



DOI:10.12404/j.issn.1671-1815.2308931

引用格式:李志昂,肖小玲,周绍发.基于改进的YOLOv5s小目标船舶遥感图像检测[J].科学技术与工程,2025,25(2):657-666.

Li Zhiang, Xiao Xiaoling, Zhou Shaofa. Small target ship remote sensing image detection based on improved YOLOv5s[J]. Science Technology and Engineering, 2025, 25(2): 657-666.

基于改进的YOLOv5s小目标船舶遥感图像检测

李志昂,肖小玲*,周绍发

(长江大学计算机科学学院,荆州434023)

摘要 遥感图像中船舶目标具有多尺度特性、背景多变及气象复杂等特点,导致小目标船舶检测存在精度低,出现误检,漏检等情况。针对上述情况,提出了一种基于YOLOv5s的小目标船舶检测改进模型。首先,为解决船舶检测中尺度变化和背景多变问题,引入了适应空间特征融合(adaptive structure feature fusion, ASFF)模块,其次,为减少检测网络的计算量和参数量引入了BoTNet注意力机制,然后为提升网络整体的检测精确度,使用了EIou边框损失函数,最后为保证网络整体的轻量化引入了Slim-neck颈部网络。实验显示,在主要数据集LEVIR-Ship上,相较于基准YOLOv5s, mAP@0.5提升了7.1%达到了81.3%,参数量降低了0.44M,计算量降低了0.6GFLOPs,权重降低了0.9M。本文方法在各项关键指标中表现更为优秀,实现了复杂环境下高精度的小目标船舶检测。在验证数据集McShips上进行对比实验。实验表明,本文方法依然表现更为优秀,验证了所提方法具有普适性。

关键词 船舶检测;YOLOv5s;小目标检测;BoTNet注意力机制

中图分类号 TP391.4 TP751.1; **文献标志码** A

Small Target Ship Remote Sensing Image Detection Based on Improved YOLOv5s

LI Zhi-ang, XIAO Xiao-ling*, ZHOU Shao-fa

(School of Computer Science, Yangtze University, Jingzhou 434023, China)

[Abstract] Ship targets in remote sensing images have multi-scale characteristics, changeable backgrounds, and complex meteorological characteristics, which lead to low accuracy, false detection, and missed detection of small target ships. In response to the above situation, an improved small-target ship detection model based on YOLOv5s was proposed. First, in order to solve the problems of scale changes and background variability in ship detection, the ASFF (adaptive spatial feature fusion) module was introduced. Secondly, in order to reduce the calculation amount and parameter amount of the detection network, the BoTNet attention mechanism was introduced, and then in order to improve the overall network to improve the detection accuracy, the EIou border loss function was used, and finally the Slim-neck network was introduced to ensure the overall lightweight of the network. Experiments show that on the main data set LEVIR-Ship, compared with the benchmark YOLOv5s, mAP@0.5 increased by 7.1% to 81.3%, the number of parameters is reduced by 0.44 M, the calculation amount is reduced by 0.6GFLOPs, and the weight was reduced by 0.9 M. The proposed method performs better in various key indicators and achieves high-precision small target ship detection in complex environments. Comparative experiments are conducted on the verification data set McShips. The experiments show that the proposed method still performs better, verifying the universal applicability of the proposed method.

[Keywords] ship detection; YOLOv5s; small target detection; BoTNet attention mechanism

中国是一个拥有广阔海域边境线的海洋大国,在海洋主权安全、海洋资源探索、海洋环境保护以及海域内情报探测等领域中,快速且准确地检测船舶是非常重要的。在数据来源上,一般有卫星合成孔径雷达(synthetic aperture radar, SAR)图像,高分可见光遥感图像。由于SAR可全天时、全天候,多

维度成像,已经成为近年来的研究热点。但随着大量光学传感器的广泛使用,高空间分辨率的可见光图的数据量快速增加。由于其在数据获取方式上成本可控,方便快捷等优点。船舶检测使用遥感光学遥感图像已经成为一种新趋势^[1]。

检测方法上,大致可以分为传统方法检测,机

收稿日期:2023-11-14; 修订日期:2024-10-21

基金项目:国家自然科学基金(61771354)

第一作者:李志昂(2000—),男,汉族,湖北武汉人,硕士研究生。研究方向:深度学习。E-mail:2022710688@yangtzeu.edu.cn。

*通信作者:肖小玲(1973—),女,汉族,湖南邵阳人,博士,教授。研究方向:智能信息处理和网络安全。E-mail:xxl@yangtzeu.edu.cn。

投稿网址:www.stae.com.cn

器学习与深度学习 3 种方法,其中传统方法与机器学习方法需要手动的图像特征提取等复杂的数据处理过程,并且在环境发生变化时,不能快速适应,常常导致检测结果不理想^[2]。深度学习方法凭借着其提取图像特征快速、检测精度高、泛化能力强等优点,已经在遥感图像检测领域广泛使用。以区域建议网络(region proposal network, RPN)为代表的二阶段检测方法(two stage)将检测任务拆分为分类和回归 2 个部分。在应用上, Jiang^[3]使用 Faster R-CNN 网络将大场景 SAR 图像中的船舶切片。然后,应用 FNLM 滤波来降噪并增强目标切片的结构信息。提出了一种优化的 Chan-Vese 模型,在 SSDD 数据集上进行了一系列实验,结果表明,该方法不仅可以准确地提取观测船的轮廓,而且可以减少模型的计算时间。赵等^[4]使用 MASATI (mAritime sATellite imagery) 海洋及湖泊船舶光学遥感数据集,实现了快速实时精准检测,其 mAP 达到了 80%。在克服环境问题上, Yuan 等^[5]在数据预处理部分,将原始图像进行去雾处理后再放入 Faster R-CNN 网络,但是精确度的提升是以庞大的计算量作为代价的,并且还出现了图片颜色失真的情况。以上的二阶段检测方法多数在精确度上确实有提升,但是检测速度和模型大小都不太理想,对于后续部署实时检测并不友好。而一阶段检测算法在保证精确度的情况下速度更快,其代表算法有 YOLOv3^[6]、YOLOv4^[7]、YOLOv5^[8] 和 RetinaNet^[9] 等算法。叶等^[10]使用改进的 YOLOv3 网络实现了雾天海洋图像船舶检测。Zhu 等^[11]通过改进 RetinaNet 的特征模块和损失函数解决了船舶检测中存在的任意方位、宽高比大、排列密集等问题。Liu 等^[12]将 YOLOv5 网络特征提取过程与 GhostbottleNet 算法相融合,大大提高了网络的精确性和实时性。化嫣然等^[13]使用优化后的 YOLOv3-Tiny 实现了较为精确的遥感图像中大目标检测。以上多数检测方法要么是使用参数量多的大模型保证高精度,要么牺牲精确度来实现模型的轻量。在数据集使用上,多数都是基于高分辨率下的遥感图像(通常为 10 m 及以下),很少有人使用中等分辨率的遥感图像检测(通常为 10 m 以上)。不同的分辨率在不同场景有着各自的优点。高分辨遥感图像凭借其图像质量优秀,且图像中的目标多为大目标,对于网络的检测精度具有质的提升,但是高分辨遥感图像覆盖的区域并不广阔,效率相对较低。中等分辨率遥感图像覆盖的区域比相同大小的高分辨图像要宽得多,在中等分辨率遥感图像中船舶都是小目标,在有限资源下,检测区域更广阔,更有利于快速检测船舶。

