

基于 SimAM 注意力机制的 DCN-YOLOv5 水下目标检测

刘向举, 刘 洋, 蒋社想

安徽理工大学 计算机科学与工程学院, 安徽 淮南 232001

摘 要:目的 针对水下环境复杂, 水下目标因光线折射导致的目标边界模糊或外观、形状可能会发生非刚性形变, 使水下目标检测困难的问题, 提出了一种基于 SimAM 注意力机制的 DCN-YOLOv5 水下目标检测方法。方法 首先, 采用 YOLOv5 所使用的双向金字塔网络(BiFPN, Bi-directional Feature Pyramid Network)在多个尺度上提取和融合特征信息, 从而提高目标辨别的准确度; 其次, 针对水下目标的外观、形状变化问题, 将 C3 模块中的 CBS 模块结合可变形卷积(DCN, Deformable Convolution Network), 提出 DBS 模块并组成 D3 模块替换部分 C3 模块, 以适应水下目标的外观、形状变化; 同时, 融入加权注意力机制(SimAM), 自适应地调节模型的关注度, 进一步在复杂场景下增强特征表达能力; 最后, 考虑目标边界模糊, 为改善目标定位精度, 采用 WIoU(Wise-IoU)损失函数来替换交叉熵损失, 能够更好地适应不同目标类型和尺寸的特点, 提高算法鲁棒性。结果 实验结果表明: DCN-YOLOv5 可以达到 87.57% 的平均精度(mAP), 检测效果优于 YOLOv5 网络和其他经典网络, 平均每张图像的识别时间仅为 24.5 ms。结论 通过实验结果可以证明模型在检测精度明显提升的同时兼顾检测的实时性, 对水下目标检测用于实际用途有着一定的参考价值。

关键词:水下目标检测; SimAM 注意力机制; 可变形卷积; WIoU

中图分类号: TP391.41; TP183 **文献标识码:** A **doi:** 10.16055/j.issn.1672-058X.2025.0002.009

DCN-YOLOv5 Underwater Target Detection Based on SimAM Attention Mechanism

LIU Xiangju, LIU Yang, JIANG Shexiang

School of Computer Science and Engineering, Anhui University of Science and Technology, Anhui Huainan 232001, China

Abstract: Objective Given the complex underwater environment, the target boundary may be blurred or the appearance and shape of the underwater target may be non-rigidly deformed due to light refraction, which makes underwater target detection difficult. A DCN-YOLOv5 underwater target detection method based on the SimAM attention mechanism was proposed. **Methods** Firstly, the bi-directional feature pyramid network (BiFPN) used by YOLOv5 was used to extract and fuse feature information on multiple scales to improve the accuracy of target recognition. Secondly, to address the variations in appearance and shape of underwater objects, the CBS module in the C3 module was combined with the deformable convolution network (DCN), and the DBS module was proposed. The DBS module was used to form the D3 module and replace part of the C3 module to adapt to the changing appearance and shape of the underwater targets. At the same time, the weighted attention mechanism was integrated to adaptively adjust the attention of the model and further improve the feature expression ability in complex scenes. Finally, considering the fuzzy boundary of the target and to improve the target positioning accuracy, the WIoU (Wise-IoU) loss function was used to replace the cross-entropy loss,

收稿日期: 2023-03-05 **修回日期:** 2023-05-18 **文章编号:** 1672-058X(2025)02-0063-08

基金项目: 安徽省重点实验室项目(ZKSYS202204)。

作者简介: 刘向举(1978—), 男, 黑龙江哈尔滨人, 硕士, 副教授, 从事物联网、智能控制和软件定义网络研究。

通讯作者: 刘洋(1999—), 男, 安徽铜陵人, 硕士研究生, 从事目标检测研究。Email: 910744805@qq.com。

引用格式: 刘向举, 刘洋, 蒋社想. 基于 SimAM 注意力机制的 DCN-YOLOv5 水下目标检测[J]. 重庆工商大学学报(自然科学版), 2025, 42(2): 63-70.

LIU Xiangju, LIU Yang, JIANG Shexiang. DCN-YOLOv5 underwater target detection based on SimAM attention mechanism[J].

Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(2): 63-70.

which can better adapt to the characteristics of different target types and sizes and improve the robustness of the algorithm.

Results Experimental results showed that DCN-YOLOv5 achieved an average precision (mAP) of 87.57%, outperforming YOLOv5 and other classical networks, with an average identification time of only 24.5 ms per image.

Conclusion The experimental results demonstrate that the model significantly improves detection accuracy while ensuring real-time detection, providing valuable insights for the practical use of underwater target detection.

