和微调是一种广泛使用的深度学习范式,可以解决训练数据不足的问题,并提高模型精度。在 YOLOv8n-fish 的训练阶段,采用 YOLOv8n 模型的部分预训练权重可以加速模型的收敛速度和精度^[20]。这是因为改进前后的模型在主干网络和输出端共享大量权重。然而,由于缺乏 YOLOv8n 的预训练权重值,改进后模型的部分卷积层需要随机初始化,并且较小的数据集会影响模型的收敛性。因此,最有效的方法是采用常见物体数据集(Common objects in context, COCO)、视觉对象分类数据集(Pascal visual object classes, VOC)等大型数据集对 YOLOv8n-fish 进行预训练,产生官方权重,然而在大型数据集上需要大量的训练时间和训练资

源,这对硬件条件产生更高要求。

如图5所示,为了在有限的硬件条件下,进一步提高模型精度,本研究提出了一种二阶段微调方法。在第一个微调阶段,采用 YOLOv8n 的官方权重和辅助数据集来产生第一阶段预训练权重^[21]。第一个微调阶段获取的预训练权重是对模型中所有层进行微调,进而获得最佳模型性能。第一个微调阶段使用的辅助数据集和第二个微调阶段使用的目标数据集具有一定的相似性,这可以有效地提升改进后模型在第二个微调阶段中学习鱼类图像的纹理、边缘和形状等细节特征。在第二个微调阶段中,利用第一阶段预训练权重和目标数据集对改进后的模型进行训练,并得到最终的目标模型权重。

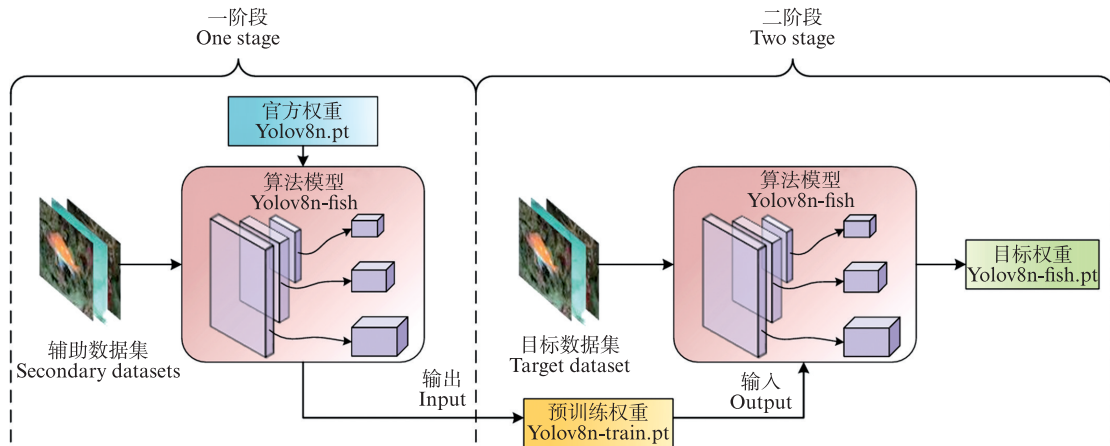


图5 二阶段微调结构图

Fig. 5 Structure of the two-stage fine-tuning

1.5 评价指标

针对识别速度采用指标帧率,帧率达到25帧/s就可以实时识别。模型的轻量化程度采用参数量、浮点运算量、模型内存占用量这3种指标来衡量。针对模型的识别精度采用3种评价指标^[22]:精确率(Precision, P)、召回率(Recall, R)、平均精度(Mean average precision, mAP₅₀),具体计算公式如下:

$$P = \frac{T_p}{T_p + F_p} \quad (6)$$

$$R = \frac{T_p}{F_p + F_N} \quad (7)$$

$$I_m = \sum_{i=1}^k \frac{\int_0^1 P(R) dR}{k} \quad (8)$$

式中:采用交并比(Intersection over union, IoU)评价目标检测算法的性能,通过计算预测边界框与真实边界框之间的重叠程度来衡量检测效果。 T_p 为IoU>0.5的检测框数量,即模型识别鱼类正确的数量; F_p 为IoU≤0.5的检测框数量,即模型识别鱼类错误的数量; F_N 为模型未识别到鱼类的数量; I_m 为模型平均精度,是包含P、R的关系函数; $P(R)$ 为dR时的精确率; k 为模型检测出目标的类别数。

