

改进 YOLOv8 的水下目标检测算法

袁 泉^{1,2}, 杨清泉^{1,2}, 袁亚隆^{1,2}, 刘凤娟^{1,2}

(1.重庆邮电大学 通信与信息工程学院,重庆 400065;

2.重庆邮电大学 通信新技术应用研究中心,重庆 400065)

摘要:针对水下光线衰减、散射等影响导致水下目标检测效果不佳的问题,提出一种基于 YOLOv8 的水下目标检测框架 ERMS-YOLOv8,提升水下目标检测性能。主干网络采用高效视觉 transformer 网络(efficient vision transformer, EfficientViT),增强模型对水下生物的特征提取能力,减少特征信息丢失;Neck 部分采用高效重参数化广义特征金字塔网络(reparameterized generalized-directional feature pyramid network, RepGFPN),增强模型对水下生物高层语义和低级空间特征的提取和融合能力,使得模型获取更加丰富的特征信息;引入混合局部通道注意力机制(mixed local channel attention for object detection, MLCA),使得模型同时融合通道信息、空间信息、局部通道信息和全局通道信息,增强了模型的表征能力;引入可扩展交并比损失函数(scalable intersection over union loss, SIoU),提升模型对目标边界信息的提取能力,从而进一步提高检测精度。实验结果表明,改进后的算法在 UPRC2021 和 DUO 数据集上 mAP 值分别达到 83.9%和 84.4%,与基准 YOLOv8 算法相比都有提高,在水下目标检测中具有优越的性能。

关键词:YOLOv8;EfficientViT;重参数化广义特征金字塔网络;注意力机制;损失函数

中图分类号:TP391.41

文献标志码:A

文章编号:1673-825X(2025)05-0729-12

Improved underwater target detection algorithm for YOLOv8

YUAN Quan^{1,2}, YANG Qingquan^{1,2}, YUAN Yalong^{1,2}, LIU Fengjuan^{1,2}

(1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China;

2. Research Center for New Communication Technology Applications, Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China)

Abstract: To address the poor performance of underwater object detection caused by light attenuation and scattering, this paper proposes an enhanced underwater object detection framework based on YOLOv8, named ERMS-YOLOv8, aiming to improve detection accuracy. The backbone is replaced with an efficient vision transformer (EfficientViT) to strengthen feature extraction of underwater organisms and reduce information loss. The neck adopts a reparameterized generalized-directional feature pyramid network (RepGFPN) to enhance the fusion of high-level semantic and low-level spatial features, enabling richer feature representation. A mixed local channel attention for object detection (MLCA) is introduced to integrate channel, spatial, local, and global channel information, thereby boosting the model's representational capacity. Additionally, a scalable intersection over union loss (SIoU) is employed to improve boundary prediction accuracy. Experimental re-

收稿日期:2024-12-06 修订日期:2025-09-08 通讯作者:袁 泉 quanyuan@cqupt.edu.cn

基金项目:重庆市教委科学技术研究项目(KJZD-M202400603)

Foundation Item: Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD-M202400603)

sults demonstrate that the proposed method achieves mAP values of 83.9% on the UPRC2021 dataset and 84.4% on the DUO dataset, outperforming the original YOLOv8 and exhibiting superior performance in underwater object detection.

Keywords: YOLOv8; EfficientViT; RepGFPN; attention mechanism; loss function

0 引言

随着经济与科技的发展,人们越来越重视海洋资源开发,水下目标检测一直是其中一个重要方向^[1]。与陆地环境不同的是,水下环境更加复杂。

基于深度学习的目标检测算法是计算机视觉领域的重要研究方向之一,主要分为两阶段检测方法和单阶段检测方法^[2]。两阶段目标检测主要包括区域卷积神经网络(regions with convolutional neural networks, R-CNN)^[3]、快速区域神经网络(fast regions with convolutional neural networks, Fast R-CNN)^[4]、更快区域卷积神经网络(faster regions with convolutional neural networks, Faster R-CNN)^[5]等方法;单阶段目标检测的代表是 YOLO(you only look once)算法^[6]和 SSD(single shot multibox detector)算法^[7]。

目前,水下目标检测通常使用单阶段目标检测方法。文献[8]提出了一种基于改进 SSD 的水下目标检测算法,将 SSD 中卷积神经网络替换为 ResNet,利用深度可变形卷积模块进行特征提取。文献[9]针对弱光条件下的水下生物,提出了 YOLOv8s-underwater 方法,引入了限制对比度自适应直方图均衡以及快速空间金字塔池化模块,极大缓解了弱光水下光线衰弱的问题。文献[10]针对水下目标的模糊和遮挡问题,提出了一种基于 RepViTS-YOLOX 的水下目标检测算法,引入空间和通道重构模块与上下文解耦头,对水下遮挡目标的定位和分类更加精准。文献[11]提出了 WCA-YOLOv8 方法,设计了特征融合模块和下采样增强模块,不仅获得了更高的检测准确性,还可达到更快的检测速度。

近年来深度学习算法在水下目标检测中发挥着重要的作用,但是面对水下复杂的环境仍然存在实时性较差、检测精度较低等问题。本文以 YOLOv8 为基础,提出一种具有更高精度并适用于实时检测的 ERMS-YOLOv8 新算法,该算法主要改进如下:

1) 将 YOLOv8 的主干网络替换为高效视觉 transformer 网络(efficient vision transformer, EfficientViT),融入高效的多尺度注意力机制,增强模型对水下目标全局特征的提取能力,在一定程度上减少计算量,满足水下检测的实时性要求;

2) 引入一种重参数化广义特征金字塔网络

(reparameterized generalized-directional feature pyramid network, RepGFPN) 来替换特征融合网络(Neck)网络中原有的路径聚合网络(path aggregation network, PANet),增强特征图之间的信息交流和整合能力,提升模型对目标在不同尺度和分辨率下的感知能力;

3) 引入混合局部通道注意力(mixed local channel attention, MLCA)机制,以提高网络的表达效果,从而满足水下目标检测的检测性能;