Keywords: underwater target detection; SimAM attention mechanism; deformable convolutions; WIoU

1 引言

水下目标检测是一项利用图像处理和计算机视觉技术的过程,用于自动识别和定位水下环境中的目标。与陆地或空中目标检测相比,水下目标检测面临独特的挑战和困难。由于水下环境的复杂性和不可控性,水下图像通常受到光线衰减、散射、噪声等多种因素的影响,导致图像质量差且细节模糊。传统的水下目标检测方法通常依赖于手工设计的特征提取器和分类器,如滤波器、纹理分析和模型匹配。然而,由于水下图像的不确定性和多样性,传统方法的性能通常不能满足实际需求,特别是在复杂场景和目标边界模糊的检测方面表现较差。

近年来,相较于传统的基于特征工程的方法,深度学习技术可以自动地从数据中学习更高级别的特征表示,从而提高目标检测的准确性和鲁棒性。R-CNN 系列^[1-3]算法通过候选区域的生成和对每个候选区域进行分类的两阶段流程来进行目标检测。SSD^[4]算法通过在不同尺度和不同层次上预测目标的位置和类别来实现目标检测。YOLO 系列^[5-7]算法也是一种单阶段目标检测算法,通过将图像拆分成多个网格并通过预测目标在网格中的位置来实现目标检测。这些深度学习算法在目标检测领域表现出了明显的优势,因此得到了广泛的应用。有研究者针对水下目标模糊及背景复杂的特点,在常用算法的基础上做了大量的针对性改进。其中,禹志鹏等^[8]改进了 UCM 算法的直方图拉伸,使处理后的水下图片对比度增强效果更好,但在图像增强的同时也增大了模型的计算量,使检测不能保证实时性;郝琨等^[9]将 DenseNet 思想引入 YOLOv3 并在主干网络中引入密集块,提出了 Den-YOLOv3 算法,使模型可以缩小检测时间,但针对水下模糊图像检测时的检测精度还不够理想;乔美英等^[10]提出在特征金字塔(FPN, Feature Pyramid Network)上采样过程中利用卷积及横纵向池化挖掘上下文信息,并引用并行空洞卷积及高效通道注意力机制(ECANet)来捕捉空间与通道的特征信息,使 Faster R-CNN 在水下目标检测中针对携带特征信息少的小目标精度有着明显的提高,但模型并不能满足水下目标检测的实时性要求,且在引入 IoU 时并未考虑预测框和标注框相互包含和完

全无交集的极端情况;强伟等^[11]在可变性卷积的基础上融合了深度可分离卷积思想,提出一种深度可分离的卷积结构,使模型对模糊目标检测精度有所提升,但实时性较差;叶赵兵等^[12]在 YOLOv3-SPP 的基础上增加了网络预测尺度,以提高水下小目标的检测占比,引入了 CIoU 边界框回归损失,以提高水下小目标的定位精度,使用 K-Means++ 聚类算法筛选出最优的 anchor 框,在缩短检测时间的同时明显提高了检测精度,但对极端环境下的误检率及漏检率较高;Li 等^[13]在 Faster R-CNN 的结构基础上改变卷积层的结构,使用 ReLU、Inception 和 HyperNet 构建块,将模型轻量化,缩小了模型的计算量并提高了检测能力,但在对复杂环境及目标受背景干扰时检测效果不理想。

上述研究对水下目标检测都有着重大意义,但其中部分算法虽在精度上有所提升,但未考虑检测实时性问题,导致计算量大、计算速度缓慢;部分算法虽兼顾了实时性,但在精度上并没有明显的提升;部分算法虽可以在提升精度的同时兼顾实时性,但模型的抗干扰效果不理想。对此,本文提出了一种改进的水下目标检测方法—基于 SimAM 加权注意力机制的 DCN-YOLOv5 水下目标检测方法。该方法首先将可变形卷积结合 D3 模块使模型具有更好的抗干扰效果,使模型在水下复杂环境仍可以很好地区分目标与背景;在模型中引入 SimAM 加权注意力机制,在检测时将更多的权重分配给更重要的目标,提升模型的检测精度;最后使用 WIoU 损失函数替换交叉熵损失,减少在计算损失值时的计算量且可以考虑预测框和标注框相互包含和完全无交集的极端情况。对该方法在水下真实环境下的数据集进行了验证,在水下复杂环境下,不仅检测精度有了明显提升,还保证了检测的实时性。

2 相关工作

2.1 YOLOv5 模型

YOLOv5 由 Glenn Jocher 于 2020 年提出,是一种高效、快速且准确率高的目标检测方法,通过单个神经网络直接预测图像中目标的位置和类别,实现了单阶段检测的策略。相比于传统两阶段方法,YOLOv5 的核心思想在于简化了目标检测流程,将检测过程高度集成,并通过优化网络结构和训练策略来提高准确性和速

度。通过这种端到端的回归方式,YOLOv5 在目标检测任务中取得了较好的性能和鲁棒性,成为了目标检测领域的重要算法之一。本文采用的是 2021 年 10 月提出的 YOLOv5 6.0 版本,其主要网络结构如图 1 所示。