2 识别结果与分析

2.1 训练环境与超参数设置

实验软硬件配置:GPU为NVIDIA GTX4070 Ti, 12 GB显存,操作系统为Win11,深度学习框架为Pytorch2.0.1, Python版本为3.11, CUDA版本为11.8, Cudnn版本为8.5.5。超参数配置:输

入图像大小为640像素×640像素,批处理为48,最大学习率为0.01,并逐渐下降,动量设置为0.937,优化器为随机梯度下降,训练轮数为500轮。

2.2 数据集

原始鱼类图像数据集^[23]是在水下遥控车辆(Remote operated vehicle, ROV)上的车载摄像机系统记录的不同水域的水下视频中获得的,其中一些高质量鱼类图像数据被用于研究鱼类识别和分割任务。由于控制ROV车辆与水下识别站相比更加容易受到水下生物的攻击和其他不可控因素干扰,因此该数据集中仅包含1 329幅高质量鱼类和其他图像。本研究选用原始鱼类图像数据集中的1 200幅鱼类图像作为二阶段微调方法中第一个微调阶段的辅助数据集,将其标注为aquatic life,再根据8:1:1的比例分为训练集、验证集和测试集^[24]。其中训练集、验证集和测试集分别为960、120和120幅水下鱼类图像。

为了防止两个微调阶段使用的数据集之间出现数据泄露问题,第二个微调阶段的目标数据集需区别于第一个阶段的辅助数据集。因此,对原始鱼类图像内的1 200幅石斑鱼图像进行重新标注为aquatic life,并从其他鱼类数据集(WildFish20)中补充400幅水下鱼类图像。对于目标数据集中鱼类的标注原则为降低标注区域内的背景占比,这是因为背景中的无用特征会影响模型的识别性能。经过重新标注后的数据集也被划分为训练集(80%)、验证集(10%)和测试集(10%)。为了进一步提升模型的泛化性能,随机对划分好的数据集进行数据增强(水平翻转、

垂直翻转、加适量噪声等),数据增强后的训练集、验证集和测试集分别为6 400、800和800幅水下鱼类图像。

2.3 二阶段微调实验对比实验

为了验证本研究提出的二阶段微调方法的有效性,分别使用原始鱼类图像官方数据集、目标数据集和二阶段微调方法在YOLOv8n模型上进行对比实验。如表2所示,经过对官方数据集重新标注和扩充后,模型的精确率P、召回率R和 mAP_{50} 分别提高了0.5%、0.7%和0.2%,这说明对官方数据集重新标注是一种正向优化。与单一

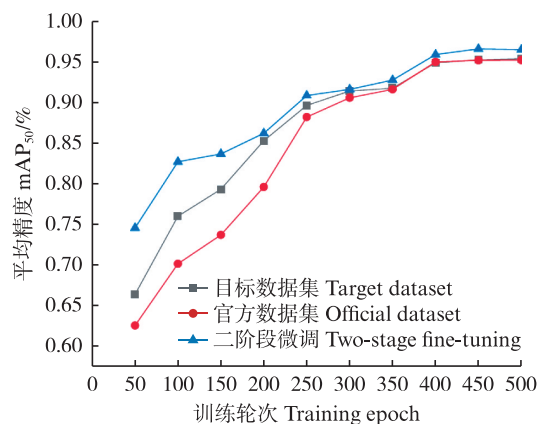
数据集的训练模型相比,通过二阶段微调方法训练出来的模型展现出更高的识别精度。其中, mAP_{50} 高达96.4%,较官方数据集和目标数据集分别提升1.3%和1.1%。由图6可知,本研究提出的二阶段微调方法较官方数据集和目标数据集具有更快的收敛速度。

2.4 模型消融实验

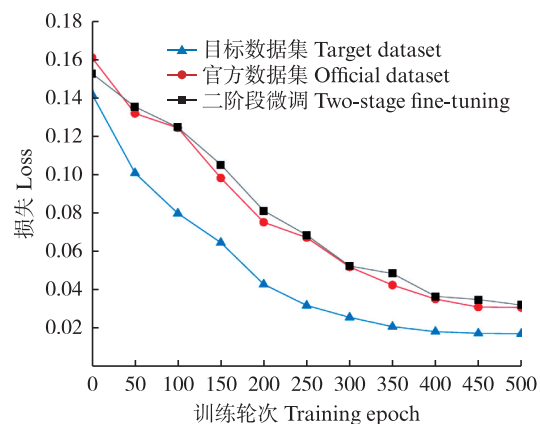
为了验证改进模型的性能,对每一个改进策略做消融实验并加以分析。在同一数据集、训练环境和超参数下进行并结合评价指标,消融实验结果见表3。

