计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2025,34(6):188−195 [doi: 10.15888/j.cnki.csa.009862] [CSTR: 32024.14.csa.009862]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于双分支卷积网络的水下目标检测①

王信诚, 朱　明

(中国科学技术大学 信息科学技术学院 自动化系, 合肥 230026)
通信作者: 王信诚, E-mail: wxc2018@mail.ustc.edu.cn

摘　要: 水下目标检测是水下作业中不可或缺的重要技术. 针对水下图像中背景复杂、待检测目标大小形状不同及存在重叠与遮挡等问题, 本文提出了一种基于双分支卷积网络的水下目标检测算法. 首先, 采用两个并行卷积神经网络作为骨干网络, 其中一个分支引入 ECA 注意力机制, 另一个分支采用可形变卷积, 以提高模型的特征提取能力. 其次, 使用 AFF 模块有效融合两个分支提取到的特征. 最后, 采用 PANet 金字塔结构作为颈部网络, 实现多尺度特征融合, 同时增加高分辨率检测头, 以进一步提高对小目标的敏感性. 本文在公开水下数据集 RUOD 上进行对比实验, 结果表明, 本文的改进算法在 RUOD 数据集上的 *mAP*50 达到了 86.8%, 相较于基准 YOLOv8n 模型提升了 2.7%, 并且相比于同规模的其他常见目标检测模型表现更优.

关键词: 双分支; 水下目标检测; 注意力机制; 可形变卷积; 特征融合

引用格式: 王信诚,朱明.基于双分支卷积网络的水下目标检测.计算机系统应用,2025,34(6):188–195. http://www.c-s-a.org.cn/1003-3254/9862.html

Underwater Target Detection Based on Dual-branch Convolutional Network

WANG Xin-Cheng, ZHU Ming

(Department of Automation, School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China)

Abstract: Underwater target detection is an essential technology in underwater operations. To address the challenges posed by complex backgrounds, varying target scales, and the presence of overlapping and occluded objects in underwater images, this study proposes an underwater target detection algorithm based on a dual-branch convolutional network. First, this study employs two parallel convolutional neural networks as the backbone, with one branch integrating the ECA attention mechanism and the other utilizing deformable convolutions to enhance the model's feature extraction capability. Next, it utilizes the AFF module to effectively fuse the features extracted from both branches. Finally, the study adopts the PANet pyramid structure as the neck network to achieve multi-scale feature fusion while incorporating a high-resolution detection head to further improve sensitivity to small targets. Comparative experiments conducted on the publicly available underwater dataset RUOD show that the improved algorithm achieves an *mAP*50 of 86.8% on the RUOD dataset, which is an enhancement of 2.7% over the baseline YOLOv8n model. Moreover, this model outperforms other common target detection models of similar scale.

Key words: dual-branch; underwater target detection; attention mechanism; deformable convolution; feature fusion

1　引言

地球海洋的面积占地球表面积的 70% 以上, 是人类未来发展与探索的重要领域. 这些资源不仅能够满足人类日益增长的需求, 缓解陆地资源枯竭带来的紧

① 基金项目: 科技创新特区计划 (20-163-14-LZ-001-004-01)
收稿时间: 2024-11-01; 修改时间: 2024-12-03, 2024-12-19; 采用时间: 2024-12-24; csa 在线出版时间: 2025-04-28
CNKI 网络首发时间: 2025-04-29

张局势, 还能推动科技创新和产业进步.

水下机器人是当前常用的海洋探测设备, 广泛应用于各种危险和复杂的水下探索任务. 其中, 水下目标检测技术是其关键的探测手段. 一个高效的水下目标检测技术可以显著提升探索任务的工作效率. 然而, 水下采集到的影像往往会由于水质和光线散射等因素的影响而存在色偏和模糊等问题. 此外, 水下生物通常个体较小且分布密集, 这对水下目标检测技术的性能和鲁棒性提出了更大的挑战. 现如今, 随着深度学习技术的出现和发展, 越来越多的学者选择将基于深度学习的目标检测方法应用于水下目标检测领域, 其中单阶段检测方法和双阶段检测方法是目前最常用, 也是效果最好的两种目标检测方法.

单阶段检测方法的推理和识别速度通常更快, 常用于实时性能有更高要求的场景. 其主要原理是在单一的前向传播中通过生成的锚框或网格, 直接完成目标的位置检测和类别检测. 单阶段目标检测的代表方法有 SSD[1]、FSSD[2]和 YOLO 系列[3–6]等. 双阶段目标检测方法则是区别于单阶段目标检测方法一次传播即完成检测的特点, 而是将目标检测过程分为生成候选区域、对候选区域进行具体分类和细致化定位两个阶段. 相较于单阶段检测方法而言, 双阶段检测方法通常能够提供更高质量的目标检测结果, 但是在训练和推理时间上则需要更多的花费. 双阶段目标检测的代表方法有 R-CNN[7]、Fast R-CNN[8]、Faster R-CNN[9]、Mask R-CNN[10]等.

