

改进 YOLOv8 的水下目标检测^①

周昕¹, 李远禄², 吴明轩¹, 范小婷¹, 王键翔¹

¹(南京信息工程大学 自动化学院, 南京 210044)

²(江苏省大气环境与装备技术协同创新中心, 南京 210044)

通信作者: 李远禄, E-mail: yuanlu_li@163.com



摘要: 针对水下目标检测中的尺度不一、重叠遮挡目标的漏检问题, 提出了一种改进的 YOLOv8 水下目标检测算法。首先, 在主干网络中引入可变形卷积 (deformable convolution network, DCN), 通过卷积核自适应形变的机制, 提高模型对重叠遮挡目标的特征提取能力; 其次, 设计了一种空洞卷积空间金字塔模块 (atrous spatial pyramid faster, ASPF), 扩大输出特征图的感受野, 提高模型对水下多尺度目标的感知能力; 最后, 对损失函数进行改进, 优化模型的训练过程并提高定位精度。将改进算法在 URPC 数据集上进行实验, 结果表明改进算法的检测精度达到了 87.3%, 相较于原始算法 YOLOv8 提高了 3.4%, 同时能够精准检测水下多尺度、重叠遮挡目标。

关键词: 水下目标检测; YOLOv8; 深度学习; 可变形卷积; WIoU

引用格式: 周昕, 李远禄, 吴明轩, 范小婷, 王键翔. 改进 YOLOv8 的水下目标检测. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9680.html>

Underwater Target Detection via Improved YOLOv8

ZHOU Xin¹, LI Yuan-Lu², WU Ming-Xuan¹, FAN Xiao-Ting¹, WANG Jian-Xiang¹

¹(School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China)

²(Jiangsu Collaborative Innovation Center for Atmospheric Environment and Equipment Technology, Nanjing 210044, China)

Abstract: An improved YOLOv8 algorithm for underwater target detection is proposed to prevent missed detection of objects with different scales and overlapping occlusion. Firstly, deformable convolutions are introduced into the backbone network (deformable convolution network, DCN) to improve the feature extraction capability of the model by means of the adaptive deformation mechanism of convolution kernels. Secondly, a module combining atrous convolution and spatial pyramid, termed ASPF, is designed to expand the receptive field of the output feature map and improve the perception ability of the model for detecting underwater targets of multiple scales. Finally, the loss function is improved to optimize the training process of the model and improve detection accuracy. The improved algorithm is tested on the URPC data set, and the results show that its detection accuracy reaches 87.3%, which is 3.4% higher than that of the original YOLOv8 algorithm. Moreover, it can accurately detect underwater targets with different scales and overlapping occlusion.

Key words: underwater target detection; YOLOv8; deep learning; deformable convolution network (DCN); WIoU

海洋占据着地球表面积的绝大部分, 蕴藏着丰富的石油、天然气、矿产和水产等资源, 深入探索海洋可以帮助我们更好地开发和利用这些资源。近年来, 我国基于“提高海洋资源开发能力、发展海洋经济、保

护海洋生态和建设海洋强国”等海洋战略做出重要部署, 加强对海洋的探索。

对于海洋的探索离不开水下目标检测技术, 其无论是在资源开发、海底捕捞、还是军事行动中都发挥

^① 基金项目: 国家自然科学基金 (61671010); 青蓝工程 (2191091900101)

收稿时间: 2024-04-24; 修改时间: 2024-05-29; 采用时间: 2024-06-04; csa 在线出版时间: 2024-09-27

着重要的作用. 现如今, 随着深度学习技术的飞速发展, 越来越多的学者将其应用于水下目标检测领域, 研究方法可以分为两类: 两阶段检测算法和单阶段检测算法.

两阶段检测算法采用分阶段的方式实现对目标的检测, 首先生成候选区域, 然后对这些区域进行细致的分类和边界框回归. 通过分阶段的方式, 两阶段目标检测算法能够在保持高精度的同时, 有效处理大规模图像数据, 代表算法有 R-CNN^[1]、Fast R-CNN^[2]、Faster R-CNN^[3]和 Cascade R-CNN^[4]等. 杨婷等^[5]在 Faster R-CNN 的基础上进行了部分改进, 首先采用自动色阶增强水下图像; 然后, 采用双向特征融合增强信息流动; 最后, 引入柔性非极大值抑制算法调整候选目标区域, 相关改进在水下数据集 URPC 上取得了良好的检测精度. 乔美英等^[6]提出了一种改进的 Faster R-CNN, 在特征金字塔网络 (feature pyramid network, FPN) 的上采样过程加入注意力机制捕捉上下文信息并在损失函数中引入线性回归增益系数提高算法对多尺度水下目标的检测精度.

