YOLO-Vortex: 基于漩涡聚合网络的水下目标 检测模型

常鲁宁1, 苑春苗1, 杨清永2*

¹天津工业大学软件学院,天津 ²天津中德应用技术大学软件与通信学院,天津

收稿日期: 2025年1月11日; 录用日期: 2025年2月12日; 发布日期: 2025年2月21日

摘要

水下目标检测在海洋探索、生态保护和水下机器人导航等领域具有重要应用。然而,由于水下环境的复杂性,如光照不均匀、悬浮颗粒干扰和低对比度图像,传统的目标检测方法在水下环境中的表现往往不尽如人意,尤其是面对数据中的噪声问题。为了解决这一问题,本研究提出了一种基于YOLOv7的改进模型用于水下目标检测。我们将YOLOv7作为基线模型,针对其在水下环境中的不足之处,对模型的关键模块进行了优化。具体而言,我们提出了一种漩涡聚合网络模块来破坏噪声数据,并在此过程前引入了空间注意力机制,帮助网络更好地关注重要特征,并抑制不相关的噪声;针对下采样过程中可能存在的信息丢失问题,我们提出了空间到深度池化模块(STD-MP),通过将空间特征转换为深度特征,结合最大池化操作完成下采样过程;最后,我们对损失函数进行了优化。实验结果表明,我们的模型相比于基准模型提升了4.2%的mAP。

关键词

水下目标检测,YOLOv7,漩涡聚合网络,噪声特征扰乱

YOLO-Vortex: Underwater Target Detection Model Based on Vortex Aggregation Network

Luning Chang¹, Chunmiao Yuan¹, Qingyong Yang^{2*}

¹School of Software, Tiangong University, Tianjin ²School of Software and Communication, Tianjin Sino-German University of Applied Sciences, Tianjin

Received: Jan. 11th, 2025; accepted: Feb. 12th, 2025; published: Feb. 21st, 2025

*通讯作者。

Abstract

Underwater object detection has important applications in fields such as ocean exploration, ecological protection, and underwater robotics navigation. However, due to the complexity of the underwater environment, including uneven lighting, interference from suspended particles, and low-contrast images, traditional object detection methods often perform suboptimally in underwater scenarios, particularly when dealing with noisy data. To address this issue, this study proposes an improved model based on YOLOv7 for underwater object detection. We use YOLOv7 as the baseline model and optimize its key modules to overcome its limitations in underwater environments. Specifically, we introduce a vortex aggregation network module to disrupt noisy data, incorporating a spatial attention mechanism before this process to help the network better focus on important features and suppress irrelevant noise. To tackle the issue of potential information loss during downsampling, we propose the Space-To-Depth Pooling (STD-MP) module, which converts spatial features into depth features and combines them with max pooling for downsampling. Finally, we optimize the loss function. Experimental results show that our model achieves a 4.2% improvement in mAP compared to the baseline model.

Keywords

Underwater Target Detection, YOLOv7, Vortex Aggregation Network, Noise Characteristic Disturbance

Copyright © 2025 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <u>http://creativecommons.org/licenses/by/4.0/</u>

CC O Open Access

1. 引言

水下目标检测技术在现代海洋科学研究、海洋资源开发、生态环境保护及军事领域中扮演着至关重 要的角色。例如,在水下机器人导航、海底地形测绘、海洋生物监测以及水下救援任务中,精确的目标 检测能力是实现这些应用的基础。然而,由于水下环境的特殊性,光线传播受限、图像质量下降、光照 不均匀、悬浮颗粒干扰以及多变的水下色彩等问题给目标检测带来了巨大的挑战。传统的目标检测方法, 如基于人工特征的检测器,往往无法在这种复杂环境中取得令人满意的效果。

近年来,随着深度学习技术的迅猛发展,基于卷积神经网络的目标检测方法取得了显著的进展。其中,YOLO 系列模型凭借其实时性和较高的检测精度,成为当前目标检测领域的代表性方法之一。然而,现有的 YOLO 模型主要针对常规场景设计,在面对水下复杂环境时,其性能可能受到限制。为了解决这一问题,研究者们逐渐关注如何对 YOLO 模型进行适应性改进,以提升其在水下环境中的检测效果。

本研究旨在探索并改进 YOLOv7 [1]模型在水下目标检测中的应用表现。我们将 YOLOv7 作为基线 模型,结合水下环境的特点,对模型的关键模块进行了优化,以增强模型在复杂背景中的目标检测能力。 通过在水下目标检测数据集上的实验验证,我们的改进方法显示出显著的性能提升。

