

DOI:10.11918/202201030

多尺度水上船舶目标视觉检测

黄靖¹, 汤宁¹, 文元桥^{2,3}, 郭玉滨¹, 朱立夫¹, 肖长诗^{2,3}

(1. 武汉理工大学 计算机与人工智能学院, 武汉 430063; 2. 武汉理工大学 国家水运安全工程技术研究中心, 武汉 430063;
3. 武汉理工大学 智能交通系统研究中心, 武汉 430063)

摘要: 水上交通场景环境复杂, 通过普通光学摄像设备获取的水面图像, 面临着视觉目标清晰度低、尺度多样化等问题, 使得可见光视觉信号里中、小尺度目标检测相对困难。为服务于各类智慧海事应用, 提出了一个旨在提高复杂水域背景下多尺度水上船舶目标检测性能的算法(multi-scale ship object detection, MS-SOD)。该算法基于当前计算机视觉技术中主流的单阶段目标检测模型框架, 在其主干网络中嵌入卷积注意力模块, 来优化船舶特征提取能力; 在多尺度特征融合网络中引入富含细节信息的浅层特征, 并使用跨阶段局部残差结构, 来优化多尺度船舶特征的融合机制; 同时, 使用焦点损失函数, 来优化模型的学习过程; 并设计自适应锚框聚类算法优化先验锚框, 以提高多尺度船舶目标检测能力。为验证提出算法的有效性和实效性, 在构建较大规模水上船舶目标数据集的基础上, 开展了广泛实验验证。结果表明: 提出的算法在测试数据集上的检测准确度超过了各主流的对比方法; 特别是对于大、中、小各尺度船舶目标的检测精度, 相对于主流的 YOLOv4 算法, 提出的算法分别提升了 11.3%、6.0% 和 10.5%。

关键词: 多尺度船舶; 目标检测; 深度学习; 注意力机制; 特征融合

中图分类号: TP399 **文献标志码:** A **文章编号:** 0367-6234(2024)05-0103-11

Multi-scale visual detection for waterborne ship targets

HUANG Jing¹, TANG Ning¹, WEN Yuanqiao^{2,3}, GUO Yubin¹, ZHU Lifu¹, XIAO Changshi^{2,3}

(1. School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430063, China;

2. National Engineering Research Center Water Transport Safety, Wuhan University of Technology, Wuhan 430063, China;

3. Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, China)

Abstract: The water transportation environment poses challenges in terms of complexity, making it difficult to achieve clear and diverse visual target detection in water surface images captured by conventional optical cameras. This difficulty is particularly prominent when detecting medium- and small-scale objects in visible light visual signals. For the development of smart maritime applications, we proposed a multi-scale ship object detection (MS-SOD) algorithm to improve the performance of multi-scale ship object detection in complex waters. MS-SOD is built based on the mainstream framework of one-stage object detection models. The convolutional block attention module is embedded into its backbone network to optimize the ability of ship feature extraction. The shallow features with rich detailed information are added to the multi-scale feature fusion network, and cross-stage-partial residual structure is used to enhance the fusion mechanism of multi-scale ship object features. Additionally, a focal loss function is employed to optimize the training process of the model, and an adaptive anchor clustering algorithm is designed to optimize the prior anchor and improve the detection capability for multi-scale ship objects. Extensive experiments are conducted on a self-built large-scale ship object dataset to validate the effectiveness and efficiency of the proposed MS-SOD algorithm. Experimental results show that the accuracy of MS-SOD outperforms various mainstream comparative methods on test dataset. Especially, compared with the YOLOv4 algorithm, the detection accuracy of large-, medium-, and small-scale ship objects improve by 11.3%, 6.0%, and 10.5%, respectively.

Keywords: multi-scale ship; object detection; deep learning; attention mechanism; feature fusion

随着交通强国和海洋强国战略的推进, 水上交通行业蓬勃发展, 推进水上交通智能化发展能有效提高水路运输效率。其中, 使用计算机视觉

(computer vision) 相关技术对航道及海域内水上船舶目标进行智能感知与检测, 是推动水运行业智能化发展的重要方向。

收稿日期: 2022-01-07; 录用日期: 2022-04-01; 网络首发日期: 2024-03-20

网络首发地址: <http://kns.cnki.net/kcms/detail/23.1235.T.20240320.1235.004.html>

基金项目: 国家自然科学基金(52072287); 浙江省科技计划(2021C01010)

作者简介: 黄靖(1977—), 男, 副教授, 硕士生导师; 文元桥(1975—), 男, 教授, 博士生导师; 肖长诗(1974—), 男, 教授, 博士生导师