单阶段检测算法是目标检测技术中的一种高效的方法, 旨在通过一个单一的网络对目标直接进行定位和分类, 不需要预先生成候选框. 这类算法的核心特点在于它们将目标检测问题转化为一个直接的回归问题, 通过一次前向传播同时完成目标的定位和分类. 相较于两阶段检测算法, 单阶段检测算法在保持较高精度的同时, 显著提高了算法的运行速度, 如 SSD^[7]、FCOS^[8]和 YOLO^[9,10]等. 鉴于单阶段算法优越的性能, 许多学者将其应用于水下场景: 张琳等^[11]提出了一种轻量化的 FCOS 算法, 通过引入自适应空间特征融合模块并改进损失函数有效提高了算法在水下鱼类公共数据集上的检测精度. 曹建荣等^[12]对 YOLOv5 算法进行改进, 通过水下暗通道先验算法增强水下图像并引入 ECA 注意力机制提高模型的特征提取能力, 从而提高检测精度. 辛世澳等^[13]针对复杂水下环境对 YOLOv7 进行改进, 通过结合 BiFormer 注意力机制提高水下目标检测精度并减少计算量. 施克权等^[14]采用 Vision Transformer 作为 YOLOv8 的主干网络, 并结合 EMA 注意力机制提高模型的特征提取能力, 改进算法在 URPC 数据集上检测精度达到了 83.2%.

随着检测算法的不断改进, 其在水下目标检测的优势得以体现. 但是, 仍然存在水下多尺度、重叠遮挡

目标检测精度低的问题. 针对此问题并结合 YOLOv8 模型的不足之处, 本文提出一种改进的 YOLOv8 水下目标检测算法. 首先, 在主干网络引入可变形卷积 DCN, 加强模型对重叠遮挡目标的特征提取; 其次, 设计了一种空洞卷积空间金字塔模块 ASPF, 通过空洞卷积捕获不同尺度的目标上下文信息, 提高模型对多尺度目标的感知能力; 最后, 对损失函数进行改进, 通过 WIoU 损失优化模型训练过程并提高检测精度.

1 YOLOv8 网络结构

YOLOv8 的网络结构包括输入端 (Input)、主干网络 (Backbone)、颈部网络 (Neck) 和检测端 (Head) 共 4 个部分组成, 网络结构如图 1 所示.

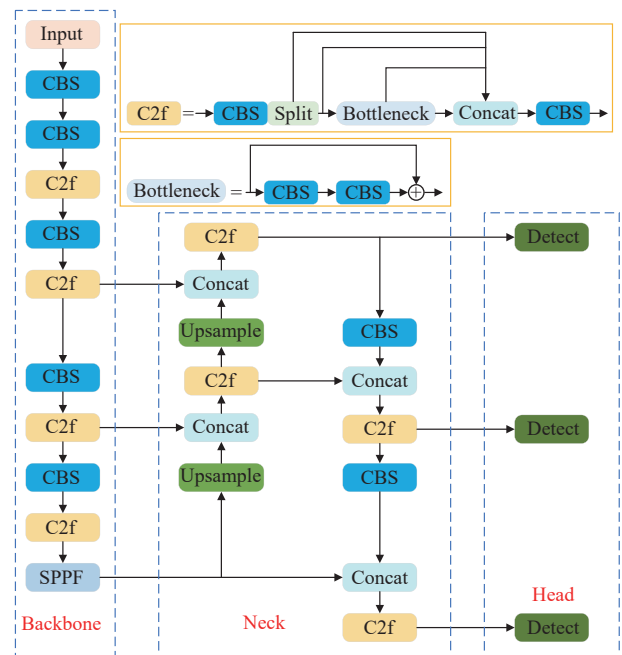


图 1 YOLOv8 网络结构

(1) 输入端: YOLOv8 的输入端采用自适应锚框计算、自适应图片缩放、自适应灰度填充和马赛克数据增强. 自适应锚框根据不同的数据集, 初始设定锚框的尺寸. 在网络的训练过程中, 通过比较初始化锚框和数据集的真实锚框的差异, 进行反向传播, 优化网络参数. 自适应图片缩放针对数据集中图片的尺寸不同, 将原始图像统一尺寸为 640×640 大小, 能够提高模型的推理速度. 自适应灰度填充根据图像的内容和特性自动调整或填充图像的灰度级别, 以改善图像的质量或增强图像特征. 马赛克数据增强则采用随机裁剪、随机

缩放和随机分布的策略对图像进行拼接, 扩充了样本的数量.

(2) 主干网络: YOLOv8 的主干网络由 CBS、C2f 和 SPPF 模块组成. 其中, CBS 模块由普通卷积层 Conv、批处理归一化层 BN 和激活函数 SiLu 构成. C2f 模块则是仿照 YOLOv7^[10] 的 ELAN 结构, 旨在通过更多的分支跨层连接, 丰富模型的梯度流, 使得模型具有更强的特征表达能力. SPPF 模块通过一个标准卷积层并串行 3 个 5×5 尺寸的最大池化下采样 (Maxpool) 层对输入特征图进行降维, 然后对降维后的特征图进行融合, 再进行一次升维操作, 输出后的特征图包含多个尺度目标的特征信息.

(3) 颈部网络: Neck 网络的核心是路径聚合网络 (path aggregation network, PANet), PANet 是自顶向下和自底向上的双向特征融合网络^[15]. 底层特征具有丰富的位置信息, 高层特征具有丰富的语义信息, PANet 将底层特征和高层特征进行融合, 使最终输出的特征同时包含位置信息和语义信息.

(4) 检测端: YOLOv8 的检测端采用了当下流行的解耦合检测头, 这是因为检测任务需要对目标进行定位和分类, 定位任务关注目标的边缘信息, 而分类任务更关注目标的纹理信息, 解耦头将坐标信息和分类信息提取出来, 通过不同的网络分支学习并输出, 减少计算复杂度, 增强模型的泛化能力.