水下目标检测领域已有很多研究. 例如, Bao 等[11]提出了一种并行高分辨率水下目标检测网络, 通过结合高分辨率网络、改进的激活函数和感受野增强模块, 显著提升了复杂水下场景中模糊和小目标的检测能力; Zhang 等[12]提出了一种新型 YOLO 网络 CGC-YOLO, 通过引入 CSPCBAM 模块增强模型提取复杂特征的能力, 以提高模糊对象检测性能, 同时使用 Cluster-NMS 在训练中保留被遮挡目标, 来提高对遮挡目标的识别效果. Yi 等[13]提出了 USSTD-YOLOv8n, 是一种改进的 YOLOv8n 算法, 通过引入 CARAFE 上采样和上下文引导块, 显著提升了水下小目标检测的性能和普适性, 且相较于 YOLOv8n 在多个数据集上均取得了更好的表现. Sun 等[14]提出了一种基于 MobileViT 和 YOLOX 的水下目标检测模型, 通过引入 DCA 机制来提高模型对全局特征的提取能力, 在保持较高检测精度的同时显著减少了参数量, 使其满足水下无人平台轻量化的需求. Lei 等[15]提出了一种改进的 YOLOv5 模型, 通过结合 Swin Transformer 和优化的多尺度特征融合方法, 使模型能够更关注水下的模糊目标, 使其在复杂水下环境下也有不错效果. Guo 等[16]提出了 UW-YOLOv8 模型, 其采用全新的 FBiFPN 结构和 LC2f 模块, 不仅有效减少了模型的计算量, 还提高了模型的检测性能. Zhou 等[17]基于 YOLO 系列模型, 并采用双边注意力机制和重采样方法, 来解决水下数据集中的噪声和类别不平衡问题, 有效地提高了复杂海洋环境中的目标检测精度.

随着目标检测算法的不断发展, 其在水下目标检测领域发挥了重要作用. 然而, 当下的水下目标检测任务仍面临许多挑战: 大多数水下图片模糊不清、对比度低, 水下待检测目标尺度较小且分布密集, 同时目标间相互遮挡的现象也较为普遍. 这些问题导致水下图像特征提取困难, 提取到的特征信息耦合严重, 从而影响算法的识别率. 针对这些挑战, 本文提出了一种基于双分支卷积网络的水下目标检测方法. 首先, 在 YOLOv8n 骨干网络的基础上, 引入另一路并行的 CNN 骨干网络, 采用双分支卷积网络实现图像的特征提取; 其次, 两个骨干分支分别使用 ECA 注意力机制[18]和可形变卷积, 用于改善模糊目标和遮挡目标的特征提取效果; 同时使用 AFF[19]特征融合模块, 联合两个分支所提取的不同特征, 以实现不同分支的特征融合; 最后采用 PANet[20]金字塔结构实现不同尺度下的特征融合, 进一步提高水下目标的检测精度.

2 相关工作

2.1 YOLOv8 网络结构

YOLOv8 是 Ultralytics 公司于 2023 年提出的一种轻量级单阶段目标检测网络结构, 主要结构包括: 输入端 (Input)、骨干网络 (Backbone)、颈部网络 (Neck) 和检测头 (Head) 共 4 部分, 其网络结构图如图 1 所示. 输入端的主要工作是接收输入图像并进行包括缩放尺寸、自适应填充和数据增强等在内的预处理操作. 首先, 对数据集中图像进行统一的尺度缩放, 将其统一调整为 640×640 的分辨率; 接着, 采用自适应图像填充改善图像边界, 保持正确的长宽比例; 最后, 通过混合增强等技术扩充数据集. 骨干网络的功能是提取各种图像特征. YOLOv8 的骨干网络除了常见的 CBS 模块外

还引入了 C2f 模块. 该模块借鉴了 YOLOv7[21]中的 C3 模块残差结构, 能获取更多的梯度流信息. 骨干网络的尾部使用 SPPF[22]模块, 实现特征图的池化与融合操作. 颈部网络负责特征融合和增强, 采用了路径聚合网络结构, 通过自底向上和自顶向下的双向路径实现不同层次特征的结合. 检测头部分则用于决策并产生最终的检测结果, 采用无锚框和解耦检测头, 通过多次池化和卷积操作实现分类与回归.

