

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

袁红春 李春桥

(上海海洋大学信息学院, 上海 201306)

摘要: 水下生物目标检测是实现水下机器人自动化捕捞的关键性技术。针对水下生物目标检测任务中存在的目标重叠、遮挡以及目标尺度小而导致的误检、漏检等问题,提出了一种基于改进 YOLO v8n 的水下生物目标检测算法 FDC - YOLO v8。首先,在主干网络中使用融合可变形卷积网络的 FDC 模块,以增强模型特征提取能力,提升其提取特征的丰富度。其次,引入融合分数阶傅里叶变换和空间注意力机制的 FrSAConv 模块,进一步分离多样目标特征,增强模型对多种特征的感知能力。最后,引入 Wise - IoU 损失函数作为模型边界框损失函数,以更好地解决目标不平衡以及尺度差异的问题。使用 RUIE 数据集进行实验,水下生物包括海胆、海星、海参、扇贝。实验结果表明,改进后的 FDC - YOLO v8 的平均精度均值达到 85.3%,较基准模型提升 2.6 个百分点,推理速度达到 769 f/s,在目标重叠、遮挡以及小尺度目标的水下生物目标检测中有更好的表现。

关键词: 水下生物识别; 目标检测; YOLO v8n; Wise - IoU; 可变形卷积网络; 分数阶傅里叶变换

中图分类号: TP393 文献标识码: A 文章编号: 1000-1298(2024)11-0140-07

OSID:



Research on FDC - YOLO v8 Underwater Biological Object Detection Method Improved by Deformable Convolution

YUAN Hongchun LI Chunqiao

(School of Information, Shanghai Ocean University, Shanghai 201306, China)

Abstract: Underwater biological target detection is a crucial technology for achieving automation in underwater robotic fishing. Aiming to address issues such as object overlap, occlusion, and false detections, missed detections caused by small object scales in underwater biological object detection tasks, an underwater biological object detection algorithm, FDC - YOLO v8 was proposed based on an improved YOLO v8n. Firstly, the FDC module was incorporated, which utilized deformable convolution networks in the backbone network to enhance the model's feature extraction capability and enrich the diversity of extracted features. Secondly, the FrSAConv module, integrating fractional Fourier transform and spatial attention mechanism, was introduced to further separate diverse object features and enhance the model's perceptual ability towards various features. Finally, the Wise - IoU loss function was introduced as the bounding box loss function to better address issues related to object imbalance and scale differences. The experiments were conducted by using the RUIE dataset, which included four types of underwater organisms: echinus, starfish, holothurian, and scallops. Experimental results demonstrated that the improved FDC - YOLO v8 achieved an mAP of 85.3%, a 2.6 percentage points improvement over the baseline model. The inference speed can reach 769 frames per second, showcasing better performance in underwater object detection of marine organisms with challenged such as object overlap, occlusion, and small-scale objects.

Key words: underwater biological recognition; object detection; YOLO v8n; Wise - IoU; deformable convolutional network; fractional Fourier transform

收稿日期: 2024-01-31 修回日期: 2024-03-07

基金项目: 国家自然科学基金项目(41776142)

作者简介: 袁红春(1971—),男,教授,博士,主要从事人工智能应用等研究, E-mail: hcyuan@shou.edu.cn

0 引言

水下生物目标检测是水产养殖智能化战略的重要组成部分^[1],是实现水下机器人自动化捕捞的关键技术^[2-3],同时也可以为水产养殖生物计数^[4]做技术支持。然而,水下环境的复杂性,如光照变化、散射和吸收等,以及水下图像采集的不便性对水下生物目标检测提出了重大挑战^[5]。为此,本文致力于开发一种高效且鲁棒的水下目标检测模型,实现对水下生物的准确识别。