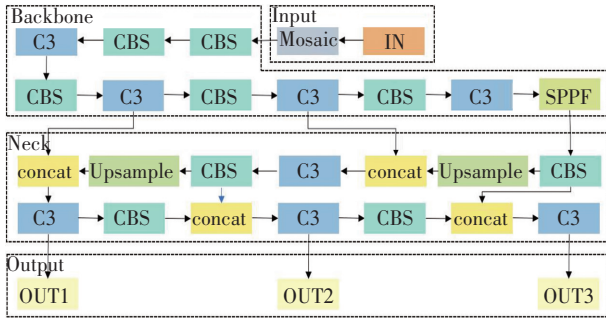


图 1 YOLOv5 网络结构

Fig. 1 The network structure of YOLOv5

YOLOv5 网络主要包含 4 个基本组件,分别为 Input、Backbone、Neck 和 Head。Input 部分采用了几项关键的数据增强和处理技术,其中 Mosaic 数据增强通过将 4 张随机选取的图像拼接成一张大图像,从而增加训练样本的多样性和复杂性;自适应锚框计算根据目标的大小和位置,动态调整锚框的大小和位置,自适应图像尺寸处理缩放是针对不同大小的目标,动态调整输入图像的尺寸以满足 Backbone 的输入尺寸要求。首先,在对数据完成了预处理及图像增强等一系列操作之后,图像被输入到 Backbone 模块中,这部分网络会对经过处理后的图像进行特征提取;然后将提取后的特征输入到 Neck 模块中进行特征融合处理,生成大、中、小 3 种类型的特征信息;最终,这些经过提取融合后的特征信息被输入 Head 模块,进行检测之后输出最终的结果。

Backbone 部分主要由 CBS、C3、SPPF 等模块组成。其中 CBS 模块是模型的基本卷积模块,分别由卷积操作、正则化操作和激活操作 3 部分组成。CBS 模块的目的是提取输入的特征并进行特征增强,CBS 结构如图 2 所示。C3 模块由 CBS 模块及残差结构组成,其结构如图 3 所示。首先,CBS 模块对输入进行卷积操作、正则化操作和激活操作;接着,残差结构模块包含两个卷积层,它将输入特征与原始特征相加,实现残差特征的转移,并保持输出特征的深度不变。这样可以降低额外参数的引入,提高网络的性能。SPPF (Spatial Pyramid Pooling Fusion) 模块是 YOLOv5 中的关键组件之一,用于改进特征提取和感受野的处理。SPPF 模块通过空间金字塔池化和特征融合来增强目标检测的性能。SPPF 模块的主要目的是捕获不同尺度的特征信息,以及将输入特征图分割成多个子区域,并对每个子区域进行池化操作。这些池化操作的尺度可以覆盖不

同大小的感受野。SPPF 模块通过采用多个不同大小的池化核大小,如 1×1 、 2×2 、 4×4 等,对输入特征图进行池化操作,以捕获不同尺度下的局部信息,SPPF 结构如图 4 所示。



图 2 CBS 模块结构

Fig. 2 The structure of CBS module

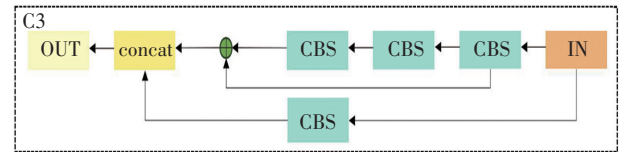


图 3 C3 模块结构

Fig. 3 The structure of C3 module

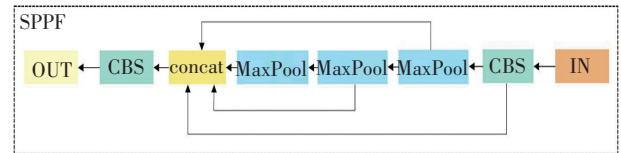


图 4 SPPF 模块结构

Fig. 4 The structure of SPPF module

在 Neck 部分,采用了双向金字塔网络 (Bi-directional Feature Pyramid Network, BiFPN) 结构,使得网络适用于多尺度输入,然后通过自上而下的信息传播方式,从高分辨率到低分辨率逐层传递和聚合特征。在 Head 部分,从 Neck 输出的特征图中,使用大、中、小 3 种不同尺度的预测头对 Neck 精炼出来的特征信息进行解耦,每个预测头分别对应一个尺度的目标框。每个预测头通常由一系列卷积层构成,以捕获不同尺度下的目标信息,最终预测头会生成一系列锚框 (Anchors),这些锚框用于表示不同尺度下可能存在的目标位置和大小。在主要的部分结束后,通过对预测头的输出进行解码和后处理操作,最后得到最终的目标框。再将预测头输出的类别概率与锚框位置信息结合起来,通过非极大值抑制 (NMS) 方法,过滤掉重叠度高的预测框,生成最终的检测结果。

2.2 注意力机制

注意力机制^[14]是一种在神经网络中被广泛使用的机制,它模拟了人类视觉和认知过程中的“关注”和“集中”注意力能力。注意力机制被广泛应用于计算机视觉、自然语言处理等领域。在目标检测领域中,经多方实验研究表明:在网络模型中添加注意力机制对增强目标模型的表征能力,降低无效目标及目标背景的干扰有着一定的正向作用,从而可以使神经网络有选择地聚焦于输入数据的特定部分,以提取重要信息并进行更准确的预测。