2. 相关工作

近年来,水下目标检测领域取得了显著进展,研究者们通过提出多种增强策略和改进模型来应对水下环境中的挑战。Lin [2]等人提出了一种名为 RoIMix [2]的增强方法,该方法通过在不同图像之间进行交

互,从中提取图案并将其混合在一起,以此来模拟重叠、遮挡和模糊对象的情况,从而构建了一个具有 更好泛化能力的模型。Fan [3]等人提出了一种结合特征增强和锚点细化的水下检测框架,该框架采用复 合连接主干来增强特征表示,并引入感受野增强模块以利用多尺度上下文特征,从而提高检测精度。Chen [4]等人提出了一种新颖的样本加权超网络(SWIPENET)和课程多类 Adaboost (CMA)的鲁棒训练范例,以 应对水下目标检测中的样本不均衡问题。Chang [5]等人将 Cascade-RCNN 和 Deformable Convolution Network应用于水下目标检测,并探讨了不同尺寸图像下两种检测算法的表现。Li [6]等人提出了一种基于联 合自监督去模糊和改进的空间变换网络的高精度水下目标检测模型,使用多任务学习架构将自监督去模 糊子网络与检测子网络相结合,以输出更为清晰的特征。Song [7]等人提出了 boosting R-CNN,两阶段水 下探测器由三个关键组件组成: RetinaRPN 区域提议网络、概率推理管道以及硬样本挖掘方法, 以提高检 测的准确性。Jain [8]等人引入了改进的 BiSkFPN 机制,使 EfficientDet 在对抗性噪声下具有更强的鲁棒 性,Walia [9]等人则开发了一个包含不同水下环境的自定义数据集,并提出了一种有效的海洋碎片检测 方法。Dai [10]等人提出了门控跨域协作网络(GCC-Net),通过实时 UIE 方法、跨域特征交互模块和门控 特征融合模块来解决水下环境中的低可见度和低对比度问题。Dai [11]等人还提出了一种边缘引导的表示 学习网络 ERL-Net, 通过边缘线索实现更具判别性的表示学习和特征聚合。Zhou [12]等人提出了调幅随 机扰动和涡旋卷积网络 AMSP-UOD, 专为解决复杂水下环境中的非理想成像因素而设计。Fu [13]等人提 出了残差特征转移模块(RFTM),以应对 DFUI 图像与水下图像间严重退化斑块的映射问题。Liu [14]等人 提出了即插即用的水下联合图像增强模块(UnitModule),旨在通过无监督学习损失改善探测器的输入图像 质量。

这些研究展示了在水下目标检测领域的多样化创新,涵盖了从数据增强、特征提取到模型架构优化 的广泛领域,推动了水下目标检测技术的发展。

3. 方法

3.1. YOLO-Vortex 整体框架

在水下目标检测任务中,由于环境复杂且噪声较多,检测模型不仅需要保持高精度,还需兼顾实时性。YOLO 系列模型因其检测速度快、精度高而广受关注,它将目标检测任务转化为目标区域预测和类别预测的回归问题。与传统方法如 Faster R-CNN [15]相比,YOLO 将目标区域和类别预测集成在单一的神经网络中,大大提升了检测效率。YOLOv7 进一步优化了该系列模型,通过高效层聚合网络(ELAN)、级联模型缩放策略以及卷积重参数化技术,在保持高效检测的同时显著提升了精度。基于 YOLOv7 在速度与精度上的平衡优势,本文选择其作为基线模型,并针对水下目标检测的数据噪声问题进行了改进,提出了基于 YOLOv7 的漩涡聚合水下目标检测网络 YOLO-Vortex。

我们设计了一个漩涡聚合网络模块,用于扰乱噪声特征,同时结合空间注意力机制,帮助网络更好 地关注重要特征并抑制无关噪声。YOLOv7 主要由骨干网络、检测头和预测模块组成。输入模块对图像 进行统一缩放,骨干网络通过 CBS、Elan 和 MPConv 层实现多尺度特征提取与增强;检测头采用基于 PAFPN 的结构,确保不同层次特征的有效融合;预测模块则通过 RepConv 结构进行最终目标的输出预 测。我们将漩涡聚合网络模块分别嵌入到骨干网络的 Elan 模块和检测头的 Elan-N 模块中,得到了 Vortex-Elan 和 Vortex-Elan-N 模块。此外,考虑到采样过程中可能导致信息丢失的问题,我们设计了空间到深度 池化模块(STD-MP),将空间特征转换为深度特征,有效解决了信息丢失的问题,并用 STD-MP 模块替换 了 YOLOv7 中的 MPConv 模块。与此同时,本文还改进了损失函数,加速了模型的收敛速度。我们提出 的水下目标检测网络 YOLO-Vortex 网络结构如图 1 所示。





