

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

PAW-YOLOv7: 河道微小漂浮物检测算法

栾庆磊, 常昕昱, 吴叶, 邓从龙, 史艳琼, 陈梓华

引用本文:

栾庆磊, 常昕昱, 吴叶, 等. PAW-YOLOv7: 河道微小漂浮物检测算法[J]. *光电工程*, 2024, 51(4): 240025.

Luan Q L, Chang X Y, Wu Y, et al. PAW-YOLOv7: algorithm for detection of tiny floating objects in river channels[J]. *Opto-Electron Eng*, 2024, 51(4): 240025.

<https://doi.org/10.12086/oe.2024.240025>

收稿日期: 2024-01-25; 修改日期: 2024-03-12; 录用日期: 2024-03-12

相关论文

基于ZYNQ的轻量化YOLOv5声呐图像目标检测算法及实现

赵冬冬, 谢墩翰, 陈朋, 梁荣华, 沈伊, 郭新新

光电工程 2024, 51(1): 230284 doi: 10.12086/oe.2024.230284

局部和全局特征融合的太阳能电池片表面缺陷检测

陶志勇, 何燕, 林森, 易廷军, 张尧晟

光电工程 2024, 51(1): 230292 doi: 10.12086/oe.2024.230292

基于级联稀疏查询机制的轻量化火灾检测算法

张小雪, 王雨, 吴思远, 孙帮勇

光电工程 2023, 50(10): 230216 doi: 10.12086/oe.2023.230216

基于改进YOLOv5s的无人机图像实时目标检测

陈旭, 彭冬亮, 谷雨

光电工程 2022, 49(3): 210372 doi: 10.12086/oe.2022.210372

更多相关论文见光电期刊集群网站 



光电工程
Opto-Electronic Engineering

<http://cn.ojournal.org/oe>



 OE_Journal



Website

DOI: 10.12086/oe.2024.240025

PAW-YOLOv7: 河道微小漂浮物检测算法

栾庆磊^{1,2}, 常昕昱^{1,2}, 吴叶^{1,2}, 邓从龙^{1,2*},
史艳琼^{1,2}, 陈梓华^{1,2}

¹安徽建筑大学机械与电气工程学院, 安徽合肥 230601;

²安徽省工程机械智能制造重点实验室, 安徽合肥 230601

摘要: 河道漂浮物检测对于船舶自动驾驶以及河道清理有着重大意义, 但现有的方法在针对河道漂浮物目标尺寸小且互相遮挡、特征信息少时出现检测精度低等问题。为解决这些问题, 本文基于 YOLOv7, 提出了一种改进模型 PAW-YOLOv7。首先, 为了提高网络模型对小目标的特征表达能力, 构建了小目标物体检测层, 并将自注意力和卷积混合模块 (ACmix) 集成应用于新构建的小目标检测层; 其次, 为了减少复杂背景的干扰, 采用全维动态卷积 (ODConv) 代替颈部的卷积模块, 使网络具有捕获全局上下文信息能力; 最后, 将 PConv (partial convolution) 模块融入主干网络, 替换部分标准卷积, 同时采用 WIoU (Wise-IoU) 损失函数取代 CIoU, 实现网络模型计算量的降低, 提高网络检测速度, 同时增加对低质量锚框的聚焦能力, 加快模型收敛速度。实验结果表明, PAW-YOLOv7 算法在本文利用数据扩展技术改进的 FloW-Img 数据集上的检测精度达到 89.7%, 较原 YOLOv7 提升了 9.8%, 且检测速度达到 54 帧/秒 (FPS), 在自建的稀疏漂浮物数据集上的检测精度比 YOLOv7 提高了 3.7%, 能快速准确地检测河道微小漂浮物, 同时也具有较好的实时检测性能。

关键词: YOLOv7; 漂浮物检测; 混合卷积自注意力机制; 全维动态卷积; Wise-IoU 损失函数

中图分类号: TP391; X522

文献标志码: A

栾庆磊, 常昕昱, 吴叶, 等. PAW-YOLOv7: 河道微小漂浮物检测算法 [J]. 光电工程, 2024, 51(4): 240025

Luan Q L, Chang X Y, Wu Y, et al. PAW-YOLOv7: algorithm for detection of tiny floating objects in river channels[J]. *Opto-Electron Eng*, 2024, 51(4): 240025

PAW-YOLOv7: algorithm for detection of tiny floating objects in river channels

Luan Qinglei^{1,2}, Chang Xinyu^{1,2}, Wu Ye^{1,2}, Deng Conglong^{1,2*}, Shi Yanqiong^{1,2}, Chen Zihua^{1,2}

¹School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei, Anhui 230601, China;

²Anhui Province Key Laboratory of Intelligent Manufacturing of Construction Machinery, Hefei, Anhui 230601, China

Abstract: Detection of floating debris in rivers is of great significance for ship autopilot and river cleaning, but the existing methods in targeting floating objects in the river with small target sizes and mutual occlusion, and less feature information lead to low detection accuracy. To address these problems, this paper proposes a small target object detection method called PAW-YOLOv7 based on YOLOv7. Firstly, in order to improve the feature expression ability of the small target network model, a small target object detection layer is constructed, and the self-attention

收稿日期: 2024-01-25; 修回日期: 2024-03-12; 录用日期: 2024-03-12

基金项目: 安徽省科技重大专项 (202203a05020022); 安徽省研究生教育质量工程项目 (2022cxcsj147)

*通信作者: 邓从龙, dengconglong@ahjzu.edu.cn.

版权所有©2024 中国科学院光电技术研究所

and convolution hybrid module (ACmix) is integrated and applied to the newly constructed small target detection layer. Secondly, in order to reduce the interference of the complex background, the Omni-dimensional dynamic convolution (ODConv) is used instead of the convolution module in the neck, so as to give the network the ability to capture the global contextual information. Finally, the PConv (partial convolution) module is integrated into the backbone network to replace part of the standard convolution, while the WIoU (Wise-IoU) loss function is used to replace the CIoU. It achieves the reduction of network model computation, improves the network detection speed, and increases the focusing ability on the low-quality anchor frames, accelerating the convergence speed of the model. The experimental results show that the detection accuracy of the PAW-YOLOv7 algorithm on the FloW-Img dataset improved by the data extension technique in this paper reaches 89.7%, which is 9.8% higher than that of the original YOLOv7, the detection speed reaches 54 frames per second (FPS), and the detection accuracy on the self-built sparse floater dataset improves by 3.7% compared with that of YOLOv7. It is capable of detecting the tiny floating objects in the river channel quickly and accurately, and also has a better real-time detection performance.

Keywords: YOLOv7 model; floating object detection; self-attention and convolution hybrid module; omni-dimensional dynamic convolution; WIoU loss function

1 引言

在计算机视觉领域, 目标检测被视为一个核心的研究方向, 而随着深度学习技术的不断发展, 这一领域在近几年已经取得了空前的成果。该技术的应用潜力日益扩大, 包括但不限于辅助驾驶、智能机器人、智能交通系统、救援活动以及运动数据分析等多个方面。然而, 在面对道路标识、河道小目标障碍物、河道漂浮物识别、海上救援等小目标场景时, 由于场景复杂、目标像素少(像素点数小于 32×32)、目标间有遮挡等, 会出现漏检误检以及精度低的问题。因此, 针对小目标检测的研究对于提高各检测领域的识别精度和效率有着重大的意义。

目前基于深度学习的目标检测方法可以分为两阶段算法^[1]和单阶段算法^[2]。在两阶段目标检测算法中, 首先识别含有目标的候选区域, 随后用于对目标进行分类和定位, 代表算法有 Fast R-CNN^[3] (Fast Region-based CNN) 和 Faster R-CNN^[4] (Faster Region-based CNN)。尽管两阶段目标检测算法具有相对较高的检测精度, 但对于小目标检测来说体量大、耗时长。与两阶段目标检测方法相比, 单阶段算法没有区域建议网络, 而是直接在输入图像上生成相应的预测框来检测目标, 单阶段算法中的代表性算法有较早时期的 SSD^[5] (Single Shot MultiBox Detector) 和近年来发展迅速的 YOLO^[6-8] (You Only Look Once) 系列算法。基于直接预测分类和定位, 单阶段目标检测算法模型参数量小、检测速度更快, 在小目标检测场景下保证较高精度的同时有更好的实时表现。