通信作者: 黄靖, huangjing@whut.edu.cn

然而,基于常规视觉感知设备(例如,普通光学摄像机)捕获的水面图像易受天气、光照、水域环境干扰,面临目标清晰度低、尺度多样化等问题,使得视觉信号中的中、小尺度目标检测相对困难,进而影响当前基于计算机视觉技术智能感知船舶目标方法的实用效果和应用价值。因此,实现高效、精准的多尺度水上船舶目标视觉检测显得至关重要,其能为智能无人船、船舶辅助驾驶、海事自动化监管等诸多下游任务提供有力的技术支撑与保障,且对发展智能水上交通战略亦有重要意义^[1]。

近年来,随着深度学习(deep learning)的迅猛发展,基于深度学习的目标检测算法^[2-8]层出不穷、大放异彩,逐渐成为当前计算机视觉中实现目标检测任务的主流方法^[9]。其中,针对多尺度目标检测的任务需求,诸多经典的深度学习目标检测算法(如 YOLO^[7]、R-CNN^[8]等)亦有相应的改进。然而,这些方法主要聚焦于多尺度目标检测通用模型、算法的设计,缺乏对于特定应用领域相关问题的关注,从而使得这些现有模型、算法部署于特定应用场景后,其检测效果和性能仍存在不足与挑战。例如,国内复杂的内河通航环境中日夜航行着各种大小、形态的船舶,能否对这些具有不同视觉信息的船舶进行有效的特征表达,直接关系到水上多尺度船舶目标检测的效果。为此,本文开展了这方面的研究与探索工作,其工作贡献归纳如下:

1) 针对中国内河水域场景复杂的特点,关注不同尺度船舶目标视觉信号(特别是船舶小目标),构建融合多尺度目标的视觉表征机制,解决通用目标检测算法对于多尺度船舶目标检测效果不理想的现状。

2) 提出了一种面向多尺度水上船舶目标的检测方法(multi-scale ship object detection, MS-SOD),旨在提高复杂水域背景下多尺度水上船舶目标的检测精度。

3) 为验证 MS-SOD 方法的有效性和实效性,在构建较大规模水上船舶目标数据集的基础上,开展了广泛实验验证。与此同时,该方法也被部署于长江海事智能监管平台,进而在中国内河实际应用场景中得到了必要实测。

1 相关工作

目前,基于深度学习的主流目标检测算法大多基于锚框(anchor-based)设计,其主要表现为两类模型框架,即以 R-CNN^[8]系列为代表的双阶段算法和以 SSD^[6]、YOLO^[7]系列为代表的单阶段算法。针对多尺度目标检测问题,R-CNN^[8]系列中的经典模型 Faster R-CNN^[10]借助 K 均值聚类(K-means clustering)算法,设计了关联待检目标的一类重要模

型部件——锚框(anchor),并使用区域提议网络(region proposal network, RPN)来生成不同尺度的锚框,从而在高维特征图(feature maps)上实现多尺度目标检测,但由于其在特征提取层上未区分尺度,使得中、小尺度目标易丢失细节信息,以致 Faster R-CNN 算法的多尺度目标检测能力不够突出;而 YOLO^[7]系列中的新晋模型 YOLOv4^[11]则使用 FPN-PAN 结构进行多尺度特征融合,并通过设置多检测头的方式实现分尺度检测,但其在特征提取上对关键特征的关注不够,特征融合不够充分,且损失函数特性单一,以致 YOLOv4 算法的多尺度目标检测效果仍有不足;YOLOv5-X 在 YOLOv4 的基础上增大模型规模,提高网络深度及宽度,特征提取能力有一定增强,检测性能有所提升,但模型的参数量与计算量也大幅增加,实时性较低。

特别是,当使用这两类支持多尺度目标检测的经典模型在水域场景中开展船舶目标检测实验时,其在构建的较大规模水上船舶目标数据集上的检测性能并不理想。Faster R-CNN 算法对各类船舶检测的平均精度 V_{mAP} (mean average precision, mAP) 是 49.6%,其中大、中、小尺度目标检测精度分别为 55.6%、29.5%、12.1%;YOLOv4 算法的 V_{mAP} 是 50.9%,其大、中、小 3 种尺度目标检测精度依次为 59.6%、27.8%、13.0%。由此可见,这些通用目标检测模型和算法对于中、小尺度船舶目标的检测效果相对较弱,其精度仅接近于大尺度目标检测结果的一半,故而导致较低的 V_{mAP} 。