表2 二阶段微调实验结果
Tab. 2 Results of two-stage fine-tuning

数据集 Datasets	精确率 Precision/%	召回率 Recall/%	平均精度 $mAP_{50}/\%$
官方数据集 Target dataset	97.4	97.5	95.1
目标数据集 Official dataset	97.9	98.2	95.3
二阶段微调 Two-stage fine-tuning	98.6	98.5	96.4



(a) 二阶段微调对比实验精度曲线图
Accuracy plot of the two-stage fine-tuning comparison experiment



(b) 二阶段微调对比实验损失曲线图
Loss profile of the two-stage fine-tuning comparison experiment

图6 二阶段微调对比实验曲线图

Fig. 6 Two-stage fine-tuning comparison experiment curve graph

表3 消融实验结果
Tab. 3 Results of ablation experiments

模型 Models	参数量 Params	浮点运算量 FLOPs/GB	内存占用量 Model size/MB	平均精度 $mAP_{50}/\%$
YOLOv8n	4.85×10^6	8.9	8.7	96.4
YOLOv8n-C2f-DualConv	2.53×10^6	7.1	6.3	96.6
YOLOv8n-EffQAFPN	5.04×10^6	9.5	6.3	96.9
YOLOv8n-fish	2.72×10^6	7.3	5.8	97.47

由表3可知, YOLOv8n的颈部网络嵌入EffQAFPN结构在保证模型复杂度基本不变的情况下, mAP_{50} 提升了0.5%。在YOLOv8n上嵌入

EffQAFPN网络取得识别精度提升的因素是: EffQAFPN结构作为一种全新的网络,旨在学习特征图中的有用信息。因此,改进后的模型可以

在水下环境中学习到鱼类目标的重要特征并抑制背景这类无用特征的提取。虽然 EffQAFPN 结构是一种轻量级的模块,但是也会增加模型的参数量和浮点运算量。引入 C2f-DualConv 模块,这一策略使得模型的识别精度有所上升的同时,模型的参数量、浮点运算量和模型内存占用量大幅度下降。

由综合消融实验可知,本研究采用的 EffQAFPN 结构和 C2f-DualConv 模块这两处改进对模型的识别精度和轻量化产生积极作用。YOLOv8n-fish 较 YOLOv8n 在 mAP_{50} 提升 1.07%,模型识别精度高达 97.47%;在模型复杂度方面,参数量、浮点运算量和模型内存占用量分别仅为 YOLOv8n 的 56.1%、82% 和 66.7%。

2.5 模型改进前后目标识别结果对比

为了验证改进后模型面对复杂场景下的鱼类目标识别能力,随机选取目标数据集中的 6 幅测试集图像进行性能对比验证。图 7 为 YOLOv8n 和 YOLOv8n-fish 分别对复杂弱光及干扰物场景下的鱼类识别结果。其中,图 7 的第 1 列图是 YOLOv8n 的识别结果,第 2 列是 YOLOv8n-fish 的识别结果,红色数字代表置信度分数,置信度分数通常通过模型的输出层生成,表示模型对特定类别的预测置信度。从识别结果可知,YOLOv8n 对复杂弱光及干扰物场景的鱼类识别存在明显的 4 处漏检和 3 处误检(漏检用黄色的框标注,误检用白色的框标注),预测框的置信度分数最高为图 7-11 中的 0.95,但是模型对小目标的预测置信度分数不高。图 7 中 YOLOv8n-fish 对复杂弱光及干扰物场景下的目标鱼类的平均置信度分数高于 0.9,预测框可以准确地定位水下鱼类大小目标。这是因为 YOLOv8n-fish 中的 DualConv 结合组卷积和异构卷积的优势,在降低模型复杂度的基础上,使得模型在原始特征图上学习到全局信息。本研究基于跨阶段局部网络(Cross stage partial network, CSPNet)的多路径残差设计出的 C2f-DualConv 模块可以使得模型在学习额外的梯度信息的基础上,加强多级特征融合能力。此外,在颈部网络和头部网络之间引入 EffQAFPN 模块,通过改善特征图中有用鱼类特征