2 改进的 YOLOv8 框架

YOLOv8 算法共有 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l、YOLOv8x 这 5 个版本, 本文选取 YOLOv8n 作为基准模型, 该模型具有参数量少, 检测速度快等优点. 在此基础上, 本文提出一种改进 YOLOv8 水下目标检测算法. 针对水下重叠遮挡目标的漏检问题, 将设计的 DCN_C2f 模块替换主干网络中的 C2f 模块; 针对水下多尺度目标设计空洞卷积空间金字塔模块 ASPF 中的模块替换 SPPF 模块; 同时, 通过改进损失函数, 加速模型收敛, 改进后的 YOLOv8 网络结构如图 2 所示.

2.1 可变形卷积 DCN

传统的卷积神经网络采用固定尺寸和形状的卷积核来提取特征信息, 这种设计假设目标物体具有一定的规则性和均匀性. 然而在水下场景中, 生物由于生活习性通常会呈现出多样的尺寸和形状, 使用常规的卷

积核难以有效捕捉这些复杂特征, 导致特征提取不准确或不完整, 尤其是目标出现重叠遮挡的情况.

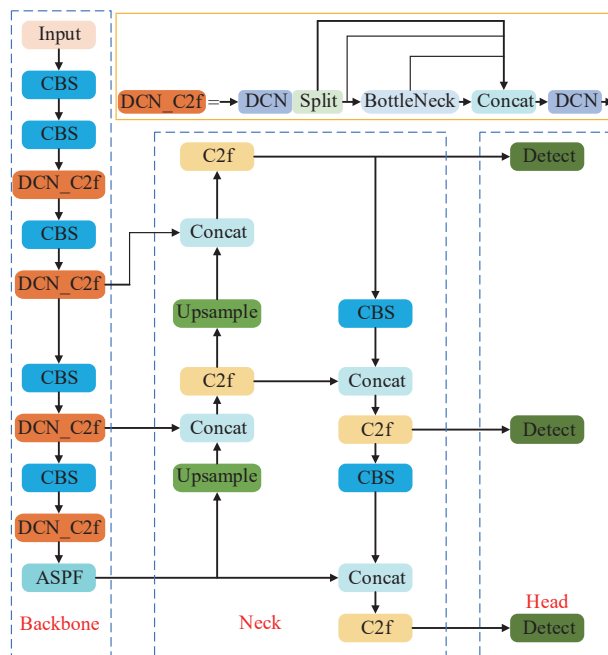


图 2 改进的 YOLOv8 网络结构

为了克服这些局限性, 本文采用可变形卷积 (deformable convolutional network, DCN)^[16] 提高模型对重叠遮挡目标的特征提取能力. 可变形卷积通过引入可学习的偏移量, 使卷积核能够自适应地调整其形状和采样位置, 卷积核不再是固定的矩形形状, 而是随着不同的阶段、不同的特征图, 根据不同的目标学习最优的卷积核结构. 这种设计不仅能够提高对物体形状和尺度变化的适应性, 还增强了模型在复杂视觉环境下的特征提取能力, 其示意图如图 3 所示.

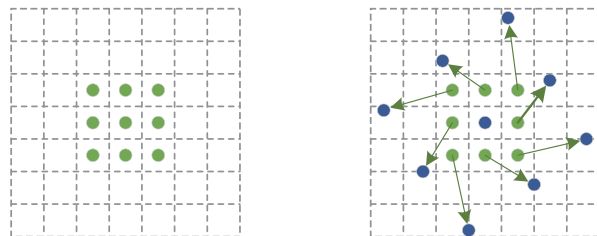


图 3 可变形卷积示意图

标准的卷积核沿着输入特征图的规则网格进行滑动, 其采样点的位置是固定的, 而可变形卷积为这些采样点引入一个额外的可学习的偏移量 offset. 偏移量的生成通过对输入特征图添加一个卷积层, 在网络模型的学习过程中使卷积核的采样点发生偏移, 集中于感

兴趣区域或目标,其计算公式如式(1)所示:

$$y(P_0) = \sum_{P_n \in R} w(P_n) \cdot x(P_0 + P_n + \Delta P_n) \quad (1)$$

其中, $y(P_0)$ 是输出特征图在位置 P_0 上的值, $w(P_n)$ 是卷积核在位置 P_n 的权重, $(P_0 + P_n + \Delta P_n)$ 是采样点偏移后的位置, 同时采样点坐标为 $R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$.

由于在卷积操作中加入了偏移量 ΔP_n , 采样点的位置一般为一个浮点数, 并不对应输入特征图上的实际像素位置, 因此还需要通过双线性插值得到偏移后的像素值, 计算公式如下:

$$x(p) = \sum_q G(q, p) \cdot x(q) \quad (2)$$

其中, $p = p_0 + p_n + \Delta p_n$ 表示输入特征图中的采样点偏移位置, q 为 p 周围的 4 个整数点, $G(q, p)$ 为一个二维的双线性插值核, 可以被分解为两个一维核:

$$G(q, p) = g(q_x, p_x) \cdot g(q_y, p_y) \quad (3)$$

其中, $g(a, b) = \max(0, 1 - |a - b|)$.