为了提高水上航道场景中船舶目标的检测性能,国内外一些学者在通用目标检测模型的基础上开展了进一步的算法优化设计。例如,Chen 等^[12]、王新立等^[13]、聂鑫等^[14]、Liu 等^[15]针对水上船舶的目标特点,对通用目标检测算法进行改进,提高了对船舶目标的检测能力;Shao 等^[16]提出显著性感知网络,利用基于显著区域检测的全局对比辅助检测算法获得更准确的船舶位置;Kim 等^[17]利用连续帧图像中目标边界框的交并比召回漏检的目标,并对目标预测概率进行贝叶斯融合,提高类别预测精度。上述方法虽然提高了对水上船舶目标的检测能力,但对多尺度船舶目标检测均未做讨论。潘慧等^[18]在 YOLOv3 算法基础上,增加网络预测尺度,并使用判别相关分析强化多尺度融合,提高了模型对船舶的检测能力,然而,引入额外检测尺度后,计算量增大,实时性下降;Han 等^[19]针对 YOLOv4 模型进行改进,使用恒等平行模式提高残差结构的特征提取能力,并使用扩张卷积强化空间金字塔池化(spatial pyramid pooling, SPP)结构增大感受野,提高空间信

息获取能力,改善小尺度船舶的漏检问题,但是其改进模块增大了模型的参数量与计算量,以致检测速度受限。

上述针对水上船舶目标检测的算法都在一定程度上提升了船舶目标的检测性能,但是他们主要针对中大型尺度目标进行检测优化,对于多尺度水上船舶目标的检测未做深入分析,且部分算法在实时性上有所欠缺。在实际场景中,受水上船舶目标的特点和监管水域复杂环境影响,视觉感知设备通常获取的远距离目标尺度小、近距离目标尺度大,且目标尺度变换程度较大。因此,使用这些目标检测算法对水域中、小尺度船舶目标进行检测时易出现漏检、误检等情况,同时考虑到实际应用需求,船舶目标检测算法还存在着相关的实时性要求。为此,本文开展了面向多尺度船舶检测问题的研究与探索工作。

2 融合多尺度视觉表征的水上船舶目标检测模型

在计算机视觉中,船舶检测可看作是目标检测

技术在特定应用领域中的一项具体任务。为更高效地开展本文的研究,我们基于通用目标检测方法的前沿成果,着力构造一种适应于水域场景中多尺度船舶目标检测的算法模型。首先选用目前在检测精度和运行速度上都具有一定技术优势的一类通用目标检测算法——YOLOv4^[11]作为本文研究的依托架构,然后根据水域场景中各类船舶目标的视觉特点,设计出多尺度视觉表征提取机制,并将其作为核心部件融入船舶目标检测模型中,形成一个适用于水域应用场景下能精准、高效检测各类船舶目标的具体方法,即多尺度水上船舶目标检测(MS-SOD)算法模型,与此同时进一步优化了该算法模型的学习策略。

2.1 模型结构

聚焦水上船舶目标的多尺度视觉表征,重点设计了检测算法模型的 Backbone、Neck 和 Head 3 个关键组件,相应提出了适应于水域场景中多尺度船舶目标检测的 MS-SOD 算法,其模型结构见图 1。

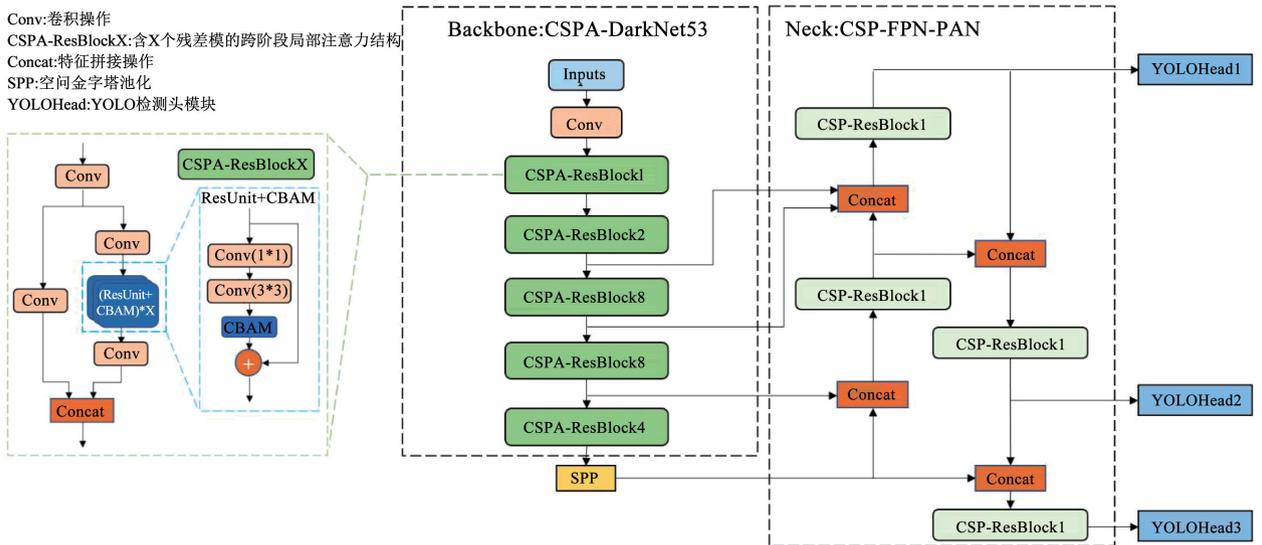


图1 MS-SOD 模型结构

Fig. 1 Structure of MS-SOD

从模型结构的设计角度看,MS-SOD 算法具有如下特点:

1) 为提高检测模型对关键有效特征的提取能力,MS-SOD 在其 Backbone 的残差单元中嵌入基于注意力机制的 CBAM^[20] (convolutional block attention module) 模块,构成 CSPA-ResBlockX (CSPA 即 cross-stage-partial attention, X 表示残差单元的数量) 结构,来加强主干网络对关键特征的注意力,并抑制冗余特征,进而使检测模型能够更多地关注船舶目标相关的重要视觉特征信息,而忽略背景等冗

余信息。

2) 为改善多尺度特征的表征能力,MS-SOD 在其 Neck 部件强化 Backbone 中浅层细节信息的引入,通过使用跨阶段局部 (cross-stage-partial, CSP)^[21] 残差结构优化多尺度特征融合,进而丰富各种尺度信息的提取。如图 1 所示,在 Neck 中增加来自主干网络浅层的特征输入通道^[22],使检测模型尽力保留中、小尺度目标的特征提取,并在特征拼接之后,将各种尺度信息共同汇入 CSP-ResBlock 结构,同时优化特征融合过程^[23]。

3) 为确保检测模型的训练效果, MS-SOD 在其 Head 中使用 Focal Loss^[24] 作为损失函数优化损失分配, 和缓解模型学习过程中存在的难易样本分布不均问题。此外, 引入 Mish 激活函数, 促进梯度信息传递, 避免训练陷入梯度饱和及局部最优^[25]。

4) 为提升检测算法对于多尺度目标的回归能力, MS-SOD 设计了一个针对水上船舶目标的自适应锚框聚类算法, 来优化模型中预设锚框的尺寸, 进而提高检测算法对多尺度船舶目标的检测与定位能力。

2.2 注意力残差结构

如图 1 所示, MS-SOD 在主干网络残差块的残差单元内嵌入了基于注意力机制的 CBAM 模块, 以此提升检测模型主干网络的特征提取能力。

具体而言, 如图 2 所示, 在残差块的残差单元 (ResUnit) 中, 主路通过 1×1 卷积和 3×3 卷积进行特征提取后, 依次通过通道注意力^[26] (channel attention, CA) 和空间注意力 (spatial attention, SA) 的注意力系数权重加权, 进一步提取有效的特征信息, 再与侧路的原始输入进行加和, 组成新的注意力残差单元 (即 ResUnit + CBAM 模块)。残差单元中的 CBAM 将通道注意力模块与空间注意力模块按顺序组成链路, 利用通道注意力机制在通道维度上学习有效的通道特征信息, 同时使用空间注意力机制在各通道层特征平面上学习有效的平面特征信息, 而后形成的混合注意力机制模块, 结合原有的残差结构, 使其能更加高效地提取关键特征。由此得到的注意力残差单元组成 MS-SOD 中新的主干网络 CSPA-Darknet53, 可以促进模型学习更多关键特征信息, 同时抑制冗余无关信息, 提高对目标的特征提取能力。

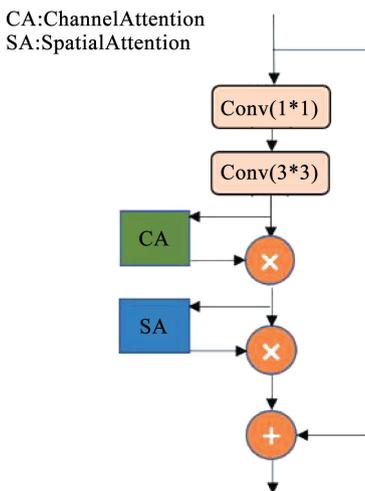


图 2 ResUnit + CBAM 模块结构

Fig. 2 Structure of ResUnit + CBAM module

2.3 优化策略

YOLOv4 模型训练所用的损失函数由 3 部分组成, 分别是边界框偏移损失 L_{bbox} 、置信度损失 L_{obj} 和分类损失 L_{cls} , 其中置信度损失和分类损失使用二元交叉熵 (binary cross entropy, BCE) 进行计算, 边界框损失则使用完全交并比 (complete intersection over union, CIoU) 计算。

模型训练时存在大量的难易样本分布不均的情况, 易分类样本占多数, 难分类样本占少数。而 BCE 损失计算方法对于预测概率大于 0.5 的易分类目标仍然给予部分不可忽略的损失^[24], 大量的易分类样本在损失中占据主导地位。因此, 使用这种方式训练出来的模型对易分类样本预测容易, 对难分类样本仍然难以预测, 训练收益较低。