的传播能力,使得模型更加关注鱼类的位置和形状,改善了干扰物的影响,提升了 YOLOv8n-fish 在不同场景下的鱼类识别能力。综上识别可视化对比结果可知,YOLOv8n-fish 模型提升了识别效果,能够适应不同复杂水下场景的鱼类目标识别任务。

2.6 不同模型对比实验

为了进一步验证 YOLOv8n-fish 的先进性,本研究选取 SSD、YOLOv5n-P6、YOLOv6n、YOLOv8n、YOLOv7-tiny 和 Faster R-CNN 算法进行对比实验。对比实验在同一数据集、训练环境和超参数下进行,对比实验结果如表 4 所示。Faster R-CNN 作为一种经典的二阶段目标识别算法,其识别速度最慢,模型复杂度最高,无法适用于复杂水下环境的鱼类识别任务,并且很难部署在资源受限的水产养殖智能化设备上;SSD 算法和 YOLO 系列算法同为优秀的一阶段目标识别算法,识别速度和模型复杂度较 Faster R-CNN 有一定程度的提升。然而 SSD 的计算量和内存占用量分别为 273.74 GB 和 90.6 MB,同样很难部署在移动端设备上。值得注意的是,YOLOv8n-Fish 算法在平均识别精度上表现最佳,比 SSD、YOLOv5n-P6、YOLOv6n、YOLOv8n、YOLOv7-tiny 和 Faster R-CNN 算法分别高出 18.5%、5.87%、3.95%、2.37%、12.28% 和 24.13%。此外,YOLOv8n-Fish 的模型轻量化程度最高,其参数量、浮点运算量和内存占用量大小分别为 2.72×10^6 、7.3 GB 和 5.8 MB,由于 C2f-DualConv 模块通过结合 3×3 卷积、 1×1 异构卷积和多条特征图路径以共享特征图中的全局信息,同时全新的二阶段微调方法给改进后的模型提供了大量且有用的特征信息。改进后模型的识别速度为 121 帧/s,仅落后 YOLOv5n-P6 的识别速度。这直接证明了本研究设计的 EffQAFPN 网络可以使得改进后模型学习到鱼类有用特征的深层空间信息,在推理阶段通过将多分支融合成一个卷积层来提高识别精度。实验结果表明,YOLOv8n-fish 模型在面对不同复杂水下场景的识别精度最佳的同时,模型复杂度最低,识别速度满足实时识别,适用于复杂水下场景下的鱼类识别任务。