基础的可变形卷积能够针对目标的形状进行特征提取, 但是对于尺度较小的目标则会提取到较多无用的特征信息, 同时计算量较大, 影响模型的性能. 因此, 本文采用文献[17]提出的 DCNv3 进一步加强模型对水下目标的特征提取, 相较于基础的可变形卷积, 对 DCNv3 进行了如下改进.

(1) 神经元共享权值: 在基础的可变形卷积中神经元具有独立的线性投影权值, 导致模型的参数也随着采样点总数线性增长, 影响了模型效率. DCNv3 则是借鉴深度可分离卷积^[18]的概念, 将卷积权值分解为深度方向和采样点方向, 深度方向由原始位置感知, 采样点方向设置为共享权值, 这样的设计可有效减少参数量, 提高模型效率.

(2) 引入多组机制: 多组机制最早应用于多头自注意力机制 (multi-head self attention, MHSA)^[19], 其结合自适应空间融合, 从而学习到更丰富的信息. DCNv3 将空间融合过程分为多个组, 每个组均有独立的采样偏移量和调节因子, 因此单个卷积层上的不同组可以具有不同的空间聚集模式, 为模型提供更强的特征表示.

(3) 归一化处理: 将基础可变形卷积中的 Sigmoid 替换为 Softmax 进行归一化处理, 避免训练不稳定的情况.

DCNv3 的计算公式如下所示:

$$y(p_0) = \sum_{g=1}^G \sum_{k=1}^K w_g m_{gk} x_g(p_0 + p_k + \Delta p_{gk}) \quad (4)$$

其中, G 表示分组数量; 对于第 g 组, $w_g \in R^{C \times C'}$ 表示与位置无关的组投影权重, $C' = C/G$ 表示分组尺度, m_{gk} 表示第 g 组中第 k 个调节因子, 用 Softmax 函数进行归一化; $x_g \in R^{C' \times H \times W}$ 表示第 g 组的输入特征映射. Δp_{gk} 表示第 g 组中采样位置 p_k 对应的偏移量.

2.2 空洞卷积空间金字塔模块

在检测任务中, 输出特征图通常需要采用不同大小的感受野来处理不同尺度的目标, 确保模型对多尺度的目标具有良好的识别能力. 感受野的大小影响着模型对输入图像中不同尺度目标的感知能力, 较小的感受野有助于捕捉图像中的局部细节信息, 对尺度较小的目标检测效果更好, 但是如果感受野过小, 可能会忽略重要的上下文信息, 导致检测性能下降. 较大的感受野则能够覆盖更广泛的区域, 捕获更多的上下文信息, 有助于模型理解较大尺度物体和场景的整体结构, 但过大的感受野则难以捕捉到足够的局部细节特征, 也会影响检测结果. 因此, 在检测任务中通常需要结合不同尺度的感受野提高模型对不同尺度目标的感知能力.

YOLOv8 算法仍然延续使用 YOLOv5 的 SPPF 结构整合不同尺度的特征, 如图 4 所示.

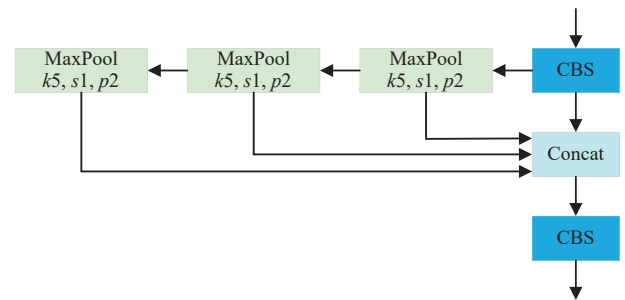


图 4 SPPF 模块结构

首先通过一个尺寸为 1×1 的 CBS 模块输出感受野为 1×1 的特征图, 然后通过连续 3 个尺寸为 5×5 , 步距为 1, 填充为 2 的最大池化下采样操作扩大输出特征图的感受野, 每个最大池化下采样操作进一步扩大输出特征图的感受野, 最后通过拼接操作融合多尺度的信息, 将输出特征图组合成一个多尺度的特征表示. 但是连续的池化操作会导致特征图的分辨率降低, 局部信息丢失, 不利于检测任务. 这是因为池化的主要目的

是下采样和特征压缩,本质上是在减少数据的空间维度.即使采用填充操作,池化后的输出特征仍然是一个下采样版本的输入特征,因为最大池化操作的窗口覆盖的每个区域只产生一个输出值.

针对 SPPF 结构的不足之处,本文设计了一种空洞卷积空间金字塔 (atrous convolution spatial faster, ASPF) 模块,能够在融合多个尺度特征的情况下通过空洞卷积避免信息丢失,其结构如图 5 所示.