为了提升训练效果, 使模型获得更好的检测性能和泛化能力, 解决数据集训练时存在的样本不平衡问题, MS-SOD 使用 Focal Loss 函数, 降低易分类样本的损失比重, 同时提高难分类样本的损失比重, 使得模型进一步专注于难分类样本的学习。Focal Loss 函数描述为

$$L_{\text{FL}}(p_i) = -\alpha_i (1 - p_i)^\gamma \log(p_i) \quad (1)$$

式中: FL 为 Focal Loss 简称; 当预测类别与真值类别相同时, p_i 取值为预测概率 p , 调制因子 α_i 取值 α ; 反之, p_i 取值 $1 - p$, α_i 取值 $1 - \alpha$; γ 为权重因子。

对于边界框损失, MS-SOD 模型保留 CIoU 损失函数, 该损失函数同时考虑预测框与真值框之间的多个关联因素: 重叠面积、中心点距离、目标框长宽比, 如此可以更好地反映预测框与真值框的差异, 提高模型的边界框回归能力。CIoU 损失公式为

$$L_{\text{CIoU}} = 1 - \left(S_{\text{IoU}} - \frac{d^2}{c^2} - \frac{v^2}{1 - S_{\text{IoU}} + v} \right) \quad (2)$$

式中: S_{IoU} 为预测框 (B_{pre}) 与真值框 (B_{gt}) 之间的交并比, d 为 B_{pre} 与 B_{gt} 组成的最小外接矩形的对角线长度, c 为两目标框的中心点之间的欧氏距离, v 为衡量两目标框长宽比一致性的参数, 具体为

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w_{\text{gt}}}{h_{\text{gt}}} - \arctan \frac{w_{\text{pre}}}{h_{\text{pre}}} \right)^2 \quad (3)$$

式中: w_{gt} 、 h_{gt} 为真值目标框的宽和高, w_{pre} 、 h_{pre} 为预测框的宽和高。

将上述 Focal Loss 用于计算置信度损失及分类损失, CIoU 函数用于计算边界框回归损失, 则 MS-SOD 模型总的损失函数定义为

$$L_{\text{total}} = L_{\text{bbox}} + L_{\text{obj}} + L_{\text{cls}} = \lambda_{\text{bbox}} \sum_{i=0}^{k \times k} \sum_{j=0}^m I_{ij}^{\text{obj}} L_{\text{CIoU}}(B_{\text{pre}}, B_{\text{gt}}) + \sum_{i=0}^{k \times k} \sum_{j=0}^m I_{ij}^{\text{obj}} L_{\text{FL}}(C_{ij}) +$$

$$\lambda_{\text{no_obj}} \sum_{i=0}^{k \times k} \sum_{j=0}^m I_{ij}^{\text{no_obj}} L_{\text{FL}}(C_{ij}) + \sum_{i=0}^{k \times k} \sum_{j=0}^m I_{ij}^{\text{obj}} \sum_{c \in C_{\text{class}}} L_{\text{FL}}(p_{ij}^c) \quad (4)$$

式中: k 为划分图像后的单行网格数, $k \times k$ 为所有待检测的网格数; m 为每个网格产生的候选框个数; i 为网格编号; j 为候选框编号; λ_{bbox} 为边界框坐标定位损失的权重参数; $\lambda_{\text{no_obj}}$ 为候选框无目标情况下置信度损失的权重参数; I_{ij}^{obj} 和 $I_{ij}^{\text{no_obj}}$ 表示以网格 i 为中心的 j 个候选框中是否存在目标的参数,两参数取值相反,当候选框存在目标时 I_{ij}^{obj} 取1、 $I_{ij}^{\text{no_obj}}$ 取0,反之 I_{ij}^{obj} 取0、 $I_{ij}^{\text{no_obj}}$ 取1; C_{ij} 为候选框的预测置信度; c 为预测类别; C_{class} 为类别编号集合; p_{ij}^c 为以网格 i 为中心的 j 个候选框中的目标在 c 类别上的预测分数。

2.4 自适应锚框聚类

原生YOLOv4算法使用预设锚框机制回归目标检测框,预设锚框的尺寸设置与训练效果直接相关。针对水上船舶目标的视觉特点,对构建的较大规模水上船舶目标数据集中所有目标使用自适应锚框聚类算法(adaptive anchor clustering, Adapt-Anchor)进行处理,优化预设锚框尺寸参数。该算法描述见算法1,其通过K-means聚类算法聚类出含 N_a 个簇类中心的集合 K ,再以 K 与尺寸数值集合 D 计算出的适应度 f 为基准,使用遗传算法迭代获得更高的适应度 f_g ,得到优化的簇类中心集合 K_g ,最后得到的锚框参数与真值目标的匹配度更高。

算法1 自适应锚框聚类算法

输入:训练数据集 $D = \{[w_1, h_1], [w_2, h_2], \dots, [w_n, h_n]\}$,IoU 阈值 S_{IoU} ,聚类簇数 N_a ,进化代数 g