传统的目标检测方法^[6-7]主要依赖于区域选择、手动特征提取和分类器分类,此类方法手工提取特征的泛化能力较差、结果误差较大,具有一定局限性。随着深度学习的兴起,神经网络展现出强大的特征提取和拟合能力。基于深度学习的目标检测方法可分为双阶段目标检测和单阶段目标检测。其中双阶段目标检测以 R-CNN^[8]、Fast R-CNN^[9] 和 Faster R-CNN^[10] 为代表,将检测任务分为候选区域提取和分类两个步骤。LIU 等^[11]在候选区域提取阶段使用 Swin Transformer 作为主干网络,利用路径聚合网络(Path aggregation network, PAN)融合深、浅层特征图,提升特征提取能力,同时使用在线难例挖掘平衡样本,使用感兴趣区域对齐替换感兴趣区域池化,然后实现水下生物的目标检测。SONG 等^[12]使用 MSRCR 算法对 90 种水下生物数据进行增强,并结合经过预训练的 Mask R-CNN 算法,通过迁移学习学习水下生物特征以进行检测。这些双阶段水下目标检测算法虽然效果得到了提升,但是其中的候选区域提取阶段消耗较高时间成本,且仍需进行目标分类和精确定位,导致目标检测速度较慢。单阶段目标检测则以 YOLO^[13-17] 和 SSD^[18] 为代表,直接在输入图像上得到目标的位置坐标和类别概率。SUNG 等^[19]基于 YOLO v1 模型^[13],融合卷积网络,对鱼类图像和随机剪裁的水下图像进行训练,虽然取得 93% 的精度,但检测速度只有 16.7 f/s,不能满足实时检测需求。冯娟等^[20]基于改进的 YOLO v4 模型,运用 MobileNet v2 和深度可分离卷积块替换原网络结构、结合多尺度视网膜与色彩恢复增强算法以及 K-means++ 聚类优化锚点框尺寸等方法,解决了水下复杂环境下单环刺螠洞口的实时准确识别问题。HU 等^[21]提出一种融合多方向边缘检测的 SSD 算法,增强了对海胆多刺边缘的特征提取。CHEN 等^[22]在 DSSD^[23] 的基础上提出了 SWIPENet 算法,该算法利用多个超特征图来提升小目标检测,并引入了一个样本加权损失函数来降低噪声影响。李健源等^[24]提出了一种基于

YOLO v5s 的改进模型模型,将 Swin Transformer (STR)作为检测层来解决水下图像模糊的问题,利用 Efficient channel attention(ECA)减少参数量同时提升精度,并引入 FIOU 损失函数来优化性能。翟先一等^[25]在 YOLO v5 算法的基础上提出了一种海参识别的目标检测算法,首先使用带色彩恢复的多尺度视网膜增强算法对图像进行预处理,然后引入了卷积注意力机制,并增加了检测头,以增强对小尺度目标的检测能力。郑荣才等^[26]基于改进的 YOLO v7 模型,运用 Stem 模块、MRep 模块和 CBAM 模块,在轻量化模型的同时,提升对鲑鱼的检测精度。水下生物目标检测数据集存在两方面问题:由于水下环境比较复杂,一些较为聚集的水产生物会存在目标重叠、遮挡等问题。水下设备拍摄距离和拍摄角度,会导致数据形成以小尺度目标为主的多尺度情况,同时会存在一些图像样本具有模糊、重影等质量不高的问题。上述单阶段目标检测算法,都是使用标准卷积操作进行特征提取,卷积单元对输入的特征图在固定位置进行采样,限制了模型对多尺度、不同方向和位置信息的适应性,检测结果仍存在误检、漏检现象。

为此,本文基于 YOLO v8n 模型进行改进,在主干网络(Backbone)中使用融合可变形卷积(Fusion deformable convolutional, FDC)特征提取模块来提高模型的特征提取能力。在模型主干网络和颈部网络(Neck)之间,使用 FrSACConv 模块分离不同类型目标特征,以期进一步增强模型对多样特征的感知能力。最后引入新的边界框损失函数 Wise-IoU^[27]以更好地解决目标不平衡以及尺度差异等问题。

1 基于可变形卷积改进的 YOLO v8 水下生物目标检测算法

YOLO v8 分为 YOLO v8n、YOLO v8s、YOLO v8m、YOLO v8l、YOLO v8x 共 5 个规模的模型,其参数量和计算量随之增大。考虑到算法精度、检测速度等因素,本文选择 YOLO v8n 作为本文的基准模型。主要改进点有:使用融合可变形卷积网络的 DCNv2_C2f 模块替换主干网络中的后 3 个 C2f 模块,使其能够学习到更丰富的图像特征;引入分数阶傅里叶变换和空间注意力机制相融合的特征增强模块,将其添加到颈部网络的初始位置,用于更好地分离不同尺度目标特征;引入 Wise-IoU 作为模型的边界框损失函数,提高其回归精度,减少误检、漏检。改进后的 FDC-YOLO v8 模型结构如图 1 所示。

