



DOI:10.12404/j.issn.1671-1815.2403743

引用格式:董从鑫,刘庆华.基于改进 YOLOv8 的船只目标检测算法[J].科学技术与工程,2025,25(12):5093-5102.

Dong Congxin, Liu Qinghua. Ship target detection algorithm based on improved YOLOv8[J]. Science Technology and Engineering, 2025, 25(12): 5093-5102.

基于改进 YOLOv8 的船只目标检测算法

董从鑫,刘庆华*

(桂林电子科技大学信息与通信学院,桂林 541010)

摘要 针对传统船只目标检测算法的精度较低、船只检测类别单一等问题,提出一种改进的 DGA-YOLOv8 海上船只目标检测算法。首先,网络采用可变形卷积扩大模型的感受野,通过引入可学习偏移量,使模型能够根据物体实际形状自适应调整感受野大小和形状,确保卷积区域能够精确覆盖船只物体的轮廓。其次,引入 GAM(global attention mechanism)注意力机制,使网络能够有效突出船只目标的关键特征,提高目标识别能力。实验结果表明:改进后算法的精确度和平均精度均值(mAP)达到 96.4% 和 92.2%,FPS(frame per second)为 43.55,在提升精度的同时也保证了一定的检测速度,满足了实时性检测的需求。同时与其他主流算法对比,其中包括 Faster R-CNN(faster region-based convolutional neural network)、YOLOv5s 和 YOLOv10 等。结果表明:所提算法具有更高的平均精度和更显著的分类效果。

关键词 YOLOv8; 船只目标检测; 可变形卷积; 注意力机制

中图分类号 TP391.4;

文献标志码 A

Ship Target Detection Algorithm Based on Improved YOLOv8

DONG Cong-xin, LIU Qing-hua*

(Information and Communication College, Guilin University of Electronic and Technology, Guilin 541010, China)

[Abstract] An improved DGA-YOLOv8 offshore ship target detection algorithm was proposed to tackle the issues of low accuracy and single ship detection categories that are present in traditional ship target detection algorithms. Firstly, the network was adapted to include deformable convolution, which expanded the model's receptive field. Learnable offsets were introduced, allowing the model to adaptively adjust the size and shape of the receptive field in response to the actual shape of the object, ensuring that the convolution area can precisely cover the contour of the ship object. Secondly, the incorporation of a GAM(global attention mechanism) attention mechanism enabled the network to effectively emphasize the key features of ship targets, thereby enhancing the target recognition capability. The experimental results demonstrate that the improved algorithm achieves accuracy and average accuracy mean (mAP) of 96.4% and 92.2%, respectively. An frames per second(FPS) of 43.55 is recorded, indicating not only an enhancement in accuracy but also the maintenance of a certain detection speed, thus fulfilling the requirements for real-time detection. When compared with other mainstream algorithms, such as faster region-based convolutional neural network(Faster R-CNN) and YOLOv5s, YOLOv10. The results show that the proposed algorithm exhibits higher average accuracy and significant superior classification performance.

[Keywords] YOLOv8; ship target detection; deformable convolution; attention mechanism

随着海上运输行业的不断发展,交通环境日渐变得复杂,船只失事等交通事故日趋增多。此外,由于海洋处于国家边境地带,常有船只走私、偷渡等危害国家安全利益的事态发生。因此,做好海上安防管理、船只监督与检测工作以及开展相关研究具有尤为重要的意义。

目前,常用海上船只目标检测的方法主要分为传统船只检测方法和基于深度学习的船只目标检测方法。传统船只检测方法主要包括使用合成孔

径雷达(synthetic aperture radar,SAR)和通过卫星遥感图像等方式^[1]。石洪基等^[2]优化了 SAR 图像中的船只目标,改进区域卷积神经网络输入和目标候选框提取,实现了多分辨率 SAR 图像中密集分布、沿岸分布等场景下的船只检测。王程丽等^[3]针对 SAR 船只目标样本量少的问题,提出了一种基于 VGG16 迁移学习的识别方法。对模型进行参数微调,解决过拟合和局部最优问题,能够有效地提高 SAR 船只目标的识别准确率。刘方坚等^[4]提出的

收稿日期:2024-05-21; 修订日期:2025-01-22

基金项目:广西创新驱动发展专项(桂科 AA21077008)

第一作者:董从鑫(2001—),男,汉族,广西贺州人,硕士研究生。研究方向:目标检测。E-mail:2440701788@qq.com。

*通信作者:刘庆华(1974—),女,汉族,四川南江人,博士,教授,博士研究生导师。研究方向:阵列信号处理。E-mail:qhliu@guet.edu.cn。

基于视觉显著性的 NanoDet 舰船检测方法,通过自动聚类划分场景、差异化显著性检测、优化模型学习显著性图特征实现 SAR 舰船图像的快速和高精确度检测效果。传统的方法虽然可以实现船只检测,但也有一定的缺陷,SAR 的图像容易产生各种噪点,且易受到海岸线背景等因素影响;卫星遥感方式虽然具有相对直观的光学图像,但是易受天气条件影响,且无法准确识别出目标外形的特点。

随着深度学习技术的发展,深度学习模型具有更高的检测精度和实时性能,能够准确和高效地完成船只检测任务。在深度学习中,目标检测主要分为一阶段算法和二阶段算法^[5]。二阶段算法需要先产生候选区域,再进行目标检测任务,如 R-CNN (region-based convolutional neural network)^[6]、Fast R-CNN (fast region-based convolutional neural network)^[7]和 Faster R-CNN (faster region-based convolutional neural network)^[8]等。