针对当前基于遥感图像船舶目标检测多数都是检测大型船舶,很少在复杂天气环境下检测小目标船舶。现提出一种基于 YOLOv5s 的改进网络。首先为保证后续研究的实时检测,采用 YOLOv5s 轻量化版本,为了让整个网络能够更好地学习到小目标船舶特征,在网络主干(backbone)部分加入 Bottleneck Transformers 注意力机制。为了让网络更轻量,在网络颈部(neck)采用轻量级结构: Slim-neck。针对小目标船舶尺度不一致的目标的特征提取问题,引入自适应空间特征融合(adaptive structure feature fusion, AS-FF),同时在提升精确度上使用 EIoU_loss 边界框损失函数。通过以上改进实现在广阔海域上的小目标船舶检测。

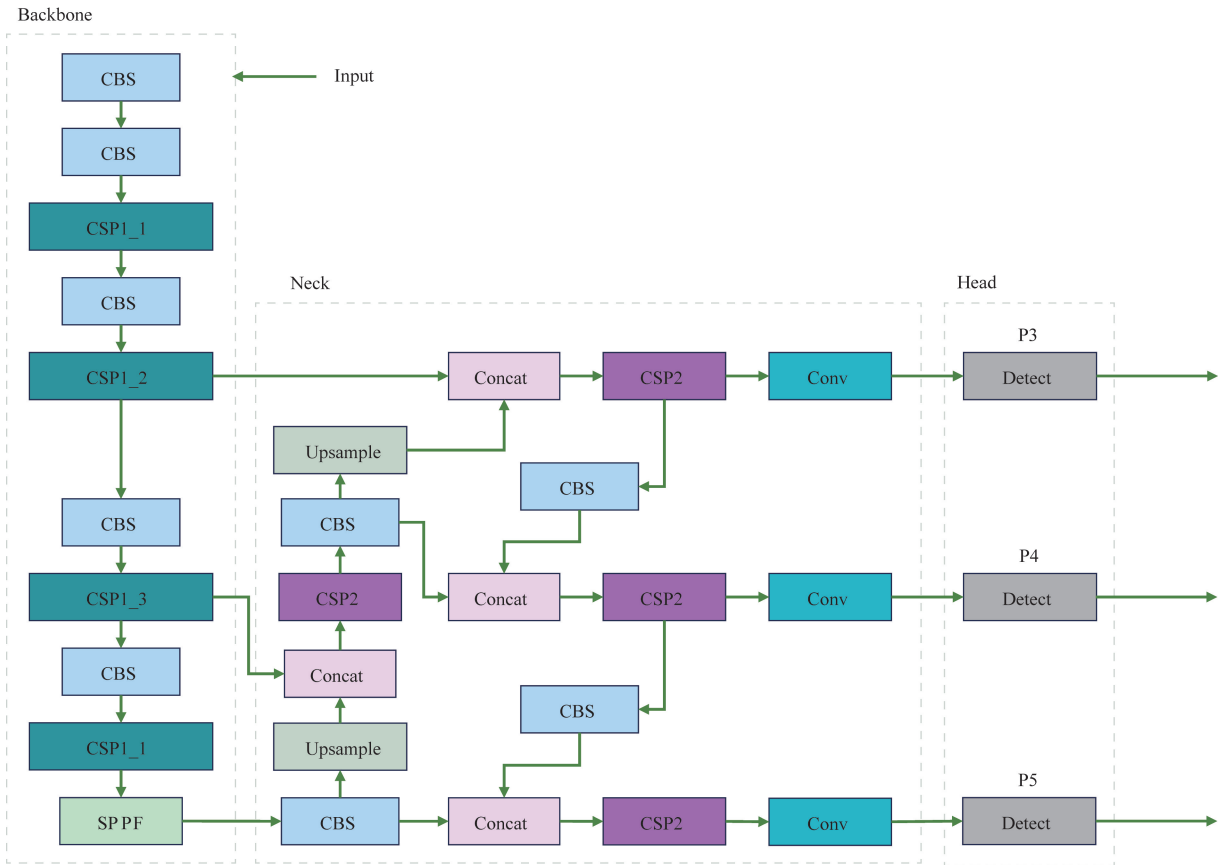
1 YOLOv5 网络

YOLO 系列网络的基本思想是将输入的特征图经过增强后,将特征图划分为 $S \times S$ 的格子(grid cells),每个格子负责对落入其中的目标进行检测,一次性预测所有各自所含目标的边界框、定位置信度、以及所有类别概率向量。相较于二阶段的目标检测方法,在保证精确度的情况下更加简捷快速。YOLOv5 目前是 YOLO 系列第 5 个更新版本,且使用较为广泛的一个网络,根据不同使用场景及需求有 YOLOv5m、YOLOv5l、YOLOv5s 等版本。选择了网络宽度小、深度低、体积小、实时性较好的 YOLOv5s 版本。YOLOv5s 网络结构大致可分为 4 个部分:输入端(input)、网络主干(backbone)、颈部网络(neck)、检测头(head)。其中输入端负责 Mosaic 数据增强,自适应锚框计算,自适应图片缩放等数据预处理功能。Backbone 是由 CBS、CSP1_X 和快速金字塔池化(spatial pyramid pooling-fast, SPPF)组成,主要负责图像卷积后的特征提取、转换等工作。Neck 由 CSP2,上采样(Upsample)和拼接模块(Concat)构成,主要负责将上采样的图像特征进行拼接后传入 head。Head 主要负责图像预测。具体的整体结构及其子结构如图 1 和图 2 所示。