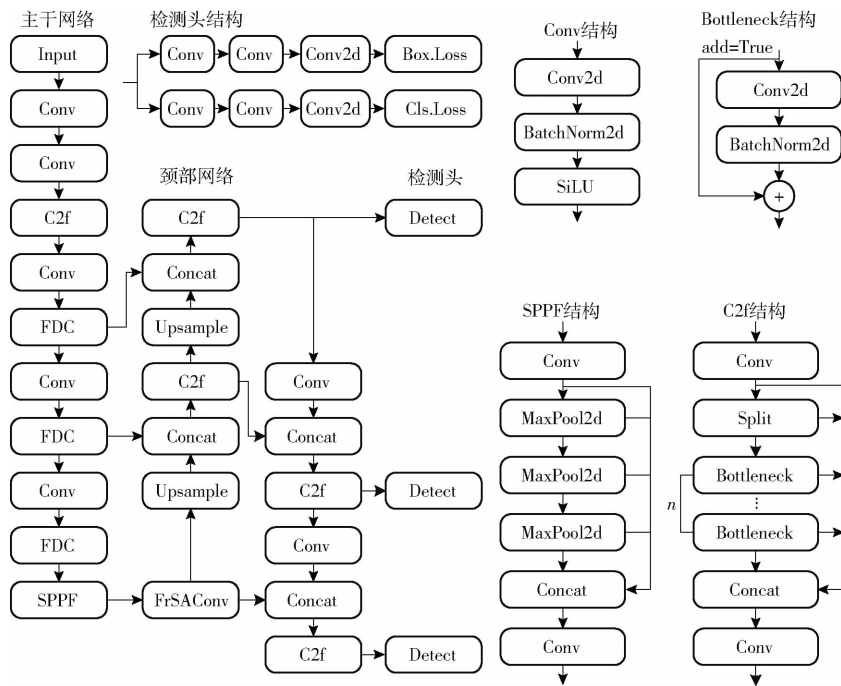


图1 FDC-YOLO v8 模型结构图

Fig. 1 FDC-YOLO v8 model structure diagram

1.1 YOLO v8 模型

YOLO v8 模型由主干网络、颈部网络和检测头组成。在主干网络中, YOLO v8 结合了 YOLO v5 的理念, 并从 YOLO v7^[17] 的 ELAN 结构中汲取灵感, 引入了 C2f 模块。该模块及其内部的 Bottleneck 模块均学习了残差特征, 丰富了梯度流, 特征提取能力得到了进一步提升。颈部网络则通过 PAN 结构^[28] 来增强网络对不同缩放尺度对象特征融合的能力。该模型检测头采用目前主流的解耦头结构 (Decoupled head), 将分类和目标检测的头部分离, 同时弃用 Anchor-Based, 使用 Anchor-Free 思想。YOLO v8 主要采用 TaskAlignedAssigner 方法^[29], 根据分类回归分数加权结果进行正负样本匹配。其 Loss 计算中使用 BCE Loss 作为分类损失, 使用 DFL (Distribution focal loss)^[30] 和 CIoU (Complete intersection over union)^[31] 作为回归损失。

1.2 FDC 特征提取模块

YOLO v8 使用的 C2f 模块的核心在于通过堆叠 Bottleneck 结构, 来实现更有效的特征提取。但水下生物图像数据存在物体重叠、遮挡、形状不规则以及小尺度等特点。以海胆为例, 其边缘多毛刺, 且数量分布密集。C2f 模块中的标准卷积核只能获取固定范围内的图像特征, 使得模型在提取多尺度、不同方向和位置的特征时具有一定的局限性, 难以有效提取海胆边缘特征。为解决这一问题, 本文引入融合可变形卷积 DCNv2 的 FDC 特征提取模块。