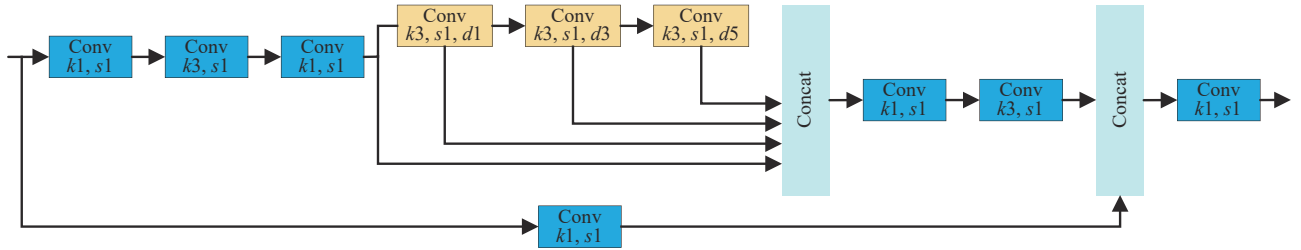


图 5 空洞卷积空间金字塔模块

从图 5 中可以看出,该结构有两个分支.其中一条支路为一个残差结构,通过一个卷积核大小为 1×1 的 CBS 模块;另一条支路首先通过 3 个 CBS 模块,使用的卷积核大小依次为 1、3、1,第 1 个 1×1 的 CBS 模块作用是降低输入特征图的通道数,减少计算量, 3×3 的 CBS 模块则是为了从空间维度提取特征信息,最后一个 1×1 的 CBS 模块则是恢复特征图的通道数并将提取到的特征进行融合和细化,通过这样的组合可以使模型在保持较低计算量的同时并有效地进行特征提取;然后再依次通过串行的 3 个尺寸为 3×3 不同膨胀率的空洞卷积 (图 5 中 d 表示空洞卷积的膨胀率) 得到不同尺度的感受野,通过通道维度拼接操作使得输出特征图能够结合不同大小的感受野从而捕获更丰富的特征信息.接着再通过 1×1 、 3×3 的 CBS 模块并与残差分支的输出特征图进行拼接;最后通过 1×1 的 CBS 模块调整通道数,与原始输入特征图通道数保持一致.其中,3 个空洞卷积的膨胀率依次为 1、3、5,在连续的空洞卷积中使用不同的膨胀率是为了防止网格效应^[20]的出现,即连续使用相同膨胀率的空洞卷积会导致卷积核只能覆盖输入特征图上的稀疏位置集合,从而忽略了其他位置的信息,这会导致特征图上的某些区域没有被有效利用,从而出现特征信息丢失的情况.因此,本文基于 HDC^[20]的思想来选择不同膨胀率的空洞卷积,避免特征信息的丢失,计算公式如式 (5) 所示:

$$M_i = \max[M_{i+1} - 2d_i, 2d_i - M_{i+1}, d_i] \quad (5)$$

其中, M_i 为第 i 层两个非零元素之间的最大距离, d_i 为第 i 层的膨胀率.该式表明选取膨胀率需要满足 $M_i \leq K$,

K 为卷积核的尺寸,即两个非零元素之间的最大距离需要小于或等于对应层的卷积核尺寸.对于膨胀率为 $[1, 3, 5]$,尺寸为 3×3 的空洞卷积,通过计算可知 M_1 、 M_2 的值均为 3,满足上述条件.

2.3 损失函数改进

在计算机视觉领域,通常使用交并比 (intersection over union, IoU) 来度量模型预测值与真实标签值之间的相似性. IoU 的值越高,则表明模型的预测值与标签值的相似性越高,预测结果越好.在训练中,通常将 IoU 作为边界框损失函数,提高模型性能.

YOLOv8 中的边界框损失函数采用 CIOU, CIOU 损失不仅考虑了预测框与真实框之间的 IoU,还考虑了它们中心点的距离以及长宽比的一致性.这种综合评估可以更精确地指导模型优化其预测的边界框,从而在定位目标时更为准确, CIOU 的计算公式如下.

$$L_{CIOU} = L_{IoU} + \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)} + \alpha v \quad (6)$$

$$\alpha = \frac{v}{L_{IoU} + v} \quad (7)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w}{h} - \arctan \frac{w_{gt}}{h_{gt}} \right) \quad (8)$$

$$L_{IoU} = 1 - \frac{W_i H_i}{w h + w_{gt} h_{gt} - W_i H_i} \quad (9)$$

其中, α 为惩罚项,用于平衡参数; v 为长宽比度量函数,用于衡量长宽比的一致性; (W_i, H_i) 为预测框和标签框重叠部分的长和宽; (w, h) 和 (w_{gt}, h_{gt}) 分别为是预测框和标签框的长和宽; (x, y) 和 (x_{gt}, y_{gt}) 分别为预测框和标签框的中心位置; L_{IoU} 为预测框和标签框的交并比,如

图 6 所示.