过程:函数 $F_{\text{AdaptAnchor}}(D, N_a, S_{\text{IoU}}, g)$

根据 D 随机生成 N_a 个簇类中心: $K = \{k_1, k_2, \dots, k_{N_a}\}$

repeat

 令 $C_i = \emptyset (1 \leq i \leq N_a)$

 for $j = 1, 2, \dots, n$ do

 计算目标样本 $x_j: [w_j, h_j]$ 与各簇中心 $k_i (1 \leq i \leq N_a)$ 的 S_{IoU} 值,找出最大 S_{IoU} 所在的簇 λ ,将样本 x_j 划入相应的簇: $C_\lambda = x_j \cup C_\lambda$;

 end for

 for $i = 1, 2, \dots, N_a$ do

 计算新的簇类中心: $\mu_i = \frac{\sum x \in C_i}{|C_i|}$

 if $\mu_i \neq k_i$ then

 更新簇中心 $K: k_i = \mu_i$

 end if

end for

until 当前簇类中心 K 中元素均未更新

计算簇类中心 K 适应度: $f = F_{\text{fitness}}(K, D, S_{\text{IoU}})$

repeat

 随机生成进化因子 V

 进化锚框: $K_g = K * V$

 计算新适应度: $f_g = F_{\text{fitness}}(K_g, D, S_{\text{IoU}})$

 if $f_g > f$ then

 更新簇类中心: $K = K_g$

 更新适应度: $f = f_g$

 end if

until 完成迭代次数 g

输出:簇类中心 $K = \{k_1, k_2, \dots, k_{N_a}\}$

3 实验结果与分析

3.1 水上船舶数据集

依据国内外水上交通目标种类特点,通过实验采集与网络收集的方式整理出共14 220张水面可见光图片,由此建立了一个较大规模的水上船舶目标数据集。在该数据集中,水上船舶目标被分为六类:邮轮、集装箱船、散货船、帆船、岛礁、其他船。图3描述了各类的样例,其中岛礁是作为训练负样本保留。此外,表1统计了数据集中各类水上船舶目标的数量。



图3 船舶类别示例

Fig.3 Examples of ship categories

表1 水上船舶目标数据集统计信息

Tab.1 Statistics of waterborne ship object dataset

类别	数量
邮轮	1 447
集装箱船	871
散货船	3 121
帆船	3 350
岛礁	5 910
其他船	8 843
总计	23 542

按照 COCO 数据集标准,对全部水上船舶目标进行了尺度分类,即依据水上船舶目标在图片中覆盖像素面积大小,分为小目标、中目标和大目标 3 类不同尺度,其中将像素面积小于 32×32 的目标定义为船舶小目标,将像素面积大于 32×32 且小于 96×96 的目标定义为船舶中目标,将像素面积大于 96×96 的目标定义为船舶大目标。依据该尺度划分标准,对水上船舶目标数据集中的目标进行了尺度统计,结果见表 2。

表 2 船舶目标尺度统计信息

Tab.2 Scale statistics of ship object dataset

目标类型	尺度大小	数量	数量占比/%
小目标	$< 32 \times 32$	2 430	10.32
中目标	$32 \times 32 \sim 96 \times 96$	6 745	28.65
大目标	$> 96 \times 96$	14 367	61.03
总计		23 542	100.00

3.2 实验环境与评估指标

硬件配置:处理器 Intel Xeon 8176 双处理器、内存 314 GB、图形处理器 Nvidia RTX 2080Ti $\times 4$ 。

软件配置:Ubuntu 20.04、Python 3.8、PyTorch 1.8、CUDA 11.1。

本文模型 MS-SOD 包含一些关键模块参数,其中,基于自适应锚框聚类算法设计的锚框尺寸为: $24 \times 17, 57 \times 30, 63 \times 83; 160 \times 47, 157 \times 122, 404 \times 88; 217 \times 275, 533 \times 183, 510 \times 329$ 。此外,总损失函数中的权重参数 $\lambda_{\text{bbox}}, \lambda_{\text{no_obj}}$ 及 Focal Loss 中的 α, γ 设置见表 3。

表 3 关键参数设置

Tab.3 Setting of key parameters

模块名称	关键参数	默认值
总损失函数	权重参数 λ_{bbox}	0.05
	权重参数 $\lambda_{\text{no_obj}}$	0.50
Focal Loss	调制因子 α	0.25
	权重因子 γ	2.00

实验中,各模型参数初始权重使用随机初始化方式,并使用随机梯度下降(SGD)优化器优化梯度更新。其他超参数设置一致,见表 4。

表 4 超参数设置

Tab.4 Setting of hyper-parameters

超参数	默认值
输入尺寸	640×640
Batch-size	8
初始学习率	0.01
梯度更新动量	0.937
权重衰减	0.000 5
IoU 阈值	0.20