河道微小漂浮物检测主要面临水面漂浮物的尺寸小、特征信息少、分散不均、背景干扰严重等问题。针对小目标尺寸小、特征信息少的问题, 马梁等人^[9]提出了一种基于动态选择机制的轻量化特征提取模块, 提出了基于自适应特征加权融合的 FPN (feature pyramid networks) 模块, 利用分组卷积的方式对所有特征通道分组且组间互不影响, 从而增加图像特征表达的准确性, 解决了过多的下采样可能导致小目标消失的问题; Zhang 等人^[10]采用 Res3Unit 结构对 YOLOv7 的骨干网络进行重构, 以提高网络模型架构获取更多非线性特征的能力, 并在特征融合区与探测头的连接处使用了 RFLA 模块中的高斯感受野方案, 以提高网络模型对图像中小目标的感受野, 解决了原网络小目标感知能力弱、特征提取不充分等问题; 针对目标分散不均、目标尺度差异大的问题, 陈旭等人^[11]基于 YOLOv5s 模型引入了可增大感受野的残差空洞卷积模块来提高空间特征的利用率, 并设计了一种特征融合模块 SCAM, 通过局部特征自监督的方式提高细节信息利用率, 提高了目标尺度差异大的小目标检测精度; 针对背景干扰严重、目标被遮挡的问题, 陆康亮等人^[12]提出了一种融合空间掩膜预测与点云投影的多目标跟踪算法, 将实例分割掩膜提取模型处理时序图像数据投影到对应的点云图像中, 获得最终的三维目标跟踪点云图像, 减少了遮挡产生的不利影响; Xiao 等人^[13]以 YOLOv7 为基础, 设计了一个 C3GhostV2 特征提取模块并引入无参数注意力 (SimAm) 机制, 增强了对目标特征的提取能力, 抑制了复杂背景对目标的干扰; Sun 等人^[14]在 YOLOv7

基础上设计并引入了 BiFusionNet 网络, 该网络更好地聚合了高层语义特征和低层语义特征, 并在此基础上引入坐标关注机制, 使网络在不增加模型复杂度的情况下关注更多关键特征, 提高了网络对小目标的检测能力。

尽管上述前人的研究已经有效地提高了各个领域的小目标图像检测的准确性, 但针对河道漂浮物实时检测的研究较少, 且在河道环境中, 由于漂浮物的密集和多样性, 上述方法的检测精度较低。此外, 不同环境光线条件下的背景干扰会影响特征提取, 进而降低检测的精度。YOLO 自 2016 年推出以来已广泛用于基于深度神经网络的对象识别和定位实时系统。YOLOv5 也被认为是一个经典版本, 而 YOLOv7 则在 YOLOv5 的基础上将模型重新参数化再引入网络架构, 并提出了一种新的高效层聚合网络架构 ELAN, 并提出了包含辅助头的训练方法。这使得 YOLOv7 成为 5 FPS 至 160 FPS 范围内最先进的物体探测器。因此, 本文在模型参数量小、检测速度更快的单阶段目标检测算法 YOLOv7 的基础上, 提出了一种改进的 PAW-YOLOv7 模型, 用于解决复杂情况下的河道微小漂浮物检测任务。针对漂浮物目标小的特点, 首先构建一种新的小目标检测网络, 通过增加新的特征输出层, 来获取更多的位置信息, 并配合 ACmix^[15] 自注意力和卷积混合模块, 增强其检测能力, 降低对小目标的漏检及误检。其次, 针对河道漂浮物背景复杂的特点, 将颈部标准卷积替换为全维动态卷积 (ODConv^[16]), 增加网络的全局上下文信息捕获能力, 进一步提升漂浮物检测的准确度; 最后, 考虑到网络的参数量, 提出 PC-ELAN 模块, 通过 PConv^[17] 部分卷积模块使网络轻量化, 减少模型参数, 加快推理速度, 同时替换损失函数为 WIoU^[18], 增加对低质量锚框的聚焦能力, 加快模型收敛速度。本文提出的河道微小漂浮物检测算法在保证检测精度的前提下, 简化了网络结构, 提高了检测速率, 满足漂浮物的实时检测要求。

2 YOLOv7 模型

YOLOv7 算法在 2022 年提出, 是目前 YOLO 系列最先进的算法之一, 它结合了众多优秀的先进思想。YOLOv7 采用重参数化卷积、模型缩放、动态标签分配等策略, 相较于大多数已知的目标检测算法, YOLOv7 在检测的准确性和速度方面显示出优势。这

一算法由三部分构成: 主干网络 (Backbone)、颈部网络 (Neck) 以及头部 (Head)。网络模型先预处理输入的图像, 将处理结果发送给主干网络, 主干网络的工作主要是负责提取图像的特征, 主干网络提取的图像特征再交由颈部网络做融合处理, 得到大、中、小三种尺寸的特征; 最后将融合后的特征发送到头部检测后输出结果。

其中主干网络由多个卷积层组成, 包括高效层聚合网络 (efficient layer aggregation networks, ELAN) 和最大池化卷积模块 MP 组成, 用于图像特征的提取; 颈部网络, 也被称为特征增强网络, 主要构成为 SPPCSPC、ELANW、CBS 和 MP 模块。SPPCSPC 用来增大感受野, ELAN-H 比 ELAN 多了两个输出进行拼接操作。这种网络通过双向特征融合机制, 即自下而上和自上而下的融合方法, 有效地将表层的详细信息与深层的语义信息结合起来。预测模块由重参数化卷积 RepConv 和检测头组成, 对颈部网络输出的加强特征层进行解耦, 实现对目标物体的定位和分类。YOLOv7 网络结构图如图 1 所示。

3 YOLOv7 网络模型改进

3.1 小目标检测层

在原 YOLOv7 网络结构中, 经过主干网络特征提取后在特征融合网络中进行特征融合, 最后在检测头输出 20×20、40×40、80×80 三个不同尺度的特征图。由于河道微小漂浮物的目标尺寸小、光照对场景目标干扰较大, 在检测过程中, 原 YOLOv7 对小目标检测的能力较弱, 出现大量的漏检误检, 检测精度不高, 本研究为了提升网络在识别较小目标方面的性能, 构建了新的特征融合模块, 增添了一个具有 160×160 输出尺寸的小目标检测层。该层的输入来自主干网络中第一个 ELAN 模块经过卷积后的特征图和上一个特征融合网络, 经过上采样得到 160×160 的特征图, 最后经过 ELAN-H 模块融合特征后输出一个具有高纬度信息的特征图, 该特征图感受视野小, 包含丰富的位置信息, 能有效降低小目标检测的漏检误检率。引入小目标检测层后使 YOLOv7 网络的目标检测尺寸范围增大, 提高了河道漂浮物检测的鲁棒性。图 5 标志处即为构建的新的的小目标检测网络。