在注意力机制中,通过计算每个输入元素的权重来确定其相对重要性。这些权重可以根据任务需求和上下文动态地调整,以反映不同输入元素对当前任务的贡献。注意力机制会计算一个权重向量,其中每个元素表示对应输入元素的重要性。然后,这些权重被应用于输入数据的不同部分,以进行加权操作或者生成加权的特征表示。

在计算机视觉中,注意力机制能够帮助模型集中关注于特定区域或特征,提高任务的准确性和效率。注意力机制在目标检测、图像分类、图像分割和生成式模型等任务中都有广泛应用。通过引入注意力机制,模型可以自动选择图像中最具代表性和关键性的区域进行处理,从而更好地理解图像内容并提高任务的准确性和鲁棒性。

2.3 IoU 损失函数

在目标检测模型中,IoU 损失函数是一种目标定位的度量指标,广泛应用于目标检测和语义分割等任务中。通过计算预测边界框与真实边界框之间的交并比,来衡量它们的重叠程度。

为了计算 IoU,在预测和真实边界框之间进行二进制掩模操作。首先,计算两个边界框的交集面积和并集面积,然后将交集面积($B \cap G$)除以并集面积($B \cup G$)得到 L_{IoU} ,如式(1)所示:

$$L_{IoU} = \frac{B \cap G}{B \cup G} \quad (1)$$

L_{IoU} 的范围在 0 到 1 之间,代表了预测与真实边界框的重合度。在使用 IoU 损失函数时,通常会使用 1 减去 L_{IoU} 作为实际的损失函数。这是因为通常希望得到最小化与真实边界框的差异,而不是最大化 L_{IoU} 。在两框不相交的情况下, L_{IoU} 始终为 0,因此很难评判回归框的好坏。

3 DCN-YOLOv5 算法

3.1 整体网络架构

为了解决水下复杂环境中的检测精度与实时性无法兼顾的问题,本文提出了一种基于 SimAM 注意力机制的 DCN-YOLOv5 水下目标检测模型,其网络结构如图 5 所示。

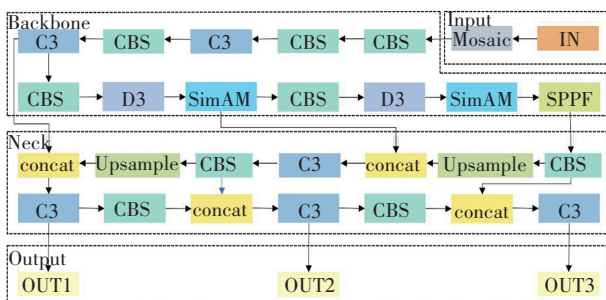


图 5 DCN-YOLOv5 网络结构

Fig. 5 The network structure of DCN-YOLOv5

该模型基于 YOLOv5 网络,并在此基础上提出了几个改进点,包括将可变形卷积与 CBS 模块结合可以更好地提取非刚性形态目标特征的 DBS 模块,并将其与 C3 模块结合,提出一种基于可变形卷积的 D3 模块替代网络中的部分 C3 模块,增加 SimAM 加权注意力机制以及引用一种边界框回归损失函数 WIoU。

首先,水下目标检测中,由于水流等因素导致目标形变或者图像中存在一些非刚性的目标形态,传统的卷积操作无法很好地适应目标的变化。对此,本文提出了一种 D3 模块。D3 模块结合了 C3 模块和 DCN (Deformable Convolutional Networks) 的思想,在保持特征提取能力的同时,允许卷积核在空间维度上发生形变,从而更好地捕获目标的非刚性形态,更好地适应水下目标的形状不规则特性,提升模型对水下目标的感知能力。

其次,水下环境中存在大量的噪声、散射、光照变化等因素,这些因素可能干扰目标的检测。对此,本文在模型中增加了 SimAM^[15] 注意力机制。SimAM 是一种自注意力机制,能够自动选择图像中与目标相关的特征,抑制无关特征,从而增强网络对目标的关注能力,提高抗干扰能力,减少误检和漏检。SimAM 注意力机制还能够自适应地学习目标之间的相似性信息,并加权调节特征的重要性。通过引入 SimAM 注意力机制,可以提升网络对水下目标的关注程度,增强对重要目标的检测能力,从而提高整体的目标检测精度。

最后,对于水下目标检测,目标可能因为水流等因素导致形状模糊,传统 YOLOv5 算法使用的 CIoU 作为损失函数计算并不适用。对此,本文引用一种新的边界框回归损失函数 WIoU (Weighted Intersection over Union) 代替 YOLOv5 的 CIoU 计算方法。WIoU 考虑了目标边界框的位置和大小之间的权重关系,能够更准确地度量目标的重叠程度。