3.2. 漩涡聚合网络模块(Vortex)

水下环境的复杂性源于各种规律性和非规律性的退化因素,如海洋生物活动、人类活动和洋流运动 等,这些因素会生成不可预测的噪声模式,从而对模型在感知和建模水下退化场景时造成极大挑战。卷 积及其变体在特征提取中起着关键作用,但在存在噪声干扰或复杂场景时,其效果往往受限。如何有效 区分背景特征和目标对象特征,成为影响检测精度的主要难题。

为应对这些问题,我们设计了一种创新性的 Vortex 模块,以减轻噪声干扰,提升网络在水下环境中 的适应性。我们的核心思路是在破坏噪声数据的同时,加入空间注意力机制来帮助网络更好地关注重要 特征,抑制不相关的噪声。虽然随机打乱噪声特征能够有效破坏噪声数据,但有用特征也会因此受到影 响,可能导致信息丢失。对此,我们通过在破坏噪声操作前引入空间注意力机制,有效地保留关键信息。 为了克服传统卷积核在较大尺寸卷积时的参数共享问题,我们在注意力机制中采用了分组卷积,从而为大尺寸卷积核提供了更精确的注意力权重。该创新提升了模型在复杂水下环境中区分背景与目标特征的能力,显著增强了水下检测的准确性。



Figure 2. Vortex module structure 图 2. Vortex 模块结构

在图 2 中,我们展示了 Vortex 模块的完整结构。Vortex 模块首先通过自定义的 Attention 模块进行特征提取,Attention 模块通过空间注意力机制计算输入特征的加权分布,并结合感受野空间特征,生成增强的特征图。接着,如果处于训练阶段,Broken 模块会对特征图进行随机扰动,通过打乱部分通道的顺序来增加特征的多样性。然后,扰动或未扰动的特征图会被分割成多个 slice 块,并对每个 slice 块进行卷积操作,卷积后的结果将其在通道维度进行拼接,然后通过批量归一化和激活函数处理,为了保证正确的特征语义信息的完整性,我们利用残差连接将原始关联特征 X 和得到的特征串联起来,得到最终的输出 Fout,该方法可以更好地适应特征衰减和噪声的影响,从而在梯度优化器的指导下获得理想退化场景的完整、正确的特征。转换可以表示如下:

$$F_{brk} = Broken(Attention(F_{in}))$$
⁽¹⁾

$$F_i^{conv} = Conv2d(F_i), i = 1, 2, \cdots, N$$
⁽²⁾

$$F_{cat} = Concat \left(F_1^{conv}, F_2^{conv}, \cdots, F_N^{conv} \right)$$
(3)

$$F_{out} = Concat \left(Attention(F_{in}), SiLU(BN(F_{cat})) \right)$$
(4)

其中 *F*₁, *F*₂,…, *F_N* 表示将 *F_{brk}* 分割成多个 slice 块, Attention 是我们引入的空间注意力机制, Broken 模块 是我们的噪声扰动策略,为增强模型的鲁棒性与特征学习的多样性, Broken 模块在训练期间通过在指定 维度 dim 上对输入张量 *x* 进行随机打乱来增加特征多样性。具体过程如下:

随机生成一个排列张量 perm,并通过以下公式得到最终的打乱索引:

$$perm = \frac{\dim_len}{group}$$
(5)

$$indices = cat\left(\left[perm \times group + i\right]_{i=0}^{group-1}\right)$$
(6)

DOI: 10.12677/csa.2025.152033

其中,dim_len为待打乱维度的长度;group为将输入张量再指定维度上划分的组数;i表示每个组内的偏移量,用于确保各组的打乱索引正确。

使用上述索引对输入张量 x 进行打乱:

$$x_{shuffled}[i] = x [indices[i]]$$
⁽⁷⁾

$$y = Gather(x, dim, indices)$$
(8)

其中, *x* 是输入张量, 包含待处理的特征; *indices* 是经过打乱后的索引, 指定了如何重新排列 *x* 中的元素; *y* 是打乱后的输出张量, *dim* 表示在此维度上执行打乱操作。