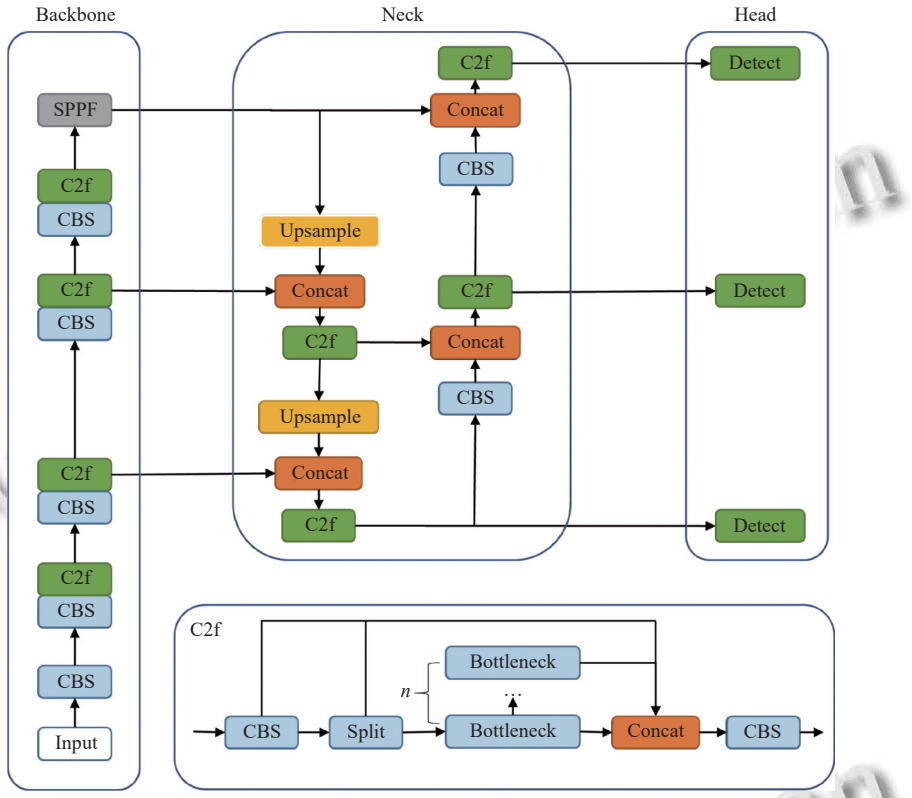

图 1 YOLOv8 网络结构图

2.2 本文工作

为了克服水下目标尺度较小、目标被遮挡以及水下图像对比度低等局限性, 提高水下目标的检测精度, 本文贡献如下: 1) 以 YOLOv8n 作为骨干网络, 增加高分辨率检测和特征融合分支, 加强网络对小目标的检测能力. 2) 结合双分支联合学习的思想, 采用双分支骨干结构实现图像特征提取, 同时加入注意力机制和可形变卷积操作, 以加深网络对模糊目标和遮挡目标的感知. 3) 采用 AFF 模块联合两个不同的骨干分支, 实现两个骨干分支下不同特征的融合. 实验证明, 本文算法性能, 在公开的 RUOD 数据集上优于 YOLOv8n 等常见的几种目标检测模型.

3 算法设计和实现

YOLOv8 有 n、s、m、l 和 x 共 5 个版本, 本文选取 YOLOv8n 作为基准模型, 该模型参数量更小, 且检测速度更快. 本文在 YOLOv8n 网络结构基础上, 结合多分支特征融合思想, 提出一种基于双分支卷积网络的水下目标检测算法, 具体网络结构如图 2 所示. 骨干网络采用双分支并行的方式, A 分支在原先骨干中引入 ECA 注意力机制, 使用 C2f_ECA 模块, 可以在保持计算效率的同时提升模型对于关键特征的感知能力; B 分支则引入可变形卷积 DCN[23], 提高模型对遮挡目标和不同尺度目标的特征提取能力; 同时增加高分辨率的 P2 分支, 并采用 AFF 模块实现不同分支间特征的融合, 增强模型对不同类型特征的感受能力; 颈部网络则采用 PANet 特征金字塔结构, 融合骨干网络的输出特征; 头部网络采用解耦头结构 (decoupled-head).

3.1 C2f_ECA 模块

注意力机制可以帮助模型更好地捕捉关键信息, 让网络在处理输入数据时集中注意力于重要部分, 忽

略不重要的部分, 让模型无论在效率还是效果上都更上一层楼. 同时通过动态调整模型权重的方式, 改变模型对不同特征的关注程度, 使得模型更加聚焦于高权重的重要信息. 具体而言, 模型可以学习到每个通道特征的重要性, 并据此调整不同通道的输出权重, 从而使得网络更加关注对当前任务有利的特征. 本文在骨干网络的 A 分支中引入 ECA 注意力机制, 以增强不同通道间特征的相互关系. ECA 注意力机制关注通道于注意力权重之间的关系, 是一种具有不错效果的通道注意力机制, 其模块的实现机制如图 3 所示.