一阶段算法可以同时完成分类和定位任务,YOLO 系列以其优越的性能被广泛应用于目标检测领域。2016 年,Redmon 等^[9-10]提出了 YOLOv1、YOLOv2 算法,在此基础上,后续研究人员们提出了 YOLOv3^[11]、YOLOv4^[12]、YOLOv5、YOLOv7^[13]等算法。李静等^[14]优化了 YOLOv3 模型中的 K-means 聚类算法,使聚类中心之间的距离最大化,实现全局最优结果,显著提升了船只目标检测的准确率,整体 mAP (平均精度均值)提高了 9.31%。Liu 等^[15]基于 YOLOv5 算法,结合 GhostbottleNet 特征提取技术,优化特征提取过程,实现更高精度和实时性的船舶检测。Wang 等^[16]提出了一种基于 YOLOv4 的船舶目标检测算法,通过优化锚框尺寸、调整模型结构和应用 Mixup 数据增强技术,提高了检测的准确性和实时性。Zhang^[17]提出了一种基于 CBAM (convolutional block attention module)-YOLOv8 的舰船目标检测算法,使用空间深度卷积提高特征利用率,加入 CBAM 注意力融合特征信息,并采用 MPDIU 损失函数,实现了较高的精度和检测速度。

上述研究中,R-CNN 系列算法虽然检测精度较高,但检测速度较低,不适用于实时检测任务,同时模型体积偏大,不利于在硬件上部署;使用 YOLO 系列算法的研究方案在应用于遥感图像处理时,无法准确识别到目标的外观形状等特点,同时对于船只较小或者图像分辨率较低时,难以准确检测,对船只的检测类别单一。因此,设计一种精度更高、分类效果更加显著的海上船只目标检测算法具有重要意义。

鉴于此,提出一种基于改进 YOLOv8 的船只目标检测算法 DGA-YOLOv8,网络中使用可变形卷积 DCNv2 (deformable ConvNets v2)^[18],通过引入可学习的偏移量,使感受野不再是固定的矩形形状,能根据物体形状进行调整,调整后的卷积区域能够更好地覆盖物体形状周围,实现感受野的扩大。同时引入 GAM (global attention mechanism)^[19]注意力机制,在特征提取后进行注意力的调整,丰富特征信息,提高检测精度。

1 DGA-YOLOv8

1.1 网络结构

由于检测需要同时兼顾精度与速度,以 YOLOv8 系列中最轻量级的 YOLOv8n 模型为基线模型,并在此基础上加以改进,改进后的网络结构如图 1 所示。

将可变形卷积 DCN 和 GAM 注意力机制加入网络中,设计 DGA-YOLOv8 算法,对比原始的 YOLOv8 网络,DGA-YOLOv8 进行以下改进。

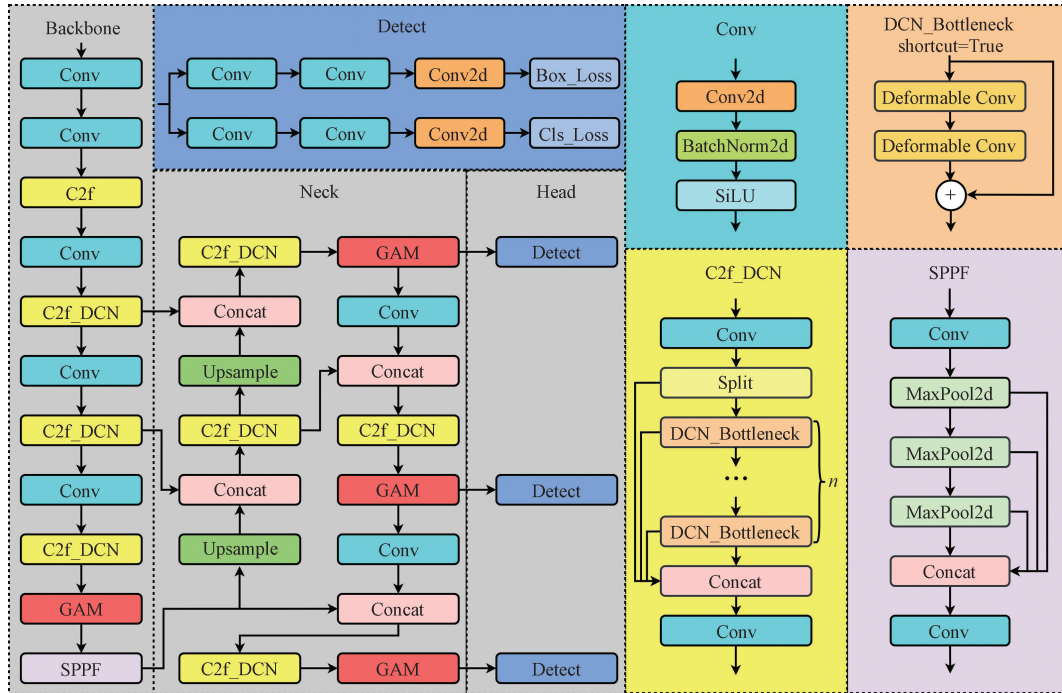
(1) 引入可变形卷积替换 Bottleneck 结构当中的传统卷积方式,设计 C2f_DCN 模块,将原始结构当中的 Backbone 和 Neck 部分结构的 C2f 模块改为 C2f_DCN 模块,使网络在应对不同尺度、形状的目标时能展现出更好的应对能力。将 C2f_DCN 模块放置于不同位置时,进行相关实验,如单独更改 Backbone 部分和单独更改 Neck 部分的 C2f 模块,最终可以产生最优的效果,如图 1 所示。

(2) 在 Backbone 结构中融入注意力机制,将 GAM 注意力模块添加到 SPPF 结构前,增强主干网络的特征提取的能力;同时在 Neck 与 Head 的连接部分中,分别在 3 个 Detect 检测头前加入 GAM 注意力,增强了特征融合的能力,提高了细节信息的利用率。

在 YOLOv8 中加入可变形卷积 DCN 和 GAM 注意力机制后,网络的参数量和计算量分别增加 2.3 M、1.2 GFLOPs (每秒十亿次的浮点运算数),但其网络宽度和深度基本保持不变,同时增加量在网络中的占比有限,目的是通过引入少量的额外参数和计算量来实现模型的精度与性能提升。

1.2 C2f_DCN 模块

可变形卷积通过传统卷积的基础上引入可学习的偏移量来增加网络对船只物体形状的变化适应性,偏移量通过传统卷积计算得出,使得卷积核的采样位置可以根据特征图的内容进行动态调整。相比于传统卷积方式,可以更灵活地适应输入特征的变化,如图 2 所示。通过引入可变形卷积,网络模型可以更好地适应不同尺度、形状和角度的船