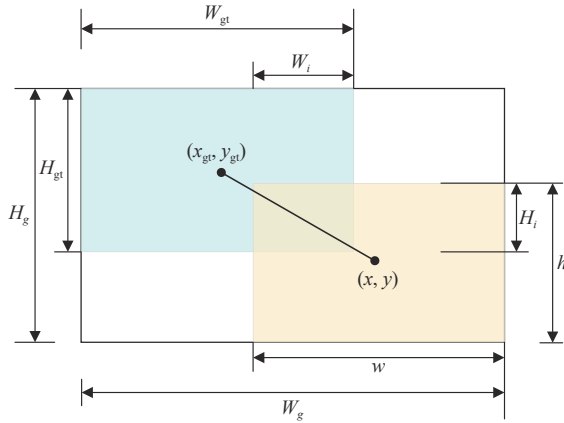


图 6 预测框和真实框的 IoU

虽然 CIoU 损失函数考虑了边界框的长宽比一致性,但是在模型反向传播更新梯度的过程中,预测框的长宽 w, h 相对于 v 的梯度为一对相反数,导致在模型训练过程中 w 和 h 不能同时变化,这会限制模型的性能.同时,由式 (8) 可知,若预测框和标签框的长和宽相同时, v 的值为 0, 惩罚项消失,这是不合理的;此外,若数据集中存在低质量的样本时, CIoU 由于其固定的惩罚机制会导致模型难以从低质量样本中学习到有效的信息,从而影响检测精度.

因此,本文使用损失函数 $WIoU^{[21]}$ 代替 CIoU 损失. $WIoU$ 通过动态非单调聚焦的梯度增益分配策略,能够根据样本的质量动态调整损失函数,降低高质量样本的竞争优势,减轻由低质量样本产生的不利梯度影响,提高模型的检测精度和鲁棒性.同时,根据距离度量得到具有两层注意力机制的 $WIoU$,其计算公式如下:

$$L_{WIoU} = R_{WIoU} \cdot L_{IoU} \quad (10)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (11)$$

其中, R_{WIoU} 为 $WIoU$ 的距离度量函数; (W_g, H_g) 分别表示最优包围框的长和宽; $*$ 表示分离操作,避免出现 CIoU 损失函数中预测框的长宽无法同时变化的情况,减少对模型训练时的干扰.

3 实验分析

3.1 数据集来源

URPC 数据集来源于全国水下机器人大赛,由水下抓取机器人上的拍摄设备捕获水下图像,该数据集

有 4 个目标类别:海胆 echinus、海参 holothurian、扇贝 scallop 和海星 starfish,共 7600 张图片,按照 8:2 的比例对数据集进行划分,其中训练集 6080 张,验证集 1520 张.

3.2 实验环境及参数设置

本节所有实验均基于集成开发环境 PyCharm 操作,在 GPU 型号为 NVIDIA GeForce RTX 2080Ti (显存为 11 GB) 的服务器上运行,操作系统为 Ubuntu 16.04,实验用到的编程环境为 Python 3.8,框架为 PyTorch 1.8.1,实验超参数设置如表 1 所示.

表 1 实验超参数设置

参数	值
优化器	SGD
初始学习率	0.01
权重衰减系数	0.0005
动量	0.937
batch_size	16
epoch	100

3.3 实验评价指标

为了定量分析本文算法对水下目标的检测性能,采用平均精度 (mean average precision, mAP) 作为本章实验的评价指标. mAP 的值为所有目标类别 AP 值的平均, AP 的值是精确率 (P) 和召回率 (R) 的 PR 曲线所占据的面积,面积越大,精确度越高.表达式如下:

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

$$AP = \int_0^1 P(R) dR \quad (14)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (15)$$

其中, TP 为真阳性,表示正样本被模型预测为正; FP 为假阳性,表示负样本被模型预测为正. FN 为假阴性,表示正样本被模型预测为负; N 为数据集中的类别数目.

3.4 消融实验

为了验证各模块:可变形卷积 DCNv3、空洞卷积空间金字塔模块 ASPF 及 $WIoU$ 损失对最终检测性能的影响,在 URPC 数据集上进行消融实验.结果如表 2 所示,其中 \surd 表示使用该模块,“—”表示不使用该模块.

表2 模块消融实验

YOLOv8	DCNv3	ASPF	WIoU	AP (%)				mAP (%)
				echinus	holothurian	scallop	starfish	
√	—	—	—	92.3	77.8	77.9	87.6	83.9
√	√	—	—	92.8	80.5	81.2	86.6	85.3
√	√	√	—	92.9	83.0	81.5	88.2	86.4
√	√	√	√	93.5	83.7	82.8	89.3	87.3

从表2中可以看出,基准模型YOLOv8在主干网络添加可变形卷积DCNv3模块后,mAP从83.9%提升至85.3%,这是因为可变形卷积DCNv3能够针对水下遮挡目标更好地提取目标特征,提高检测精度;然后将SPPF模块替换为设计的空洞卷积空间金字塔ASPF模块后,mAP从85.3%进一步提升到86.4%,通过采用不同膨胀率的空洞卷积整合多尺度的特征能够提高模型对多尺度目标的感知能力,进一步提高检测精度;最后,采用WIoU损失后,mAP提升至87.3%。此外,为了直观分析WIoU损失与CIoU损失的优劣,将二者的损失函数曲线图进行对比(如图7所示)。