2 YOLOv5s 网络的改进

2.1 BoTNet 注意力机制

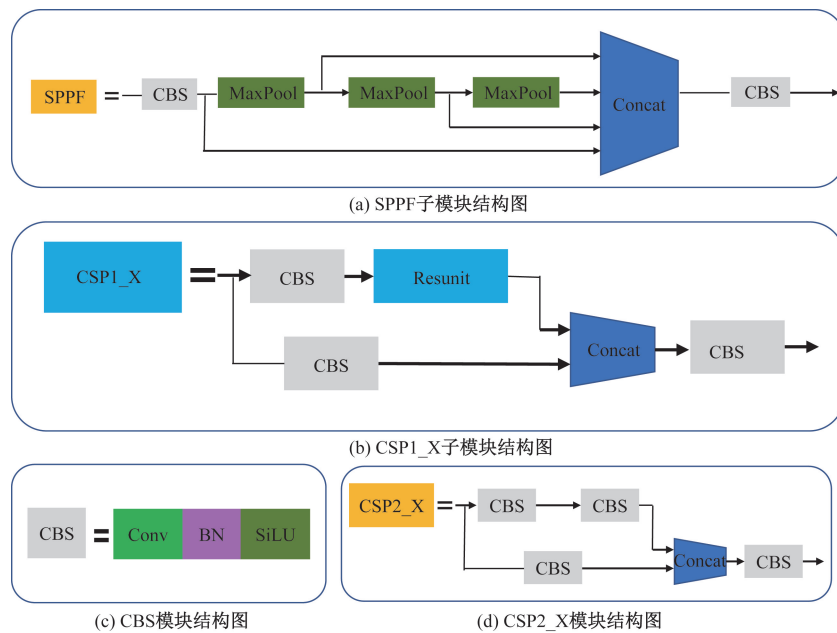
注意力机制在目标检测领域已经应用非常广泛^[14],其主要思想是借鉴了人类视觉在面对大量信息的处理决策过程。人类视网膜不同的部位具有不同程度的信息处理能力,且人类大脑对于事物关注程度取决于它的重要性,所以为了合理利用有限的视觉信息处理资源,需要将有限的处理资源用在重要信息部分。



Input 为输入图像;Backbone 为网络主干;Neck 为颈部网络;Head 为检测头;CBS 为由卷积层 + 批量归一化 + SiLU 激活函数组成的模块; CSP1_x 为 Cross Stage Partial(CSP)连接的第 1 个子模块中的第 x 个部分;SPPF 为一种空间金字塔池化结构;Upsample 为上采样;CSP2 为 cross stage partial(CSP)连接的第 2 个子模块;Concat 为拼接模块;Conv 为卷积层;Detect 为不同尺度的检测头

图 1 YOLOv5s 的网络主体结构图

Fig. 1 YOLOv5s network main structure diagram



MaxPool 为最大池化层;Concat 为相加模块;Resunit 为残差单元;Conv 为卷积层;BN 为批量归一化;SiLU 为一种激活函数

图 2 YOLOv5s 子模块的结构图

Fig. 2 Structure diagram of YOLOv5s submodule

为了使改进的 YOLOv5s 网络在参数量和计算量上没有大幅增加的情况下,有效地提升检测的精度。提出了在主干网络中添加 BoTNet (bottleneck transformers) 注意力机制,它是一种概念上简单但功能强大的主干架构^[15],主要思想是将自注意力机制融入目标检测等多个计算机视觉任务。仅通过在 ResNet 的最后 3 个瓶颈块中用全局自注意力替换空间卷积,而不进行其他更改。从而达到增加网络整体精确度,减少或者不带来额外的计算量和参数量。具体结构如图 3 所示。

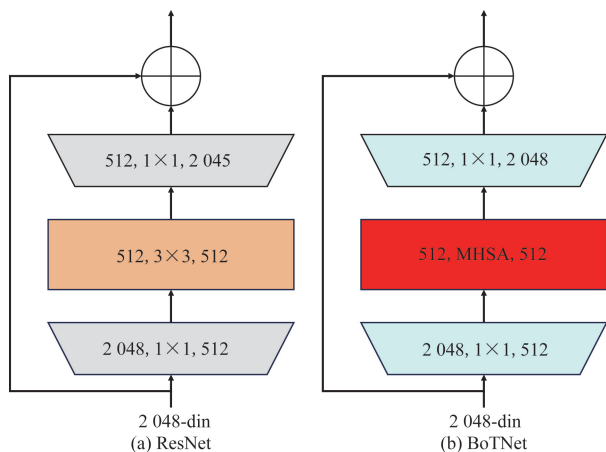


图 3 BoTNet 与 ResNet 结构差异图

Fig. 3 Diagram of the structural differences between BoTNet and ResNet

其中多头自注意力 (multi-head self-attention, MHSA) 的计算公式公式为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (1)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (2)$$

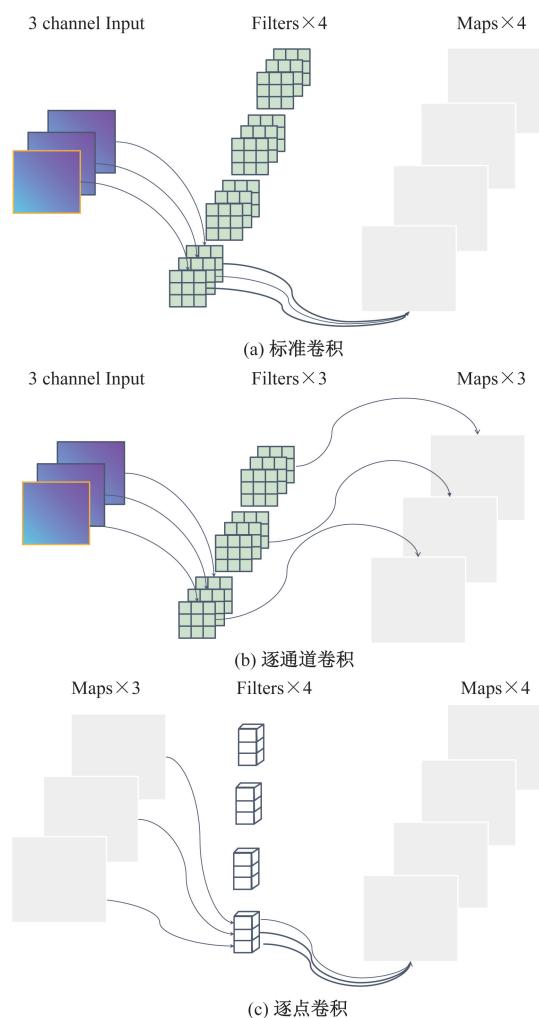
$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)\mathbf{W}^O \quad (3)$$