Backbone 为骨干网络; Conv 为卷积; C2f 为跨阶段双分支特征融合模块; C2f_DCN 为本文设计模块, 其中 DCN 为可变形卷积网络; GAM 为全局注意力机制; SPPF 为空间金字塔池化快速模块; Neck 为颈部网络; Concat 为拼接; Upsample 为上采样; Head 为头部网络; Detect 为检测模块; Conv2d 为二维卷积; Box_Loss 为锚框损失; Cls_Loss 为类别损失; BatchNorm2d 为 2 维批量归一化; SiLU 为激活函数; Split 为分割; DCN_Bottleneck 为本文设计模块, 其中 Bottleneck 为瓶颈模块; Deformable Conv 为可变形卷积; shortcut = True 为使用残差连接; MaxPool2d 为 2 维最大值池化

图 1 DGA-YOLOv8 网络结构
Fig. 1 Network structure of DGA-YOLOv8

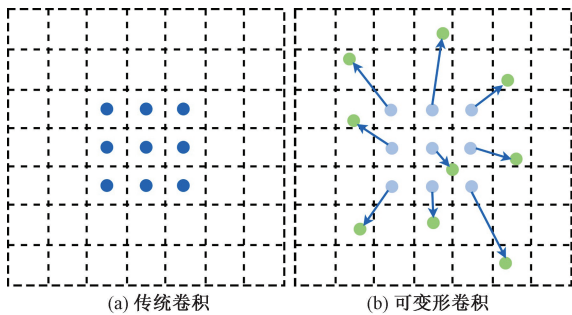


图 2 传统卷积与可变形卷积采样过程

Fig. 2 Traditional convolution and deformable convolution sampling processes

只物体, 有利于提高网络模型在船只检测中的性能和精度。

设输入特征图为 x , w 为卷积核, 大小为 3×3 , y 为输出特征图, 可变形卷积的计算公式为

$$y(p_0) = \sum_{p_n \in R'} w(p_n)x(p_0 + p_n + \Delta p_n) \quad (1)$$

式(1)中: p_0 为特征图上的每个像素点的位置; p_n 为卷积核窗口中其他像素点与中心像素点的相对位置; Δp_n 为位置偏移量; R' 定义了卷积核大小以及卷积核周围像素点的偏移, 可表示为

$$R' = \{(-1, -1), (-1, 0), \dots, (0, 0), (1, 1)\} \quad (2)$$

在 YOLOv8 网络结构中, C2f 模块中使用的传统卷积方式只能对规则的矩形区域进行特征提取, 虽然在处理简单或规则的数据时效果良好, 但针对海上复杂目标时无法完全拟合船只目标的形状和姿态。因此, 在处理尺度不同、形状复杂的目标时, 原本的网络具有一定的局限性, 当出现噪声或者背景干扰时, 漏检或错检的可能性会提升。

为提高模型对不同船只物体的特征提取能力和对复杂目标的适应能力, 利用可变形卷积设计 C2f_DCN 模块, 将原本 C2f 模块中的卷积方式用可变形卷积替换, 其结构如图 3、图 4 所示。首先对输入的特征图进行卷积操作改变通道数, 再通过 Split 操作切分特征, 接着通过堆叠多个可变形卷积模块来扩大网络的感受野。使用可变形卷积可以自适应地调整采样点的位置, 从而不断调节感受野的尺寸和大小, 使网络能够更好地适应输入特征的几何变化, 捕获到更多的对象信息。

1.3 GAM 注意力机制

在网络中引入注意力机制来增强网络对船只

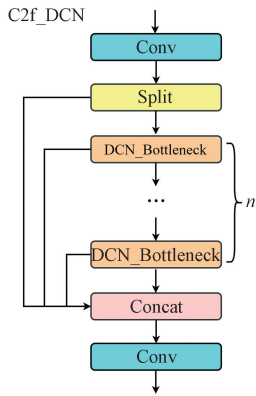


图3 C2f_DCN 模块
Fig. 3 C2f_DCN module

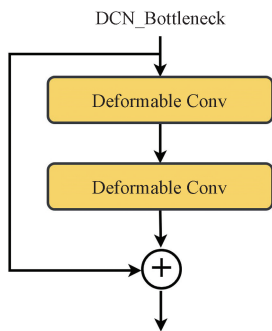


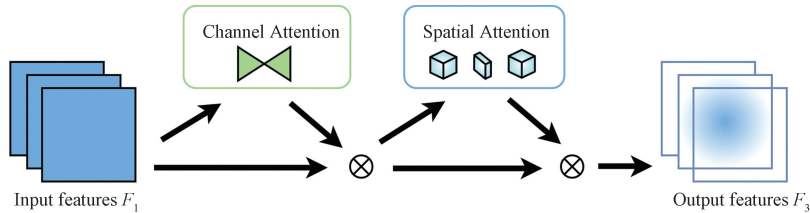
图4 DCN_Bottleneck 结构
Fig. 4 DCN_Bottleneck structure

的特征提取能力。注意力机制可以有效帮助模型聚焦于输入图像中最有信息量的部分,并忽略掉无关的背景信息,使模型能获取更多有效的特征。在网络模型中加入 GAM 注意力机制,有助于模型更好地捕捉全局信息,提高特征的表达能力。其模块结构如图 5 所示,其结合了通道注意力机制和空间注意力机制两部分,不仅能够关注通道信息,同时还重视空间信息,能够在减少信息弥散的情况下,同时放大全局维度交互特征^[19]。输入特征图为 F_1 , 先后经过通道注意力模块和空间注意力模块计算生成通道权重和空间权重,将通道权重和空间权重应用到原始特征图上,得到输出特征图 F_3 。

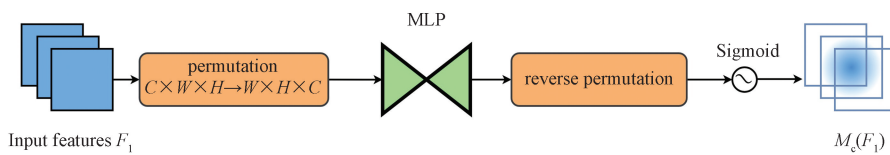
通道注意力子模块结构如图 6 所示,对于给定的输入特征图 F_1 , 将特征图从原本的 $C \times W \times H$ (其中, C 为通道数, W 为宽度, H 为高度) 进行维度转换为 $W \times H \times C$, 将维度转换后的特征图经过多层感知机 MLP, 再转换回原来的维度, 之后经过 Sigmoid 激活函数输出。