DCNv2 相对于标准卷积, 其卷积核相同, 但在

卷积操作时对采样点位置进行偏移操作。以图 2 中海胆为例, 假设卷积核为单通道 3×3 , 图 2 中标准卷积图白色的点对应的位置为标准卷积核及其采样点的对应位置, 标准卷积是将卷积核对应位置采样点进行加权求和, 只能提取卷积核范围内的像素特征, 需要多次卷积才能提取到海胆边缘特征。而 DCNv2 则是对采样点的位置添加一个偏移量 Δp_n 来对采样点进行偏移, 同时添加一个权重系数 Δm_n , 来控制采样点的关注度, 去除无关的上下文信息。其公式为

$$y(\mathbf{p}_0) = \sum_{\mathbf{p}_n \in \mathbf{R}} w(\mathbf{p}_n) x(\mathbf{p}_0 + \mathbf{p}_n + \Delta \mathbf{p}_n) \Delta m_n \quad (1)$$

式中 \mathbf{p}_0 ——卷积核移动后对应采样区中心位置坐标

\mathbf{p}_n ——卷积核中其他采样点与中心采样点相对位置

$x(\mathbf{p}_0 + \mathbf{p}_n + \Delta \mathbf{p}_n)$ ——取出该坐标采样点函数

$w(\mathbf{p}_n)$ ——卷积核中 \mathbf{p}_n 位置对应权重

图 2 中可变形卷积图白色的点为卷积核, 橙色的点为偏移后的采样点位置。DCNv2 的变形操作, 通过学习来进一步调整采样位置, 从而使模型能够

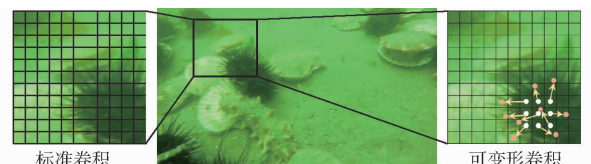


图2 标准卷积与可变形卷积对比

Fig. 2 Standard convolution and deformable convolution

自适应地提取目标特征。

本文提出的 FDC 特征提取模块融合 DCNv2 的可变形原理,通过自适应地学习采样点的偏移位置,以提高卷积操作对目标特征的提取效能。通过深度学习获取的自适应位置特征信息,FDC 能够更灵活地适应数据特征分布。与标准卷积相比,FDC 具有较快的特征提取能力。FDC 特征提取模块的设计采用堆叠的 FDCB 结构,该结构使用残差结构将输入和提取到的特征进行融合。在特征提取的过程中,FDCB 结构兼顾标准卷积和 DCNv2 的特性,使得该模块在不同尺度、方向和位置上能更全面地提取特征,同时可以避免特征丢失。FDC 模块使用残差结构将每层提取到的特征都在最后一层进行特征融合,进而得到不同层次的细节特征,从而提高模型特征提取的准确性。FDC 模块和 FDCB 结构框图如图 3 所示。

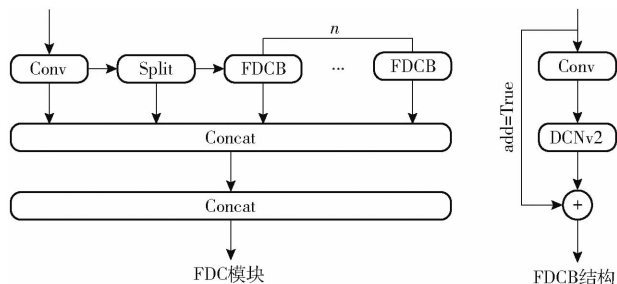


图 3 FDC 模块和 FDCB 结构框图

Fig. 3 FDC module and FDCB structure diagram

1.3 FrSAConv 特征增强模块

经过 FDC 特征提取模块,模型成功提取出不同尺度、方向和位置上的丰富特征,并且这些特征被融合在一起。尽管模型在进行检测和识别任务时能够从中获取不同特征,但其对这些特征的感知能力仍显不足。为了使模型更充分地利用这些特征,本文引入 FrSAConv 特征增强模块,以提高模型对多样特征的感知能力。

FrSAConv 引入分级阶傅里叶变换 (Fractional Fourier transform, FrFT) 算法。通过 FrFT,模型可以将时域信息转换为频域信息,将图像特征进行转换,从而将不同类型的目标特征从频域信息中更好地分离出来。随后 FrSAConv 将分离出来的特征与常规卷积提取到的特征进行融合,并利用空间注意力机制 (Spatial attention, SA) 对融合后的多尺度特征进行聚焦,然后进行卷积操作以及 softmax 归一化处理。为了避免过多卷积可能导致梯度消失,将归一化结果和原先提取特征结合并通过卷积操作之后,与初始输入信息进行融合,组成残差结构。FrSAConv 结构如图 4 所示。