3.2 基于可变形卷积的 D3 模块

3.2.1 可变形卷积

深度学习领域中,为了更好地捕捉和建模非刚性形态,Dai 等^[16]提出了可变形卷积(DCN),DCN 通过添加可学习的偏移量的卷积层和全连接层,改进了传统的固定卷积核。这种改进使得模型能够根据不同物体或结构的形状变化来调整卷积核的大小和形状。通过降低模型的复杂性,DCN 能够更好地适应图像中的目标,并提高对非刚性变形和遮挡等复杂情况的处理能力。Zhu 等^[17]提出的 DCNv2 通过更全面地整合可变形卷积的特性,并引入扩展可变形建模范围的调节机制,进一步提高了可变形卷积模型的建模能力。相比于 DCN,DCNv2 不仅能够调整感知输入特征的偏移量,而且可以调节来自不同空间位置输入特征的幅度,这使得 DCNv2 在处理自适应能力有限问题方面表现出

更好的性能。DCNv2 的计算公式如式(2)一式(4)所示。对于一个参数为 1 的 3×3 卷积核,有

$$\mathbf{R} = \{x | x \in (-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\} \quad (2)$$

式(2)中, $\mathbf{R}$ 的每个元素 x 表示卷积核中每个位置相对于中心点的偏移量信息。这些偏移量将用于调整输入特征图的采样位置,以便更好地适应不同的特征分布。通过使用 $\mathbf{R}$ 矩阵中的偏移量信息,DCNv2 可以实现更灵活和精确的字段调节感知,从而提高模型的自适应能力。

$$y(p_0) = \sum_{p_k \in \mathbf{R}} \omega(p_k) \cdot x(p_0 + p_k + \Delta p_k) \cdot \Delta m_k \quad (3)$$

式(3)中,偏移量和调制参数是通过额外的卷积层来进行学习的。输入特征图经过这个卷积层后,会得到每个采样点位置的偏移量 Δp_k 和调制参数 Δm_k 。其中, p_0 表示中心采样点位置, p_k 表示采样点位置, Δp_k 表示采样点位置 k 相对于中心位置的偏移量,而 Δm_k 则代表采样点位置 k 的调制参数。通过学习得到不同采样点位置的偏移量和调制参数,DCNv2 能够实现更加准确和自适应地感知区域调整。

$$y(k) = \sum_{i=1}^{n_k} x(p_{ki} + \Delta p_k) \cdot \frac{\Delta m_k}{n_k} \quad (4)$$

式(4)中,每个采样点的位置 p_{ki} 可以表示为第 k 个有效采样中第 i 个像素点的采样位置, n_k 表示有效采样位置对应的像素点个数。为了得到精确的采样值,需要通过双线性插值法来计算 $x(p_{ki} + \Delta p_k)$ 。这样可以根据偏移量 Δp_k 调整采样点位置,并获得更准确的像素值。

3.2.2 DBS 模块

本文融合 DCNv2 思想与 CBS 模块,提出 DBS 模块,其结构如图 6 所示。

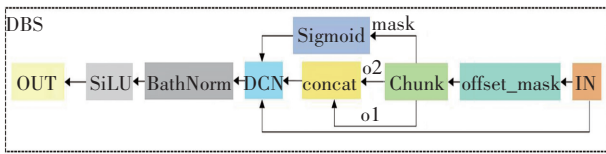


图 6 DBS 模块结构

Fig. 6 The structure of DBS module

首先,输入经过偏移掩码卷积,得到的连续数据分割为 o1、o2、mask 的 3 个单元块。其中,o1 和 o2 单元块经连接操作,mask 单元块经过 sigmoid 激活函数操作。它们与原始输入一起经过可变形卷积操作,经过归一化操作及 SiLU 激活函数得到最终的输出。DBS 模块针对非刚性目标的检测精度有所提升。

其中,偏移掩码卷积能够对输入特征图进行自适应空间变换,使得卷积核能够更好地适应输入特征的局部结构。同时,偏移掩码卷积与传统卷积相比,对捕捉非刚性目标的形状和姿态变化有着更好的效果,可以更好地适应这些形状和姿态的变化,从而提高检测的精度。对 mask 单元块使用 sigmoid 激活函数去代替 SiLU 激活函数可以将输出映射到 0 和 1 之间,以便后

续使用这些值作为可变形卷积操作的掩码,这样更适合做出二元决策。

3.2.3 D3 模块

本文将 DBS 模块替换掉 C3 模块中用于提取局部特征的 CBS 模块,提出 D3 模块,其结构如图 7 所示。这样一来,D3 模块能够更好地处理水下目标因光线折射导致的非刚性形变和位置变化,从而提高检测精度。通过学习可变形卷积的参数,模型能够根据目标的具体形状和位置信息,动态调整卷积核的采样位置,从而更准确地捕捉目标的特征。这种改进能够提高目标检测模型在水下场景中的性能,使其更适应水下目标的复杂性和多样性。

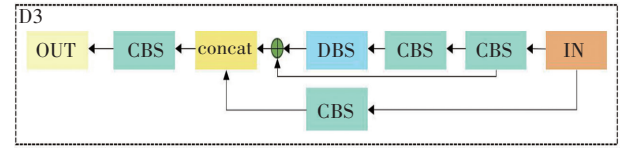


图 7 D3 模块结构

Fig. 7 The structure of D3 module

3.3 嵌入 SimAM 的注意力机制

在复杂、模糊和光照条件不佳的水下环境中,目标的关键信息可能会被噪声和干扰因素混淆,从而使得结果的精度较低。为了改善这一问题以及考虑水下复杂环境中目标检测的实时性,本文在模型中嵌入了 SimAM 注意力机制。SimAM 是一种轻量级的注意力机制,关键优势在于其强大的特征增强表征能力和高效的计算性能,其结构如图 8 所示。