最终,输出特征图 y 与输入形状相同,但其通道数据已通过随机索引打乱,从而为后续网络层提供 了新的特征表示。这种随机打乱方法有效增强了模型对特征的学习能力,减小了因过拟合而导致的泛化 能力下降,特别在处理噪声数据和复杂场景时显著提升了模型的鲁棒性和准确性。

Attention 模块通过动态权重学习和感受野特征提取的结合,增强了模型在卷积过程中的空间感知能力。给定输入特征图 $X \in \mathbb{R}^{B \times C \times H \times W}$,其中 B, C, H, W分别表示批次大小、通道数、高度和宽度。通过平均池化和逐点卷积操作生成动态权重图 $W \in \mathbb{R}^{B \times C \times K^2 \times H' \times W'}$,其中 K 是卷积核大小,H'和W' 是下采样后的高度和宽度。

$$W = softmax(reshape(AvgPool(X) \times Conv(X)))$$
(9)

通过深度卷积操作,在输入特征图上生成感受野特征 $F \in \mathbb{R}^{B \times C \times K^2 \times H' \times W'}$:

$$F = SiLU\left(BN\left(Conv\left(X\right)\right)\right) \tag{10}$$



图 3. Vortex-Elan 和 Vortex-Elan-N 模块结构

感受野特征与动态权重进行逐元素相乘,得到加权特征 $Y \in \mathbb{R}^{B \times C \times K^2 \times H' \times W'}$: $Y = F \times W$ (11)

将加权特征 Y 重排以匹配原始空间维度,并通过最终卷积层生成输出特征图 $Z \in \mathbb{R}^{B \times C' \times H \times W}$:

Z = Conv(rearrange(Y))(12)

该过程通过动态调整每个空间位置的感受野,增强了模型的空间注意力,尤其在处理复杂场景时, 能够提升特征提取的效果。

如图 3 所示, Vortex-Elan 和 Vortex-Elan-N 模块整体结构和原始 Elan 和 Elan-N 模块相似,我们将 Elan 和 Elan-N 模块中的部分普通卷积模块替换为了我们设计的 Vortex 模块,提高了模型的抗噪声干扰 能力。

3.3. 空间到深度池化模块(STD-MP)

在水下目标检测中,由于水下环境的复杂性,传统目标检测网络的下采样模块在处理低分辨率图像 时往往存在信息丢失的问题,尤其是对噪声较为敏感,进而影响检测性能。YOLOv7 中的下采样模块主 要由最大池化和卷积组成,通过池化操作减少特征图的尺寸,并通过卷积层进行特征压缩。虽然该方式 有效地降低了计算复杂度,但在处理低分辨率水下图像时,可能会丢失部分关键的细节特征,尤其是在 处理噪声较大的图像时,检测效果会显著下降。

针对这一问题,我们提出了一种基于空间到深度卷积的下采样改进方法,旨在保留更多的空间信息 并提升水下目标检测的精度。其核心思想是通过将空间信息转换为深度信息来保留更多的特征,从而提 高对目标的检测能力,尤其是在处理水下图像噪声和模糊目标时。该方法的优势在于,通过空间到深度 的转换,能够在不增加过多计算量的情况下保留更多的细节信息,特别是对低分辨率水下图像的模糊目 标具有更好的识别能力。





受到 SPDConv [16]模块的启发,我们将 SPDConv 模块应用于 MPConv 下分支中,用以替代原有的 3 × 3 卷积核。如图 4 所示, SPDConv 模块通过将空间特征转换为深度特征,在特征图上每隔一个像素提取一个值,从而生成四个特征图,且不会丢失任何信息。随后,这些特征图在通道维度上进行拼接,使输出通道数量扩展至原来的 4 倍,之后对拼接后的特征图进行卷积操作,最终生成无信息丢失的二倍下



采样特征图。我们将改进后的 MPConv 模块命名为空间到深度池化模块(STD-MP),如图 5 所示。

Figure 5. STD-MP module structure 图 5. STD-MP 模块结构

3.4. 损失函数

在计算机视觉领域,交并比(Intersection over Union, IoU)是一种常用的指标,用于衡量模型预测边界 框与真实边界框之间的相似程度。IoU 的值越高,表示预测结果与真实标签的吻合度越高,模型性能也越 优秀。在模型训练过程中,IoU 通常被用作边界框的损失函数,以帮助提升模型的表现。在 YOLOv7 中, 采用了 CioU [17]。作为边界框的损失函数。CIoU 损失不仅结合了预测框与真实框之间的 IoU,还综合考 虑了二者中心点的距离以及宽高比的一致性。通过这样的多维度评估,模型能够更精准地优化边界框的 位置,从而提升目标定位的准确性。CIoU 的具体计算公式如下所示。