式中: $\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ 为得到的注意力的值; \mathbf{Q} 、 \mathbf{K} 、 \mathbf{V} 分别为查询量 (query)、键 (key) 和值 (value); d_k 为 key 的维度; \mathbf{W}_i 为参数矩阵; head_i 为第 i 个检测头的注意力得分; \mathbf{W}^O 为参数矩阵; Concat 为拼接操作; h 为检测头个数; $\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ 为多头自注意力机制的值。

2.2 Slim-neck

船舶检测的最终目的是实时检测,那么网络的体量大小和结构就应该更轻量。在 YOLOv5s 的颈部网络使用了 Slim-neck 结构^[16]。它主要由广义稀疏卷积技术 (generalized-sparse convolution, GSConv) 实现。其中 GSConv 是在标准卷积 (standard convolution, SC) 和深度可分离卷积 (depthwise separable convolution, DSC) 的基础上进行的改进。DSC 相对

于 SC 的优势是大幅度减少卷积的参数和计算量。但是缺点也很明显,输入图像的通道信息在计算过程中被分离。这一缺点将会导致 DSC 的特征提取和融合能力比 SC 低得多,最终网络会因学习了不完善的图像特征而导致精确度下降。而 GSConv 以较低的时间复杂度尽可能保留通道信息连接,保证了模型的准确性和运行速度之间的平衡。GS bottleneck、VoV-GSCSP 是在 GSConv 基础上衍生的一系列轻量的颈部网络,具体结构变化如图 4 和图 5 所示。



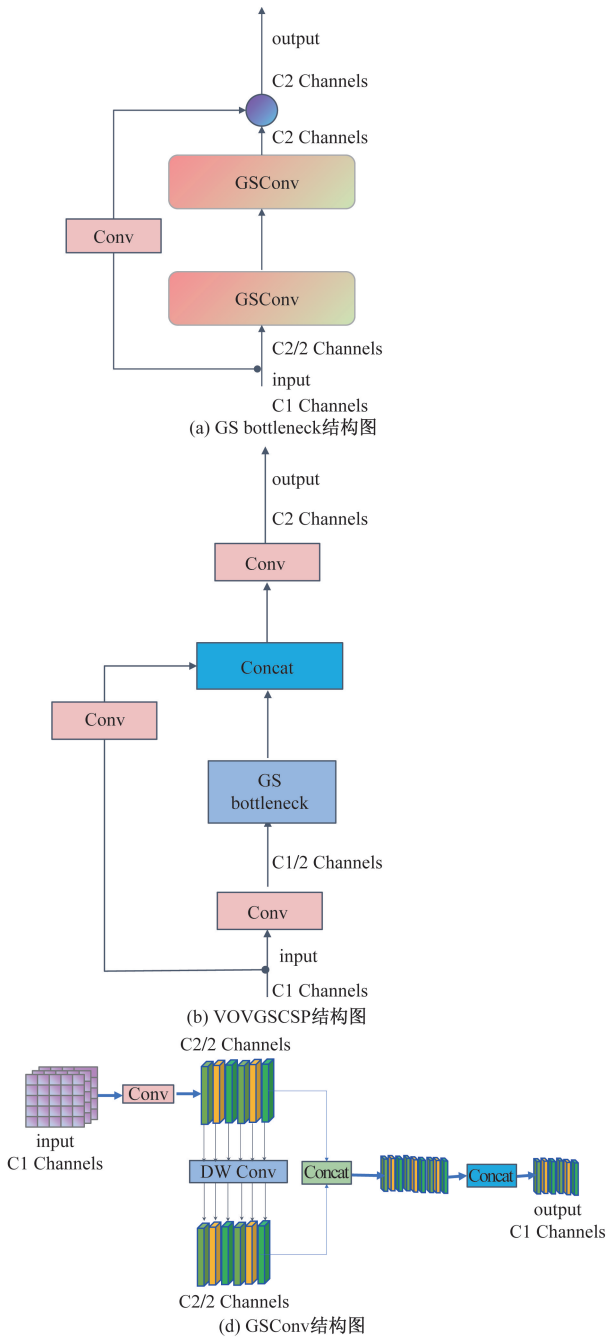
3 channel Input 为彩色图像的 3 通道输入; Filters 为滤波器; Maps 为输出的特征图
图 4 标准卷积和深度可分离卷积

Fig. 4 Standard convolution and depthwise separable convolution

其中 DSC 由图 4(b) 逐通道卷积和图 4(c) 逐点卷积 2 个阶段组成。

2.3 EIou 边界框回归函数

边界框回归是决定目标定位性能的关键步骤,也是在目标检测领域中衡量一个网络性能重要指



Conv 为标准卷积; input 为输入; c1 channels 为有 c1 个通道; c2 channels 为有 c2 个通道; output 为输出

图 5 Slim-neck 的组成结构图

Fig. 5 Slim-neck composition structure diagram

标之一。其主要的思想就是计算预测框与真实框的交并比值(intersection over union, IoU), IoU 越大说明检测效果越好。但是 IoU 仅考虑预测框与真实框的重叠面积, 没有考虑二者的位置与尺度信息, 导致网络效果不好。并且无法处理特殊情况——当 2 个框不相交时, $IoU = 0$, $loss = 0$, 导致梯度无法回传, 网络无法训练。为此引入了 EIoU^[17] 替换 YOLOv5s 使用的 CIoU^[18], 它考虑了边界框回归的重叠面积、中心点距离、纵横比, 一定程度上解决了

此前的 GIoU、DIoU 的缺点。CIoU 损失 (L_{CIoU}) 与 EIoU 损失 (L_{EIoU}) 的计算公式为

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (4)$$

$$L_{EIoU} = L_{IoU} + L_{dis} + L_{asp} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \quad (5)$$

式中: IoU 为预测框与真实框的交并比值; ρ 为两个中心点之间的欧式距离; α 为权重函数; v 为用来度量纵横比的相似性; 边框损失函数包含 3 个部分: 重叠损失 (L_{IoU})、中心距离损失 (L_{dis}) 和宽高损失 (L_{asp}); b^{gt} 、 w^{gt} 、 h^{gt} 和 b 、 w 、 h 分别为实框和预测框的中心点、边宽、边长; c^2 、 c_w^2 、 c_h^2 分别为真实框和预测框最小外接矩形的对角线、边宽、边长。