空间注意力子模块结构如图 7 所示, 将经过通道注意力子模块的输出特征图 F_2 依次经过两次 7×7 的卷积进行特征信息融合, 最后再通过 Sigmoid 输出。

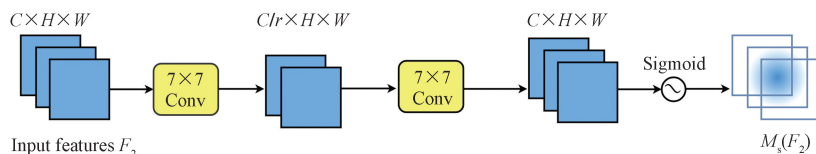
整个过程的定义为



Input features 为输入特征; Channel Attention 为通道注意力; Spatial Attention 为空间注意力; Output features 为输出特征
图5 GAM 注意力模块
Fig. 5 GAM attention module



permutation 为维度转换; MLP 为多层感知机; reverse permutation 为维度反转换; Sigmoid 为激活函数
图6 通道注意力模块
Fig. 6 Channel attention module



Conv 为卷积

图7 空间注意力模块

Fig. 7 Spatial attention module

$$F_2 = M_c(F_1) \otimes F_1 \quad (3)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (4)$$

式中: M_c 为通道注意力模块; M_s 为空间注意力模块; \otimes 表示逐元素相乘。

2 实验环境配置

本次实验在 Windows10 操作系统下进行, GPU 使用 NVIDIA TITAN RTX, 显存为 24 GB, 采用 Pytorch 深度学习框架进行模型训练, 运行环境为: CUDA 11.1、Python 3.8.18、Pytorch 1.9.0。

2.1 实验参数配置

基于对数据集特性、计算资源和硬件环境的综合考虑, 将训练过程中的具体参数设置为如表 1 所示。为了满足算法结构并兼顾计算效率, 将图像大小设置为 640×640 。训练 200 次足以使模型达到稳定的性能, 避免了欠拟合或过拟合。批量大小选择为 16, 避免内存溢出; SGD 在深度学习中有良好的性能, 因此选择 SGD 作为优化算法。初始学习率和学习率衰减因子设为 0.01 以加快初期训练并防止出现局部最优解。权重衰减值设为 0.0005 以简化模型; 动量值设为 0.937 可促进模型快速收敛。

2.2 数据集介绍

本次实验所使用的数据集为自制的海上船只数据集, 其主要包括公开的 SeaShips 海上船只数据集 7 000 张、网络搜集图片、航道实拍船只图片等共计 11 081 张图片, 具体数据分布如表 2 所示。

表 1 实验环境

Table 1 Experimental environment

| 参数 | 配置 |
|---------|------------------|
| 图像大小 | 640×640 |
| 迭代次数 | 200 |
| 批量大小 | 16 |
| 优化器 | SGD |
| 初始学习率 | 0.01 |
| 学习率衰减因子 | 0.01 |
| 权重衰减 | 0.0005 |
| 动量 | 0.937 |

表 2 数据集分布

Table 2 Dataset distribution

| 船只类型 | 训练集/张 | 验证集/张 | 总计/张 |
|-------|-------|-------|--------|
| 矿砂船 | 1 508 | 168 | 1 676 |
| 散装货船 | 1 325 | 148 | 1 473 |
| 普通货船 | 1 188 | 133 | 1 321 |
| 集装箱货船 | 1 052 | 117 | 1 169 |
| 渔船 | 1 154 | 129 | 1 283 |
| 客运船 | 378 | 43 | 421 |
| 邮轮 | 795 | 89 | 884 |
| 大型货轮 | 1 513 | 169 | 1 682 |
| 帆船 | 202 | 23 | 225 |
| 其他类型 | 858 | 89 | 947 |
| 总计 | 9 973 | 1 108 | 11 081 |

2.3 评价指标

为了验证网络模型的性能, 本次实验主要采用准确率 (precision, 记为 P)、召回率 (recall, 记为 R) 和平均精度均值 mAP (mean average precision) 对模型检测性能进行评估。其中准确率是指正确预测的样本 (TP) 在全部预测为正的样本结果中 (TP + FP) 所占的比例, 其计算公式为

$$P = \frac{TP}{TP + FP} \quad (5)$$

召回率则是指正确预测的样本数量 (TP) 在实际为正的样本数量 (TP + FN) 中所占的比例, 其计算公式为

$$R = \frac{TP}{TP + FN} \quad (6)$$

对于每个类别来说, P - R 曲线与坐标轴所围成的面积即为该类别的平均精度值 (AP), 将各类别的 AP 取平均便得到 mAP, 即代表整个数据集中所有类别 AP 的均值, 两者的计算公式分别为

$$AP = \int_0^1 P(R) dR \quad (7)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (8)$$

式中: N 为数据集中样本的类别数; AP_i 为第 i 个类别的平均精度值。

mAP 值作为模型性能的核心指标, 综合考虑了精确度和召回率, 更能够提供一个全面且直观的评估。mAP@0.5 为在 IoU (intersection over union) 阈值设为 0.5 时计算得到的模型的平均精度均值, mAP@0.5:0.95 为在 IoU 阈值由 0.5~0.95 的区间 (步长为 0.05) 内变化时, 计算得到的模型的平均精度均值。

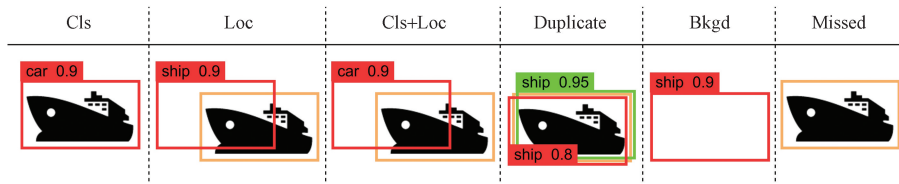
为了更全面地评估算法性能, 使用 TIDE (toolkit for identifying detection and segmentation errors)^[20], 一个用于分析目标检测和实例分割算法中错误来源的框架和工具包。在 TIDE 中定义 6 种错误类型, 如图 8 所示。