4) 将损失函数用 SIoU 替代 CIoU(CompleteIoU),利用不同尺度的辅助边界,加速实现边框的回归。

1 YOLOv8

1.1 YOLOv8 算法原理

YOLOv8 算法建立在 YOLO 系列算法成功的基础上,结合 Darknet-53 网络结构思想并改进 YOLOv5 中 C3 模块为 C2f 模块,抛弃了前几代模型的 Anchor-Base 部分,提供了一个全新的 SOTA 模型^[12]。YOLOv8 可以完成检测、分类、分割任务^[13]。YOLOv8 被分为 n、s、m、l、x 这五种模型大小,它们主要区别在于模型大小和检测精度上的平衡,而 n、s、m 是在目标检测中常用的三种模型。

1.2 YOLOv8 网络结构

YOLOv8 的结构分为四个部分:输入端、主干网络(Backbone)、Neck 和检测头(Head)输出,其网络结构如图 1 所示。YOLOv8 在 Backbone 和 Neck 部分参考了 YOLOv7 的 ELAN 设计思想,最重要的改进是将 YOLOv5 的跨阶段部分连接模块(cross stage partial, CSP)换成了梯度流更丰富的跨阶段部分融合模块(cross stage partial network fusion, C2f)。C2f 模块由 2 个卷积-批归一化激活模块(CONv-BN-SiLU, CBS)以及多个 BottleNeck 构成,而 CBS 由一个卷积层、一个批量归一化层以及激活函数构成^[14]。C2f 的结构如图 2 所示。

1.3 YOLOv8 损失函数

YOLOv8 的损失函数主要用于训练神经网络,使其能够准确预测图像中物体的位置和类别。损失函数通常包括坐标回归损失、分类损失以及边界框损失等部分。

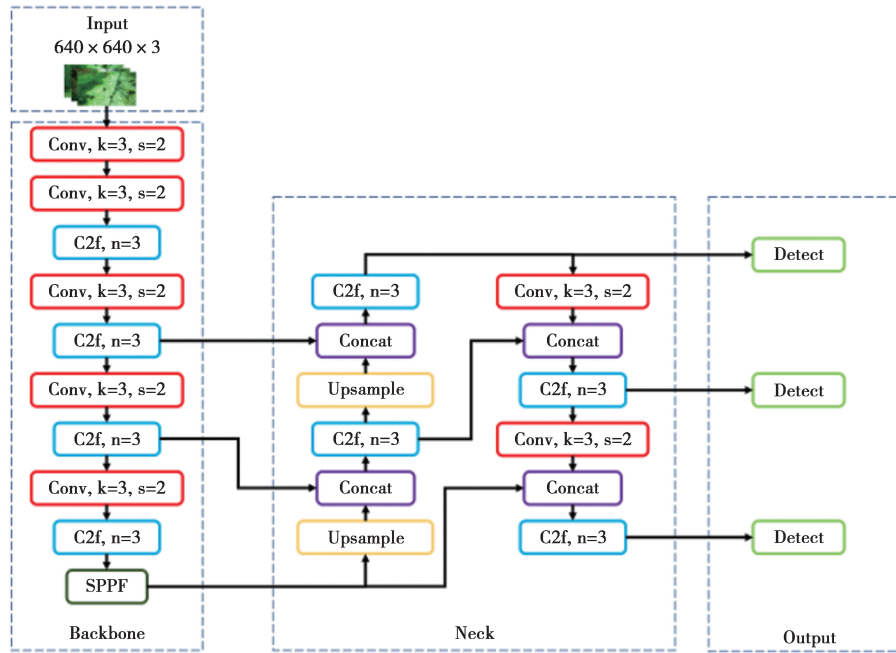


图 1 YOLOv8 网络结构

Fig.1 Network structure of YOLOv8

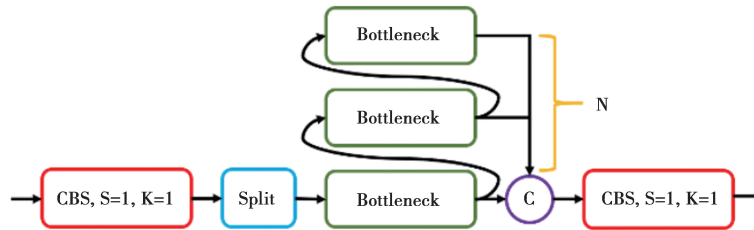


图 2 C2f 模块结构

Fig.2 Module structure of C2f

分类损失使用二元交叉熵 (BCE) 判断“是否为 此类”,并输出置信度,表示为

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \times \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (1)$$

式(1)中: y_i, p_i 分别代表样本标签的积分值和样本预测为正类的概率; N 代表样本总数。

由于去掉了对象损失,在输出中也去掉了“对象置信度”,因此对直接输出的各个类别的“置信度分数”求最大值,将其作为 anchor 框的置信度。

在回归损失中,使用分布焦点损失 (distribution focal loss, DFL) 将预测位置迅速聚焦到目标位置附近^[15],表示为

$$D_{DFL}(S_i, S_{i+1}) = -((y_{i+1} - y) \log(S_i) + (y - y_i) \log(S_{i+1})) \quad (2)$$

式(2)中: S_i, S_{i+1} 分别为网络输出的预测值、临近预

测值; y, y_i, y_{i+1} 分别为标签的实际值、标签积分值和临近标签积分值。

当预测位置聚集在目标附近时,通过 CIoU 损失函数 f_{CIoU} 进一步精确位置 P_{CIoU} 为

$$P_{CIoU} = I_{IoU} - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha \nu \quad (3)$$

$$L_{Loss, CIoU} = 1 - P_{CIoU} \quad (4)$$

式(3)–(4)中: $\frac{\rho^2(b, b^{gt})}{c^2}$ 为预测框 b 和真实框 b^{gt} 中心点欧氏距离的归一化值; α 为正权衡参数; ν 为长宽比一致性参数; I_{IoU} 为交并比 (intersection over union, IoU) 值^[16]。交并比是一种用于衡量目标检测算法精度的常见指标,在评估模型预测边界框与真实边界框之间重叠程度时非常有用,具体计算方式为交集面积与并集面积之比。其中,交集面积定义为模型预测的边界框与真实边界框重叠部分的面积。