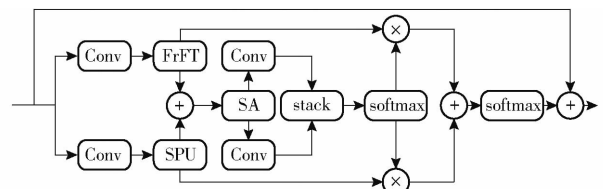


图 4 FrSAConv 结构框图

Fig. 4 FrSAConv structure diagram

1.4 Wise-IoU 损失函数

损失函数通过衡量真实值与预测结果的差异来加速网络收敛,提高网络精度。目标检测领域,模型通常使用预测框与真实框之间的交并比 (Intersection over union, IoU) 来衡量损失,是目标检测损失函数的重要组成部分。YOLO v8n 模型采用 CIoU 作为边界框损失函数,该损失函数把预测框与目标框的重叠面积、中心点距离、长宽比等影响因素都融入到损失计算中,但 CIoU 是以高质量的锚框标注为前提,对于手工标注中的一些质量较低的锚框仍然不能很好地收敛。

为了解决这一问题,采用 Wise-IoUv2^[27] 作为检测框回归损失函数。Wise-IoUv1^[27] 根据距离度量构建距离注意力,得到具有两层注意力机制的 Wise-IoUv1。而 Wise-IoUv2 是参照 Focal Loss 设计的一种针对交叉熵的单调聚焦机制的损失函数,有效降低了简单示例对损失值的贡献。使得模型能够聚焦于困难示例,提升了分类性能。其公式为

$$L_{WIoUv2} = L_{IoU}^* \gamma L_{WIoUv1} \quad (\gamma > 0) \quad (2)$$

式中 L_{WIoUv1} —— Wise-IoUv1 的值

γ —— 指数幂

L_{IoU}^* —— 单调聚集系数

梯度增益 L_{IoU}^* 在模型训练中会随着 L_{IoU} 的减小而减小,进而导致收敛速度在后期变慢。因此引入 L_{IoU} 的均值作为归一化因子。引入后公式为

$$L_{WIoUv2} = \left(\frac{L_{IoU}}{\overline{L_{IoU}}} \right)^\gamma L_{WIoUv1} \quad (3)$$

式中 $\overline{L_{IoU}}$ —— 归一化因子

2 实验与结果

2.1 实验环境与衡量指标

本文所用实验环境为:操作系统为 Ubuntu 16.04.4 LTS,处理器为 AMD Ryzen 9 5950X 16Core Processor,显卡为 NVIDIA GEFORCE RTX 3090,显存 24 GB,基于 Python 3.8.18 和 Pytorch 1.13 搭建实验环境。实验使用 YOLO v8n 作为基准网络模型,设置超参数批尺寸为 32,训练周期为 300,初始学习率为 0.01。

采用精确率(Precision)、召回率(Recall)、F1值(F1 score)来评价模型性能,平均精度(Average precision, AP)和平均精度均值(Mean average precision, mAP)来对模型检测精度进行评价,采用参数量来衡量模型大小,采用帧率(Frame per second, FPS)衡量检测速度,采用浮点运算次数衡量模型复杂程度,采用计算量衡量模型执行时间。

2.2 数据集

为了检测水下目标,选取大连理工大学与獐子岛集团水下机器人联合实验室采集的水下图像RUIE作为实验数据集。该数据集共包含4757幅图像,由4种不同的水下生物组成,分别是扇贝(scallop)、海参(holothurian)、海胆(echinus)和海星(starfish)。由于该数据集存在部分图像标记有误或未标记现象,经数据预处理生成新的包含3765幅图像的数据集。实验按照比例8:1:1将数据集划分为训练集、验证集和测试集,随机选取3024幅图像作为训练集,选取378幅图像作为验证集,选取363幅图像作为测试集。为了便于使用其他模型进行对比实验,本文对数据集原有数据标注格式进行了转换,并保存YOLO标注格式、Voc标注格式以及coco标注格式的3份数据。其中YOLO标注格式的数据集用于YOLO v8模型的训练测试,Voc标注格式和coco标注格式的数据用于其他对比模型的训练测试。