3.2 PC-ELAN 模块

本文将部分卷积模块替换主干部分 ELAN 模块中

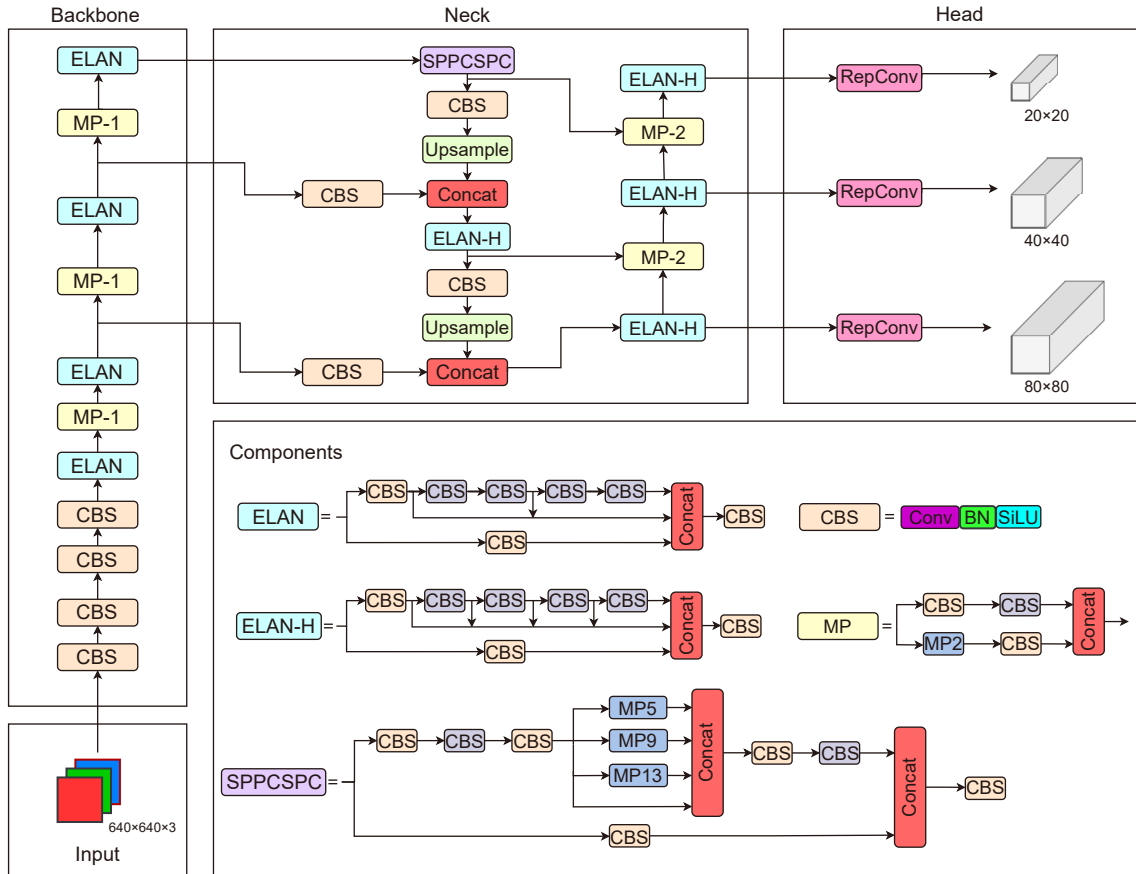


图 1 YOLOv7 网络结构图
Fig. 1 YOLOv7 network structure

的部分卷积核为 3×3 大小, 步长为 2 的卷积, 构成新模块 PC-ELAN, 以达到轻量化网络, 提高网络精度的效果。PC-ELAN 网络结构图如图 5 中模块部分所示。

ELAN 模块主要由 VoVNet^[19] 和 CSPNet^[20] 组成, 是一种高效的层聚合网络, 利用模块中的堆叠结构可以优化整个网络的梯度长度。ELAN 解决了当模型扩展时 (即模型达到一定深度并继续堆叠模块时), 由于精度增益降低导致的整体网络精度下降的问题。但是, ELAN 在参数数量和计算量上并非最优解, 因此本文引入部分卷积 (PConv) 替换 ELAN 模块中 3×3 卷积层的卷积核。相较于常见的卷积 DWConv^[21]、GhostNet^[22] 等轻量化网络虽然可以减少网络的参数量, 但是其存在频繁的内存访问, 对应的通道数会成倍增加从而导致延迟的增大并降低整体的运算速度和检测精度。PConv 同时减少参数量和内存访问次数, 能更有效地提取空间特征, 弥补了传统标准卷积的低效率。

PConv 仅对输入的部分通道应用常规卷积进行空间特征提取, 其余通道保持不变, 对于连续或常规内

存访问只计算第一个或最后一个连续通道, 将其视为整个特征图的代表。

$$r = \frac{c_p}{C}, \tag{1}$$

$$F = h \times w \times k^2 \times c_p^2, \tag{2}$$

$$F = h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p, \tag{3}$$

式 (1) 中: C 为常规卷积通道数量, c_p 为 PConv 通道的数量。式 (2) 和式 (3) 中: F 为每秒浮点运算次数 (FLOPs), h 为特征图的高度, w 为特征图的宽度, k 为卷积核的宽度和高度。

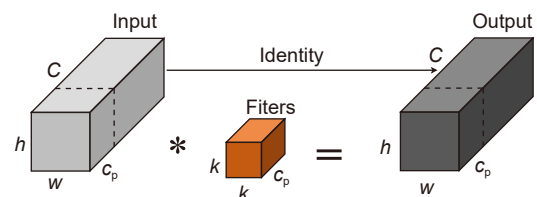


图 2 PConv 结构图
Fig. 2 PConv structure diagram

r 值一般设置为 1/4, 即 c_p 为输入通道 C 的 1/4, 代入式 (2), PConv 比传统的 Conv 的计算量减少了 15/16; 代入式 (3), PConv 比传统 Conv 的访问量减少了 3/4。

由此可见, 在 ELAN 模块中引入部分卷积构成的 PC-ELAN 网络, 其计算量和内存访问量都有大幅度的减少, 同时优化了网络的整体结构, 推理精度和速度都有所提升。

3.3 全维动态卷积模块

传统卷积也称静态卷积, 采用固定的内核大小和步长来卷积输入特征图, 无论输入数据如何变化, 权重都保持不变, 在每个位置都采取相同的权重参数进行卷积运算, 导致参数冗余。此外, 传统卷积只考虑局部像素关系, 无法捕捉全局背景信息。为了解决这些问题, Yang 等人^[23]提出了一种条件卷积 CondConv, 而后 Chen 等人^[16]提出了动态卷积 Dynamic Convolution, 通过引入一种可学习的核权重机制, 实现了对多个卷积核的线性加权。这些加权值与输入数据直接相关, 从而赋予了动态卷积对不同输入的依赖性。简而言之, 根据不同的输入数据, 动态卷积采用不同的卷积核。之后对这些不同的卷积核, 进行注意力加权。动态卷积可定义为

$$y = (\alpha_{w1} W_1 + \dots + \alpha_{wn} W_n) \times x, \quad (4)$$

其中: x 和 y 分别表示输入特征和输出特征, W_n 表示第 n 个卷积核。动态卷积采用单个注意力, 输入对于输出卷积核有相同的注意力值, 但其他三个维度 (卷积核的空间维度、输入通道维度以及输出通道维度) 都被忽略了。

为了考虑全维度, 本文引入了全维动态卷积 (ODConv), 创新性地运用了多维注意力机制和并行处理方法, 在卷积层的每一部分都从四个不同的维度分析和学习注意力值。这个全新的方法考虑了卷积核的四个方面: 输入通道的数量、核的感受范围、核的输出通道数量, 以及卷积核的总数。这些注意机制相辅相成, 它们在卷积核中的应用增强了 CNN 的特征提取能力。ODConv 的输出可以表示为

$$y = (\alpha_{w1} \odot \alpha_{f1} \odot \alpha_{c1} \odot \alpha_{s1} \odot W_1 + \dots + \alpha_{wn} \odot \alpha_{fn} \odot \alpha_{cn} \odot \alpha_{sn} \odot W_n) \times x, \quad (5)$$

其中: x 和 y 分别表示输入特征和输出特征, α_{wn} 是第 n 个卷积核的注意力标量, α_{fn} 、 α_{cn} 、 α_{sn} 分别代表输出通道维度、输入通道维度、空间维度的注意力标量。ODConv 的计算过程如图 3 所示。

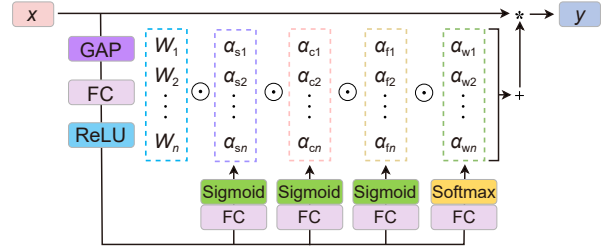


图 3 ODConv 的计算过程

Fig. 3 The process of calculating ODConv

ODConv 采用了四种类型的互补注意力机制, 这些机制逐步增强了卷积操作, 使其在位置、通道和核等不同维度上对输入具有区分性。这种方法能够有效捕捉更丰富的上下文信息, 从而提高性能。因此, 使用较少的卷积核的 ODConv 能显著增强网络对河道漂浮物的特征提取能力。