积;并集面积定义为模型预测的边界框和真实边界框总覆盖的面积,即两个边界框各自的面积之和减去交集面积。IoU 取值范围在 0 到 1 之间,如果接近于 1,表示模型预测的边界框与真实边界框高度重叠,预测准确性较高;如果接近于 0,表示模型预测的边界框与真实边界框几乎没有重叠,预测准确性较低。 α 和 ν 的具体计算为

$$\nu = \frac{4}{\pi^2} \times \left(\arctan\left(\frac{w^{gt}}{h^{gt}}\right) - \arctan\left(\frac{w}{h}\right) \right)^2 \quad (5)$$

$$\alpha = \frac{\nu}{(1 - I_{IoU}) + \nu} \quad (6)$$

式(5)—(6)中: w 和 h 分别为预测框的宽和高; w^{gt} 和 h^{gt} 分别为真实值的宽和高。

2 YOLOv8 算法改进

2.1 整体框架

本文以 YOLOv8 为基础模型,提出了 ERMS-

YOLOv8 目标检测方法,专注于改善由于水下目标模糊导致检测精度低的问题。ERMS-YOLOv8 的框架如图 3 所示。首先,替换主干网络为 EfficientViT,增强模型对水下生物的特征提取能力,减少特征信息丢失;其次,在 Neck 部分采用 RepGFPN 网络模块,增强模型对水下生物高层语义和低级空间特征的提取和融合能力,从而使得模型获取更加丰富的特征信息;随后,引入 MLCA 注意力机制,使得模型同时融合通道信息,空间信息、局部通道信息和全局通道信息,增强模型的代表能力;最后,引入 SLoU 损失函数,提升模型对目标边界信息的提取能力,从而进一步提高检测精度。ERMS-YOLOv8 的输入为水下图片,首先在 EfficientViT 进行特征提取,得到 4 张不同尺度大小的特征图;然后,将这 4 张特征图送入 RepGFPN 的 Neck 中,通过特征融合提取特征信息的表征能力;最后,通过解耦检测头对目标进行分类和定位,最终得到水下目标检测结果。

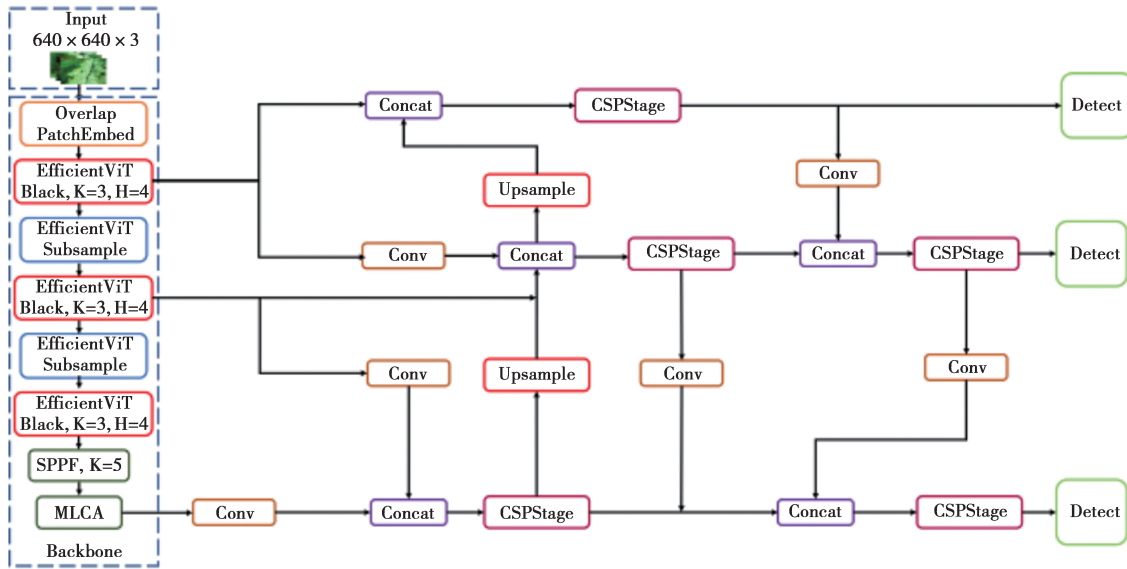


图 3 ERMS-YOLOv8 的网络结构

Fig.3 Network Structure of ERMS-YOLOv8

2.2 EfficientViT 主干网络及原理

YOLOv8 的传统主干网络结构由卷积层和残差模块相叠加构成。由于自身模型的计算冗余会限制训练和推理速度,因此 YOLOv8 不适合在水下这种复杂场景下部署。同时,传统的主干网络在处理跨尺度信息方面存在一定的局限,难以准确捕捉到水下目标的全局特征。为了解决上述问题,文献[17]提出一种高效视觉 Transformer 网络:EfficientViT,其主干网络结构如图 4 所示。

三明治布局块在前馈网络层(feed forward network, FFN)之间应用单个内存绑定多头注意力机制层(multi-head self-attention, MHSA),减少了 MHSA 中内存受限带来的时间成本,应用更多的 FFN 层来允许不同信道之间进行通信,从而提高了内存效率。具体来说,三明治布局应用了单一自注意力层 φ_i^A 进行空间混合,该层夹在 FFN 层的 φ_i^F 之间,计算公式为

$$X_{i+1} = \prod_{i=1}^N \varphi_i^A \times (\varphi_i^A \times (\prod_{i=1}^N \varphi_i^F(X_i))) \quad (7)$$

式(7)中, X_i 表示第 i 块的完整输入特征,布局将 X_i

变化为 X_{i+1} , 在单一自注意力层 φ_i^A 前后分别有 N 个 FFN 层的 φ_i^F 。该设计减少了模型中自注意力层造

成的内存时间损耗,并采用了更多的 FFN 层来实现不同特征通道之间的高效通信。

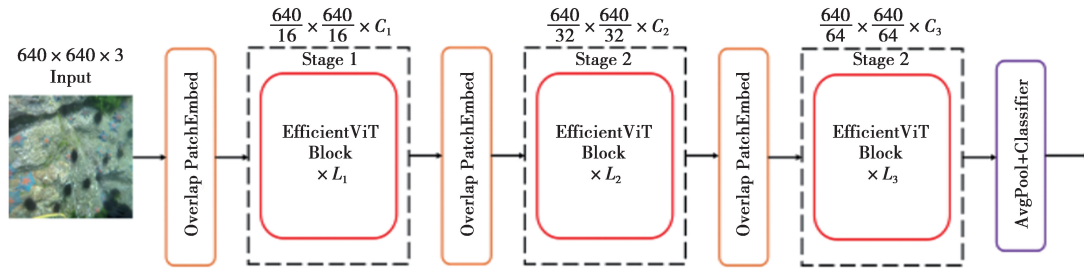


图 4 EfficientVit 主干网络结构

Fig.4 Backbone network structure of EfficientVit

为了解决 MHSA 中注意力头冗余导致的计算效率低下的问题,文献[17]参考 Efficient CNNs 的设计,提出了级联注意力模块(CGA)为每个头部提供完整特征的不同分割,从而显式地分解头部之间的注意力计算。这种注意力可以表示为