2.3 FDC-YOLO v8与原模型对比

从表1可见,改进后的FDC-YOLO v8与原模型YOLO v8n在水下生物目标检测方面均表现出良好的效果,对各种水下生物可以进行有效检测,基本满足任务需求。FDC-YOLO v8模型在各分类的平均精度上均比原模型高,其中海胆、海星、海参、扇贝的平均精度均值分别提升1.7、0.9、6.9、0.7个百分点,总的平均精度提升2.6个百分点。这可能是FDC-YOLO v8模型采用DCNv2模块,通过对采样点增加偏移量而提取到更丰富的信息,同时使用FrFT来强化各尺度特征,导致模型特征提取和特

表1 FDC-YOLO v8与YOLO v8n性能对比

Tab.1 Performance comparison between FDC-YOLO v8 and YOLO v8n %

生物种类	YOLO v8n				FDC-YOLO v8			
	精确率	召回率	F1值	平均精度均值	精确率	召回率	F1值	平均精度均值
海胆	87.5	84.7	86.1	90.5	87.5	86.8	87.1	92.2
海星	85.3	82.0	83.6	88.5	84.6	83.5	84.0	89.4
海参	79.6	59.6	68.2	67.6	79.2	66.0	72.0	74.5
扇贝	87.2	74.4	80.3	84.2	87.0	75.2	80.7	84.9
平均	84.9	75.2	79.7	82.7	84.6	77.9	81.1	85.3

征融合能力得到增强,同时使用Wise-IoU解决了目标不平衡以及尺度差异的问题。

此外FDC-YOLO v8模型在各生物上准确率同YOLO v8模型比较接近,但其召回率较YOLO v8模型提升明显。其中海胆、海星、海参、扇贝分别提升2.1、1.5、0.4、0.8个百分点,平均召回率提升2.7个百分点。F1值也分别提升1、0.4、3.8、0.4个百分点,平均F1值提升1.4个百分点,模型均衡性得到提升。本文提出的FDC-YOLO v8n与原模型相比,在RUIE数据集上具有更出色的检测效果,并且模型具有更高的精度和鲁棒性。

2.4 FDC-YOLO v8与其他方法对比

为了验证本文提出模型的有效性,在相同环境下,对水下目标检测领域应用较广泛的几个算法,对本文数据集作了对比训练,算法训练对比结果如表2所示。通过表2可知,本文提出的FDC-YOLO v8n模型在检测多种类别的水产品时,平均精度均值达到85.3%,高于其他对比模型。同时,FDC-YOLO v8n模型帧率也处于较高水平,检测速度快。综合来看,本文提出的算法能更好地平衡精度和速度,更具有应用价值。

表2 对比实验结果

Tab.2 Comparative experiments results

模型	平均精度/%				平均精度均值/%	帧率/(f·s ⁻¹)
	海胆	海星	海参	扇贝		
Faster R-CNN	90.2	87.6	69.3	76.8	80.9	30
YOLO v3-tiny	89.4	85.7	70.1	71.4	81.6	909
YOLO v5n	90.9	88.1	62.2	80.5	80.4	400
YOLO v6n	90.5	88.8	68.2	82.1	82.4	833
SSD	84.2	73.1	44.5	51.9	63.4	148
FDC-YOLO v8	92.2	89.3	74.9	84.9	85.3	769

2.5 FDC-YOLO v8模型消融实验

为了评估不同改进对模型性能的优化程度,设置5组消融实验,实验结果如表3所示。从表3可以看出,使用单一改进都会提升模型检测精度。当同时使用3种改进方法时,算法检测精度达到最高,比初始基准模型提升2.6个百分点。模型大小方面,改进后模型参数量以及计算量与原模型处于同一量级。综合来看,本文提出的水下生物目标检测模型具有更好的检测效果。

为了直观对比两个模型的实验结果,随机抽取1幅图像,对该图像进行检测,检测效果如图5所示。对比检测结果表明,YOLO v8n模型存在漏检误检的现象,而FDC-YOLO v8模型检测结果则更完整。同时YOLO v8n对小尺度目标的检测均标记正确。分