ODCBS 模块由三个部分组成。第一个部分是 ODConv, 它有助于在整个内核空间进行全面的特征提取。第二个部分是批量归一化 BN, 它可以缓解梯度消失和爆炸的问题。最后一个部分是 SiLU 激活函数, 它能促进平滑的梯度流动, 进一步提高模型的稳定性。

3.4 混合卷积自注意力机制模块

在小目标检测领域有两大挑战。第一, 小目标的像素小, 特征表示较弱, 携带的特征信息有限; 第二, 小目标多存在于复杂背景中, 容易与背景混淆, 通过标准的特征提取方法具有很大的挑战性。为了解决这些问题, 可以在特征提取过程中集成注意力机制。注意力机制能够使网络关注特定区域、学习分布模式、重新校准注意力并强调关键位置。

传统的卷积利用整个特征图的共享权重来关注局部的感受野信息, 通过共享权重, 可以降低参数量并提升训练的效率。另一方面, 自注意力模块利用相关像素对之间的相似函数, 通过动态计算注意力权重, 基于输入特征上下文进行加权平均操作。这种方法赋予了注意力模块足够的灵活性, 使其能够适应性地关注不同区域, 捕获更多特征, 能有效地减少复杂背景干扰的影响。

卷积和自注意力模块拥有不同的互补性质, 一种研究方案是将自注意力模块作为卷积模块的增强, 如 SENet^[24]、CBAM^[25]。另一种是将自注意力模块和卷积结合在一个整体中, 如 AA-ResNet^[26]。卷积和自注意力这两个模块在很大程度上依赖于相同的 1×1 卷

积运算, 因此提出了 ACmix 混合注意力机制模块, 它结合了传统卷积和自注意力模块的优点, 以最小的参数量将两者紧密的集成起来。ACmix 模块结构图如图 4 所示。

ACmix 分为两个阶段, 首先通过三次 1×1 的卷积投影, 将输入特征重组为 N 个片段的子特征集, 再将中间特征按照卷积和自注意力的方式分别复用和聚合, 最后将这两种技术产生的输出进行加权求和, 权重由两个可学习的标量决定, 如式 (6):

$$F_{out} = \alpha F_{conv} + \beta F_{att}, \quad (6)$$

其中: F_{out} 为最终输出, F_{conv} 和 F_{att} 分别代表卷积和自注意力的输出, 本文 α 和 β 设置为 1。

ACmix 的主要作用是对特征图进行通道注意力机制的调整。如图 5 所示, 在新引入的小目标检测层的第一个特征层后增加 ACmix 模块, 增强模型对远处小目标的特征感知和位置信息, 减少对干扰背景的关注。

3.5 损失函数改进

在 YOLOv7 中, 所使用的损失函数类型涵盖了边界损失、置信度损失以及分类损失。其中, 边界损失的计算采用了 CIoU 损失函数:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \left(\frac{v}{(1 - IoU) + v} \right) v, \quad (7)$$

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

其中: b 和 b^{gt} 表示预测框与真实框的中心点, ρ 为两个框中心点的欧式距离。 c 是两个框对角线的距离, w^{gt} 、 h^{gt} 代表真实框的宽度和高度, w 、 h 代表预测框的宽度和高度。由公式, CIoU 是基于预测框和真实框的宽高比进行计算的, 当两者宽高比一致时, 惩罚项就失效了。而现有的损失函数 GIoU^[27]、DIoU^[28]、CIoU、SIoU^[29] 也都是基于交并比或变体, 它们通过引入几何惩罚项来解决 IoU 损失的梯度消失问题。然而, 这些损失函数都没有考虑到训练数据中存在低质量样本的情况, 这可能会干扰模型的学习过程并降低定位性能。

因此本文采用 WIoU loss 来替换原 YOLOv7 的损失函数。WIoU 损失函数通过动态非单调聚焦机制, 用“离群度”替代 IoU 对锚框进行质量评估, 并提出了合理的梯度增益分配策略, 提高模型的定位能力。WIoU 包含三个版本, WIoUv1、WIoUv2、WIoUv3, 其中 WIoUv2 在 WIoUv1 上加入了单调聚焦系数, WIoUv3 在 WIoUv2 上加入了动态非单调聚焦系数。

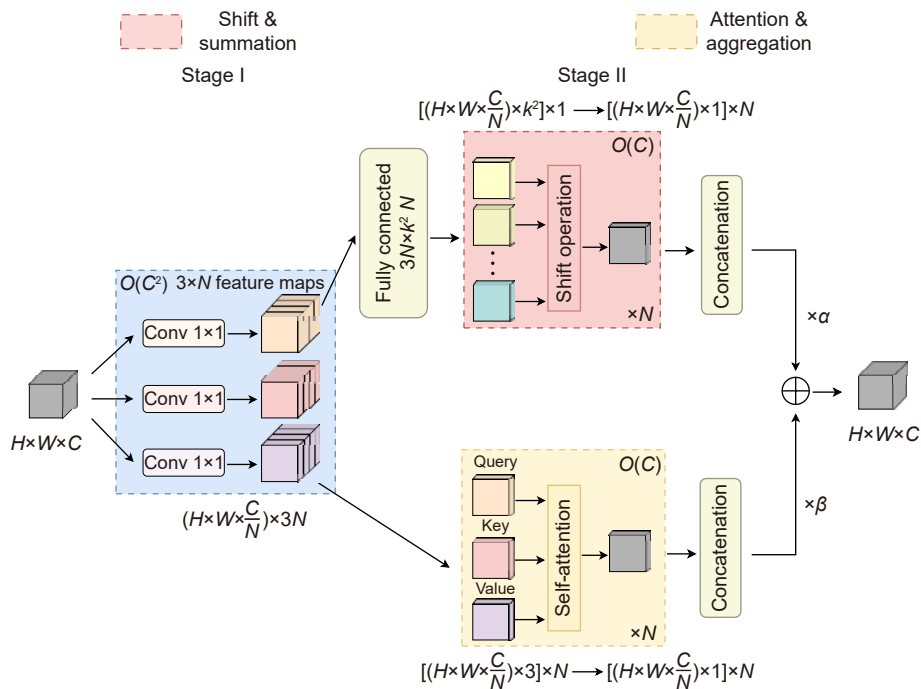


图 4 ACmix 结构示例图
Fig. 4 ACmix structure diagram

WIoU 系列损失函数的方程如式 (9) ~ (11) 所示:

$$L_{WIoUv1} = R_{WIoU} L_{IoU} = \exp \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*} L_{IoU}, \quad (9)$$

$$L_{WIoUv2} = L_{IoU}^{\gamma} L_{WIoUv1} = \left(\frac{L_{IoU}^*}{L_{IoU}} \right) L_{WIoUv1}, \quad (10)$$

$$L_{WIoUv3} = \frac{\beta}{\delta \alpha^{\beta - \delta}} L_{WIoUv1}, \beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty], \quad (11)$$

其中: W_g 、 H_g 代表最小边框的宽和高。*表示将 W_g 、 H_g 从计算图中分离, 防止 R_{WIoU} 产生阻碍收敛的梯度。 $\overline{L_{IoU}}$ 是动量的运行平均值, β 是离群值, 定义离群值是为了描述锚框的质量。离群值越小, 意味着锚点质量越高。相应地, 较小的梯度增益会分配给这类高质量的锚框, 从而使边界框回归更集中于平均质量的锚框。