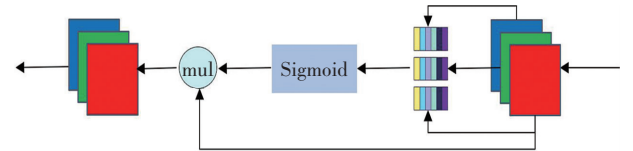


图 8 SimAM 模块结构

Fig. 8 The structure of SimAM module

首先,SimAM 通过自适应地学习并利用目标间的相似性信息,精准计算特征之间的相似性度量,以确定不同特征的权重。这使得 SimAM 能更有针对性地关注与目标检测任务相关的重要特征,并调节这些特征的重要性。SimAM 的特征加权如式(5):

$$F_A(i, j, c) = F(i, j, c) \cdot A(i, j) \quad (5)$$

式(5)中, $F(i, j, c)$ 是输入特征图的一个元素, $A(i, j)$ 是对应的注意力权重。式中将输入特征图的每个元素乘以对应的注意力权重,从而得到加权特征图 F_A 。这个过程不需要额外的参数,只需要一个元素级别的乘法操作,计算量较小,相较于传统注意力机制的效率更高。其次,SimAM 的设计较轻量级,所有的运算都极为高效,无需额外的大量参数或复杂的计算步骤。这意

味着,在提升模型性能的同时,SimAM 能保持超高的计算效率。SimAM 的特征归一化如式(6):

$$A(i,j)=\frac{\exp(S(i,j))}{\sum(\exp(S(i,j)))} \quad (6)$$

式(6)中, $S(i,j)$ 是特征相似性矩阵的一个元素,式中将特征相似性矩阵归一化为一个概率分布,这个概率分布可以表示每个特征的重要性。这个过程可以使模型更有针对性地关注与目标检测任务相关的重要特征,突出了 SimAM 用于增强目标检测模型的特征表示能力。本文采用 SimAM 注意力机制可以改善目标的关键信息被干扰因素混淆的问题,在保持实时性检测的前提下,一定程度提高了检测精度,并通过消融实验证明。

3.4 WIoU

在水下目标检测中,因训练数据中难以避免地包含低质量示例,所以如距离、纵横比的几何度量都会加剧对低质量示例的惩罚从而使模型的泛化性能下降。传统的损失函数不能在锚框与目标框较好地重合时削弱阻碍对象检测模型泛化的惩罚,过多地干预训练使模型没有很好的泛化能力。对此,本文使用 WIoU 代替 YOLOv5 的 CIoU, WIoU 提出了一个动态的非单调 FM,可以降低高质量锚框的竞争力和掩盖低质量例子的影响。WIoU 的动态非单调聚焦机制使用“离群度”替代 CIoU 对锚框进行质量评估,并提供了明智的梯度增益分配策略。该策略在降低高质量锚框竞争力的同时,也减小了低质量示例产生的有害梯度。这使得 WIoU 可以聚焦普通质量的锚框,并提高检测器的整体性能且使模型的泛化能力得到了有效的提高。

在计算速度上, YOLOv5 引入 WIoU 所增加的计算成本主要在于聚焦系数的计算、IoU 损失的均值统计。在实验条件相同时, WIoU 因为没有对纵横比进行计算反而有更快的速度。在损失函数对比实验中, WIoU 的计算时长仅为 CIoU 的 87.7%。

4 仿真实验与结果分析

4.1 实验配置与评价指标

实验采用 2020 年全国水下机器人(湛江)大赛提供的水下真实图像作为数据集,数据集具有绿色色调图片及正常色调图片,多样性的数据集对水下目标检测算法的准确性和鲁棒性具有考验。数据集包含海胆、海参、扇贝和海星 4 类目标共 6 555 张图片,其中随机选取 4 775 张图片作为训练集,1 000 张图片作为验证集,800 张图片作为测试集。

实验使用 Windows11 操作系统, Intel core i7-12700H 处理器, 16 GB 内存, 显卡型号为 NVIDIA GTX 3070Ti, 8GB 显卡。

实验使用 PyTorch 深度学习框架搭建了 DCN-YOLOv5 模型进行训练,训练过程中的超参数设置如表 1 所示。

表 1 超参数设置

Table 1 Settings of hyperparameters

超参数	值
epoch	100
batch size	8
动量	0.9
权重衰减系数	0.000 5
初始学习率	0.01

实验采用各类的平均精准度(mAP)和平均每张图像的计算时间作为指标来评估模型的性能,其中平均精度与准确率及召回率相关。上述指标的具体计算方法如所式(7)一式(10):

$$P=\frac{T_p}{T_p+F_p} \quad (7)$$

$$R=\frac{T_p}{T_p+F_N} \quad (8)$$

$$A_p=\int_0^1 P dR \quad (9)$$

$$L_{mAP}=\frac{1}{N}\sum_{i=1}^N A_p(i) \quad (10)$$

式(9)中, P 表示准确率(precision, P),表示模型预测正确的正样本(T_p)与模型预测的总正样本(T_p+F_p)之比; R 表示召回率(recall, R),表示预测正确的正样本与总标签数(T_p+F_N)之比; A_p 表示 P - R 曲线所围成的面积; N 表示类别个数; L_{mAP} 表示各类的平均精准度, L_{mAP} 越大表示模型的性能越好。