图7中蓝色曲线表示CIoU,红色曲线表示采用WIoU,在采用WIoU损失后,可以看出模型在50个epoch左右趋于收敛,且收敛效果更好,CIoU损失将

近100个epoch才趋于收敛;这表明了通过采用WIoU损失改进YOLOv8,可以优化网络的训练过程。

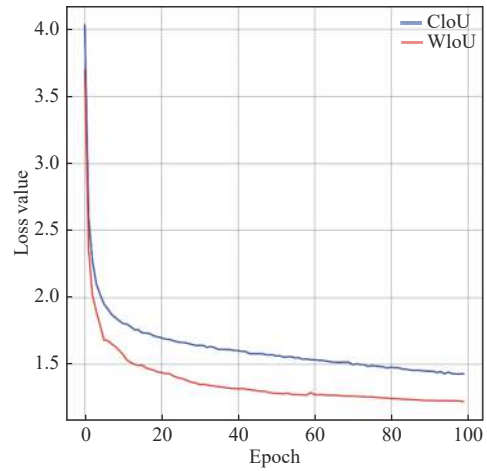


图7 Loss曲线收敛情况对比

3.5 不同算法检测效果对比

(1) 直观对比

为了直观分析本文改进算法的检测效果,选取两组具有代表性的水下图像,对不同算法的检测性能进行对比,如图8、图9所示。

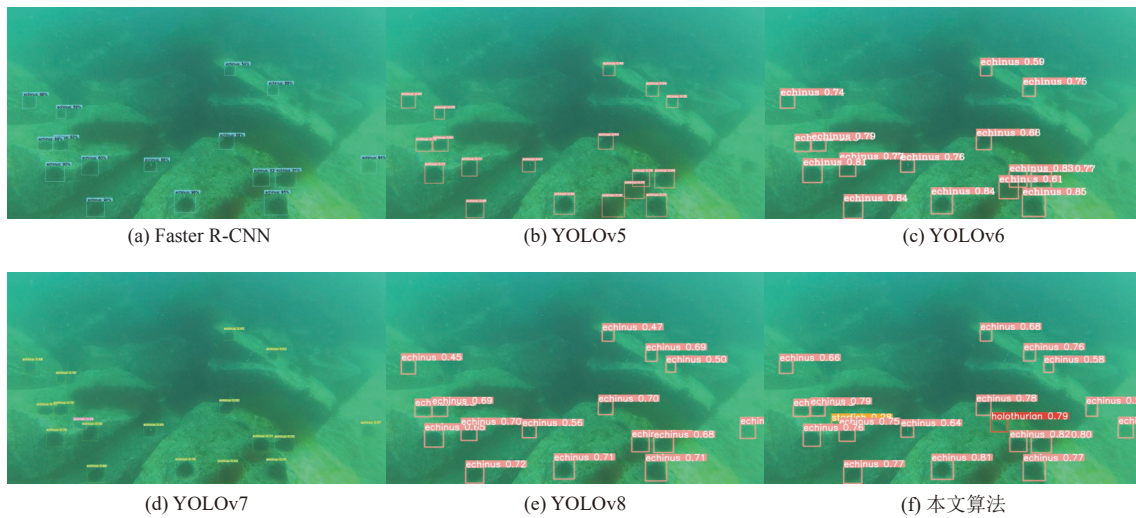


图8 水下多尺度目标检测效果

图8为多尺度水下目标检测效果;图9为重叠遮挡目标检测效果,该图中底部岩石背面有一只海星被遮挡住了大部分,不易分辨;(a)-(f)分别是Faster R-CNN、YOLOv5、YOLOv6、YOLOv7、YOLOv8和本文算法的检测效果图。

从图8、图9中可以看出,图8中包含近景和远景

目标,目标尺度不一,相较于其他检测算法,本文算法能准确检测出场景中不同尺度的目标。图9中底部岩石后面有一只海星被遮挡住了大部分,相较于其他检测算法,本文算法可以准确检测出被遮挡的海星。综上所述,本文算法针对多尺度、被遮挡的水下目标有着良好的检测效果,验证了本文算法的有效性。

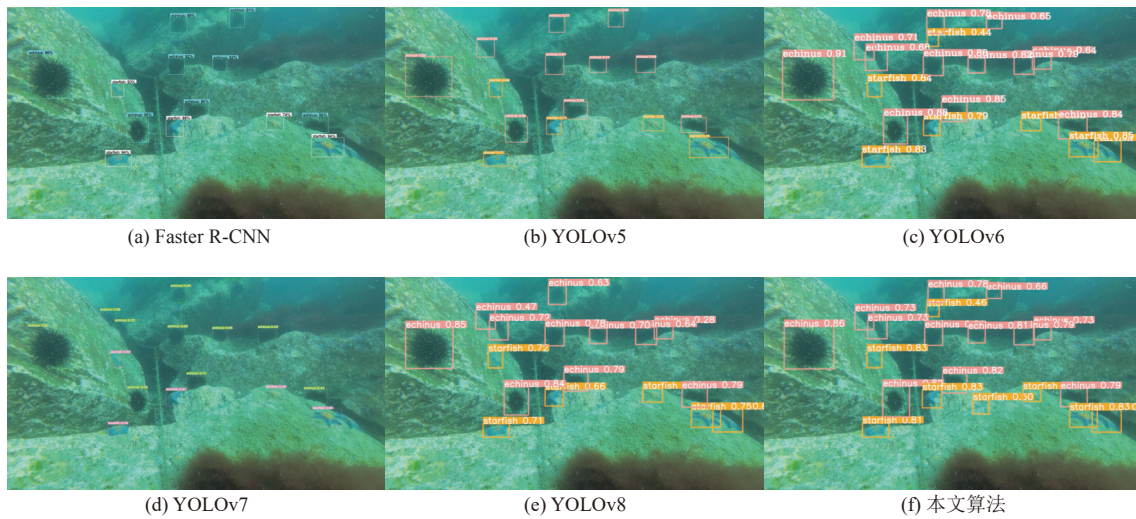


图9 水下遮挡目标检测效果

(2) 定量分析

为了定量分析本文算法的有效性,将本文算法的检测结果与当前的主流目标检测算法 Faster R-CNN、YOLOv5、YOLOv6、YOLOv7、YOLOv8 等进行对比分析,采用 AP 和 mAP 作为算法优劣的评价指标,表 3 为各种检测算法在 URPC 数据集上的 AP 值和 mAP 值对比。