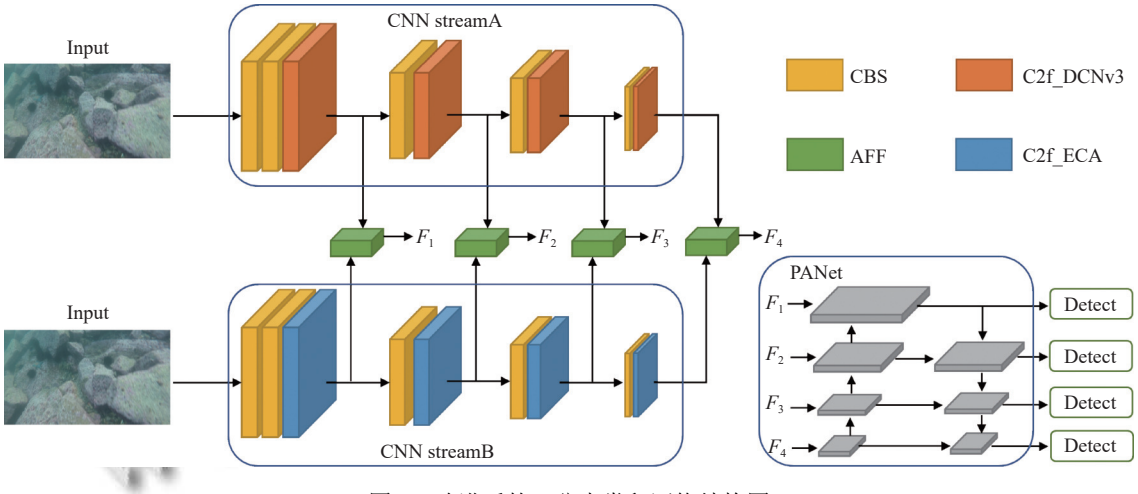

图 2 改进后的双分支卷积网络结构图

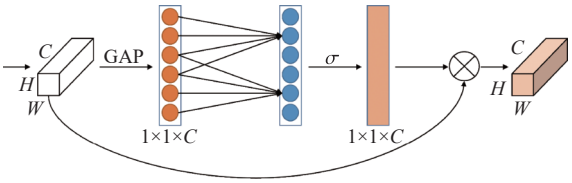

图 3 ECA 注意力示意图

ECA 注意力机制利用一维卷积在不同通道间实现高效交互, 在保持特征图维度不变的情况下, 从中提取各通道间信息的相互依赖关系. 这种保持维度不变的方法在降低了计算复杂度的同时, 还能够使模型更好地关注重要特征, 从而提高模型的整体特征表达能力. 首先进行全局平均池化 (global average pooling, GAP) 操作, 使输入的特征图维持通道数不变; 然后进行卷积核大小为k的一维卷积操作, 并通过 Sigmoid 激活函数得到各通道的权重, 如式 (1) 所示:

$$\omega = \sigma(C1D_k(y)) \tag{1}$$

其中, ω为各通道的权重, σ为激活函数, $C1D$为一维卷积操作; 最后采用对应元素相乘的方式将计算得到的结果与原始输入的特征图进行融合, 得到最终输出的特征图.

整个 ECA 注意力机制的交互影响范围, 即一维卷积核的大小k, 与输入特征图的通道数C有关, k与C之间有如下映射关系:

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{Odd} \tag{2}$$

其中, γ和b分别设置为 2 和 1, Odd表示离结果最近的奇数值, ψ表示映射关系.

本文结合 ECA 注意力模块, 设计了 C2f_ECA 模块替代原有的 C2f 模块, 将 C2f 模块中的 Bottleneck 结构替换为新的 ECA_Block 结构, 在原有的 CBS 结构之后加入了 ECA 模块, 使得改进后的 C2f_ECA 模块可以同时兼顾有效降低模型计算量和实现特征的局部跨通道交互两个特点.

3.2 可形变卷积 DCN

常规的卷积操作是将特征图划分成多个相同的区域, 使这些区域大小与卷积核大小相同, 并对每个区域进行卷积操作, 这种方式下, 卷积的区域在特征图上的大小和位置是固定的. 这种固定的卷积方法在面对形状和尺寸较为复杂的物体时, 实际使用效果不佳, 会导致特征提取不准确. 而水下环境的待检测目标, 通常具有大小不一、形态多样的特点, 这使得常规的卷积操作在水下目标检测任务中发挥的作用十分有限.

为应对这种情况, 本文在骨干网络 B 分支中采用可形变卷积 DCNv3 (deformable convolutional network)

来替换 C2f 模块中的部分常规卷积操作, 以加强模型对不同大小、不同形状物体的特征提取能力. 可形变卷积在感受野中引入了可学习的偏移量, 不断学习和变化的偏移量会将感受野区域变成与物体实际形状相近的区域, 而不再仅是一个固定大小矩形区域. 具体的示意图如图 4 所示, 其中图 4(a) 为标准卷积, 图 4(b)–(d) 为可形变卷积.