$$\tilde{X}_{ij} = \text{Attn}(X_{ij}W_{ij}^Q, X_{ij}W_{ij}^K, X_{ij}W_{ij}^V) \quad (8)$$

$$\tilde{X}_{i+1} = \text{Concat}[\tilde{X}_{ij}]_{j=1:h} W_i^P \quad (9)$$

式(8)–(9)中: X_{ij} 表示输入特征 X_i 的第 j 次分裂; \tilde{X}_{ij} 为输出特征; h 为注意力头部的总数; W_{ij}^Q 、 W_{ij}^K 、 W_{ij}^V 分别是映射将输入特征分割不同子空间的投影层; W_i^P 是将连接输出特征投影与输入维度一致的线性层;Attn 和 Concat 分别为注意力计算函数和合并数组函数。

2.3 RepGFPN

特征金字塔网络大多数采用卷积神经网络

(convolutional neural networks, CNN) 作为主干网络来提取多尺度特征。研究发现,相比于 CNN,特征金字塔网络(feature pyramid network, FPN)更强调高层次语义和低层次空间信息交换。因此,在目标检测模型中一般使用 FPN。

图 5 展示了三种特征金字塔网络的设计。图 5 中,虚线框表示设计的每个 FPN 层。传统 FPN 引入了自上而下的路径来融合多尺度特征。考虑到单向信息流的限制,文献[18]提出了 PANet 网络结构,增加了一个额外的自下而上的路径聚合网络。为了实现高效的跨尺度连接以及加权特征图融合,文献[19]提出了 BiFPN 网络结构,去除只有一条输入边的节点,并在同一层次上从原始输入添加额外的边。为了增加跨层和跨尺度的连接,文献[20]提出了 GFPN 结构。

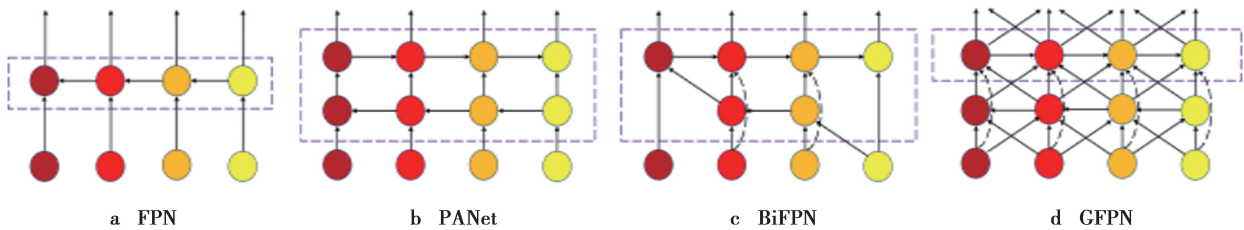


图 5 特征金字塔网络演化设计

Fig.5 Evolutionary design of feature pyramid network

GFPN 虽然能够充分交换不同层次的信息,但是带来了大量额外的上采样和下采样计算量。为了进一步提高目标检测的能力,本文引入了 RepGF-PN^[21],结构如图 6 所示。

RepGFPN 网络结构在 GFPN 的基础上减少了额外的上采样操作,通过 5 个 CSPStage^[22] 模块接受

更多节点输入,同时融合来自不同相邻层的不同尺度图像特征,在保证一定计算量的同时,增强特征复用及表达能力,提高目标检测的精准度。CSPStage 模块通过引入递归操作以及重参数化机制,不断优化特征的融合过程减少模型推理的时间。CSPStage 模块如图 7 所示。

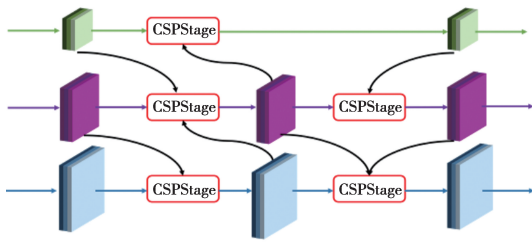


图 6 RepGFPN 结构

Fig.6 Structure of RepGFPN

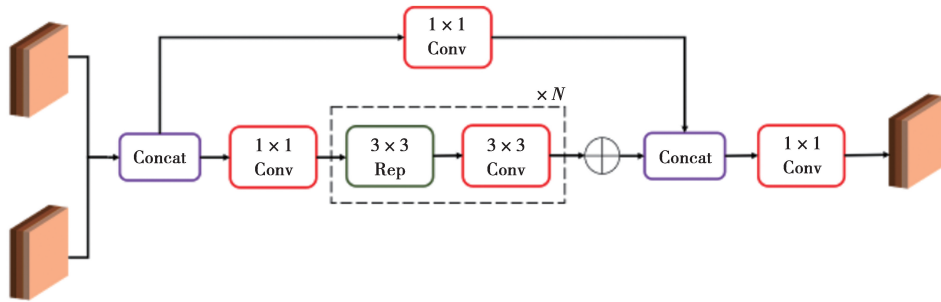


图 7 CSPStage 结构

Fig.7 Structure of CSPStage

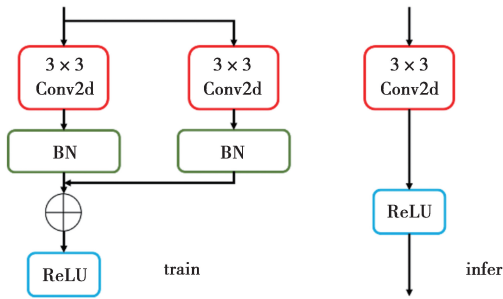


图 8 Rep 结构

Fig.8 Structure of Rep

2.4 MLCA 注意力机制

注意力机制是计算机视觉中应用最广泛的组成部分之一,它可以帮助神经网络突出重要元素,抑制无关元素。现有的注意力模块主要集中在通道领域,这两种注意力机制与人类大脑中基于特征和基于空间的注意力完全对应。然而,绝大多数通道注意力机制只包含通道特征信息,忽略了空间特征信息,而空间特征信息对于构建空间注意力图又至关重要。对空间特征信息的忽略导致模型表示效果或目标检测性能较差。空间注意力模块往往复杂且昂贵,一些注意力过程虽然包含空间方面的信息,但计算量和参数要求太大,无法直接将这信息纳入轻量级网络模型。文献[24]提出了一种轻量级的混合局部通道注意力(MLCA)模块来提高目标检测网络的性能。该模块可以同时包含信道信息和空间信