改进后的 PAW-YOLOv7 网络结构图如图 5 所示, 其中红色虚线方框为改进后的各个模块。

4 实验结果与分析

4.1 实验数据集

为了验证模型在河道微小漂浮物目标上的检测效果, 本研究使用了 FloW-Img^[30]数据集, 和自建数据集。FloW-Img 数据集是一个关于城市水域漂浮垃圾检测的数据集。它由欧卡智船联合 AI 专家、图灵奖获得者 Yoshua Bengio 的 Mila 实验室、清华大学和西北工业大学的研究团队共同发布, 包括 2000 张图片。单个类别为 bottle。FloW-Img 包含了不同场景、不同天气和光照条件以及遮挡等条件。FloW-Img 数据集中分辨率小于 32×32 像素的目标检测物占总数据集的 55%, 小目标占比较多, 可以充分验证模型针对河道微小漂浮物的检测效果。将此数据集按照 6:2:2 的比例拆分成训练集、验证集和测试集。

由于使用的数据集规模较小, 而被检测的河道漂浮物规模也很小, 为了减少样本过少的问题, 防止过

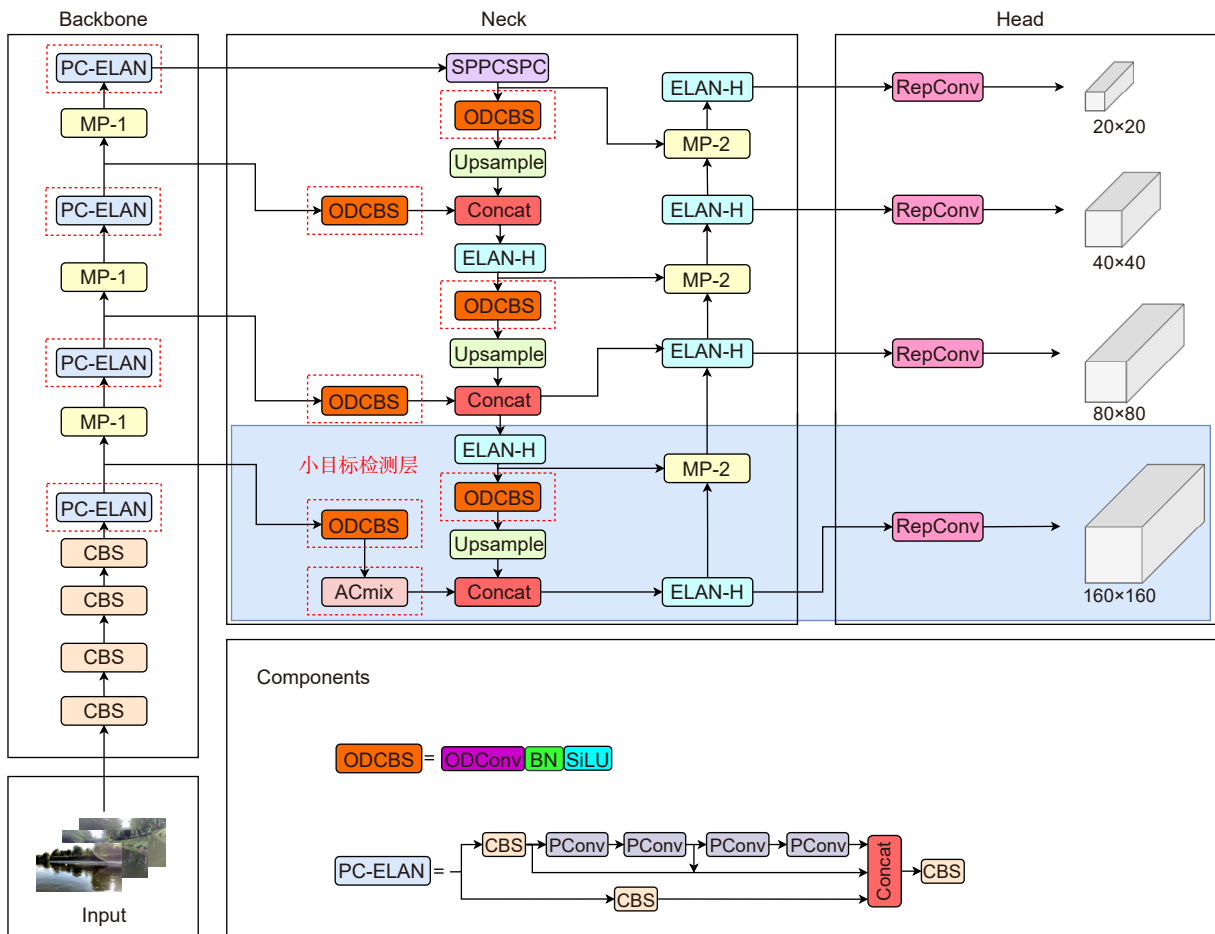


图 5 PAW-YOLOv7 网络结构图
Fig. 5 PAW-YOLOv7 network structure diagram

度拟合, 同时为了防止同一张图片进入不同数据集, 本文采用了 MixUp^[31] 和 Mosaic^[7] 数据增强方法对训练集增强, 使样本总量扩充到 3000 张图片, 增强后的数据集如图 6 所示, 并将图像裁切为 640 像素×640 像素作为输入图像。

由于河道漂浮物数据集较为匮乏, 本文自建微小漂浮物数据集以检测本文算法的泛用性。数据集包括了 2006 张河道图片, 尺寸为 4000 像素×3000 像素, 单个类别为 bottle。该数据集包含了不同的光照条件, 不同的背景复杂度, 其中分辨率小于 32 像素×32 像素的目标检测物占总数据集的 73%, 共 3213 个目标。将图像裁切为 640 像素×640 像素作为输入图像, 并按照 6 : 2 : 2 的比例拆分成训练集、验证集和测试集。

本文所采用数据集的尺度分布情况如图 7 所示。

4.2 评价指标与实验参数设置

为了准确验证本文所提算法的优越性, 采用平均精度 (mAP)、模型计算量 (floating-point operations per second, FLOPs)、每秒帧率 (frame per second, FPS)、准确率 (Precision, P)、召回率 (Recall, R) 作为评价指标。其中 P 和 R 可以通过公式计算得出:

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

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

其中: TP 代表识别出的真正例的数量, 而 FP 代表错误地将负例识别为正例的数量, FN 则是将正例错误地识别为负例的数量。FPS 表示模型每秒处理的图像帧数, 它用于评估模型的检测效率, 模型的 FPS 值越高, 其检测速度就越快; FLOPs 表示使用每秒十亿次的浮点运算, FLOPs 越大表示占用的计算资源越多。

本文实验在 Windows10 的操作环境下进行, 显卡采用 GeForce RTX 4090, 显存为 24 G, CUDA 的版本为 11.8; 使用 Python3.11 计算机语言, 在 Pytorch 深度学习开发框架下编写程序。Batch size 设为 16, epoch 设为 300 轮, 初始学习率为 0.01, 动量参数设置为 0.937, 超参数均为默认。

4.3 消融实验

为验证所提算法改进策略的有效性, 本文以 YOLOv7 为基础模型, 在不同数据集上进行不同改进策略之间的消融实验, 其中 Head (小目标检测层), PC-ELAN, ODCBS, ACmix, WIoU 为本文提出的



图 6 不同数据扩充方法的结果
Fig. 6 Results of different data expansion methods

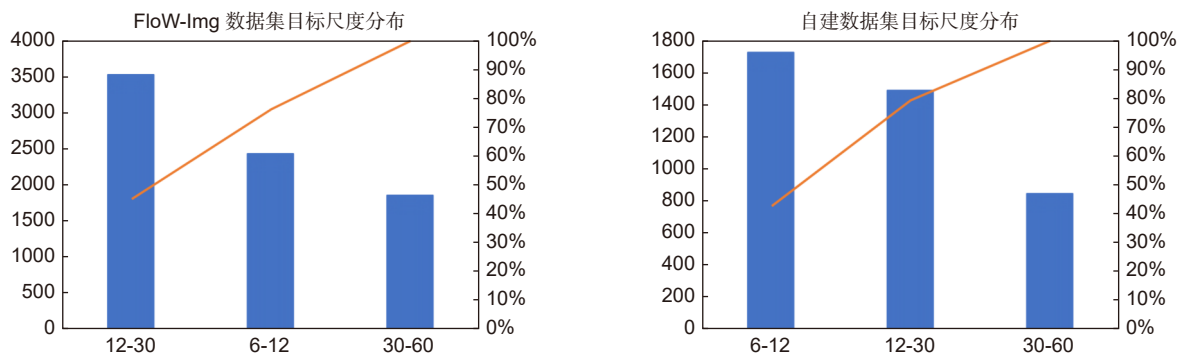


图 7 数据集目标尺度分布
Fig. 7 Object scale distribution of the dataset

改进策略, 表中“√”表示引入此改进策略。其中, 单独添加 ACmix 模块时, 由于未添加小目标检测层, 此时该模块添加在第一个特征融合层中。在 FloW-Img 数据集上的消融实验结果如表 1 所示。