$$L_{CloU} = L_{loU} + \frac{\left(x - x_{gt}\right)^{2} + \left(y - y_{gt}\right)^{2}}{\left(W_{g}^{2} + H_{g}^{2}\right)} + \alpha \nu$$
(13)

$$\alpha = \frac{\nu}{L_{loU} + \nu} \tag{14}$$

$$\nu = \frac{4}{\pi^2} \left(\arctan \frac{w}{h} - \arctan \frac{w_{gt}}{h_{gt}} \right)$$
(15)

$$L_{loU} = 1 - \frac{W_i H_i}{wh + W_{gl} H_{gl} - W_i H_i}$$
(16)

其中, α 是一个惩罚项,用于对参数进行平衡; ν 表示长宽比度量函数,用于评估预测框与真实框在宽高比例上的一致性; (*W_i*,*H_i*)表示预测框与真实框重叠区域的宽度和高度; (*w*,*h*)和(*w_{gt}*,*h_{gt}*)分别代表预测框和真实框的宽度与高度; (*x*,*y*)和(*x_{gt}*,*y_{gt}*)分别为预测框和真实框的中心坐标; *L_{loU}*则为预测框与真实框的交并比。这些定义如图 6 所示。



Figure 6. IoU between the predicted box and the true box 图 6. 预测框与真实框的 IoU

尽管 CloU 损失函数在设计上考虑了边界框的长宽比一致性,但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 的梯度是一对相反数。这种特性会导致训练过程中 w 和 h 无法同时调整, 从而限制了模型性能的提升。此外,根据公式(15),当预测框与真实框的长宽相同时, v 的值为 0,使惩 罚项失效,这显然是不合理的;另外,在存在低质量样本的数据集中,CloU 的固定惩罚机制会使模型难 以从这些样本中有效学习,从而对检测精度产生负面影响。

因此,本文采用 WIoU [18]损失函数来替代 CIoU 损失。WIoU 通过一种动态非单调聚焦的梯度增益 分配策略,根据样本质量动态调整损失权重,减弱高质量样本的竞争优势,同时缓解低质量样本带来的 不良梯度影响,从而提升模型的检测精度与鲁棒性。此外,WIoU 基于距离度量引入了双层注意力机制, 其计算公式如下所示:

$$L_{\rm WIoU} = R_{\rm WIoU} \cdot L_{\rm IoU} \tag{17}$$

$$R_{\text{WIoU}} = \exp\left(\frac{\left(x - x_{gt}\right)^{2} + \left(y - y_{gt}\right)^{2}}{\left(W_{g}^{2} + H_{g}^{2}\right)^{*}}\right)$$
(18)

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。

4. 实验

4.1. 数据集

在本研究中,我们采用了TrashCan [19]数据集中的TrashCan-Instance版本来对模型进行训练和验证。 该数据集由日本海洋地球科学技术机构(JAMSTEC)管理,图像来源于 J-ED [20]数据集。TrashCan-Instance 数据集共包含 7212 张图像,其中 6065 张用于训练,1147 张用于验证,涵盖 22 类目标物体,包括垃圾、 ROV (遥控水下机器人)以及多种海洋动植物。数据集中的图像涵盖了各种水下场景,目标物体在不同程度的遮挡、模糊以及复杂的光照条件下均有分布。这些图片捕捉了水下环境中的高度不规则目标形态和 复杂背景,主要用于识别水下的垃圾和其他废弃物。为了适应模型的输入需求,我们对图像进行了预处 理,并将所有输入图像的大小调整为 640×640。通过该数据集的训练,能够有效评估模型在不同水下场 景下的性能,并确保其在水下目标检测任务中的泛化能力。

4.2. 实验环境及参数设置

实验环境配置如下:显存大小为 16 GB,操作系统为 Windows 10 的 64 位专业版,深度学习框架使用 PyTorch 1.11, CUDA 版本为 11.4,编程语言为 Python 3.8。训练时,输入图片大小统一为 640 × 640。 实验超参数设置如表 1 所示。

Table 1. Experimental hyperparameter settings 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	300