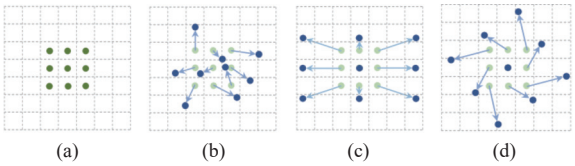

图 4 可形变卷积示意图

传统卷积的计算方式如式 (3) 所示, 其中 p_0 是输出特征图的每个点, 对应卷积核中心点, p_n 则是该点在对应卷积核范围内的每个偏移量.

$$y(p_0)=\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n) \tag{3}$$

而可变形卷积则在式 (3) 的基础上为每个点引入了一个偏移量 Δp_n, 该偏移量通常是非整数.

$$y(p_0)=\sum_{p_n\in R} w(p_n)\cdot x(p_0+p_n+\Delta p_n) \tag{4}$$

为解决非整数偏移量带来的像素点实际位置不对应问题, 通常需要使用双线性插值来得到偏移后的像素值. 使用双线性插值的方法可以将非整数位置映射到特征图中的具体的像素点, 具体如式 (5) 所示:

$$x(p)=\sum_q \max(0,1-|q_x-p_x|)\cdot\max(0,1-|q_y-p_y|)\cdot x(q) \tag{5}$$

其中, $p=p_0+p_n+\Delta p_n$, q 为 p 周围的 4 个整数点.

总结而言, 可形变卷积可以使网络在训练过程中动态调整其感受野的大小和形状, 使其能够更好地适应不同尺度和形状的图像特征. 在面对几何结构复杂的水下图像时, 采用可形变卷积可以在训练过程中不断学习最优的采样位置, 最终提取更有意义和更具代表性的特征. 本文将可形变卷积与 C2f 模块结合, 提出 C2f_DCNv3 模块, 具体结构如图 5 所示. 将 C2f 模块中的部分常规卷积替换成可形变卷积 DCNv3, 利用可形变卷积动态调节感受野的能力, 提高模块对不同尺寸、不同形状物体的特征学习能力, 最终增强整个模型的泛化能力.

3.3 AFF 特征融合模块

为了融合两个骨干分支所提取到的特征信息, 本文采用了 AFF 特征融合模块, 将从两个不同骨干分支提取的图像特征进行融合, 以实现更丰富的特征表达. AFF 模块是 2021 年由 Dai 等[19]提出的一种基于通道注意力的特征融合机制, 通过融合不同层或分支的特征, 从而提高算法的识别精度. AFF 模块的具体结构如图 6 所示.

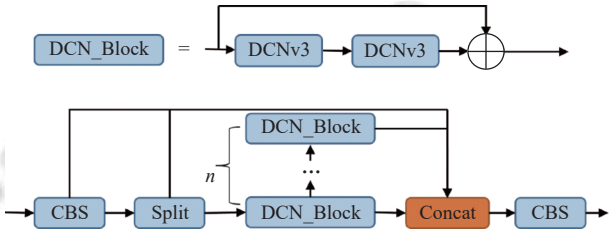

图 5 C2f_DCNv3 示意图

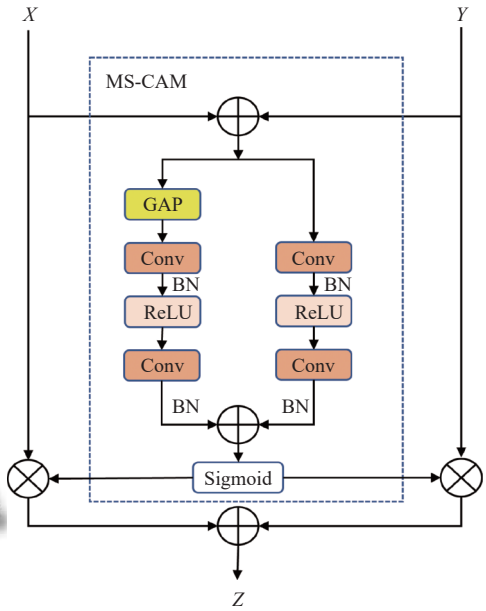

图 6 可形变卷积示意图

AFF 模块以多尺度通道注意力模块 (MS-CAM) 为核心, 该模块采用 1×1 的卷积核来对每个像素点进行单独处理, 以此关注通道的尺度问题, 并针对每个像素位置都单独计算通道交互, 同时在注意力模块内部组合局部上下文特征和全局上下文特征. MS-CAM 的结构如图 6 中的虚线部分所示. 整个 AFF 模块的整体实现流程大致如式 (6):

$$Z=M(X\oplus Y)\otimes X+(1-M(X\oplus Y))\otimes Y \tag{6}$$

首先采用对应位置相加的方法将输入的两个特征 X 和 Y 进行初步融合得到求和后的特征, 再经过多尺

度通道注意力模块得到数值在 0–1 之间的输出权重, 并将 X 和 Y 分别与输出权重进行对应位置数值相乘, 最后再次执行对应位置相加得到融合后的特征 Z.