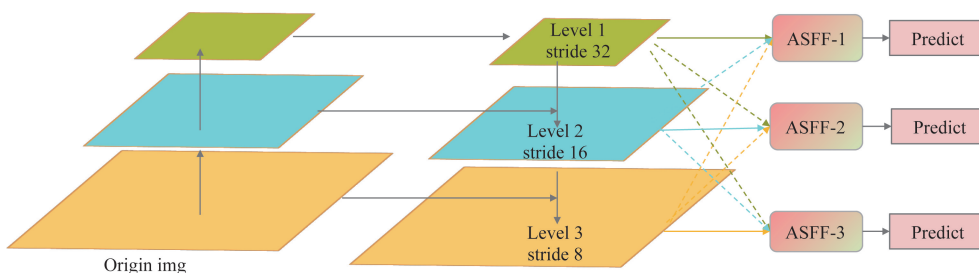
根据式 (4) 中 v 反映的纵横比的差异, 而不是宽高分别与其置信度的真实差异, 所以有时会阻碍模型有效的优化相似性。而 EIoU 在 CIoU 的惩罚项基础上的第三部分将纵横比的影响因子拆开分别计算目标框和锚框的长和宽, 有效地提高了收敛速度和定位精度。

2.4 ASFF

在目标检测领域中, 针对物体检测中尺度变化问题, 现行常用的解决方法是使用金字塔特征方法。针对遥感图像中海上的小目标船舶, 其大小和尺度变化多样, 不同特征尺度之间的一致性会导致金字塔特征方法提取的特征是有限制的。具体来说, 当检测具有特征金字塔的对象时, 采用启发式引导特征的选择: 大实例通常与上特征图相关联, 小实例与下特征图相联系。当某个对象在某一级别的特征图中被视为正相关时, 其他级别的特征图相应区的域将被视为背景。因此, 如果图像包含大小对象, 则不同级别的特征之间的冲突往往占据特征金字塔的主要部分。这种不一致性干扰了训练期间的梯度计算, 并降低了特征金字塔的有效性。自适应空间特征融合 (adaptive structure feature fusion, ASFF) 很大程度上解决了此类问题^[19]。

ASFF 的具体解决方法-恒等缩放与自适应融合: 对于某一级别的特征, 首先将其他级别的特征集成并调整为相同的分辨率, 然后进行训练以找到最佳融合。在每个空间位置, 不同级别的特征被自适应地融合。并且引入的计算开销很小、实现简单。具体的网络结构如图 6 所示。

最后各个改进在网络中位置以及总体结构如图 7 所示。



Origin img 为原始图片;Level 为特征图的级别;stride 为步长;Predict 为预测结果

图6 ASFF 结构图

Fig. 6 ASFF structure diagram

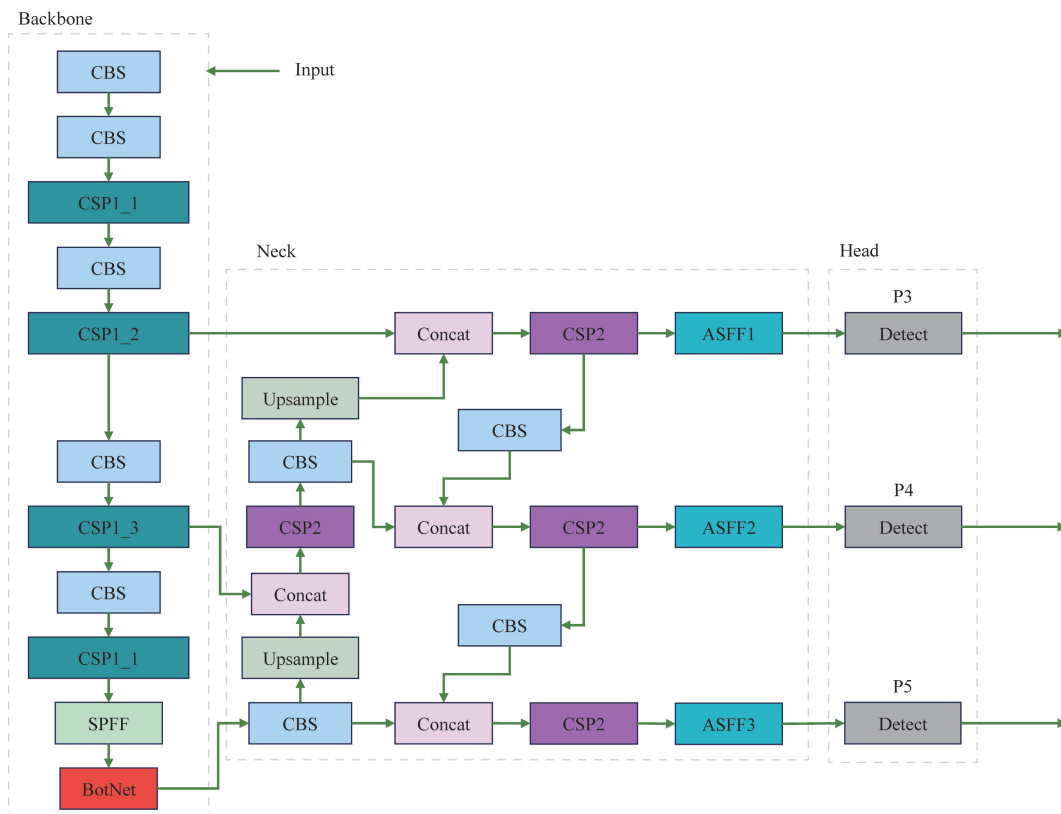


图7 改进后的 YOLOv5s 结构图

Fig. 7 Improved YOLOv5s structure diagram

3 结果与分析

3.1 数据集

本文使用的是开源数据集 LEVIR-Ship,它是由 Chen 等^[20]所提出。LEVIR-Ship 中的图像由高分一号和高分六号卫星的多光谱相机拍摄,空间分辨率为 16 m,包含 1 973 个正样本和 1 923 个负样本,覆盖了 85 个场景,如港口、海面、云层等一些复杂场景。超过 3 000 个小型船舶实例。所以 LEVIR-Ship 数据集的场景复杂性,实例数据量是非常符合复杂场景下的光感图像船舶检测。将 3 896 张图片按照 8:1:1 的比例划分为训练集、验证集、测试集。

为验证本文方法的普适性和有效性,选取了西北工业大学开源 McShips 数据集作为验证数据

集^[21]。该数据集包括 14 709 张带标注的图像,包含 6 类军舰和 7 类民用船舶,拍摄地点包含不同海域、不同天气条件和不同拍摄视角,即包含不同尺度的船舶。其中也包含一部分小目标船舶。实验中,该数据集被整体划分为民用船和军舰 2 个类别,各自包含 5 945 和 5 382 实体船舶。