模型的性能主要从船舶检测算法的实时性和精确度 2 个方面进行评估。在实时性方面,使用模型参数量与每秒千兆次级浮点运算数(GFLOPS)衡量模型的计算复杂度,使用帧率(FPS)衡量检测的实时性;在精确度方面,使用 V_{mAP} 衡量模型的总体检测精度,使用 $V_{\text{AP}_{50}}$ 衡量 IoU 阈值为 0.50 时的各类别检测精度,使用 $V_{\text{AP}_s}, V_{\text{AP}_m}, V_{\text{AP}_l}$ 衡量模型对小、中、大 3 种不同尺度船舶目标的检测精度,其中 V_{mAP} 计算方法采用 COCO 数据集常用的目标检测精度计算公式为

$$V_{\text{mAP}} = \frac{1}{10} \sum_{S_{\text{IoU}} \in \{(0.50+0.05i) | 0 \leq i \leq 9\}} \left(\frac{1}{n} \sum_{c=1}^n P_c^{S_{\text{IoU}}} \right) \quad (5)$$

式中: S_{IoU} 为交并比阈值,在 $[0.50, 0.95]$ 内间隔 0.05 取值,共 10 种取值方式, $P_c^{S_{\text{IoU}}}$ 为阈值等于 S_{IoU} 时 c 类别的检测精度, n 为类别总数,即求各类别平均精度后,再求不同 IoU 阈值下的均值,得到 V_{mAP} 。

3.3 实验结果与对比分析

3.3.1 检测性能分析

研究选用 Anchor-free 目标检测算法中的代表模型 CenterNet^[27]、FCOS^[28],以及 Anchor-based 目标检测算法中的经典双阶段检测算法 Faster R-CNN,和著名单阶段检测算法 YOLOv4、YOLOv5-X^[29] 作为实验对比的 Baseline 模型。在水上船舶目标检测数据集上,这些算法模型的对比结果见表 5。

表 5 检测精度对比

Tab.5 Comparison of detection accuracy

算法类型	阶段分类	算法模型	$V_{\text{mAP}}/\%$
Anchor-free		CenterNet	52.3
		FCOS	52.8
Anchor-based	双阶段	Faster R-CNN	49.6
	单阶段	YOLOv4	50.9
	单阶段	YOLOv5-X	60.9
	单阶段	Our MS-SOD	62.3

表 5 结果表明,本文提出的 MS-SOD 相比主流的 YOLOv4 算法在 V_{mAP} 指标上提高了 11.4%,性能略优于模型规模更大的 YOLOv5-X,且高于其他算法在水上船舶目标数据集上的检测精度。

为了进一步验证本文方法在多尺度目标检测上的性能提升,针对经典的 YOLOv4 算法、YOLOv5-X 算法与本文的 MS-SOD 算法开展了分尺度检测精度比较,其结果见表 6。

表6 多尺度船舶目标检测精度比较

Tab.6 Comparison of multi-scale detection accuracy for ship objects

模型	V_{mAP}	$V_{AP_{50}}$	V_{AP_s}	V_{AP_m}	V_{AP_l}
YOLOv4	50.9	71.8	13.0	27.8	59.6
YOLOv5-X	60.9	81.9	14.9	40.6	69.4
Our MS-SOD	62.3	83.0	23.5	33.8	70.9

表6结果表明:在多尺度水上船舶目标检测测试上,本文提出的MS-SOD算法的检测效果更好。相比YOLOv4,MS-SOD对小、中、大尺度水上船舶目标的检测精度均有提升。其中,船舶小目标检测精度 V_{AP_s} 提升了10.5%,中尺度船舶目标检测精度 V_{AP_m} 提升了6.0%,船舶大目标检测精度 V_{AP_l} 提高了11.3%;而相比YOLOv5-X,虽然 V_{mAP} 较接近,但小目标检测精度 V_{AP_s} 仍提升了8.6%。结果显示,MS-SOD模型在引入了CBAM注意力模块后,显著改善了主干网络的特征提取能力,同时通过在Neck部分引入浅层细节特征,CSP-ResBlock促进多尺度特征融合,优化各个尺度的特征输出,整体提高了模型对不同尺度目标的检测能力。

为验证检测算法模型的实时性能,开展了针对模型参数量和计算量的实验对比分析。表7描述了当前具有实时检测优势的YOLOv4、YOLOv5-X和本文提出的MS-SOD的对比结果,其表明本文提出的MS-SOD算法模型参数量减少了 2.7×10^6 ,浮点运算量降低了4GFLOPs,实时检测帧率小幅提高。在配置Nvidia RTX2080Ti图形处理器的环境中,MS-SOD实时检测帧率达到39帧/s,相同实验条件下比YOLOv4提高了7帧/s,比YOLOv5-X提高了16帧/s,其主要是因为CBAM注意力模块并未大幅增加模型参数量,并且在Neck部分使用CSP-ResBlock结构取代常规的多卷积操作,使得参数量整体减少,计算量同步下降,模型推理速度加快,算法的实时性提高。