RepGFPN 网络中 Rep 的结构如图 8 所示。该结构推理阶段的主体仅由 3x3 卷积和 ReLU 堆栈组成,在训练阶段利用多分支形式的过参数化网络,在推理过程中将多分支融合为一个分支^[23]。多分支架构的优点是利于训练,缺点是不利于推理,通过结构重参数化技术实现训练时间和推理时间结构的解耦。

息,以及局部信息和全局信息,采用一维卷积加速方法来减少计算量和参数的数量,以提高网络的表达效果。

MLCA 的工作流程:第 1 步,将输入特征图经过局部平均池化和全局平均池化处理,局部池化提取局部区域的特征信息,全局池化捕捉整个特征图的统计信息;第 2 步,双分支结构处理,对全局信息分支使用全局平均池化将特征图压缩从而保留全局信息,对局部空间信息分支通过局部平均池化将特征图划分为多个局部块从而提取局部空间信息;第 3 步,对两个分支的特征分别经过一维卷积处理,压缩通道维度,同时保持空间维度不变;第 4 步,混合注意力融合,将局部和全局信息融合,生成最终的注意力特征图。

2.5 SIoU 损失函数

YOLOv8 采用 CIoU^[25] 作为边界框回归损失函数。尽管 CIoU 在提高精度方面有优势,但是也存在一些缺点,尤其是在处理不同大小的目标框时可能表现不佳,本文引入 SIoU^[26] 作为新的边界框损失函数。

SIoU 一般由 4 部分构成,包括角度损失、距离损失、形状损失和 IoU 损失。SIoU 的定义为

$$L_{Siou} = 1 - I_{IoU} + \frac{\Delta + \Omega}{2} \tag{10}$$

式(10)中,角度损失为

$$\Lambda = 1 - 2 \times \sin^2\left(\arcsin\left(\frac{C_x}{\sigma}\right) - \frac{\pi}{4}\right) \quad (11)$$

$$\frac{C_x}{\sigma} = \sin\theta \quad (12)$$

式(11)—(12)中: σ 为预测框中点与真实框中点之间的距离; C_x 为两框中点之间的高度差。距离损失的计算为

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma t}) \quad (13)$$

$$\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2, \rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)^2, \gamma = 2 - \Lambda \quad (14)$$

形状损失为

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega t})^\theta \quad (15)$$

$$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, \omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (16)$$

式(13)—(16)中: $b_{c_x}^{gt}$ 、 $b_{c_y}^{gt}$ 、 b_{c_x} 、 b_{c_y} 分别为真实框中心

坐标以及预测框中心坐标; w^{gt} 、 h^{gt} 、 w 、 h 分别为真实框和预测框的宽和高; C_w 、 C_h 分别为真实框和预测框中心点的宽度差和高度差。

3 实验结果与分析

3.1 数据集与实验环境

本文实验使用 URPC2021 和 DUO 公开水下数据集。其中,URPC2021 数据集包含 6 575 张图片以及对应的标注文件,检测目标包含海胆、海星、海参、扇贝四种水下生物,涵盖了真实的水下检测场景,本文将该数据集按照 7:3 的比例进行划分,其中 4 585 张作为训练集,1 990 张作为验证集。DUO 数据集包含 7 782 张图片以及对应的标注文件,同样涵盖海胆、海星、海参、扇贝这四种典型的水下生物,将该数据集同样按照 7:3 的比例进行划分,5 400 张作为训练集,2 382 张作为验证集。URPC2021 和 DUO 数据集的部分图片如图 9 所示。

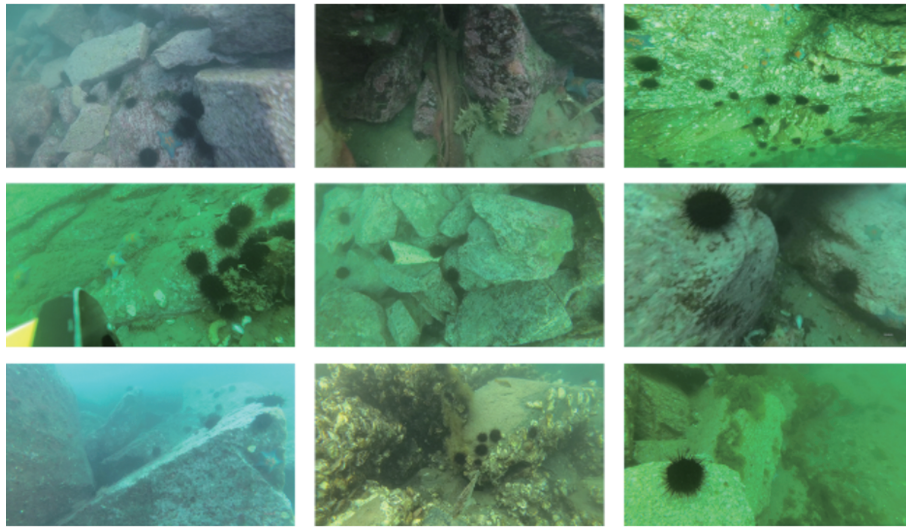


图 9 数据集的部分图片

Fig.9 Partial images of dataset

本实验的实验环境如表 1 所示。本文训练模型采用随机梯度下降(SGD)作为优化器,输入图像尺寸为 640×640,批量大小为 16,训练周期为 300 轮。

表 1 实验环境配置

Tab.1 Experimental environment configuration

实验环境	环境配置
编译环境	Python3.8
GPU	NVIDIA RTX 3080
CPU	Intel Core i9-10980XE
内存	128 G
集成开发环境	PyCharm