3.2 模型训练平台与参数设置

实验平台是主要配置环境为 Windows 10 专业版操作系统,CPU 为 6 核 12 线程 AMD R5 5600,GPU 为 NVIDIA RTX 3070,CUDA 版本为 11.6,基于 PyTorch1.12.1 框架下 Python 版本为 3.8 进行网络训练和测试。网络默认的关键参数如表 1 所示。

3.3 评价标准

为验证本文算法的检测性能和速度,探究各个改

表 1 关键参数表

Table 1 Key parameter table

参数名称	初始值	参数名称	初始值
图像大小	640 × 640	IoU 阈值	0.2
初始学习率	0.01	色调	0.015
优化函数	Adam(1 × 10 ⁻²)	饱和度	0.7
学习率动量	0.937	亮度	0.4
权重衰减系数	0.0005	Mosaic 概率	1.0

进方法的有效性, 通过对比基准 YOLOv5s 和改进方法的检测图像差异来评估模型性能。采用了目标检测领域常见的精准率 (precision, P)、召回率 (recall, R)、平均精度均值@0.5 (mean average precision, mAP) 和 10 亿次浮点运算量 (GFLOPs) 作为评价指标, 具体计算见式为

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (6)$$

$$R = \frac{T_p}{T_p + F_n} \times 100\% \quad (7)$$

$$AP = \int_0^1 P(R) dR \quad (8)$$

$$mAP = \frac{1}{n} \sum_{i=1}^m AP_i \quad (9)$$

式中: T_p 为预测正确的正样本数量; F_p 为预测错误的正样本数量; F_n 为预测错误的负样本数量; n 为数据集中的类别数。

3.4 对比实验

为证明所引入的 BoTNet 注意力机制的在所提的整个网络中优越性, 将 BoTNet 与最近热门的注意力机制进行对比, 如 SE (squeeze-and-excitation networks)^[22]、CBAM (convolutional block attention module)^[23]、SIMAM (simple, parameter-free attention)^[24]、CA (coordinate attention)^[25]。实验条件都是使用基准 YOLOv5s 网络加上不同的注意力机制, 对比的结果如图 8 和表 2 所示。

为体现提出网络的优越性, 将改进后 YOLOv5s 与基准 YOLOv5s 进行 mAP@0.5 数据对比, 其结果

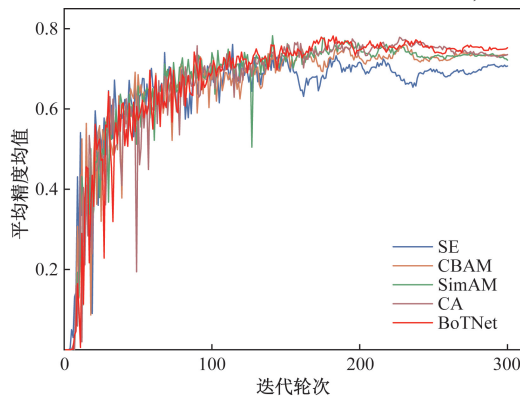


图 8 注意力机制平均精度均值对比图

Fig. 8 Attention mechanism average accuracy mean comparison chart

如图 9 所示。同时将与 CenterNet、YOLOv3、YOLOv4 等热门网络进行对比, 其中对比网络都采用默认参数设置。各项评价指标的实验结果如表 3 所示。

从图 9 中可以看出改进后 YOLOv5s 在 300 个 epoch 迭代次数, 从训练阶段的开始到结束 mAP@0.5 值全程是优于基准 YOLOv5s 的, 证明了在船舶检测中, 提出的方法更有效且对小目标检测表现更优秀。

表 2 注意力机制对比表

Table 2 Attention mechanism comparison table

名称	mAP/%	参数量/M	计算量/G	权重/M
SE	73.9	7.05	15.8	14.5
CBAM	77.5	7.01	15.8	14.5
SIMAM	76.0	7.01	15.8	14.5
CA	76.1	7.04	15.8	14.5
BoTNet	77.2	6.69	15.4	13.8

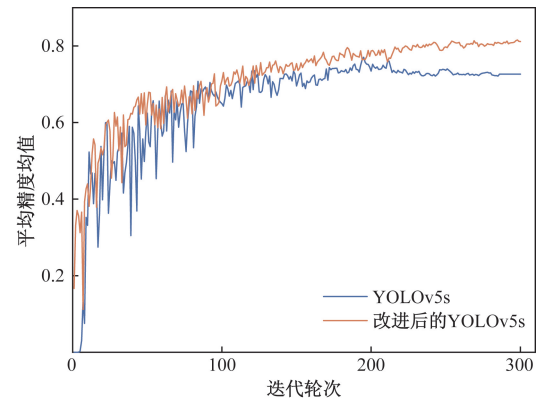


图 9 平均精度均值对比图

Fig. 9 Average accuracy mean comparison chart

表 3 不同网络的评价指标对比

Table 3 Comparison of evaluation indicators of different networks

网络	图片大小	mAP/%	精准率/%	召回率/%	参数量/M	计算量/G	权重/M
CenterNet	512 × 512	69.7	100	10.0	32.67	70.2	124
YOLOv3	416 × 416	75.9	81.4	70.9	61.50	154.5	123.4
YOLOv4	416 × 416	68.4	75.9	65.0	9.12	20.8	35.4
YOLOv5s	640 × 640	74.2	81.2	71.9	7.01	15.8	14.5
YOLOv8s	640 × 640	79.3	84.1	73.4	11.1	28.4	22.5
本文方法	640 × 640	81.3	81.2	76.8	6.57	15.2	13.6

从表 3 可以看出, CenterNet 出现了精准率为 100% 但召回率仅为 10% 的情况, 表明在预测结果时出现了没有误检但是有漏检的情况, 除 CenterNet 的精准率外。在与较新的 YOLOv8s 对比中, 改进后 YOLOv5s 在所有指标中依然是最优秀的, 实现了在参数量, 计算量和权重都下降的情况下实现了 mAP 的提升。