4.2 实验结果及分析

4.2.1 损失函数对比实验

为了验证 WIoU 损失函数在复杂水下环境中是否能减小模型的计算量,本文对比了 5 种常用的 IoU 边界框回归损失函数(DIoU、CIoU、EIoU、GIoU、SIoU)和引入的改进模型中使用的 WIoU 损失函数。实验结果如表 2 所示。通过对比实验结果,可以评估 WIoU 损失函数在复杂水下环境中的性能和计算量。

表 2 损失函数对比实验结果

Table 2 Comparison of experimental results of loss functions

Loss	$L_{mAP}/\%$	计算时间/ms
DIoU	84.87	24.5
CIoU	84.11	24.1
EIoU	84.89	24.0
GIoU	85.15	23.7
SIoU	85.21	23.3
WIoU	86.22	21.4

实验结果表明:引入 WIoU 无论是平均精准度还是每张图像的平均计算时间都优于其他 IoU 损失函数,进一步验证了模型中引入 WIoU 在水下复杂环境中目标检测的准确率和实时性的合理性和有效性。

4.2.2 注意力机制对比实验

为了验证模型增加 SimAM 注意力机制的有效性,并进一步分析水下复杂环境中不同注意力机制对目标检测性能的影响,本文选取 ECA(Efficient Channel Attention)、GAM(Global Attention Mechanism)、SE(Squeeze-and-Excitation)和 SGE(SpatialGroup-wiseEnhance)注意力模块与之进行对比实验。为了确保公平,比较不同注意力模块的性能,实验将这些注意力模块放置在网络结构的相同位置,并且保持网络的其他结构不进行任何改变,并控制一些因素的影响,例如训练参数、数据预处理方法和数据增强策略等,以确保实验结果的可靠性和可重复性。然后,使用真实水下图像数据集进行训练和评估,比较不同注意力机制模块的性能差异。通过对结果的分析 and 比较,可以得出每种注意力机制的优劣势,并确定 SimAM 在水下复杂环境下目标检测任务中的有效性。实验结果如表 3 所示。

表 3 注意力机制对比实验结果
Table 3 Comparison of experimental results of attention mechanisms

Attention	$L_{mAP}/\%$	计算时间/ms
ECA	84.34	24.6
GAM	84.11	28.8
SE	83.42	24.5
SGE	83.31	24.0
SimAM	85.57	22.9

实验结果表明:引入 SimAM 注意力机制相比于 ECA、GAM、SE、SGE 注意力模块, L_{mAP} 分别提高了约 0.89%、1.12%、1.81%、1.92%,且由于 SimAM 注意力机制计算的高效性,每张图像的平均计算时间对比其他注意力机制均高,进一步验证了 SimAM 在水下复杂环境的目标检测中对准确性和实时性的合理性和有效性。

4.2.3 消融实验

为了进一步验证本文提出的各个改进方法分别对水下复杂环境中目标检测检测精度及计算时间的影响,采用消融实验进行分析。在实验中,保持训练策略一致,每次实验修改提出的一个改进点,以评估其对检测性能的影响。最后,综合所有的改进点,加入所有改进点进行实验,以综合评估模型的性能,实验结果如表 4 所示。这样通过实验结果,可以评估每个改进点对水下复杂环境中目标检测性能的影响,并确定它们的贡献程度。

表 4 消融实验结果
Table 4 Results of ablation experiment

网络模型	$L_{mAP}/\%$	计算时间/ms
YOLOv5	84.11	24.1
YOLOv5+D3	86.33	27.2
YOLOv5+SimAM	85.57	24.9
YOLOv5+WIoU	86.22	21.4
YOLOv5+D3+SimAM+WIoU	87.57	84.5

从实验结果可以看出:经过消融实验的测试,在单独替换 D3 模块之后,由于可变形卷积中新增了用于学习偏移量的卷积层和全连接层,使得计算变得复杂,导致每张图像的平均计算时间增加了 3.1 ms,但 L_{mAP} 较原模型提升了 2.22%。在模型组单独引入 SimAM 注意力机制之后,由于 SimAM 是一种轻量级的注意力机制且有高效的计算性能,导致每张图像的平均计算时间仅增加了 0.8 ms,但 L_{mAP} 却提升了 1.46%。在单独使用 WIoU 替换原模型的 CIoU 之后,由于 WIoU 的计算成本主要在于聚焦系数的计算和 IoU 损失的均值统计,计算成本较 CIoU 要低很多,所以每张图像的平均计算时间较 CIoU 低了 2.7 ms,且 L_{mAP} 也提升了 2.11%。在将所有改进点全部融合之后,每张图像的平均计算时间较原模型仅提升了 0.4 ms,但 L_{mAP} 却提高了足足 3.46%。进一步验证了本文提出的改进方法在水下复杂环境中目标检测准确性和实时性的合理性和有效性。