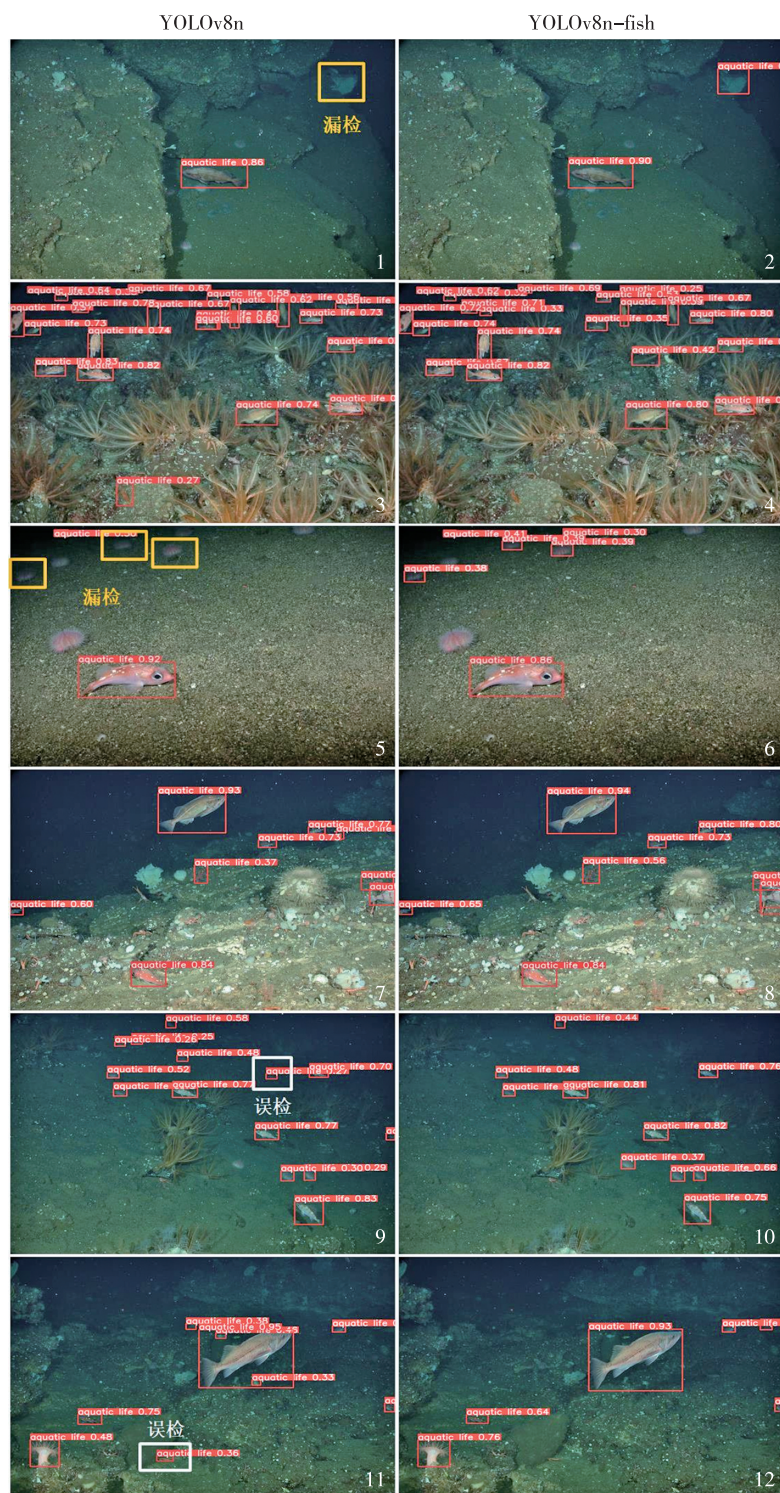


图7 弱光及干扰物场景下模型改进对比图

Fig. 7 Comparison of model improvement in low light and interference scene

表 4 先进模型对比结果
Tab. 4 Comparison of advanced models results

模型 Models	参数量 Params	浮点运算量 FLOPs/GB	内存占用量 Model size/MB	平均精度 mAP ₅₀ /%	识别速度 FPS/(帧/s)
YOLOv5n-P6	3.85×10^6	7.1	8.0	91.60	142
YOLOv6n	4.50×10^6	13.1	14.4	93.52	114
Faster R-CNN	28.05×10^6	947.3	108.0	73.34	12
YOLOv8n	4.85×10^6	8.9	6.0	95.10	83
YOLOv8n-fish	2.72×10^6	7.3	5.8	97.47	121
SSD	23.61×10^6	273.7	90.6	78.97	39
YOLOv7-tiny	6.02×10^6	13.0	11.5	85.19	88

为了更加直接验证 YOLOv8n-fish 模型在水下鱼类目标识别任务中的先进性,从目标数据集选取低光照、背景多干扰物和密集目标等复杂水下场景,并选取其他主流目标识别模型进行可视化对比。不同复杂水下场景各目标识别模型的模型可视化结果如图 8 所示。由结果可知, YOLOv8n-fish 模型的置信度分数优于其他模型,最高达到 0.90。在面对低光照和卵石等类似背景干扰物的情况下,略微出现漏检误检的情况。YOLOv5n-P6 模型虽然识别置信度分数较高,但存在漏检误检的情况,总共存在 3 处误检和 7 处漏检(漏检用黄色的框标注,误检用白色的框标注),相比而言 YOLOv5n-P6 判别相似特征的能力较差,未能有效地学习水下鱼类特征图中的多尺度特征。由图 8 可知,目前主流的一阶段和二阶段目标识别模型在面对复杂水下场景中存在不同程度的漏检和误检情况。Faster R-CNN 的鱼类识别共存在 4 处误检和 6 处漏检,而 YOLOv7-tiny 和 SSD 存在多处的误检和漏检问题,说明它们易受类似鱼类的背景干扰物影响,造成模型的识别精度下降。YOLOv7-tiny 虽然可以较准确的定位水下鱼类位置,但置信度分数不高,最高为 0.85,并且容易将类似背景特征识别为鱼类,其对复杂水下环境中鱼类特征的学习能力较差。YOLOv8n-fish 模型对于不同复杂水下场景下的鱼类目标识别置信度分数优于其他模型,精度在 0.95 以上。本研究提出更强的特征学习模块 C2f-DualConv 和多尺度特征融合网络 EffQAFPN,在此基础上,采用二阶段微调方法使得模型充分学习到不同复杂场景下的有用特征