3.2 评价指标

为了测试本文基于 YOLOv8 改进模型的效果,本文利用平均精度(mean average precision, mAP)指标,其定义为

$$P_{AP} = \int_0^1 P(R) d(R) \quad (17)$$

$$P_{mAP} = \frac{1}{n} \sum_{i=1}^n P_{AP,i} \quad (18)$$

式(17)—(18)中:AP 为精准度值中的召回率 R 以及精准度 P 相关, R 、 P 的定义分别为

$$R = \frac{P_{TP}}{P_{TP} + N_{FN}} \quad (19)$$

$$P = \frac{P_{TP}}{P_{TP} + P_{FP}} \quad (20)$$

式(19)—(20)中: P_{TP} 为当检测结果与真值之间的IoU 值超过预定义的阈值时,认为该检测为真阳性(true positive, TP),表明目标识别成功;相反, P_{FP} 表示如果IoU 低于该阈值,则检测被分类为假阳性(FP),表明识别错误; N_{FN} 为错误预测的样本,被称为真阴性(FN)。

为了验证改进的算法是否满足水下目标检测实时性的要求,本文将从模型计算量(giga floating point operations, GFLOPs)和每秒帧率(frame per second, FPS)来衡量确定。FPS 越大,表示检测速度越快。

3.3 对比实验

为了验证本文目标检测算法的有效性,本文将对比不同算法在URPC2021数据集上的表现。具体实验结果如表2所示。

对比Faster-RCNN以及不同版本YOLO系列算法在URPC2021数据集上的mAP@50、GFLOPs、

FPS、参数量评价指标可见,YOLOv8 仅以 3.0×10^8 的参数量和8.1GFLOPs的计算量,在mAP@50上达到了81.0%,成为最佳基础模型。

随后,通过对比结果分析可知,本文所提出的算法相较于基础模型YOLOv8,虽然浮点计算量和网络参数量有所增加,但在mAP@50上提高了2.9百分点,在mAP@50-95上提高了3.3百分点。同时,经过实验可以得到本文算法FPS达到了230 虽然在一定程度上牺牲了处理图像特征的时间成本,但是仍然满足实时性需求。

最后,为了验证本文算法的优越性,与一些同类算法在相同实验设备和数据集的情况下进行对比实验,包括WCA-YOLOv8^[11]算法,IEMAyoloViT^[27]算法。实验结果表明,ERMS-YOLOv8 相较于这两个算法具有一定的优越性。

为了验证ERMS-YOLOv8的泛化能力,本文对比不同算法在DUO数据集上的表现,具体结果如表3所示,通过对比分析可知,ERMS-YOLOv8 相较于基础模型YOLOv8在mAP@50上提高了2.1百分点,在mAP@50-95上提高了2.7百分点。显然,本文算法在不同的数据集上都具有较好的检测效果。

表 2 不同算法在URPC2021数据集的实验结果对比

Tab.2 Comparison of experimental results of different algorithms under URPC2021 dataset

模型	mAP@50/%	mAP@50-95/%	GFLOPs	参数量/ 10^8
Faster-RCNN	69.4	38.8	27.8	42.0
YOLOv5	80.2	46.5	7.1	2.5
YOLOv6	79.6	45.8	11.8	4.2
YOLOv8	81.0	47.1	8.1	3.0
WCA-YOLOv8	83.7	49.0	13.1	6.6
IEMAyoloViT	83.3	45.7	9.5	4.0
本文算法	83.9	49.1	9.6	4.3

表 3 不同算法在DUO数据集下实验结果对比

Tab.3 Comparison of experimental results of different algorithms under DUO dataset

模型	mAP@50/%	mAP@50-95/%	GFLOPs	参数量/ 10^8
Faster-RCNN	70.1	54.0	27.8	42.0
YOLOv5	81.3	59.0	7.1	2.5
YOLOv6	80.9	59.5	11.8	4.2
YOLOv8	82.3	62.4	8.1	3.0
WCA-YOLOv8	84.3	64.5	13.1	6.6
IEMAyoloViT	83.8	63.7	9.5	4.0
本文算法	84.4	65.1	9.6	4.3

为了验证 MLCA 的有效性,本文使用 MLCA 和一些常用的注意力机制对 YOLOv8 进行对比实验。实验结果如表 4 所示。实验结果表明,当加入 ML-

CA 时,YOLOv8 在 URPC2021 数据集上表现出优异的性能,在几乎不增加模型参数规模的前提下,提升目标检测的精度。

表 4 YOLOv8 引入不同注意力机制在 URPC2021 数据集下实验结果对比

Tab.4 Comparison of experimental results of YOLOv8 introducing different attention mechanisms under URPC2021 dataset

注意力机制	mAP@ 50/%	mAP@ 50-95/%	GFLOPs	参数量
CPCA	81.3	47.1	8.4	127 040
SimAm	81.1	47.0	8.2	0
DA	81.4	47.2	8.4	266 624
MLCA	81.4	47.3	8.2	10

3.4 消融实验

下面通过消融实验验证本文算法的优越性,消融实验结果如表 5 所示。表 5 中模型 1 将 YOLOv8 主干网络替换为 EfficientViT,相较于 YOLOv8, mAP50%提升了 0.6 百分点,mAP@ 50-95%提升了 0.4 百分点,增强了模型的整体特征提取能力。模型 2 将 YOLOv8 的颈部网络替换为 RepGFPN,相对于 YOLOv8, mAP@ 50%提升了 1.0 百分点,mAP@ 50-95%提升了 0.6 百分点,增强了多尺度特征融合,

提高了模型的多尺度检测能力,在不带来额外计算负担的条件下,实现更高的精度。模型 3 在主干网络中添加 MLCA 注意力机制,mAP@ 50%提升了 0.4 百分点,mAP@ 50-95%提升了 0.2 百分点,提高了目标信息的提取能力。模型 4 将损失函数 CIoU 替换为 SIoU,mAP@ 50%和 mAP@ 50-95%得到了小幅度提升。模型 5、模型 6、模型 7 展示了将这些改进结合在一起会使得目标检测的效果得到更大幅度的提升。

表 5 消融实验结果

Tab.5 Results of ablation experiment

模型	EfficientViT	RepGFPN	MLCA	SIoU	mAP@ 50/%	mAP@ 50-95/%	GFLOPs	参数量/10 ⁸
YOLOv8					81.0	47.1	8.1	3.0
模型 1	✓				81.6	47.5	9.5	4.0
模型 2		✓			82.0	47.7	8.5	3.3
模型 3			✓		81.4	47.3	8.2	3.0
模型 4				✓	81.3	47.2	8.1	3.0
模型 5	✓	✓			83.0	48.3	9.4	4.3
模型 6	✓	✓	✓		83.6	48.8	9.5	4.3
模型 7	✓	✓	✓	✓	83.9	49.1	9.6	4.3