4.2.4 网络模型对比实验

为了进一步验证本文所提出的 DCN-YOLOv5 算法在水下复杂环境中目标检测的优势,将本文提出的算法与其他常见算法进行了对比,对比实验结果如表 5 所示。

表 5 网络模型对比实验结果
Table 5 Comparison of experimental results of network models

网络模型	$L_{mAP}/\%$	计算时间/ms
YOLOv5	84.11	24.1
YOLOv3-SPP	75.42	33.0
SSD	68.21	55.2
Faster R-CNN	78.85	93.5
本文算法	87.57	24.5

从实验结果可以看出:本文所提出的 DCN-YOLOv5 算法在每张图像的平均计算时间上虽然较 YOLOv5 原模型有略微提升,但 L_{mAP} 却有显著提升。而对比其他常见算法,不论在 L_{mAP} 还是在每张图像的平均计算时间都有显著的优势。

5 结 论

针对水下环境场景复杂,目标边界模糊,水下目标因光线折射问题导致的外观、形状可能会发生变化而导致的检测困难问题,本文提出并实现了一种基于 SimAM 注意力机制的 DCN-YOLOv5 水下目标检测方法。通过将模型结合可变性卷积、添加注意力机制及替换 IoU 损失函数的方法对水下目标检测所遇到的环境干扰及目标模糊问题取得了精度上的提升。通过对水下真实图像数据集进行实验,验证了本文提出的算法取得了较好的性能表现,与传统方法和其他 YOLOv5 模型相比,DCN-YOLOv5 算法在水下目标检测任务上表现出更高的精准性和鲁棒性。

然而,本文仍然有一些值得改进的地方,例如需要进一步优化模型的结构及参数,扩大数据集规模,替换轻量化的网络,在保证精度的同时提升检测速度,更高效地检测水下目标。

参考文献(References):

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 580-587.
- [2] GIRSHICK R. Fast R-CNN[C]// Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
- [3] REN S, HE K, 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.
- [4] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]// Proceedings of the 14th European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [5] 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. Piscataway: IEEE Press, 2016: 779-788.
- [6] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 7263-7271.
- [7] REDMON J, FARHADI A. Yolov3: An incremental improvement[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 89-95.
- [8] 禹志鹏,白国振,刘怀周.基于 UCM 算法中直方图拉伸的水下图像增强研究[J].重庆工商大学学报(自然科学版), 2022, 39(5): 10-16.
- [9] YU Zhi-peng, BAI Guo-zhen, LIU Huai-zhou. Research on Underwater image enhancement based on histogram stretching in UCM algorithm[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2022, 39(5): 10-16.
- [9] 郝琨,王阔,王贝贝.基于改进 Mobilenet-YOLOv3 的轻量级水下生物检测算法[J].浙江大学学报(工学版), 2022, 56(8): 1622-1632.
- HAO Kun, WANG Kuo, WANG Bei-bei. Lightweight underwater biological detection algorithm based on improved mobilenet-YOLOv3[J]. Journal of Zhejiang University (Engineering Edition), 2022, 56(8): 1622-1632.
- [10] 乔美英,史建柯,李冰峰,等.改进损失函数的增强型 FPN 水下小目标检测[J].计算机辅助设计与图形学学报, 2023, 35(4): 525-537.
- QIAO Mei-ying, SHI Jian-ke, LI Bing-feng, et al. Enhanced FPN underwater small target detection with improved loss function[J]. Journal of Computer-Aided Design & Computer Grap, 2023, 35(4): 525-537.
- [11] 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(4): 747-754.
- QIANG Wei, HE Yu-yao, GUO Yu-jin, et al. Exploring underwater target detection algorithm based on improved SSD[J]. Journal of Northwestern Polytechnical University, 2020, 38(4): 747-754.
- [12] 叶赵兵,段先华,赵楚.改进 YOLOv3-SPP 水下目标检测研究[J].计算机工程与应用, 2023, 59(6): 231-240.
- YE Zhao-bing, DUAN Xian-hua, ZHAO Chu. Research on underwater target detection by improved YOLOv3-SPP [J]. Computer Engineering and Applications, 2023, 59(6): 231-240.
- [13] LI X, TANG Y, GAO T. Deep but lightweight neural networks for fish detection[C]//Proceedings of the OCEANS 2017-Aberdeen. Piscataway: IEEE, 2017: 1-5.
- [14] YANG N, QIANG Z, HUI Y. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021(452): 48-62.
- [15] YANG L, ZHANG R Y, LI L, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]// International Conference on Machine Learning. PMLR, 2021: 11863-11874.
- [16] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 764-773.
- [17] ZHU X Z, HU H, LIN S, et al. Deformable ConvNets v2: More deformable, better results[C]// Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press. 2019: 9308-9316.

责任编辑:李翠薇