3.4 新增高分辨率检测头

在 YOLOv8n 骨干网络中, 输入图片经过 4 次下采样的卷积操作, 分别产生了 4 个输出层: p2、p3、p4 和 p5, 对应的特征图分辨率依次为 160×160、80×80、40×40 和 20×20. 然而, 原始 YOLOv8n 对象检测模型仅使用了 p3、p4 和 p5 这 3 个输出层. 在颈部网络中, 采用了 PANet 结构, 以实现对骨干网络 3 个输出层的多尺度特征融合, 最终检测头将在这 3 个特征图上进行目标检测.

由于水下待检测目标通常较小, 原始图像经过多次缩小采样操作后, 其蕴含的原始目标特征信息会不断减少, 从而导致检测效果不佳. 因此, 本文在原有的 3 个输出层 (p3、p4、p5) 的基础上, 增加了高分辨率的 p2 层分支, 以增强网络对小目标的检测性能. 因为对于 p2 层而言, 它的输出特征图分辨率为 160×160, 在骨干网络中只经过了两次缩小采样的卷积操作, 因此保留了更多的图像底层特征信息. 在颈部网络中, 新增一层融合分支, 将 p2 层特征与骨干传递下来的同尺度特征进行融合, 并添加对应的检测和分类头部, 加强模型对小型待检测目标的识别和分类能力.

4 实验分析

4.1 数据集

本文实验采用了 RUOD 水下数据集, 该数据集共包含 14 000 张图片, 涵盖了鱼类、海胆、珊瑚、海星、海参、扇贝、潜水员、墨鱼、海龟和水母共计 10 种常见水下目标类别. 此外, 该数据集中还提供了真实的水下检测场景, 具有如模糊和遮挡等效果的样本. 我们按照 7:2:1 的比例对数据集进行了划分, 从中随机选取了 9 800 张图片作为训练集, 2 800 张图片作为验证集, 剩下的 1 400 张图片则作为测试集.

4.2 实验设置

实验设置的最大训练次数为 300 轮次, warmup 设置为 3 个轮次, 优化器类型使用 SGD, 权重衰减系数设置为 0.000 5, 每个训练批次中的图像数量为 16, 图片的输入尺寸设置为 640×640 分辨率, 初始学习率设置为 0.01. 同时在训练其他模型用于对比实验时, 为确保实验结果的公正性和不同模型间的可比对性, 所有实验均采用相同的参数配置. 实验中使用的电脑配置如表 1.

表 1 训练所用机器配置表

类型	型号	参数
系统	Ubuntu	22.04 LTS
处理器	i5-8400	6核
显卡	GTX 1080Ti	11 GB
内存	DDR4	32 GB

4.3 评价指标

为了验证实验结果, 并通过定量的方式分析不同方法的性能, 本文使用 $F1$ 分数 ($F1$-Score)、精确率 (Precision)、召回率 (Recall) 和平均精度均值 (mean average precision, mAP) 来作为实验的评价指标.

$F1$ 分数、精确率和召回率的定义如下:

$$F1=\frac{2\times P\times R}{P+R} \tag{7}$$

$$P=\frac{TP}{TP+FP} \tag{8}$$

$$R=\frac{TP}{TP+FN} \tag{9}$$

其中, FP 是假阳性数目、FN 是假阴性数目、TP 是真阳性数目、P 表示精确率、R 表示召回率. 精确率 P 表示在所有被模型分类为正例的样本中, 真实类别为正例的样本比例, 也就是模型预测正确的准确率; 召回率 R 表示在所有实际的正例样本中, 被模型正确检测出来的比例, 也就是所有被识别出来的正例占比. 本文实验中使用精确率 P 来评估模型预测为正例的样本中的准确程度, 使用召回率 R 来评估模型对正例的预测能力, 而 $F1$ 分数则用于综合评估模型的性能, 由 P 和 R 计算得来, 表示综合考虑两者因素之后的调和平均值. 上述 3 种指标越接近 1, 表示模型性能越好.

平均精度均值的定义如下:

$$AP=\int_0^1 P(R)\mathrm{d}x \tag{10}$$

$$mAP=\frac{\sum_{i=1}^{N}AP_i}{N} \tag{11}$$

其中, AP 表示平均精度 (average precision), N 为类别数目. mAP 的计算方式通常有两种: mAP50 和 mAP50-95. 其中, mAP50 表示在 IoU (intersection over union) 阈值为 0.5 时的检测平均精度, 而 mAP50-95 则是在 IoU 阈值范围为 0.5–0.95 之间计算的平均精度. 本文采用的指

标为 *mAP*50. *mAP* 的取值范围在 0–1 之间, 数值越接近 1, 表示检测的平均精度越高, 模型的性能也越佳.