表 3 消融实验结果

Tab. 3 Ablation experiments results

DCNv2_C2f	FrSACConv	Wise-IoU	平均精度均值/ %	参数量	浮点运算次数
			82.7	3 006 428	8.1×10^9
✓			83.9	3 126 307	8.0×10^9
	✓		84.4	3 557 340	8.5×10^9
		✓	83.4	3 006 428	8.1×10^9
✓	✓	✓	85.3	3 677 219	8.4×10^9

注: ✓表示使用此模块。

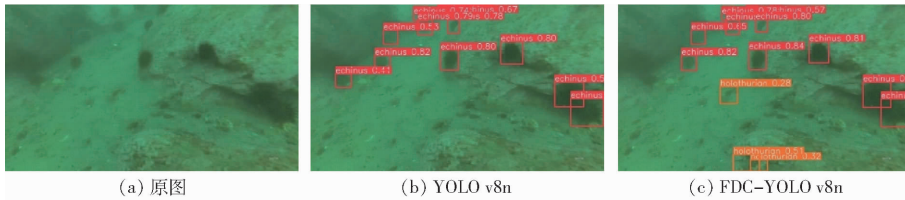


图 5 YOLO v8n 和 FDC-YOLO v8n 目标检测结果对比

Fig. 5 Comparative analysis of object detection results between YOLO v8n and FDC-YOLO v8n

型主干网络和颈部网络之间使用融合 FrFT 和 SA 的 FrSACConv 特征增强模块,将 FDC 提取到的多样特征进行分离,进一步提升模型对多尺度特征的感知能力。最后引入 Wise-IoUv2 损失函数替换了原有的 CIoU 损失函数,作为模型的边界框损失

析可知,本文提出的算法在解决水下生物存在的重叠、遮挡以及小尺度目标问题方面具有更好的效果。

3 结束语

为了解决水下生物图像存在的重叠、遮挡问题以及以小尺度为主的多尺度情况,本文提出了 FDC-YOLO v8n 目标检测算法,基于 YOLO v8n 基础框架进行改进。首先在主干网络部分,使用融合 DCNv2 的 FDC 特征提取模块,来提升模型对不同尺度、方向和位置特征的提取能力。其次,在模

函数,解决了目标不平衡以及尺度差异的问题。实验结果表明,在水下生物目标检测领域,改进的 FDC-YOLO v8n 算法较其他传统算法有较高的平均精度均值,同时其参数量和计算量均在较低水平,可满足水下设备目标检测算法部署条件。

参 考 文 献

- [1] FAYAZ S, PARAH S A, QURESHI G J. Underwater object detection: architectures and algorithms—a comprehensive review [J]. Multimedia Tools and Applications, 2022, 81(15): 20871–20916.
- [2] 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(Supp. 2): 204–209. (in Chinese)
- [3] 李辉,王晓宇,刘云,等. 融合多尺度特征和多重注意力的水下目标检测[J]. 农业工程学报, 2022, 38(20): 129–139. LI Hui, WANG Xiaoyu, LIU Yun, et al. Detecting underwater objects using multi-scale features fusion and multiple attention [J]. Transactions of the CSAE, 2022, 38(20): 129–139. (in Chinese)
- [4] 王静,李蔚然,刘业强,等. 基于计算机视觉的养殖动物计数方法研究综述[J]. 农业机械学报, 2023, 54(增刊 1): 315–329. WANG Jing, LI Weiran, LIU Yeqiang, et al. Review of vision counting methods and applications for farmed animals [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(Supp. 1): 315–329. (in Chinese)
- [5] XU S, ZHANG M, SONG W, et al. A systematic review and analysis of deep learning-based underwater object detection [J]. Neurocomputing, 2023, 527: 204–232.
- [6] CUTTER G, STIERHOFF K, ZENG J. Automated detection of rockfish in unconstrained underwater videos using haar cascades and a new image dataset: labeled fishes in the wild [C] // IEEE Winter Conference on Applications of Computer Vision Workshops, 2015: 57–62.
- [7] RIZZINI D L, KALLASI F, OLEARI F, et al. Investigation of vision-based underwater object detection with multiple datasets [J]. International Journal of Advanced Robotic Systems, 2015, 12(6): 77.
- [8] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [J]. arXiv Preprint, arXiv: 1311.2524, 2013.
- [9] GIRSHICK R. Fast R-CNN [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440–1448.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C] // Advances in Neural Information Processing Systems, 2015.
- [11] LIU J, LIU S, XU S, et al. Two-stage underwater object detection network using swin transformer [J]. IEEE Access, 2022, 10: 117235–117247.
- [12] SONG S, ZHU J, LI X, et al. Integrate MSRCR and Mask R-CNN to recognize underwater creatures on small sample datasets [J]. IEEE Access, 2020, 8: 172848–172858.