表 1 消融实验结果

Table 1 Results of ablation experiments

组别	Head	ACmix	ODCBS	WIoU	PC-ELAN	mAP/%	FLOPs/G	FPS
1						79.9	105.4	101
2	√					81.1	119.5	86
3		√				85.6	101.6	75
4			√			80.8	109.8	97
5				√		81.5	105.4	108
6					√	78.1	83.7	124
7	√	√				87.3	115.3	56
8	√	√	√			88.2	119.2	47
9	√	√	√	√		90.8	119.2	48
10	√	√	√	√	√	89.7	97.8	54

由表 1 可得, 第一组为原 YOLOv7 算法, 平均精度 (mAP) 为 79.9%。第二组为加入小目标检测层, 精度上升了 1.2%, 说明小目标检测层可以有效增加小目标尺寸的检测范围。第三组引入了 ACmix 模块使平均精度有大幅提升, 提高了 5.7%, 说明 ACmix 可以提高网络对小目标空间位置的特征感知度。第四组用全维动态卷积替换了部分标准卷积, 使 mAP 提升了 0.9%。第五组将原 CIoU 损失函数替换为 WIoU, 使网络对普通质量锚框更加关注, mAP 提升了 1.6%。第六组将 ELAN 模块替换为 PC-ELAN, 虽然精度有所损失, 但是模型的运算量下降了 20.6%, 使模型更轻量化。第七组将第一组与第二组相结合, 综合了两者的优点, 精度提升了 7.4%。第八组和第九组分别在前一组基础上加入了全维动态卷积和优化了损失函数, 使网络模型精度进一步提升, 但是运算量较高。第十组在第九组的基础上融入了部分卷积, 得到本文所提的最终算法, 平均精度 (mAP) 较原 YOLOv7 提升了 9.8%, 运算量下降了 7.2%, 同时 FPS 也能维持在 54, 能够满足实时检测, 保证了精度与速度的有效平衡。

4.4 对比实验

为了验证本文算法的优越性, 将本文算法与当下主流算法 YOLOv5, YOLOv7, YOLOv8 等进行对比, 配置环境以及训练参数均一致, 超参数均为默认设置。

且为了保证实验的公平性, 所有实验均未使用预训练模型。本文分别在河道微小漂浮物稀疏的 FloW-Img 数据集和自建河道数据集上进行算法对比实验。

由表 2 可知, 在 FloW-Img 微小漂浮物数据集的检测中, 本文 PAW-YOLOv7 算法平均精度达到了 89.7%, 检测性能优于其它算法, 模型运算量也低于大部分主流算法。在两阶段算法中, Faster R-CNN 算法比本文算法精度低了 12.9%。在单阶段算法中, SSD 算法比本文算法精度低 16.4%; YOLOv3 算法精度相比前两种有所提升, 但仍比本文算法低 4%; YOLOv5s 和 TPH-YOLOv5 在河道微小漂浮物检测中表现出优异的实时性能, FPS 最高达到了 236, 但是检测精度明显低于本文算法; YOLOv8l 算法在 FPS 方面比本文算法高出 63, 但是精度比本文算法低 3.3%, 且运算量过高, 参数量较大难以应用于河道漂浮物检测。本文算法在兼顾高精度的同时降低了模型的运算量, 且 FPS 大于 30 满足了实时检测要求, 在综合性能上优于其它算法。

表 2 FloW-Img 数据集各算法对比实验数据

Table 2 Comparative experimental data of each algorithm on FloW-Img dataset

算法	mAP/%	FPS	FLOPs/G	Params/M
SSD	73.3	71	78.4	26.3
Faster R-CNN	76.8	63	75.1	137.1
YOLOv3	85.7	125	96.3	61.5
YOLOv5s	84.1	236	64.7	10.7
TPH-YOLOv5	82.3	69	125.6	26.1
YOLOv7	79.9	101	105.4	37.2
YOLOv8l	86.4	117	155.4	43.6
PAW-YOLOv7	89.7	54	97.8	25.4

表 3 中可以看出, 本文算法在漂浮物尺寸小的自建数据集上的检测精度相较于原 YOLOv7 算法提高了 3.7%, 检测精度远超表中其它主流算法。本文算法在保证高精度的同时, 其网络模型运算量也比大多数算法低, FPS 提升到了 68, 能很好地完成实时检测任务。

由表 1、表 2 和表 3 可以得出, 本文提出的算法在复杂场景、像素低的河道漂浮物检测中性能优异, 结构更加简单, 且超过其它主流算法。由于本文算法加深了整体网络的深度, 扩大了模型结构导致本文算法的实时检测性能降低, 但依然满足实时检测要求。

表 3 自建数据集各算法对比实验数据

Table 3 Comparative experimental data of each algorithm on self-constructed dataset

算法	mAP/%	FPS	FLOPs/G	Params/M
SSD	57.9	67	78.4	26.3
Faster R-CNN	61.3	61	75.1	137.1
YOLOv3	59.7	137	96.3	61.5
YOLOv5s	66.5	242	64.3	10.7
TPH-YOLOv5	63.7	74	125.6	26.1
YOLOv7	68.1	98	105.1	37.2
YOLOv8l	65.9	106	155.4	43.6
PAW-YOLOv7	71.8	68	97.7	25.4

4.5 可视化分析

本文通过比较在各种复杂情况下的检测效果来展示 YOLOv7 和本文算法的性能。这些复杂情况包括目标距离远、受到强烈光线照射以及被遮挡等因素造成的干扰。图 8 展示了在这些不同情境下两种算法的检测结果。

从图 8 (a) 的较远场景小目标检测结果对比可以看出, YOLOv7 在检测远距离、特征不明显的小目标时会出现较多的漏检和误检情况, 例如正下方将树叶错检为漂浮物瓶子。本文算法引入了小目标检测层使目标检测尺寸范围增大, 并通过 ACmix 模块增强模型对远处小目标的特征感知和位置信息, 因此在针对漏检和误检的问题有较好表现, 同时在远距离特征不明显的小目标场景中仍有较好表现。

从图 8 (b) 的强光干扰场景检测结果对比可以看出 YOLOv7 算法仅检测出少量目标, 本文中引入的全维动态卷积技术, 能够捕捉更多的上下文信息, 从而显著增强了卷积在特征提取方面的效能。本文算法可以有效削弱强光影响。