信息。因此,改进后模型在面对低光照和卵石等类似背景干扰物的情况下,未出现漏检误检的情况。识别结果可视化实验充分验证了 YOLOv8n-fish 模型在鱼类目标识别任务中的有效性和先进性。

3 结论

为增强复杂水下弱光及干扰物环境下鱼类目标的快速准确识别,本研究基于传统 YOLOv8n 模型进行优化,提出轻量化模块 C2f-DualConv 对 YOLOv8n 的特征提取网络进行改进,并设计 EffQAFPN 网络和二阶段微调方法,旨在实现模型轻量化程度的同时,提高其在复杂水下弱光及多干扰物场景的鱼类识别精度。通过实验验证了其性能:(1)为降低模型复杂度和提升模型识别速度,基于 DualConv 和 C2f 模块设计了 C2f-DualConv;设计全新的颈部网络 EffQAFPN 以优化模型的鱼类有用特征学习能力;对比改进前后的模型, YOLOv8n-Fish 的平均精度提升了 1.37%,且参数量、浮点运算量和内存占用量大小分别为 2.72×10^6 、7.3 GB 和 5.8 MB。(2)通过对比实验和分析, YOLOv8n-Fish 在识别精度和模型轻量化上均优于 SSD、Faster R-CNN 和 YOLO 系列,识别速度达到 121 帧/s,较 YOLOv5n-P6 仅下降 21 帧/s。因此,本研究设计的水下鱼类模型易于部署在水产养殖智能化设备终端上。(3)为提升模型在复杂环境下的鱼类目标精度,提出一种二阶段微调方法。实验证明,该二阶段微调方法在不增加模型复杂度的前提下可有效提升目标模型识别精度。



图8 多种模型推理图

Fig. 8 Multi-model inference diagram

作者声明本文无利益冲突。

参考文献:

- [1] 农业农村部渔业渔政管理局, 全国水产技术推广总站, 中国水产学会. 中国渔业统计年鉴 2022[M]. 北京: 中国农业出版社, 2022.
- Fishery and Fishery Administration Bureau of the Ministry of Agriculture and Rural Affairs, National Fisheries Technology Extension Center, China Society of Fisheries. China fishery statistical yearbook 2022 [M].

Beijing: China Agriculture Press, 2022.

- [2] 刘世晶, 李国栋, 刘晃, 等. 中国水产养殖装备发展现状[J]. 水产学报, 2023, 47(11): 119615.
- LIU S J, LI G D, LIU H, et al. Current development status of aquaculture equipment in China[J]. Journal of Fisheries of China, 2023, 47(11): 119615.
- [3] 徐皓, 刘晃, 黄文超. 深远海设施养殖装备技术进展与展望[J]. 上海海洋大学学报, 2023, 32(5): 893-902.
- XU H, LIU H, HUANG W C. Advances and outlook of offshore aquaculture equipment technology[J]. Journal of Shanghai Ocean University, 2023, 32(5): 893-902.