4.3. 实验评价指标

为了定量分析本文算法在水下目标检测中的性能,本实验采用平均精度(mean Average Precision, mAP) 作为评价指标。mAP 值为所有目标类别平均精度(AP)值的平均值,而 AP 则是通过计算精确率(Precision, P)和召回率(Recall, R)之间的 PR 曲线所围成的面积来得到的。PR 曲线下的面积越大,表示模型的检测精度越高。其计算公式如下:

$$P = \frac{TP}{TP + FP} \tag{19}$$

$$R = \frac{TP}{TP + FN} \tag{20}$$

$$AP = \int_0^1 P(R) dR \tag{21}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N}$$
(22)

其中, TP 表示真阳性, 指模型将正样本正确预测为正样本; FP 表示假阳性, 指模型将负样本错误预测为正样本; FN 表示假阴性, 指模型将正样本错误预测为负样本; N 为数据集中的类别数。

4.4. 与其他方法的比较

为了证明本文所提模型的优越性,我们采用了控制变量法进行对比实验,将本文模型与当前主流目标检测器在相同的实验环境和相同的数据集上进行了性能比较。在我们的数据集上训练 200 个 epoch 后,

训练结果显示性能已经趋于稳定。因此,300个 epoch 足以完成训练过程。表 2 展示了不同模型在精确率 (*P*)、召回率(*R*)、*mAP* 以及平均检测时间上的对比结果,其中最佳结果已用粗体标出。

Model	P (%)	<i>R</i> (%)	mAP@.5	Average Detection Time (ms)
YOLOv5	73.5	65.8	66.5	12.7
YOLOv6 [21]	73.1	66.3	66.5	21.94
YOLOv7	76	65	67.6	13.7
YOLOv8	73.5	64.4	69.4	12.9
YOLOv9 [22]	74.5	64.2	69.5	26.8
YOLOv10 [23]	74.9	61.7	65.3	19.3
YOLOv11	79.9	60.7	68.6	21.1
AMSP-UOD	76.2	61.3	70.3	36.9
Ours	78	65.5	71.8	25.6

Table 2. Comparison of the detection effect of YOLO-Vortex with other models 表 2. YOLO-Vortex 与其他模型的检测效果比较

对比结果显示,本文提出的模型不仅在检测精度上表现出色,同时也满足了实时检测的速度要求。 YOLO-Vortex 在精确率(*P*)、召回率(*R*)和平均精度(*mAP*)上分别达到了78%、65.5%和71.8%,显著优于原 有算法。特别是在 *mAP* 方面,YOLO-Vortex 相比原算法提升了4.2%,实现了显著的性能提升。整体来 看,YOLO-Vortex 在水下目标检测任务中,优于表2中的其他模型,在检测速度与精度之间取得了良好 的平衡。

4.5. 不同类别的比较

如表 3 所示,YOLO-Vortex 模型在多个类别上的性能均优于 YOLOv7,尤其在平均精度 mAP@.5 和精确率 P 方面表现突出。例如,在目标类别"rov"上,YOLO-Vortex 的 mAP 提高了 1.6%,精确率 P 提高了 0.8%,召回率 R 提高了 10.3%。对于类别"trash_cup",YOLO-Vortex 的 mAP 高达 0.973,显著优于 YOLOv7 的 0.46,表明改进模型在小目标检测中表现更为卓越。

此外,YOLO-Vortex 在平衡精确率与召回率方面具有更强的能力。例如,在类别"trash_branch"上, YOLO-Vortex 的 *mAP* 达到 0.832,高于 YOLOv7 的 0.802。尽管如此,某些类别的精确率 *P* 仍有提升空间,例如 "animal_crab"的 *P* 仅为 0.156,但其 *mAP* 仍比 YOLOv7 高出 0.088,说明改进模型在复杂背景下的鲁棒性有所增强。

总体而言,YOLO-Vortex 凭借对特征噪声的有效抑制和重要特征的强化,在复杂水下场景中的检测 能力显著优于 YOLOv7。

4.6. 消融实验

为了验证各模块: 漩涡聚合网络模块 Vortex、STD-MP 模块、WioU 损失对最终检测性能的影响,我们在 TrashCan 数据集上进行了消融实验,结果如表 4 所示,其中√表示使用该模块。

从表 4 中可以看出, 当基准模型 YOLOv7 的 Elan 和 Elan-N 模块被替换为我们提出的 Vortex-Elan 和 Vortex-Elan-N 模块后, *mAP* 从 67.6%提升至 69.1%, 这主要得益于 Vortex 模块能够在水下噪声环境中更 有效地提取目标特征, 从而提高检测精度。随后,将 MP 模块替换为 STD-MP 模块后, *mAP* 从 69.1%进 一步提升到 71.3%, 这是因为 STD-MP 模块解决了网络在下采样过程中信息丢失的问题, 进一步提升了