3.5 改进算法检测效果图

为了更加清晰地展示改进算法的检测效果,本文将选取部分照片进行对比验证,检测效果如图 10 所示。图 10 中第一列为原图、第二列为 YOLOv8 的检测图、第三列为改进算法的检测图。由图 10 可知,YOLOv8 算法在进行水下目标检测时存在漏检和误检的问题,检测效果不佳。本文算法有效改善了这些问题,在保持水下检测实时性的同时,提高模型对水下目标的检测精度。

3.6 鲁棒性实验

为了验证模型的抗干扰能力,本文选用两种场

景进行鲁棒性实验。

第一个鲁棒性实验通过改变高斯噪声的标准差来对图像添加不同强度的高斯噪声,不同强度高斯噪声的图像对比如图 11 所示,实验结果如表 6 所示。在低噪声条件下(高斯噪声强度为0.01),模型的检测精准度与召回率几乎没有受到影响,说明模型具有一定的抗干扰能力;在高斯噪声强度为 0.05 的条件下,模型的检测精准度与召回率大幅下降。这是因为高斯噪声会使图像变得模糊,目标的边缘和细节变得不清晰,导致模型难以识别目标。

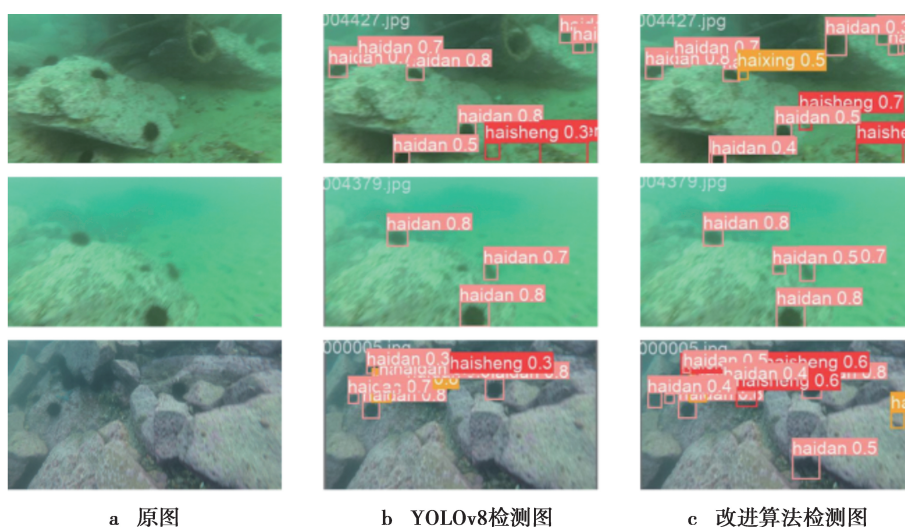


图 10 检测效果对比

Fig.10 Comparison of detection effects

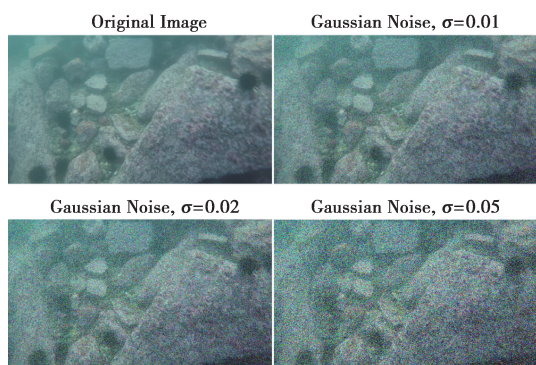


图 11 不同强度高斯噪声的图像对比

Fig.11 Comparison of images with different intensities of Gaussian noise

表 6 不同高斯噪声实验结果对比

Tab.6 Comparison of experimental results of different Gaussian noises

高斯噪声强度	mAP@ 50/%	召回率/%
0	83.9	77.3
0.01	83.7	77.0
0.02	81.2	74.5
0.05	75.4	68.5

第二个鲁棒性实验测试不同光照强度对检测精度与召回率的影响。不同强度光照的图像对比如图 12 所示。图 12 中,光照强度-0.2 表示图像变暗原图的 20%,0.2 表示变亮原图的 20%,实验结果如表 7 所示。当略微降低光照强度时,模型的检测精度与召回率几乎没有受到影响,说明模型具有一定的抗干扰能力,并且当适量提升光照强度时,检测

精度与召回率略有提升。

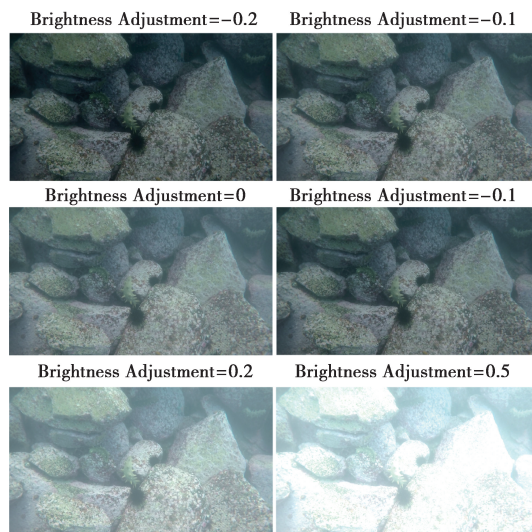


图 12 不同强度光照的图像对比

Fig.12 Comparison of images under different intensities of lighting

表 7 不同光照强度实验结果对比

Tab.7 Comparison of experimental results under different light intensities

光照强度	mAP@ 50/%	召回率/%
-0.2	83.7	77.0
-0.1	83.9	77.2
0	83.9	77.3
0.1	84.1	77.6
0.2	84.0	77.3
0.5	78.2	72.1