3 实验结果对比与分析

主要进行消融实验和算法对比实验, 在消融实验中, 针对算法的不同模块进行系统性能测试与分析, 观察到不同模块对最终性能产生的具体影响。将本文算法与现有的其他主流算法进行对比, 通过在同一数据集上进行实验, 比较不同算法在各关键指标上的表现。

3.1 消融实验

为了分析不同改进模块的有效性, 进行消融实验, 对比各个模块对检测模型性能的影响。表 3 为



分类错误 (Cls) 表示定位正确, 但分类错误; 定位错误 (Loc) 表示分类正确, 但定位错误; 分类和定位错误 (Cls + Loc) 表示分类和定位均错误; 重复检测错误 (Duplicate) 表示多次检测到目标物体; 背景错误 (Bkgd) 表示将背景检测为目标物体; 未检测错误 (Missed) 表示未检测到目标物体

图 8 TIDE 错误类型定义

Fig. 8 TIDE error type definition

表 3 模块消融实验

Table 3 Module ablation experiment

| YOLOv8n | C2f_DCN | GAM | 准确率/ 率/% | 召回 率/% | mAP@0.5/ % | mAP@0.5: 0.95/% |
|---------|---------|-----|-------------|-----------|---------------|--------------------|
| √ | — | — | 93.1 | 86.1 | 91.8 | 73.1 |
| √ | √ | — | 94.7 | 86.3 | 91.4 | 73.1 |
| √ | — | √ | 95.4 | 87.4 | 91.4 | 73.5 |
| √ | √ | √ | 96.4 | 87.3 | 92.2 | 74.1 |

注: √表示使用该模块。

以 YOLOv8n 为基线模型, 依次改进可变形卷积、全局注意力机制的方式进行消融实验的结果。

由表 3 可知, 通过改进 C2f_DCN, 模型的精确度有了良好的提升, 说明网络使用可变形卷积后, 面对复杂的几何变换时能够保持稳定的性能, 对目标特征的提取能力有了较好的提升。在加入 GAM 注意力模块之后, 对模型的精确度有了进一步的提升, 准确率和召回率分别提高了 2.3% 和 1.3%, 说明加入注意力能降低模型对目标的错检概率, 能够更好地区分背景和船只目标, 提升了模型的抗干扰能力。将 C2f_DCN 模块和 GAM 注意力机制同时加入网络时, 形成 DGA-YOLOv8 方法, 则准确率、召回率、mAP@0.5 和 mAP@0.5:0.95 分别提升了 3.3、1.2、0.4 和 1.0 个百分点, 实验结果表明, 当 C2f_DCN 和 GAM 两者一起作用于网络时, 取得了最优的效果。

图 9 为本文模型 DGA-YOLOv8 和 YOLOv8 模型在训练过程中的性能指标变化曲线, 包括锚框损失、类别损失、mAP@0.5 和 mAP@0.5:0.95 随着训练轮数的增加而变化的曲线。从图 9 中可以看出, DGA-YOLOv8 与 YOLOv8 模型相比, 其训练时的损失下降得更快, 模型所需要的收敛时间更短。mAP 的变化同样也表明本文模型性能更优于 YOLOv8 模型, 并进一步证明将可变形卷积和 GAM 注意力机制加入网络中对提高性能有重要作用。

使用 TIDE 工具包计算 YOLOv8 算法和 DGA-YOLOv8 算法的各项错误类别, 结果如表 4 所示。

表 4 TIDE 结果对比

Table 4 Comparison of TIDE results

| 模型 | Cls | Loc | Both | Duplicate | Bkg | Missed | FP | FN |
|------------|-------|-------|-------|-----------|-------|--------|-------|-------|
| YOLOv8 | 1.57 | 1.39 | 0.12 | 0.17 | 1.15 | 0.58 | 5.56 | 2.28 |
| DGA-YOLOv8 | 1.60 | 1.19 | 0.14 | 0.18 | 1.10 | 0.65 | 5.53 | 2.08 |
| 提升 | +0.03 | -0.20 | +0.02 | +0.01 | -0.05 | +0.07 | -0.03 | -0.20 |

注: 加粗数字表示该项错误类型得到改善。

从表 4 结果可以看出, 改进后的模型降低了定位和背景检测的错误率。具体而言, 使用可变形卷积能够帮助模型更好地定位目标物体, 提高检测灵活性和准确性; 注意力的引入使模型在检测过程中关注到更多重要信息, 对目标和背景的区分能力更强。此外, 改进后的模型在改善错检和漏检方面也有更出色的表现, 其加入的可变卷积和注意力机制能提升特征提取能力和抗干扰能力, 不仅减少了对目标的误判, 提升识别的准确性, 还能更全面地检测出目标, 降低了错误率。

3.2 注意力机制对比实验

为了验证 GAM 注意力机制的有效性, 在训练条件相同的前提下, 与其他注意力机制进行了对比实验, 结果如表 5 所示。

由表 5 可知, 当网络中使用 GAM 注意力机制时, 整体取得了最优的效果, 具体而言, 准确率、召回率和 mAP@0.5:0.95 分别提高了 2.3、1.3 和 0.4 个百分点, 虽然 mAP@0.5 略有下降, 但模型整体性能和效果上都得到了一定的优化。相比之下, 加入 SE (squeeze-and-excitation)^[21]、ECA (efficient channel attention)^[22] 和 CBAM (convolutional block attention

表 5 不同注意力机制比较

Table 5 Comparison of different attention mechanisms

| 模型 | 准确率/ % | 召回率/ % | mAP@0.5/ % | mAP@0.5:0.95/ % |
|---------|-----------|-----------|---------------|--------------------|
| YOLOv8n | 93.1 | 86.1 | 91.8 | 73.1 |
| + SE | 95.5 | 86.7 | 90.9 | 71.7 |
| + ECA | 93.9 | 84.9 | 91.1 | 72.5 |
| + CBAM | 92.9 | 87.2 | 91.4 | 72.6 |
| + GAM | 95.4 | 87.4 | 91.4 | 73.5 |