检测效果。最后,引入 WIoU 损失函数后, mAP 提升至 71.8%。

Table	e 3. Comparison of different categories
表 3.	不同类别的比较

分类 -		YOLO-Vortex			YOLOv7	
	P (%)	<i>R</i> (%)	mAP@.5	P (%)	<i>R</i> (%)	mAP@.5
rov	0.914	0.96	0.943	0.906	0.857	0.927
plant	0.68	0.559	0.638	0.546	0.565	0.579
animal_fish	0.78	0.458	0.598	0.768	0.498	0.584
animal_starfish	0.888	0.699	0.752	0.902	0.747	0.789
animal_shells	0.527	0.525	0.509	0.445	0.541	0.49
animal_crab	0.156	0.429	0.327	0.17	0.413	0.239
animal_eel	0.658	0.643	0.623	0.611	0.635	0.585
animal_etc	0.843	0.554	0.612	0.791	0.569	0.602
trash_clothing	0.895	0.501	0.717	0.901	0.471	0.586
trash_pipe	0.859	0.905	0.891	0.913	0.929	0.92
trash_bottle	0.925	0.731	0.735	0.909	0.692	0.727
trash_bag	0.78	0.646	0.71	0.772	0.657	0.694
trash_snack_wrapper	0.873	0.412	0.421	0.915	0.412	0.421
trash_can	0.988	0.853	0.92	0.948	0.849	0.897
trash_cup	1	0.388	0.973	0.86	0.25	0.46
trash_container	0.827	0.816	0.887	0.863	0.792	0.888
trash_unknown_instance	0.748	0.685	0.696	0.745	0.698	0.701
trash_branch	0.866	0.824	0.832	0.845	0.765	0.802
trash_wreckage	0.878	1	0.978	0.892	1	0.966
trash_tarp	0.625	0.484	0.516	0.672	0.548	0.559
trash_rope	0.782	0.743	0.74	0.718	0.704	0.704
trash_net	0.674	0.697	0.775	0.631	0.697	0.744

Table 4. Module ablation experiments 表 4. 模块消融实验

YOLOv7	\checkmark	\checkmark		
Vortex		\checkmark	\checkmark	\checkmark
STD-MP			\checkmark	\checkmark
WIoU				\checkmark
mAP@.5	67.6	69.1	71.3	71.8
Average Detection Time (ms)	13.7	24.0	22.7	25.6

DOI: 10.12677/csa.2025.152033

4.7. 检测结果对比

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,针对 YOLOv7 以及我们 改进后的算法 YOLO-Vortex 进行对比,结果如图 7 所示。



Figure 7. Comparison of detection results: The first line is an example of a YOLO-Vortex detection result, and the second line is an example of a YOLOv7 detection result

图 7. 检测结果比较:第一行为 YOLO-Vortex 检测结果示例,第二行为 YOLOv7 检测结果示例

5. 结论

本文提出了一种改进的 YOLOv7 水下目标检测算法,能够更加准确地检测水下目标。通过改进 Elan 模块、MP 模块以及引入 WIoU 损失函数,增强了模型在噪声环境下对特征的提取能力。改进后的算法相 比原 YOLOv7,检测精度提升了 4.2%。在未来的工作中,我们将专注于水下图像增强,以获取视觉质量 更高的图像,进一步提升检测精度。同时,计划设计适用于水下实时检测的轻量化算法,以减少参数量 和计算量,提升检测速度。

参考文献

- [1] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 7464-7475. <u>https://doi.org/10.1109/cvpr52729.2023.00721</u>
- [2] Lin, W., Zhong, J., Liu, S., Li, T. and Li, G. (2020) ROIMIX: Proposal-Fusion among Multiple Images for Underwater Object Detection. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, 4-8 May 2020, 2588-2592. <u>https://doi.org/10.1109/icassp40776.2020.9053829</u>
- [3] Fan, B., Chen, W., Cong, Y. and Tian, J. (2020) Dual Refinement Underwater Object Detection Network. *Computer Vision—ECCV* 2020, Glasgow, 23-28 August 2020, 275-291. <u>https://doi.org/10.1007/978-3-030-58565-5_17</u>
- Chen, L., Zhou, F., Wang, S., et al. (2020) SWIPENET: Object Detection in Noisy Underwater Images. arXiv: 2010.10006. https://doi.org/10.48550/arXiv.2010.10006
- [5] Chang, D. (2021) CDNet Is All You Need: Cascade DCN Based Underwater Object Detection RCNN. arXiv: 2111.12982. https://doi.org/10.48550/arXiv.2111.12982
- [6] Li, X., Li, F., Yu, J., et al. (2022) A High-Precision Underwater Object Detection Based on Joint Self-Supervised Deblurring and Improved Spatial Transformer Network. arXiv: 2203.04822. https://doi.org/10.48550/arXiv.2203.04822
- [7] Song, P., Li, P., Dai, L., Wang, T. and Chen, Z. (2023) Boosting R-CNN: Reweighting R-CNN Samples by RPN's Error