3.5 消融实验

为验证所提出网络改进的有效性, 对每个改进的模块进行消融实验, 实验的结果如表 4 所示。

表4 消融实验结果表
Table 4 Ablation experiment results table

改进名称	ASFF	BotNet	EIoU	Slim-neck	P/%	R/%	mAP@0.5/%	GFLOPs	权重/M
YOLOv5s	×	×	×	×	81.2	71.9	74.2	15.8	14.5
改进1	√	×	×	×	78.2	74.8	76.5	24.2	25.4
改进2	√	√	×	×	85.1	69.6	77.2	15.4	13.8
改进3	√	√	√	×	83.6	73.4	80.8	15.8	14.5
改进4	√	√	√	√	81.2	76.8	81.3	15.2	13.6

从表4可以看出,每个部分的改进都有提升效果,自适应空间特征融合(ASFF)的引入对于网络的学习尺度不一致的目标时更准确,BotNet注意力机制使网络的计算量得到了有效降低,EIoU通过提升网络的定位性能提高最终的精确度,Slim-neck整体降低或减少了网络最终的计算量和权重大小。

3.6 实例检测

为验证所提出的改进后YOLOv5s网络实际效果,选取了测试集的部分实例将提出的改进的YOLOv5s与对比的网络进行比较,同时为验证复杂环境下的检测效果,选取了黑夜下的海洋环境和云层遮挡的海洋环境,检测结果如图10和图11所示。

从图10可以看出,YOLOv5s出现误检的情况,YOLOv3凭借其网络的参数量巨大,得到了置信度最高的检测成绩,但是与改进的YOLOv5s相比没有太大优势。从图11可看出CenterNet出现了无法检测的情况,YOLOv3和YOLOv4都出现漏检的情况,只有改进的YOLOv5s没有出现误检,漏检的情况,总的来说,本文方法在实例检测中达到了更好的效果,实现了较为准确的小目标船舶检测。

3.7 基于McShips数据集实验对比

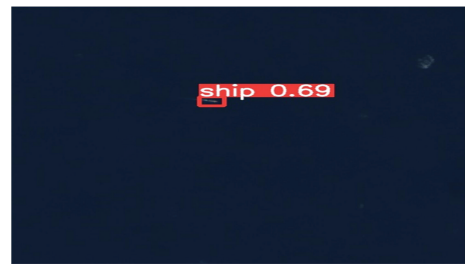
在实验条件一致的情况下,将本文方法与基准YOLOv5s、YOLOv7-tiny、YOLOv8s进行对比实验,实验结果如表5所示。

从表5可看出,本文方法在精准率、召回率、mAP评价指标上,全面由于对比网络。相较于基准YOLOv5s,mAP提升了3.9个百分点,在与YOLOv8s对比中,本文方法的mAP依旧提升了0.5个百分点。充分证明了本文方法的普适性。

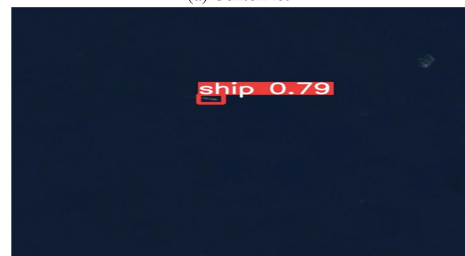
表5 在McShips上对比实验结果表

Table 5 Compare experimental result tables on McShips

网络	mAP/%	精准率/%	召回率/%
YOLOv5s	88.4	88.6	83.3
YOLOv7-tiny	76.0	81.3	70.6
YOLOv8s	91.8	92.4	84.7
本文方法	92.3	92.4	87.4



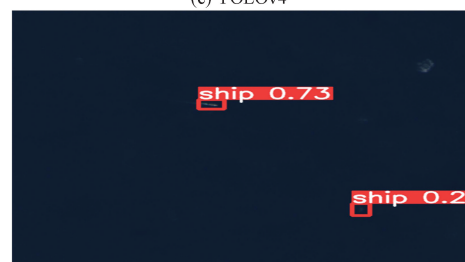
(a) CenterNet



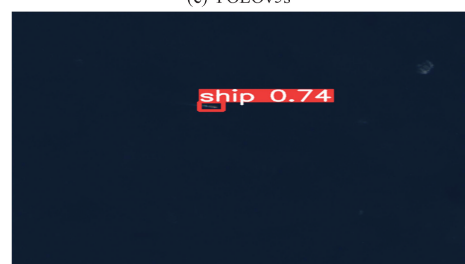
(b) YOLOv3



(c) YOLOv4



(d) YOLOv5s



(e) 改进的YOLOv5s

图10 黑夜下的海洋环境实例测试对比图

Fig.10 Marine environment example test comparison chart at night

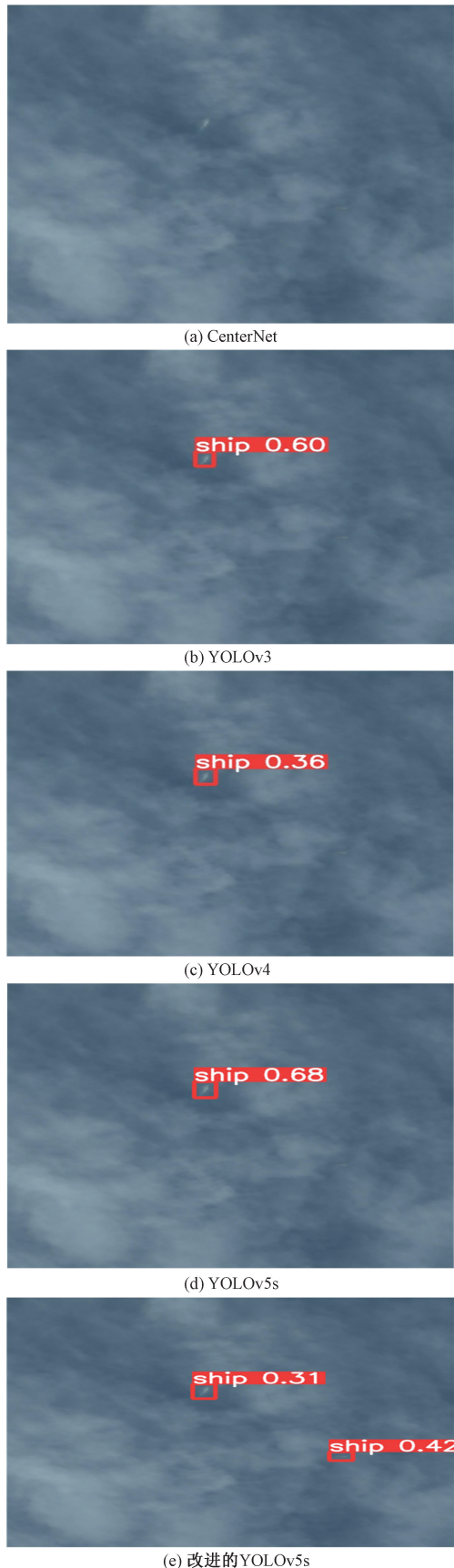


图 11 云层遮挡下的海洋环境实例测试对比图
Fig. 11 Comparison chart of test examples of marine environment under cloud cover