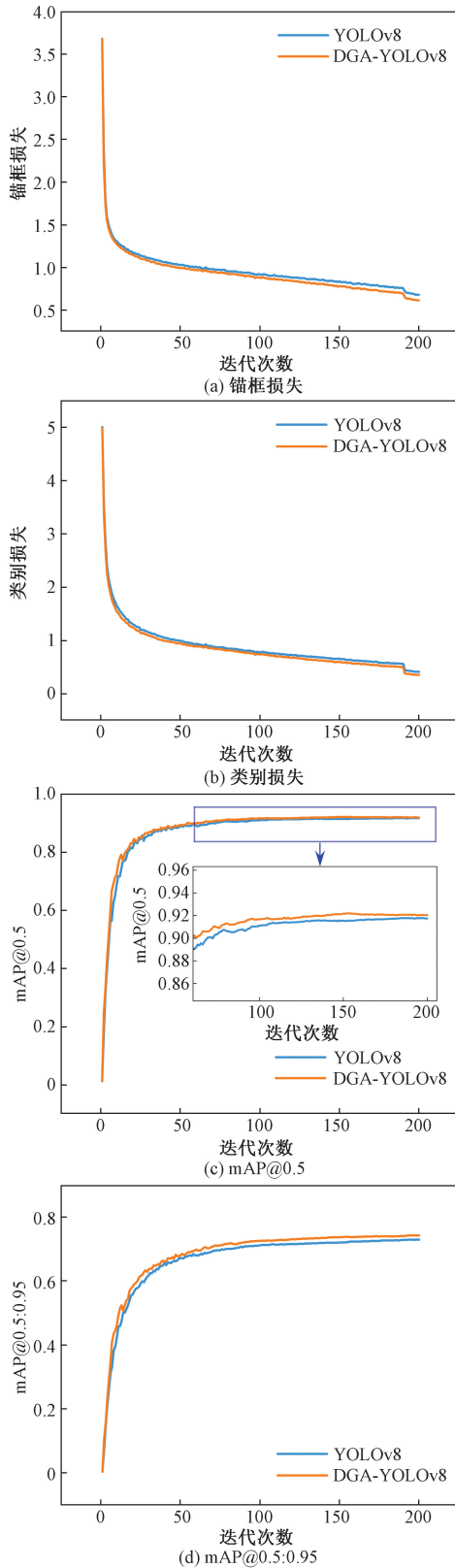


图9 训练过程分析

Fig. 9 Training process analysis

module)^[23] 注意力机制后, mAP@0.5 和 mAP@0.5:0.95 均出现下降趋势,表明模型的性能有所降低。

3.3 算法对比实验

为了进一步验证本文算法的优势,将本文算法与其他目前较为热门的算法在同一数据集上、相同的训练参数下进行对比实验,结果如表6所示。

表6 不同算法对比实验

Table 6 Comparison experiment of different algorithms

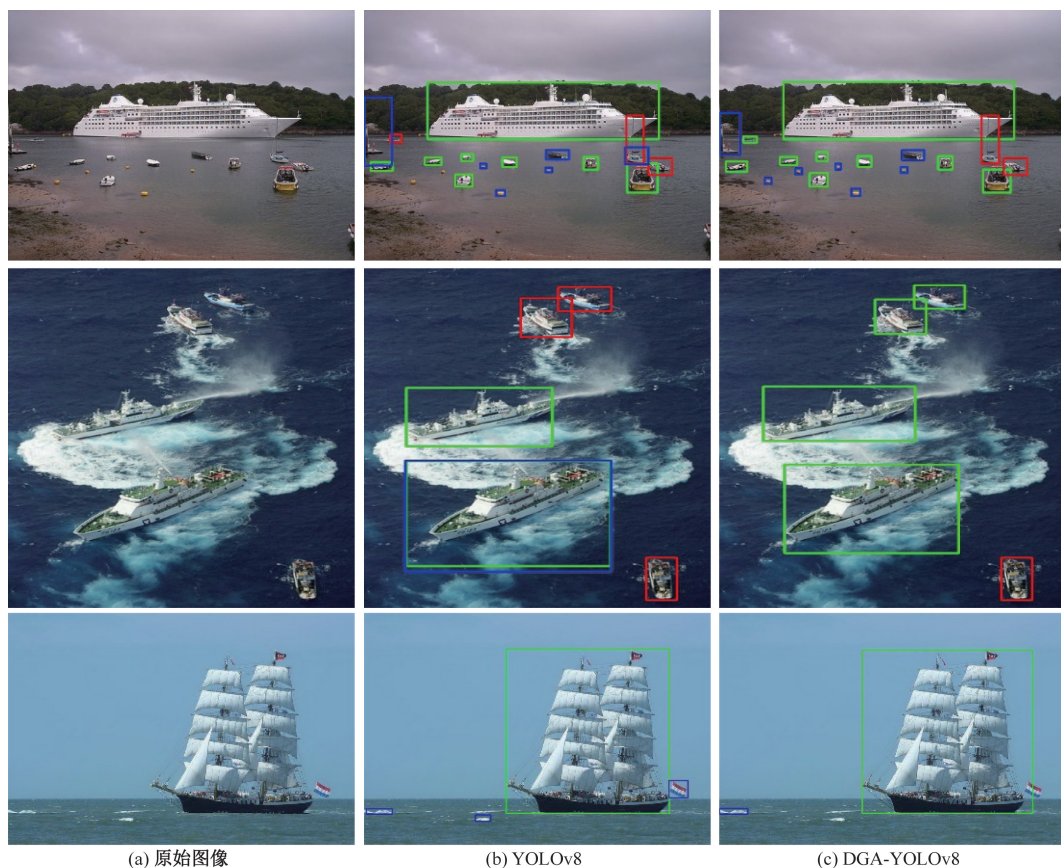
| 模型 | 准确率/ % | 召回率/ % | mAP@0.5/ % | FPS | 体积/ MB |
|-----------------------------|-----------|-----------|---------------|-------|-----------|
| Faster R-CNN | 86.5 | 77.7 | 86.9 | 8.00 | 108.00 |
| YOLOv3 | 93.1 | 86.4 | 90.4 | 6.33 | 235.00 |
| YOLOv5s | 92.4 | 89.7 | 90.2 | 27.77 | 13.80 |
| YOLOv7-tiny | 92.1 | 85.3 | 91.0 | 38.81 | 11.70 |
| YOLOv8n | 93.1 | 86.1 | 91.8 | 43.11 | 5.98 |
| CBAM-YOLOv8 ^[17] | 91.6 | 87.8 | 91.1 | 52.18 | 7.00 |
| YOLOv9s | 95.3 | 87.6 | 91.1 | 24.13 | 19.30 |
| YOLOv10n | 95.5 | 85.9 | 90.6 | 65.79 | 5.53 |
| DGA-YOLOv8 | 96.4 | 87.3 | 92.2 | 43.55 | 10.60 |