4 结束语

针对水下复杂环境导致出现水下目标模糊等情况,本文提出了一种改进 YOLOv8 的目标检测算法来提高水下目标的检测精度。本文引入了 EfficientViT 主干网络、RepGFPN 颈部网络和 MLCA 注意力机制来增强模型对水下目标生物的特征提取能力和跨尺度信息的融合能力,有效提升了对水下模糊目标的检测精度;替换模型损失函数为 SIoU,在不损失计算量的前提下,提升模型对边界信息的提取能力,有效提升了目标检测的精度。通过一系列的对比实验和消融实验,验证了所提出算法的有效性。本文方法仍有进一步改进空间,未来将通过相关技术进一步提高水下目标的检测精度,将模型部署在水下的实际场景中并得到应用。

参考文献:

- [1] 宋明龙.基于改进 SSD 的水下目标检测算法研究[D].淮南:安徽理工大学,2023.
SONG M L. Research on underwater target detection algorithm based on improved SSD[D]. Huainan: Anhui University of Science and Technology, 2023.
- [2] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 580-587.
- [4] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015: 1440-1448.
- [5] REN S Q, HE K M, 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.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 779-788.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016: 21-37.
- [8] 强伟,贺昱曜,郭玉锦,等.基于改进 SSD 的水下目标检测算法研究[J].西北工业大学学报,2020,38(04): 747-754.
QIANG W, HE Y Y, GUO Y J, et al. Research on underwater target detection algorithm based on improved SSD[J]. Journal of Northwestern Polytechnical University, 2020, 38(04): 747-754.
- [9] 陈宇梁,董绍江,孙世政,等.改进 YOLOv5s 的弱光水下生物目标检测算法[J].北京航空航天大学学报,2024,50(2): 499-507.
CHEN Y L, DONG S J, SUN S Z, et al. Enhanced YOLOv5s for low-light underwater biological-object detection[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(2): 499-507.
- [10] 陶洋,朱腾,钟邦乾,等.RepViTS-YOLOX:水下模糊及遮挡目标检测方法[J].计算机工程与应用,2024,60(13): 200-208.
TAO Y, ZHU T, ZHONG B Q, et al. RepViTS-YOLOX: detection method for blurred and occluded underwater targets[J]. Computer Engineering and Applications, 2024, 60(13): 200-208.
- [11] 李大海,李冰涛,王振东.基于改进 YOLOv8 的水下目标检测算法[J].计算机应用,2024,44(11): 3610-3616.
LI D H, LI B T, WANG Z D. Underwater object detection algorithm based on improved YOLOv8[J]. Journal of Computer Applications, 2024, 44(11): 3610-3616.
- [12] SHI K, WANG J, ZHU H, et al. Intelligent target detection of underwater optical images based on image enhancement [C]//Proceedings of Fourteenth International Conference on Information Optics and Photonics (CIOP 2023). Xi'an, China: SPIE, 2023: 1-7.
- [13] YI H, LIU B, ZHAO B, et al. Small object detection algorithm based on improved YOLOv8 for remote sensing [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024(17): 1734-1747.
- [14] 李姝,李思远,刘国庆.基于 YOLOv8 无人机航拍图像的小目标检测算法研究[J].小型微型计算机系统,2024,45(9): 2165-2174.
LI S, LI S Y, LIU G Q. Research on small target detection algorithm based on YOLOv8 UAV aerial images[J]. Journal of Chinese Systems, 2024, 45(9): 2165-2174.
- [15] LI X, WANG W, WU L, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[J]. Advances in Neural Information Processing Systems, 2020(33): 21002-21012.
- [16] YU J H, JIANG Y N, WANG Z Y, et al. Unitbox: An advanced object detection network [C]//Proceedings of

- the 24th ACM international conference on Multimedia. New York, USA: ACM, 2016: 516-520.
- [17] LIU X, PENG H, ZHENG N, et al. EfficientViT: Memory efficient vision transformer with cascaded group attention[C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 14420-14430.
- [18] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 8759-8768.
- [19] TAN M, PANG R, LE Q V. EfficientDet: Scalable and efficient object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 10778-10787.
- [20] JIANG Y Q, TAN Z Y, WANG J Y. GiraffeDet: A heavy-neck paradigm for object detection [EB/OL]. [2024-12-01]. <https://arxiv.org/abs/2202.04256>.
- [21] XU X Z, JIANG Y Q, CHEN W H, et al. Damo-YOLO: A report on real-time object detection design [EB/OL]. [2024-12-01]. <https://arxiv.org/abs/2211.15444>.
- [22] 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. Seattle, WA, USA: IEEE, 2020: 390-391.
- [23] 龙阳,肖小玲.改进 YOLOv8 的金属表面缺陷检测模型 [J].制造技术与机床, 2024(8): 187-194.
LONG Y, XIAO X L. Enhanced YOLOv8 for metal-surface defect inspection [J]. Manufacturing Technology & Machine Tool, 2024(8): 187-194.
- [24] 陈腾杰,李永安,张之好,等.基于改进 YOLOv8n+DeepSORT 的带式输送机异物检测及计数方法 [J].工矿自动化, 2024, 50(8): 91-98.
CHEN T J, LI Y A, ZHANG Z H, et al. Foreign-object detection and counting on belt conveyors via improved YOLOv8n plus DeepSORT [J]. Industry and Mine Automation, 2024, 50(8): 91-98.
- [25] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression [C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto California, USA: AAAI Press, 2020: 12993-13000.
- [26] DANG Y J, CHEN S L, MIAO H C, et al. Target detection from drone perspectives: Enhancing YOLOv5_3S with SiLU loss and SPD modules [C]//Cross Strait Radio Science and Wireless Technology Conference (CSR-SWTC). Guilin, China: IEEE, 2023: 1-3.
- [27] 施克权,李祺,隋皓,等. IEMAYOLOViT: 基于改进 YOLOv8 的水下目标检测算法 [J]. 电讯技术, 2025, 65(1): 54-62.
SHI K Q, LI Q, SUI H, et al. IEMA-yoloViT: An improved YOLOv8-based underwater object detection algorithm [J]. Telecommunication Engineering, 2025, 65(1): 54-62.

作者简介:

袁 泉,正高级工程师,硕士,主要研究方向为大数据分析、自然语言处理。E-mail:quanyuan@cqupt.edu.cn。

杨清泉,硕士研究生,主要研究方向为图像处理、目标检测。E-mail:2673507123@qq.com。

袁亚隆,硕士研究生,主要研究方向为目标检测。E-mail:1426355739@qq.com。

刘凤娟,硕士研究生,主要研究方向为目标检测。E-mail:2164319393@qq.com。

(编辑:陈文星)