- [4] 矫萌璐, 张海燕, 李欣. 基于Faster-RCNN的海洋牧场鱼类识别与分类[J]. 海洋湖沼通报, 2024, 46(3): 65-72.
JIAO M L, ZHANG H Y, LI X. Identification and classification of marine ranching fish based on the faster-RCNN [J]. Transactions of Oceanology and Limnology, 2024, 46(3): 65-72.
- [5] 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(s2): 204-209.
XIONG H T, LIN Q, XUAN K, 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).
- [6] 梅海彬, 黄政, 袁红春. 基于YOLOv7模型改进的轻量级鱼类目标检测方法[J]. 大连海洋大学学报, 2023, 38(6): 1032-1043.
MEI H B, HUANG Z, YUAN H C. A lightweight fish object detection method improved based on the YOLOv7 model[J]. Journal of Dalian Ocean University, 2023, 38(6): 1032-1043.
- [7] LIU Y, AN D, REN Y J, et al. DP-FishNet: dual-path pyramid vision transformer-based underwater fish detection network[J]. Expert Systems with Applications, 2024, 238: 122018.
- [8] ZHOU Z Y, HU Y J, YANG X F, et al. YOLO-based marine organism detection using two-terminal attention mechanism and difficult-sample resampling[J]. Applied Soft Computing, 2024, 153: 111291.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.
- [10] REDMON J, FARHADI A. YOLOv3: an incremental improvement [J]. arXiv preprint arXiv: 1804.02767, 2018: 1-6.
- [11] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [J]. arXiv preprint arXiv: 2004.10934, 2022.
- [12] LI C 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.
- [13] 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 [C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 7464-7475.
- [14] 岳凯, 张鹏超, 王磊, 等. 基于改进 YOLOv8n 的复杂环境下柑橘识别[J]. 农业工程学报, 2024, 40(8): 152-158.
YUE K, ZHANG P C, WANG L, et al. Recognizing citrus in complex environment using improved YOLOv8n [J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(8): 152-158.
- [15] ZHONG J C, CHEN J Y, 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.
- [16] 谭厚森, 马文宏, 田原, 等. 基于改进 YOLOv8n 的香梨目标检测方法[J]. 农业工程学报, 2024, 40(11): 178-185.
TAN H S, MA W H, TIAN Y, et al. Improved YOLOv8n object detection of fragrant pears [J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(11): 178-185.
- [17] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE, 2020: 1571-1580.
- [18] 肖进胜, 张舒豪, 陈云华, 等. 双向特征融合与特征选择的遥感影像目标检测[J]. 电子学报, 2022, 50(2): 267-72.
XIAO J S, ZHANG S H, CHENG Y H, et al. Remote sensing image object detection based on bidirectional feature fusion and feature selection [J]. Acta Electronica Sinica, 2022, 50(2): 267-272.
- [19] GENG X, SU Y, CAO X, et al. YOLOFM: an improved fire and smoke object detection algorithm based on YOLOv5n [J]. Scientific Reports, 2024, 14 (1): 4543.
- [20] FU H, CHENG G Q. Convolutional neural network based efficient detector for multicrystalline photovoltaic cells defect detection [J]. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2023, 45(3): 8686-8702.
- [21] WEN Y, CHEN L T, DENG Y, et al. Rethinking pre-training on medical imaging [J]. Journal of Visual Communication and Image Representation, 2021, 78: 103145.
- [22] 周鑫, 何晓新, 郑昌文. 基于图像深度学习的无线电信号识别[J]. 通信学报, 2019, 40(7): 114-25.
ZHOU X, HE X X, ZHEN C W. Radio signal recognition based on image deep learning [J]. Journal on Communications, 2019, 40(7): 114-125.
- [23] CUTTER G, STIERHOFF K, ZENG J M. Automated detection of rockfish in unconstrained underwater videos using haar cascades and a new image dataset: labeled fishes in the wild [C]//2015 IEEE Winter Applications and Computer Vision Workshops. Waikoloa: IEEE, 2015: 57-62.
- [24] LI Y T, FAN Q S, HUANG H S, et al. A modified YOLOv8 detection network for UAV aerial image recognition [J]. Drones, 2023, 7(5): 304.

An improved YOLOv8n based on lightweight model for underwater fish target detection

CAO Yu, LI Jiayang, WANG Fang

(College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China)

Abstract: In order to solve the problems of insufficient accuracy of fish target recognition in complex underwater environment, high complexity and slow recognition speed of traditional target recognition models, an underwater fish target recognition model based on improved YOLOv8n-fish was proposed. In this paper, a lightweight double-convolutional module C2f-DualConv is proposed to improve the feature learning ability of the C2f module in YOLOv8n. Based on the idea of efficient structure reparameterization, a new neck network EffQAFPN was designed to balance the recognition accuracy and speed of the target model. A two-stage fine-tuning method was used to improve the recognition accuracy of the fish target recognition model in the underwater low light and interference environment. The experimental results show that the average accuracy of the YOLOv8n-fish model in the test set is 97.47%, which is 1.07% higher than that of the traditional YOLOv8n model. However, the number of parameters, floating-point arithmetic and model memory occupation of the improved model are only 56.1%, 82% and 66.7% of the original model. The recognition speed of the YOLOv8n-fish model is second only to that of YOLOv5n-P6, which can reach 121 f/s. Experimental results show that the YOLOv8n-fish model can significantly reduce the computational cost while maintaining high recognition accuracy, and provide effective technical support for intelligent monitoring of aquaculture.

Key words: aquaculture; YOLOv8n; dual convolutional model; lightweight; deep learning