由表6可知,Faster R-CNN 和 YOLOv3 算法的FPS 值都偏低,并且模型体积都较大。YOLOv5s 算法在精度和 mAP 上略低于 YOLOv3,但其 FPS 值提高到了 27.77,同时在模型大小上,比 YOLOv3 减少了 221.2 MB,与前两者对比模型更加轻量化。YOLOv7-tiny 在精确度和 mAP 上与 YOLOv5 并没有显著的差异,FPS 上则有了较大的提升。YOLOv8n 在召回率上虽比 YOLOv5s 略低,但其精确度和 mAP 值都有了一定的提升,同时模型的体积也得到缩减,检测速度显著得到了提高。文献[17]提出的 CBAM-YOLOv8 算法虽然参数量和计算量得到了一定的降低,模型得到了轻量化,但是精度却低于原始模型。YOLOv9s 算法在准确率、mAP 方面均不及本文算法,同时检测速度较低;YOLOv10n 虽然在检测速度上达到最高,但准确率和 mAP 均略低于本文算法。所提出的 DGA-YOLOv8 模型精确度和 mAP 值提升到了 96.4% 和 92.2%,同时 FPS 也较高,模型大小在 YOLOv8n 的基础上仅增加 4.62 MB,与其他模型相比,本文模型在准确率和速度方面都有着良好的表现。

3.4 可视化分析

图10为检测过程中的 TP(绿色框)、FP(蓝色框)和 FN(红色框)的可视化结果。通过可视化结果,可以对模型检测性能的各个方面进行评估和分析。

由图10可知,原始的 YOLOv8 模型出现了较多的检错与漏检的现象,相比之下,所提出的 DGA-YOLOv8 算法在检错和漏检方面有了一定的改善。对于图中存在多数小目标船只的情况,本文算法对远处的小目标船只仍能实现正确的检测,提高了对



绿色框为 TP;蓝色框为 FP;红色框为 FN

图 10 检测结果对比

Fig. 10 Comparison of detection result

小目标的识别准确性。其次,在漏检方面,改进后的算法能够检测出区域内的更多船只,检测的覆盖率得到了提高并降低了漏检率,同时对船只类别的分类效果更为准确。此外,从图 10 中可以看出,改进后的模型可以有效地从周围的环境干扰中将船只准确识别出来,对海上周围环境杂物有着更好的分辨能力。由此可见,所采取的改进方式不仅有效提升了模型的检测准确性,降低了误检和漏检的风险,同时还增强了模型对背景的抑制能力。

为了更直观地展示改进后的模型对特征提取能力的增强,使用 Grad-CAM 技术生成热力图进行展示。热力图可以直观地反映出模型所关注于特征图的哪些区域,通过这种可视化的方式,可以更好地了解模型用于预测的图像区域,图 11 为对比改进后模型与原始模型的特征提取过程。

由图 11 可知, YOLOv8 模型对图像中的船只目标,其所关注的只有少数部分船只,且只关注到了整体目标中的局部位置,并没有对其余小目标的船只给予更多的关注。相较于基线模型,在涉及小目标船只的场景中,本文方法能够更加聚焦于所有存在目标的区域,展现出更高的关注度。对于尺寸更

小的船只目标,本文方法的检测效果也更为出色,从而实现了更为精准的目标识别与定位。针对尺寸较大的船只,本文模型能够更精准地聚焦于船体部分,确保对目标区域的有效识别。此外,从图 11 中可以看出, YOLOv8 模型在特征提取过程中受到部分背景区域的干扰。而本文方法则对背景干扰有着更好地抑制能力,能够更有效地将注意力集中在船只目标上。

4 结论

基于 YOLOv8 提出一种改进的海上船只目标检测算法。该算法使用可变形卷积并设计 C2f_DCN 模块替换原始模型中的 C2f 结构,通过引入可变形卷积,扩大了模型的感受野,使模型在特征提取时能够自适应地调整感受野大小和形状,以更好地捕捉输入数据的关键特征。此外,通过嵌入 GAM 注意力机制,网络对重要特征的提取能力得到了强化,同时能够抑制不相关的背景信息的干扰,增强对关键信息的关注能力。实验结果表明,改进后模型的精确度和 mAP 分别提高至 96.4% 和 92.2%,在原始模型的基础上都有了良好的提升;在一定程度