for Underwater Object Detection. Neurocomputing, 530, 150-164. https://doi.org/10.1016/j.neucom.2023.01.088

- [8] Jain, S. (2024) DeepSeaNet: Improving Underwater Object Detection Using EfficientDet. 2024 4th International Conference on Applied Artificial Intelligence (ICAPAI), Halden, 16 April 2024, 1-11. https://doi.org/10.1109/icapai61893.2024.10541265
- [9] Walia, J.S. and Seemakurthy, K. (2023) Optimized Custom Dataset for Efficient Detection of Underwater Trash. *Towards Autonomous Robotic Systems*, Cambridge, 13-15 September 2023, 292-303. https://doi.org/10.1007/978-3-031-43360-3_24
- [10] Dai, L., Liu, H., Song, P. and Liu, M. (2024) A Gated Cross-Domain Collaborative Network for Underwater Object Detection. *Pattern Recognition*, **149**, Article 110222. <u>https://doi.org/10.1016/j.patcog.2023.110222</u>
- [11] Dai, L., Liu, H., Song, P., et al. (2023) Edge-Guided Representation Learning for Underwater Object Detection. arXiv: 2306.00440. https://doi.org/10.48550/arXiv.2306.00440
- [12] Zhou, J., He, Z., Lam, K., Wang, Y., Zhang, W., Guo, C., et al. (2024) AMSP-UOD: When Vortex Convolution and Stochastic Perturbation Meet Underwater Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence, 38, 7659-7667. <u>https://doi.org/10.1609/aaai.v38i7.28599</u>
- [13] Fu, C., Fan, X., Xiao, J., Yuan, W., Liu, R. and Luo, Z. (2023) Learning Heavily-Degraded Prior for Underwater Object Detection. *IEEE Transactions on Circuits and Systems for Video Technology*, **33**, 6887-6896. <u>https://doi.org/10.1109/tcsvt.2023.3271644</u>
- [14] Liu, Z., Wang, B., Li, Y., He, J. and Li, Y. (2024) UnitModule: A Lightweight Joint Image Enhancement Module for Underwater Object Detection. *Pattern Recognition*, **151**, Article 110435. <u>https://doi.org/10.1016/j.patcog.2024.110435</u>
- [15] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/tpami.2016.2577031
- [16] Sunkara, R. and Luo, T. (2023) No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects. *Machine Learning and Knowledge Discovery in Databases*, Grenoble, 19-23 September 2022, 443-459. <u>https://doi.org/10.1007/978-3-031-26409-2_27</u>
- [17] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R. and Ren, D. (2020) Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34, 12993-13000. https://doi.org/10.1609/aaai.v34i07.6999
- [18] Tong, Z., Chen, Y., Xu, Z., et al. (2023) Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. arXiv: 2301.10051. https://doi.org/10.48550/arXiv.2301.10051
- [19] Hong, J., Fulton, M. and Sattar, J. (2020) Trashcan: A Semantically-Segmented Dataset towards Visual Detection of Marine Debris. arXiv: 2007.08097. https://doi.org/10.48550/arXiv.2007.08097
- [20] Japan Agency for Marine Earth Science and Technology (2018) Deep-Sea Debris Database. http://www.godac.jamstec.go.jp/catalog/dsdebris/e/index.html
- [21] Li, C., Li, L., Jiang, H., et al. (2022) YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv: 2209.02976. https://doi.org/10.48550/arXiv.2209.02976
- [22] Wang, C., Yeh, I. and Mark Liao, H. (2024) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *Computer Vision—ECCV* 2024, Milan, 29 September-4 October 2024, 1-21. https://doi.org/10.1007/978-3-031-72751-1
- [23] Wang, A., Chen, H., Liu, L., et al. (2024) YOLOv10: Real-Time End-to-End Object Detection. arXiv: 2405.14458. https://doi.org/10.48550/arXiv.2405.14458