- [13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [14] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [15] REDMON J, FARHADI A. YOLO v3: an incremental improvement [J]. arXiv Preprint, arXiv: 1804.02767, 2018.
- [16] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLO v4: optimal speed and accuracy of object detection [J]. arXiv Preprint, arXiv: 2004.10934, 2020.
- [17] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLO v7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.
- [18] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // Computer Vision-ECCV, 2016; 21 - 37.
- [19] SUNG M, YU S C, GIRDHAR Y. Vision based real-time fish detection using convolutional neural network [C] // OCEANS 2017-Aberdeen, 2017.
- [20] 冯娟,梁翔宇,曾立华,等.基于改进 YOLO v4 的单环刺螠洞口识别方法 [J]. 农业机械学报, 2023, 54(2): 265 - 274, 377.
FENG Juan, LIANG Xiangyu, ZENG Lihua, et al. *Urechis unicinctus* burrows recognition method based on improved YOLO v4 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(2): 265 - 274, 377. (in Chinese)
- [21] HU K, LU F, LU M, et al. A marine object detection algorithm based on SSD and feature enhancement [J]. Complexity, 2020; e5476142.
- [22] CHEN L, LIU Z, TONG L, et al. Underwater object detection using invert multi-class Adaboost with deep learning [C] // International Joint Conference on Neural Networks (IJCNN), 2020.
- [23] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector [J] // arXiv Preprint, arXiv: 1701.06659v1, 2017.
- [24] 李健源,柳春娜,卢晓春,等.基于改进 YOLO v5s 和 TensorRT 部署的鱼道过鱼监测 [J]. 农业机械学报, 2022, 53(12): 314 - 322.
LI Jianyuan, LIU Chunna, LU Xiaochun, et al. Fish passage monitoring based on improved YOLO v5s and TensorRT deployment [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(12): 314 - 322. (in Chinese)
- [25] 翟先一,魏鸿磊,韩美奇,等.基于改进 YOLO 卷积神经网络的水下海参检测 [J]. 江苏农业学报, 2023, 39(7): 1543 - 1553.
ZHAI Xianyi, WEI Honglei, HAN Meiqi, et al. Underwater sea cucumber identification based on improved YOLO convolutional neural network [J]. Jiangsu Journal of Agricultural Sciences, 2023, 39(7): 1543 - 1553. (in Chinese)
- [26] 郑荣才,谭鼎文,徐青,等.基于改进 YOLO v7 的鲑鱼检测模型轻量化研究 [J/OL]. 农业机械学报, 1 - 12 [2024 - 08 - 26]. <http://kns.cnki.net/kcms/detail/11.1964.S.20240816.1621.007.html>.
ZHENG Rongcai, TANG Dingwen, XU Qing, et al. Research on lightweight salmon detection model based on improved YOLO v7 [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 1 - 12 [2024 - 08 - 26]. <http://kns.cnki.net/kcms/detail/11.1964.S.20240816.1621.007.html>.
- [27] TONG Z, CHEN Y, XU Z, et al. Wise - IoU: bounding box regression loss with dynamic focusing mechanism [J]. arXiv Preprint, arXiv: 2301.10051, 2023.
- [28] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; 8759 - 8768.
- [29] FENG C, ZHONG Y, GAO Y, et al. TOOD: task-aligned one-stage object detection [C] // IEEE/CVF International Conference on Computer Vision (ICCV), 2021.
- [30] LI X, WANG W, WU L, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection [C] // Advances in Neural Information Processing Systems, 2020.
- [31] ZHENG Z, WANG P, LIU W, et al. Distance - IoU Loss: faster and better learning for bounding box regression [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993 - 13000.
- [32] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017; 764 - 773.