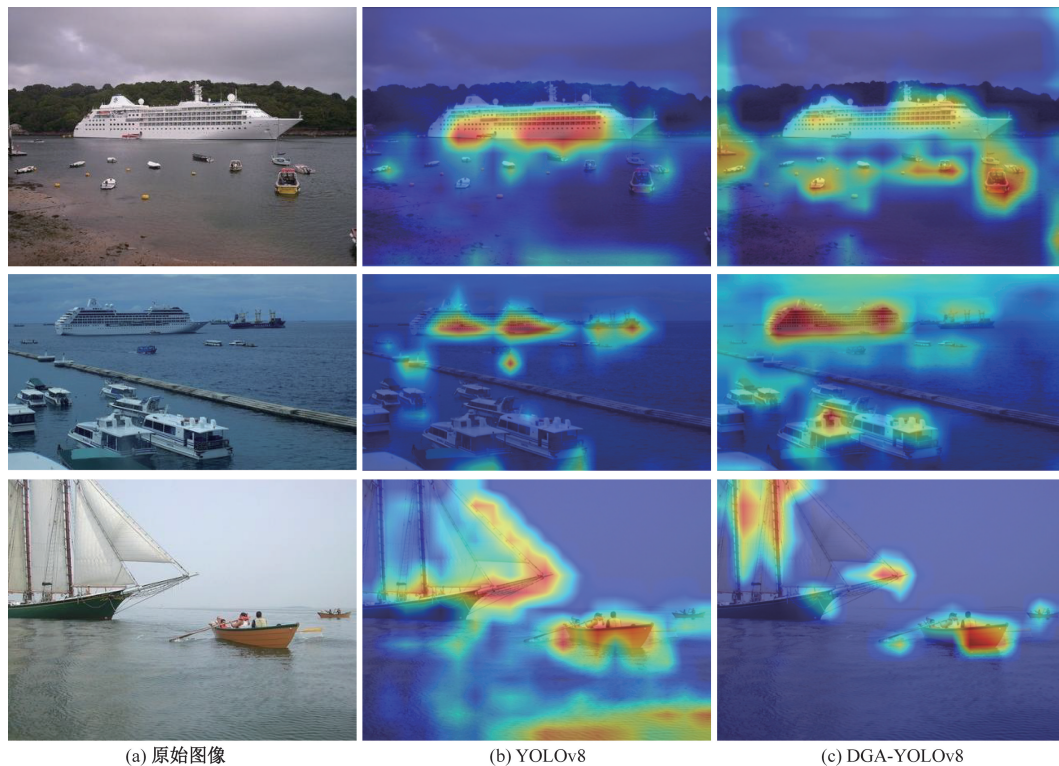


图 11 热力图效果对比

Fig. 11 Comparison of heat map effects

上兼顾了准确度和速度,同时模型大小为 10.6 MB,满足硬件平台的部署和实时检测的任务需求。然而,模型的泛化能力仍需进一步地验证,在面对不同场景的船只时,网络性能可能会受到影响,同时其检测精度还有继续提升的空间。

在未来研究中,针对模型在小目标和遮挡物检测方面中的不足,可进一步优化网络,提升这些方面的检测能力;同时探索更轻量化的模型,在减少参数数量的同时提升检测速度和效率。

参 考 文 献

- [1] 徐慧智,宋爱秋,武笑宇. 基于均匀设计的船舶目标检测深度学习模型训练方法研究[J]. 科学技术与工程, 2022, 22(25): 11241-11249.
Xu Huizhi, Song Aiqiu, Wu Xiaoyu. Research on training method of deep learning to ship target detection based on uniform design [J]. Science Technology and Engineering, 2022, 22(25): 11241-11249.
- [2] 石洪基,郎海涛,宋棋,等. 基于改进区域卷积神经网络的 SAR 图像船只检测方法[J]. 中国海洋大学学报(自然科学版), 2019, 49(S2): 185-191.
Shi Hongji, Lang Haitao, Song Qi, et al. Ship detection method in SAR images based on improved regional convolutional neural network[J]. Journal of Ocean University of China (Natural Science Edition), 2019, 49(S2): 185-191.
- [3] 王程丽,赵静,杨攀攀,等. 基于迁移学习的小样本 SAR 船只目标识别[J]. 移动通信, 2022, 46(4): 22-27.
Wang Chengli, Zhao Jing, Yang Panpan, et al. Smallsample SAR ship target recognition based on transfer learning[J]. Mobile Communications, 2022, 46(4): 22-27.
- [4] 刘方坚,李媛. 基于视觉显著性的 SAR 遥感图像 NanoDet 舰船检测方法[J]. 雷达学报, 2021, 10(6): 885-894.
Liu Fangjian, Li Yuan. NanoDet ship detection method for SAR remote sensing images based on visual saliency[J]. Radar Journal, 2021, 10(6): 885-894.
- [5] 宣以国,余成波,蒋启超,等. 基于改进 YOLOv7 的道路裂缝和坑洞检测算法[J]. 科学技术与工程, 2024, 24(17): 7205-7213.
Xuan Yiguo, Yu Chengbo, Jiang Qichao, et al. Improved YOLOv7 road crack and pothole detection algorithm[J]. Science Technology and Engineering, 2024, 24(17): 7205-7213.
- [6] 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. New York: IEEE, 2014: 580-587.
- [7] Girshick R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2015: 1440-1448.
- [8] 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, 2016, 39(6): 1137-1149.
- [9] 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. New York: IEEE, 2016: 779-788.
- [10] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//

- Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 7263-7271.
- [11] Redmon J, Farhadi A. YOLOv3: an incremental improvement [J]. arXiv Preprint, 2018: arXiv: 1804.02767.
- [12] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv Preprint, 2020: arXiv: 2004.10934.
- [13] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2023: 7464-7475.
- [14] 李静, 鲜林, 王海江. 基于YOLOv3的船只检测算法研究[J]. 成都信息工程大学学报, 2023, 38(1): 37-43.
Li Jing, Xian Lin, Wang Haijiang. Research on ship detection algorithm based on YOLOv3[J]. Journal of Chengdu University of Information Technology, 2023, 38(1): 37-43.
- [15] Liu T, Zhou B, Zhao Y, et al. Ship detection algorithm based on improved YOLOv5[C]//2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE). New York: IEEE, 2021: 483-487.
- [16] Wang B, Han B, Yang L. Accurate real-time ship target detection using YOLOv4[C]//2021 6th International Conference on Transportation Information and Safety (ICTIS). New York: IEEE, 2021: 222-227.
- [17] Zhang J. Ship target detection based on CBAM-YOLOv8[C]//International Conference on Mechatronic Engineering and Artificial Intelligence. Shanghai: SPIE, 2024: 659-665.
- [18] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks [C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 764-773.
- [19] Liu Y, Shao Z, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions[J]. arXiv Preprint, 2021: arXiv: 2112.05561.
- [20] Bolya D, Foley S, Hays J, et al. Tide: a general toolbox for identifying object detection errors[C]//Computer Vision-ECCV 2020: 16th European Conference. Glasgow: Springer International Publishing, 2020: 558-573.
- [21] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 7132-7141.
- [22] Wang Q, Wu B, Zhu P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020: 11534-11542.
- [23] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision(ECCV). New York: IEEE, 2018: 3-19.