从图 8 (c) 的小目标严重遮挡场景中, YOLOv7 在针对遮挡目标时普遍出现漏检情况, 而本文可以有效识别被遮挡目标。

由图 9 可以看出, 在自建数据集中针对极小像素目标的检测, 本文算法能准确检测到微小漂浮物, 且

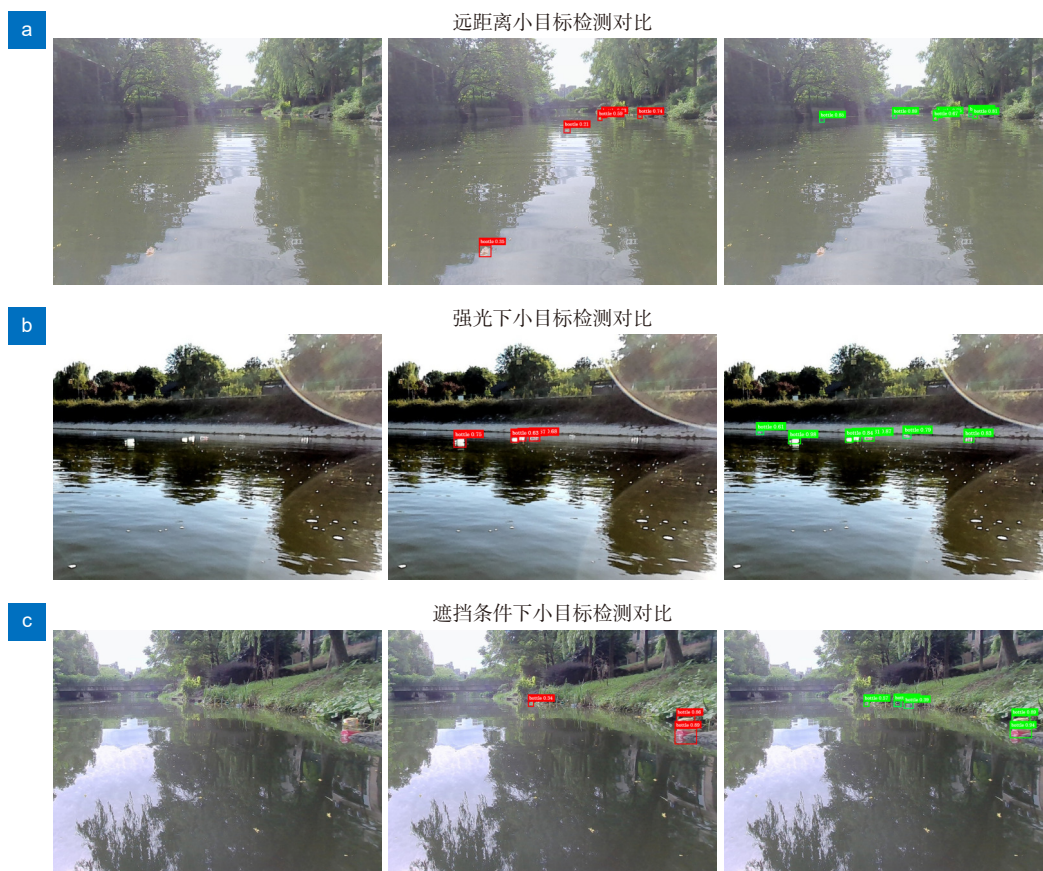


图 8 不同场景下不同算法目标检测结果。(左) 检测图片; (中) YOLOv7 模型; (右) 本文算法

Fig. 8 Target detection results of different algorithms in different scenes. Left: detection image, Center: YOLOv7 model, Right: algorithm of this paper

检测置信度高于原 YOLOv7 算法。本文算法在两个不同数据集的目标检测中均有较好表现。

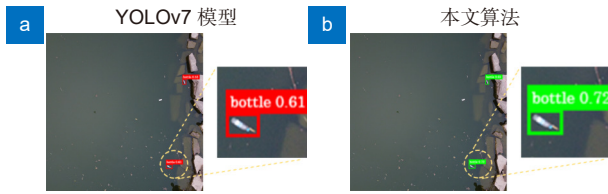


图 9 自建数据集检测精度对比

Fig. 9 Comparison of detection accuracy of self-built datasets

此外, 为了更形象地展示本文算法在检测过程中对检测目标关注度的优化效果, 将检测结果绘制成热力图的形式, 网络着重关注的区域采用暖色系绘制, 如图 10 所示。

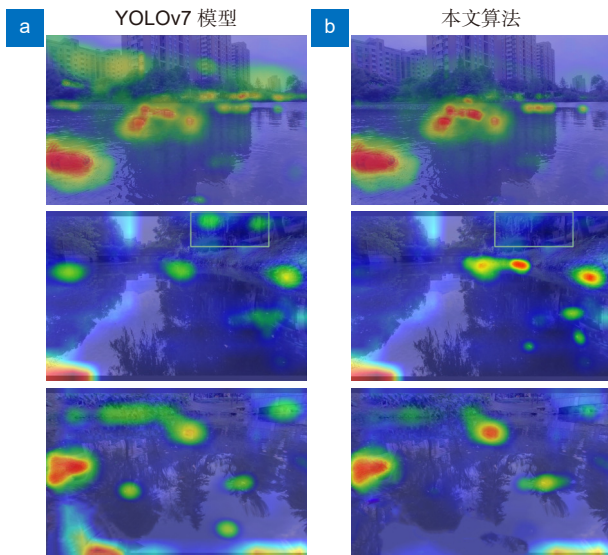


图 10 不同算法热力图对比结果

Fig. 10 Comparison results of heat maps with different algorithms

由热力图对比可得出, 在改进 YOLOv7 算法后, 网络检测更关注于检测目标和削弱复杂背景的干扰, 更加注重图像中目标位置检测信息的提取, 进而提升检测精度。

5 结论

本次研究提出了一种名为 PAW-YOLOv7 的河道微小漂浮物检测算法。首先, 针对网络对小目标感知度差, 在原有的最后一个特征融合模块后构建了一个输出尺寸为 160×160 的小目标检测层, 并在小目标检测层引入 ACmix 注意力机制模块, 增强模型对远处小目标的特征感知和位置信息, 使 mAP 提高了 7.4%。其次, 为了提高网络的特征提取能力以及对普通锚框

的关注度, 引入了全维动态卷积 (ODConv) 并替换了 WIoU 损失函数, 检测精度达到了 90.8%。最后, 在主干部分将 PConv 与 ELAN 模块结合为 PC-ELAN 模块来大幅减少网络的运算参数量, 以损失少量精度来换取网络参数量的大幅度下降。本文改进的算法在 FloW-Img 数据集上有较好的表现, mAP 提高了 9.8%, 在自建的河道漂浮物数据集上 mAP 提高了 3.7%, 表明该算法在针对河道的微小漂浮物目标检测性能有大幅提高, 减小模型体量的同时也保证了检测速度满足了实时性检测需求。下一步的工作重点是进一步探索研究领域, 旨在简化模型和减少模型参数量, 以便更有效地部署模型, 并使用多种数据集来检验其广泛适用性。

参考文献

- [1] Yan L, Yamaguchi M, Noro N, et al. A novel two-stage deep learning-based small-object detection using hyperspectral images[J]. *Opt Rev*, 2019, 26(6): 597-606.
- [2] Zhu H, Zhou S Y, Liu X, et al. Survey of single-stage object detection algorithms based on deep learning[J]. *Ind Control Comput*, 2023, 36(4): 101-103.
朱豪, 周顺勇, 刘学, 等. 基于深度学习的单阶段目标检测算法综述[J]. *工业控制计算机*, 2023, 36(4): 101-103.
- [3] Girshick R. Fast R-CNN[C]//*Proceedings of the 2015 IEEE International Conference on Computer Vision*, 2015: 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>.
- [4] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Trans Patt Anal Mach Intellig*, 2017, 39(6): 1137-1149.
- [5] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 21-37. https://doi.org/10.1007/978-3-319-46448-0_2.
- [6] Redmon J, Farhadi A. YOLOv3: An incremental improvement[Z]. arXiv: 1804.02767, 2018. <https://doi.org/10.48550/arXiv.1804.02767>.
- [7] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[Z]. arXiv: 2004.10934, 2020. <https://doi.org/10.48550/arXiv.2004.10934>.
- [8] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops*, 2021: 2778-2788. <https://doi.org/10.1109/ICCVW54120.2021.00312>.
- [9] Ma L, Gou Y T, Lei T, et al. Small object detection based on multi-scale feature fusion using remote sensing images[J]. *Opto-Electron Eng*, 2022, 49(4): 210363.
马梁, 苟于涛, 雷涛, 等. 基于多尺度特征融合的遥感图像小目标检测[J]. *光电工程*, 2022, 49(4): 210363.
- [10] Zhang Y, Sun Y P, Wang Z, et al. YOLOv7-RAR for urban vehicle detection[J]. *Sensors*, 2023, 23(4): 1801.
- [11] Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. *Opto-Electron Eng*, 2022, 49(3): 210372.
陈旭, 彭冬亮, 谷雨. 基于改进 YOLOv5s 的无人机图像实时目标