4.4 实验结果

为了验证本文改进算法在水下目标检测上的有效性, 本文采用 RUOD 数据集作为基准数据集, 在设置相同参数的前提下, 将多个不同的模型同时进行训练, 并与本文提出的改进模型进行对比, 同时在定量分析中选择精确率 *P*、召回率 *R*、*F*1 分数和 *mAP*50 作为实验的对比指标. 具体对比结果如表 2 所示. 从表 2 可以看出, 本文算法的 *mAP*50 指标达到了 86.8%, 相较于改进前的基准模型 YOLOv8n 的输出结果 84.1% 而言提高了 2.7%; 与 YOLO 系列的其他算法进行比较, 比 YOLOv5n、YOLOv9t、YOLOv10n 和 YOLO11n 分别提高了 3.0%、2.6%、2.2% 和 1.9%. 同时本文算法的 *F*1 分数也比基准模型 YOLOv8n 高了 2.1%, 相较于其他 YOLO 系列的算法均有所提高, 说明在综合考虑精确率和召回率之后, 本文算法的综合性能是最优的. 与 UW-YOLOv8[16]模型相比, 二者 *mAP*50 结果相同, 但是本文算法参数量与计算量远低于 UW-YOLOv8, 在轻量性上表现出更好的性能. 除此之外, 本文模型相对于基准模型 YOLOv8n 参数量升高了 0.4M, 计算量则增加了 4.1G, 提高的幅度均不大. 虽然使用双分支结构并引入 ECA 注意力机制和可形变卷积使得模型大小与计算量有所增加, 但提升幅度并不大, 仍属于轻量化模型, 完全能满足水下环境部署的要求.

表 2 不同算法检测结果

模型	*P* (%)	*R* (%)	*F*1 (%)	*mAP*50 (%)	Params (M)	FLOPs (G)
YOLOv5n	84.9	76.3	80.4	83.8	2.7	7.5
YOLOv8n	85.3	77.5	81.2	84.1	3.4	8.7
YOLOv9t	84.8	77.2	80.8	84.2	**2.1**	7.6
YOLOv10n	85.0	77.0	79.3	84.6	2.8	8.2
YOLO11n	85.7	77.2	81.2	84.9	2.6	**6.4**
UW-YOLOv8	86.4	**80.7**	83	86.8	16.3	23.8
本文算法	**86.5**	80.4	**83.3**	**86.8**	3.8	12.8

我们选取了部分水下图像, 并使用不同的检测模型进行对比, 以此来展示本文改进算法在不同情况下的识别效果, 具体检测效果如图 7 所示. 图 7 第 1 行展示的是常见的普通水下场景图像, 其中包含了大、中、小 3 种尺寸的水下目标. 从结果来看, YOLO 系列模型能够准确识别大、中尺寸的目标, 但少数模型会漏检图像边缘的小目标. 而本文改进的模型则成功识别出所有待检测目标. 面对遮挡场景时, 待检测目标往往被岩石等物体部分遮挡, 此时原始 YOLO 模型的检测效果明显不如改进模型. 从图 7 第 2 行的结果可以看出, YOLO 系列原始模型存在大量漏检现象, 而本文改进模型则能够识别出绝大多数被遮挡的目标.