表 3 不同算法检测精度对比

算法	AP (%)				mAP (%)
	echinus	holothurian	scallop	starfish	
Faster R-CNN	87.6	73.4	75.7	70.1	76.7
YOLOv5	91.3	75.0	79.4	84.7	82.6
YOLOv6	90.2	78.8	78.6	83.5	82.8
YOLOv7	93.1	76.7	78.6	84.9	83.3
YOLOv8	92.3	77.8	77.9	87.6	83.9
本文算法	93.5	83.7	82.8	89.3	87.3

从表 3 可以看出,本文算法的检测精度达到了 87.3%,比两阶段算法 Faster R-CNN 检测精度高了 10.6%;比单阶段算法 YOLOv5、YOLOv6、YOLOv7 和 YOLOv8 的检测精度分别 4.7%、4.5%、4.0% 和 3.4%,各项指标均取得了最高的检测精度。

4 结论与展望

本文提出一种改进 YOLOv8 的水下目标检测算法,能够准确检测水下多尺度、重叠遮挡目标.通过可变形卷积 DCNv3、空洞卷积空间金字塔模块 ASPF 和 WIoU 损失改进 YOLOv8,改进后的算法相较于 YOLOv8 检测精度提高了 3.4%.在后续的工作中,将针对水下图像进行增强,获得视觉质量更高的水下图

像,进一步提高水下目标检测的精度;同时,设计适用于水下实时检测的轻量化算法,减少参数量和计算量,提高算法的检测速度。

参考文献

- Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587.
- Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.
- Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 91–99.
- Cai ZW, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6154–6162.
- 杨婷,高武奇,王鹏,等.自动色阶与双向特征融合的水下目标检测算法.激光与光电子学进展,2023,60(6): 0610012.
- 乔美英,史建柯,李冰锋,等.改进损失函数的增强型 FPN 水下小目标检测.计算机辅助设计与图形学报,2023,35(4): 525–537.
- Liu W, Anguelov D, Erhan D, et al. SSD: Single shot MultiBox detector. Proceedings of the 14th European

- Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
- 8 Tian Z, Shen CH, Chen H, *et al.* FCOS: A simple and strong anchor-free object detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 44(4): 1922–1933.
- 9 Li CY, Li LL, Jiang HL, *et al.* YOLOv6: A single-stage object detection framework for industrial applications. arXiv:2209.02976, 2022.
- 10 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023. 7464–7475.
- 11 张琳, 葛艳, 杜军威, 等. 改进 FCOS 网络的海洋鱼类目标检测. *计算机系统应用*, 2023, 32(3): 309–315. [doi: [10.15888/j.cnki.csa.008965](https://doi.org/10.15888/j.cnki.csa.008965)]
- 12 曹建荣, 庄园, 汪明, 等. 基于 ECA 的 YOLOv5 水下鱼类目标检测. *计算机系统应用*, 2023, 32(6): 204–211. [doi: [10.15888/j.cnki.csa.009149](https://doi.org/10.15888/j.cnki.csa.009149)]
- 13 辛世澳, 葛海波, 袁昊, 等. 改进 YOLOv7 的轻量化水下目标检测算法. *计算机工程与应用*, 2024, 60(3): 88–99.
- 14 施克权, 李祺, 隋皓, 等. IEMAyoloViT: 基于改进 YOLOv8 的水下目标检测算法. *电讯技术* (在线出版). doi: [[10.20079/j.issn.1001-893x.231206005](https://doi.org/10.20079/j.issn.1001-893x.231206005)]. (2024-02-02).
- 15 田枫, 贾昊鹏, 刘芳. 改进 YOLOv5 的油田作业现场安全着装小目标检测. *计算机系统应用*, 2022, 31(3): 159–168. [doi: [10.15888/j.cnki.csa.008359](https://doi.org/10.15888/j.cnki.csa.008359)]
- 16 Dai JF, Qi HZ, Xiong YW, *et al.* Deformable convolutional networks. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 764–773.
- 17 Wang WH, Dai JF, Chen Z, *et al.* InternImage: Exploring large-scale vision foundation models with deformable convolutions. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023. 14408–14419.
- 18 Howard AG, Zhu ML, Chen B, *et al.* Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
- 19 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 20 Wang PQ, Chen PF, Yuan Y, *et al.* Understanding convolution for semantic segmentation. In: *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Lake Tahoe: IEEE, 2018. 1451–1460.
- 21 Tong ZJ, Chen YH, Xu ZW, *et al.* Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. arXiv:2301.10051, 2023.

(校对责编: 张重毅)