- 检测[J]. 光电工程, 2022, 49(3): 210372.
- [12] Lu K L, Xue J, Tao C B. Multi target tracking based on spatial mask prediction and point cloud projection[J]. *Opto-Electron Eng*, 2022, 49(9): 220024.
陆康亮, 薛俊, 陶重霖. 融合空间掩膜预测与点云投影的多目标跟踪[J]. 光电工程, 2022, 49(9): 220024.
- [13] Xiao Z, Wan F, Lei G B, et al. FL-YOLOv7: a lightweight small object detection algorithm in forest fire detection[J]. *Forests*, 2023, 14(9): 1812.
- [14] Sun Y, Yi L, Li S, et al. PBA-YOLOv7: an object detection method based on an improved YOLOv7 network[J]. *Appl Sci*, 2023, 13(18): 10436.
- [15] Pan X R, Ge C J, Lu R, et al. On the integration of self-attention and convolution[C]//*Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 805–815.
<https://doi.org/10.1109/CVPR52688.2022.00089>.
- [16] Li C, Zhou A J, Yao A B. Omni-dimensional dynamic convolution[C]//*The Tenth International Conference on Learning Representations*, 2022.
- [17] Chen J R, Kao S H, He H, et al. Run, Don't walk: chasing higher FLOPS for faster neural networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 12021–12031.
<https://doi.org/10.1109/CVPR52729.2023.01157>.
- [18] Tong Z J, Chen Y H, Xu Z W, et al. Wise-loU: bounding box regression loss with dynamic focusing mechanism[Z]. arXiv: 2301.10051, 2023. <https://doi.org/10.48550/arXiv.2301.10051>.
- [19] Lee Y, Hwang J W, Lee S, et al. An energy and GPU-computation efficient backbone network for real-time object detection[C]//*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019: 752–760.
<https://doi.org/10.1109/CVPRW.2019.00103>.
- [20] 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)*, 2020: 1571–1580.
<https://doi.org/10.1109/CVPRW50498.2020.00203>.
- [21] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 1800–1807.
<https://doi.org/10.1109/CVPR.2017.195>.
- [22] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 1577–1586. <https://doi.org/10.1109/CVPR42600.2020.00165>.
- [23] Yang B, Bender G, Le Q V, et al. CondConv: conditionally parameterized convolutions for efficient inference[C]//*Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019: 117.
- [24] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.
<https://doi.org/10.1109/CVPR.2018.00745>.
- [25] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*15th European Conference on Computer Vision*, 2018: 3–19.
https://doi.org/10.1007/978-3-030-01234-2_1.
- [26] Bello I, Zoph B, Le Q, et al. Attention augmented convolutional networks[C]//*2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019: 3285–3294.
<https://doi.org/10.1109/ICCV.2019.00338>.
- [27] Rezatofighi H, Tsai N, Gwak J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019: 658–666.
<https://doi.org/10.1109/CVPR.2019.00075>.
- [28] Zheng Z H, Wang P, Liu W, et al. Distance-loU loss: faster and better learning for bounding box regression[C]//*Thirty-Seventh AAAI Conference on Artificial Intelligence*, 2019.
<https://doi.org/10.1609/aaai.v34i07.6999>.
- [29] Gevorgyan Z. SloU loss: more powerful learning for bounding box regression[Z]. arXiv: 2205.12740, 2022.
<https://doi.org/10.48550/arXiv.2205.12740>.
- [30] Cheng Y W, Zhu J N, Jiang M X, et al. FloW: a dataset and benchmark for floating waste detection in inland waters[C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision*, 2021: 10933–10942.
<https://doi.org/10.1109/ICCV48922.2021.01077>.
- [31] Zhang H Y, Cissé M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[C]//*6th International Conference on Learning Representations*, 2017.

作者简介



栾庆磊 (1979-), 男, 硕士, 副教授, 研究方向为机器视觉、图像处理。

E-mail: qingleiluan@ahjzu.edu.cn



【通信作者】邓从龙 (1990-), 安徽马鞍山人, 助理实验师, 研究方向为图像处理、智能控制。

E-mail: dengconglong@ahjzu.edu.cn



常昕昱 (2000-), 男, 硕士研究生, 研究方向为机器视觉、图像处理。

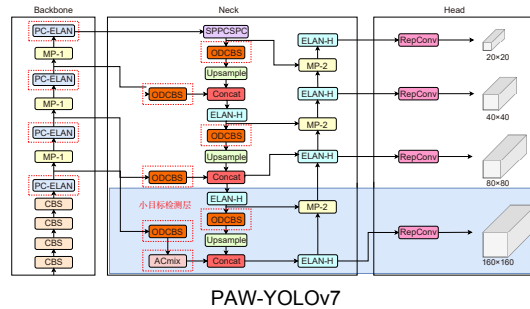
E-mail: black299@163.com



扫描二维码, 获取PDF全文

PAW-YOLOv7: algorithm for detection of tiny floating objects in river channels

Luan Qinglei^{1,2}, Chang Xinyu^{1,2}, Wu Ye^{1,2}, Deng Conglong^{1,2*}, Shi Yanqiong^{1,2}, Chen Zihua^{1,2}



Overview: In recent years, with the continuous development of deep learning technology, target detection has achieved unprecedented results in the field of computer vision and has been applied to a large number of scenarios, such as intelligent driving, rescue activities, and motion data analysis. In many target detection tasks, river float detection is of great significance for automatic ship driving and river cleaning, at present, target detection has a better performance in medium and large target detection, but the accuracy and real-time performance in the face of detection of tiny floats in the river is poor and the model volume is large. Since the detection of tiny floating objects in the river channel mainly faces the problems of small target size, little feature information, uneven dispersion, and serious background interference of floating objects on the water surface, the existing methods have a good performance in target detection of small floating objects in the river channel. Existing methods for the detection of floating objects in the river channel will face these difficulties such as low detection accuracy, leakage and false detection, bad real-time, and other problems. In order to solve these problems, this paper proposes an improved river small target detection model PAW-YOLOv7 based on YOLOv7. Firstly, in order to improve the feature expression ability of the network model for small targets, a small target object detection layer is constructed, a 160×160-size output is added, and self-attention and convolutional mixing module (ACmix) is integrated and applied to the newly constructed small target detection layer to achieve the effect of enhancing the model's feature perception and location information of distant small targets. Secondly, to reduce the interference of complex backgrounds, the new ODCBS module is constructed by using Omni-dimensional dynamic convolution (ODConv) instead of the convolution module of the neck, and the attention value is analyzed and learned from the spatial dimension of the convolution kernel, the dimension of the input channel, and the dimension of the output channel, respectively, in each part of the convolutional layer to enable the network to effectively capture richer contextual information. Finally, the PConv (partial convolution) module is integrated into the backbone network to replace part of the standard convolution, while the WIoU (Wise-IoU) loss function is used to replace the CIoU, to realize a reduction in the computation of the network model, improve the network detection speed, and at the same time, increase the low-quality anchor frames' focusing ability, and accelerate the model convergence speed. The experimental results show that the detection accuracy of the PAW-YOLOv7 algorithm on the FloW-Img dataset improved by the data extension technique used in this paper reaches 89.7%, which is 9.8% higher than that of the original YOLOv7. The detection speed reaches 54 frames per second (FPS), and the detection accuracy on the self-constructed sparse floater dataset improves by 3.7% compared with that of YOLOv7. It can quickly and accurately detect tiny floating objects in the river channel and also has better real-time detection performance. Finally, compared with the mainstream detection methods, the method in this paper has the best comprehensive effect.

Luan Q L, Chang X Y, Wu Y, et al. PAW-YOLOv7: algorithm for detection of tiny floating objects in river channels[J]. *Opto-Electron Eng*, 2024, 51(4): 240025; DOI: 10.12086/oe.2024.240025

Foundation item: Project supported by Anhui Provincial Major Science and Technology Project (202203a05020022), and Anhui Province Graduate Education Quality Project (2022xcysj147)

¹School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei, Anhui 230601, China; ²Anhui Province Key Laboratory of Intelligent Manufacturing of Construction Machinery, Hefei, Anhui 230601, China

* E-mail: dengconglong@ahjzu.edu.cn