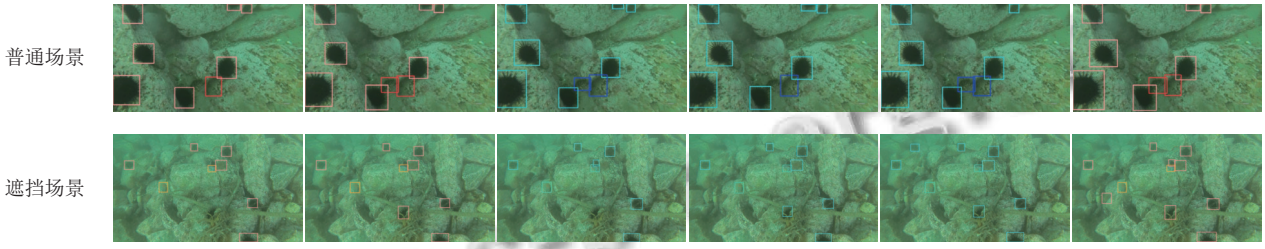

图 7 检测结果对比图

4.5 消融实验

为了分析各模块组合对模型整体性能的影响, 本文设计了消融实验, 在相同硬件和参数配置的前提下, 分别对各模块的单独和组合效果进行分析, 具体如表 3 所示. 其中“+”号表示在基准模型 YOLOv8n 的基础上使用了该模块. 由表 3 可以看出, 与原始的 YOLOv8n 相比, 单独添加 p2 高分辨率分支和 ECA 注意力机制后, *mAP*50 分别提高了 0.7% 和 0.3%, 检测效果有所提升但并不显著; 加入可形变卷积 DCN 之后, *mAP*50 提高了 1.4%, 这是由于可形变卷积对水下多尺度以及遮挡目标有着更好的特征提取效果; 在采用双分支结构, 并加入 AFF 特征融合模块之后, *mAP*50 进一步提高到了 86.8%, 相较于原始网络上升了 2.7%, 提升效果较为显著. 这说明了本文所提出的模块以及双分支的骨干结构, 对原始算法确实有改进效果.

表 3 消融实验结果 (%)

模型	*P*	*R*	*F*1	*mAP*50
YOLOv8n	85.3	77.5	81.2	84.1
+p2	84.4	77.4	80.7	84.8
+ECA	84.5	77.6	80.8	84.4
+DCN	85.7	79.3	79.3	85.5
+ECA+DCN	85.5	78.4	81.2	85.3
本文算法	**86.5**	**80.4**	**83.3**	**86.8**

5 结论与展望

本文提出一种基于双分支卷积网络的水下目标检测算法. 采用两个并行卷积网络作为骨干, 两个分支分别使用了 ECA 注意力机制和 DCN 可形变卷积, 提高骨干网络对模糊目标和遮挡目标的特征提取能力; 使用 AFF 模块实现两个不同分支间的特征融合, 帮助模型更精确捕捉目标的特征信息; 最后在颈部网络引入高分辨率分支, 并在头部位置增加对应的检测和分类模块, 提高整体模型对小型目标的检测效果. 经过多次实验验证, 本文改进后的方法在 RUOD 数据集上达到 86.8% 的 *mAP*50, 最终检测效果优于原始的 YOLOv8n 模型和其他常见的目标检测模型. 下一步的研究工作, 将在本文的基础上致力于算法的轻量化与泛用性.

参考文献

1 Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.

2 Li ZX, Yang L, Zhou FQ. FSSD: Feature fusion single shot multibox detector. arXiv:1712.00960, 2017.

3 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.

4 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525.

5 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.

6 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv:2004.10934, 2020.

7 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587.

8 Girshick R. Fast R-CNN. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448.

9 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]

10 He KM, Gkioxari G, Dollár P, *et al*. Mask R-CNN. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2961–2969.

11 Bao ZW, Guo Y, Wang JY, *et al*. Underwater target detection based on parallel high-resolution networks. Sensors, 2023, 23(17): 7337. [doi: 10.3390/s23177337]

12 Zhang Z, Tong QS, Huang XF. An efficient YOLO network with CSPCBAM, ghost, and cluster-NMS for underwater target detection. IEEE Access, 2024, 12: 30562–30576. [doi: 10.1109/ACCESS.2024.3368878]

13 Yi WG, Yang JW, Yan LW. Research on underwater small target detection technology based on single-stage USSTD-YOLOv8n. IEEE Access, 2024, 12: 69633–69641. [doi: 10.1109/ACCESS.2024.3400962]

14 Sun Y, Zheng WX, Du X, *et al*. Underwater small target detection based on YOLOX combined with MobileViT and double coordinate attention. Journal of Marine Science and Engineering, 2023, 11(6): 1178. [doi: 10.3390/jmse11061178]

15 Lei F, Tang FF, Li SH. Underwater target detection algorithm based on improved YOLOv5. Journal of Marine Science and Engineering, 2022, 10(3): 310. [doi: 10.3390/jmse10030310]

16 Guo A, Sun KQ, Zhang ZY. A lightweight YOLOv8 integrating FasterNet for real-time underwater object detection. Journal of Real-time Image Processing, 2024, 21(2): 49. [doi: 10.1007/s11554-024-01431-x]

17 Zhou ZY, Hu YJ, Yang XF, *et al*. YOLO-based marine organism detection using two-terminal attention mechanism and difficult-sample resampling. Applied Soft Computing, 2024, 153: 111291. [doi: 10.1016/j.asoc.2024.111291]

18 Wang QL, Wu BG, Zhu PF, *et al*. ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11531–11539.

19 Dai YM, Gieseke F, Oehmcke S, *et al*. Attentional feature fusion. Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 3559–3568.

20 Liu S, Qi L, Qin HF, *et al*. Path aggregation network for instance segmentation. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 8759–8768.

21 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.

22 Liu ST, Huang D, Wang YH. Learning spatial fusion for single-shot object detection. arXiv:1911.09516, 2019.

23 Dai JF, Qi HZ, Xiong YW, *et al*. Deformable convolutional networks. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 764–773.

(校对责编: 王欣欣)