4 结论

针对遥感卫星需要在广大区域内快速检测船舶的需求,使用中等分辨率数据集——LEVIR-Ship,进行小目标船舶检测任务,在 YOLOv5s 基础网络结构上,进行了一系列优化改进,在进行实验与分析后,得到以下结论。

(1) 优化了颈部网络,边框损失函数,引入了自适应空间特征融合(ASFF)以及 BotNet 注意力机制解决小目标船舶检测和网络体量问题,实验结果表明,改进后 YOLOv5s 方法实现了在降低计算量和权重大小的情况下,检测的 mAP 值得到了有效提升,相较于基准 YOLOv5s 和 YOLOv8s, mAP 分别提升了 7.1 个百分点和 2 个百分点,在云层和黑夜的复杂环境下实现了高精度小目标船舶检测。

(2) 在船舶种类大而全的开源数据集 McShips 上,进行了对比实验,结果表明本文方法表现出了更优秀的性能,验证了所提出网络的普适性和有效性。

(3) 后续研究将会收集海上更多气象环境下的船舶数据图像数据,增加数据集的多样性,并研究每个气象环境对于检测效果的影响,并思考解决不利气象环境下的模糊图片的船舶检测。

参 考 文 献

- [1] Song Z H, Sui H G, Li Y C. A survey on ship detection technology in high-resolution optical remote sensing images[J]. Geomatics and Information Science of Wuhan University, 2021, 46(11): 1703-1715.
- [2] Muhammmad Y, Wan J H, Xu M M, et al. Ship detection based on deep learning using SAR imagery: a systematic literature review [J]. Soft Computing, 2023, 27(1): 63-84.
- [3] Jiang M, Gu L, Li X, et al. Ship contour extraction from SAR images based on faster R-CNN and Chan-Vese model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-14.
- [4] 赵振强, 何水原, 梁永志. 基于 Faster R-CNN 的遥感影像船舶检测识别[J]. 测绘通报, 2021(11): 59-64.
Zhao Zhenqiang, He Shuiyuan, Liang Yongzhi, et al. Research on remote sensing image ship detection and identification based on Faster R-CNN[J]. Bulletin of Surveying and Mapping, 2021(11): 59-64.
- [5] Yuan Z, Geng J, Dai T. Ship detection from remote sensing images based on deep learning[C]// Geo-Spatial Knowledge and Intelligence: 5th International Conference. Chiang Mai: Springer, 2018: 336-344.
- [6] 袁小平, 马绪起, 刘赛. 改进 YOLOv3 的行人车辆目标检测算法[J]. 科学技术与工程, 2021, 21(8): 3192-3198.
Yuan Xiaoping, Ma Xuqi, Liu Sai, et al. An improved algorithm of pedestrian and vehicle detection based on YOLOv3 [J]. Science Technology and Engineering, 2021, 21(8): 3192-3198.
- [7] 许小华, 陈豹, 王海菁, 等. 基于改进 YOLOv4-Tiny 的河湖船

- 船目标检测算法[J]. 人民长江, 2023, 54(9): 264-271.
- Xu Xiaohua, Chen Bao, Wang Haijing, et al. Object detection algorithm for ships on rivers and lakes based on improved YOLOv4-Tiny[J]. Yangtze River, 2023, 54(9): 264-271.
- [8] Xu X, Zhang X, Zhang T. Lite-yolov5: a lightweight deep learning detector for on-board ship detection in large-scene sentinel-1 sar images[J]. Remote Sensing, 2022, 14(4). DOI: 10.3390/rs14041018.
- [9] 凡文俊, 赵曙光, 郭力争. 基于改进 RetinaNet 的船舶检测算法[J]. 计算机应用, 2022, 42(7): 2248-2255.
- Fan Wenjun, Zhao Shuguang, Guo Lizheng. Ship detection algorithm based on improved RetinaNet[J]. Journal of Computer Applications, 2022, 42(7): 2248-2255.
- [10] 叶乐, 李朝锋. 基于频域通道注意力的 YOLOv3 网络的雾天海洋图像船舶检测[J]. 上海海事大学学报, 2023, 44(2): 18-24.
- Ye Le, Li Chaofeng. Frequency channel attention-based YOLOv3 network for ship detection from foggy marine images[J]. Journal of Shanghai Maritime University, 2023, 44(2): 18-24.
- [11] Zhu M, Hu G, Zhou H, et al. Arbitrary-oriented ship detection based on RetinaNet for remote sensing images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 6694-6706.
- [12] Liu T, Zhou B J, Zhao Y S, et al. Ship detection algorithm based on improved YOLO V5[C]//6th International Conference on Automation, Control and Robotics Engineering (CACRE). Dalian: IEEE, 2021: 483-487.
- [13] 化嫣然, 张卓, 龙赛, 等. 基于改进 YOLO 算法的遥感图像目标检测[J]. 电子测量技术, 2020, 43(24): 87-92.
- Hua Yanran, Zhang Zhuo, Long Sai, et al. Remote sensing image target detection based on improved YOLO algorithm[J]. Electronic Measurement Technology, 2020, 43(24): 87-92.
- [14] Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: a survey[J]. Computational Visual Media, 2022, 8(3): 331-368.
- [15] Srinivas A, Lin T Y, Parmar N, et al. Bottleneck transformers for visual recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 16519-16529.
- [16] Li H, Li J, Wei H, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles[J]. arXiv preprint arXiv: 2206.02424, 2022.
- [17] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [18] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2021, 52(8): 8574-8586.
- [19] Liu S, Huang D, Wang Y. Learning spatial fusion for single-shot object detection[J]. arXiv preprint arXiv: 1911.09516, 2019.
- [20] Chen J, Chen K, Chen H, et al. A degraded reconstruction enhancement-based method for tiny ship detection in remote sensing images with a new large-scale dataset[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.
- [21] Zheng Y, Zhang S. Mcships: a large-scale ship dataset for detection and fine-grained categorization in the wild[C]//IEEE International Conference on Multimedia and Expo (ICME). New York: IEEE, 2020: 1-6.
- [22] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. London: IEEE, 2018: 7132-7141.
- [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). Munich: Springer, 2018: 3-19.
- [24] 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. New York: PMLR, 2021: 11863-11874.
- [25] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 13713-13722.