表7 模型参数规模和计算量对比

Tab.7 Comparison of parameter scales and calculation amounts of different detection models

模型	参数量/ 10^6	GFLOPs	帧率/(帧·s ⁻¹)
YOLOv4	59.5	133.7	32
YOLOv5-X	88.5	220.1	23
MS-SOD	56.8	129.7	39

3.3.2 后处理优化

本文提出的MS-SOD与YOLOv4一样,仍采用

NMS后处理方式对大量的预测框结果进行筛选,尤其是对预测同一目标的冗余候选框过滤,通过计算预测框之间 S_{IoU} ,根据设定的 S_{IoU} 确定需要保留的最佳结果框。若 S_{IoU} 高于阈值,则认为2预测框存在重复预测,去掉其中置信度较低的预测框,反之则认为2预测框分别预测了2个不同的目标。因而,设定不同的 S_{IoU} ,NMS对预测框的抑制效果不同,即获得的检测结果不同。针对水上船舶目标数据集,在实验中进一步分析了提出的MS-SOD算法模型在不同 S_{IoU} 下的性能变化,其结果见表8。

表8 不同 S_{IoU} 下MS-SOD模型性能对比

Tab.8 Performance comparison of MS-SOD at different S_{IoU}

S_{IoU}	V_{mAP}	$V_{AP_{50}}$	V_{AP_s}	V_{AP_m}	V_{AP_l}
0.50	61.6	83.8	23.2	33.7	70.1
0.55	61.9	83.8	23.3	33.8	70.3
0.60	62.0	83.7	23.5	33.9	70.6
0.65	62.1	83.4	23.6	33.8	70.6
0.70	62.3	83.0	23.5	33.8	70.9
0.75	62.2	82.2	23.0	33.5	70.9
0.80	61.9	80.8	22.0	32.8	70.7
0.85	60.9	78.4	20.1	30.4	70.1
0.90	58.4	73.4	17.6	25.9	68.2
0.95	47.7	57.4	14.4	18.5	56.7

表8结果表明,当 S_{IoU} 取0.70时,MS-SOD算法的平均精度 V_{mAP} 为62.3,其相比于 S_{IoU} 取0.50时,检测精度 V_{mAP} 提高了0.7%,船舶大目标检测精度提高了0.8%。

可以看出,在不增加额外计算开销的前提下, S_{IoU} 优化方法改进了NMS后处理的效果,进一步提高了MS-SOD算法在水上船舶目标数据集上的检测精度。

3.3.3 消融实验

为了进一步验证提出模型MS-SOD的有效性,开展了必要的消融实验。针对完整的MS-SOD模型,分别去除4个重要模块(即CBAM、CSP-FPN-PAN、Focal Loss、Adapt-Anchor),相应地构成4个不同的检测模型,并分别为MS-SOD-1、MS-SOD-2、MS-SOD-3和MS-SOD-4。然后,在水上船舶目标数据集上,对这4个衍生算法模型的检测精度进行了比较测试,其结果见表9。

表 9 消融实验结果

Tab.9 Results of ablation experiments

模型	CBAM	CSP-FPN-PAN	Focal Loss	Adapt-Anchor	$V_{mAP}/\%$	$V_{AP_s}/\%$	$V_{AP_m}/\%$	$V_{AP_l}/\%$
YOLOv4					50.9	13.0	27.8	59.6
MS-SOD-1		✓	✓	✓	58.8	18.5	32.9	67.4
MS-SOD-2	✓		✓	✓	57.7	13.1	28.2	66.7
MS-SOD-3	✓	✓		✓	61.1	20.4	28.1	70.1
MS-SOD-4	✓	✓	✓		60.7	20.8	28.4	69.3
MS-SOD	✓	✓	✓	✓	62.3	23.5	33.8	70.9

表 9 结果显示,在消融实验中,去除 CBAM 的 MS-SOD-1, V_{mAP} 降低了 3.5%, 表明 CBAM 在主干网络引入注意力机制,有效改善了网络对各尺度目标的特征提取能力,提高了整体检测性能;在 MS-SOD-2 中,移除了 CSP-FPN-PAN 模块,采用原始的 FPN-PAN 模块,其 V_{mAP} 下降了 4.6%, 且 V_{AP_s} 下降了 10.4%, V_{AP_l} 下降了 4.2%, 说明提出的特征融合策略对于改善多尺度特征提取是有效的,引入的浅层细节特征增强了船舶小尺度目标的特征提取,并显著提高了船舶小目标的检测精度;在 MS-SOD-3 中,移除改进损失函数 Focal Loss,其算法的检测精度小幅下降了 1.2%, 表明 Focal Loss 在训练中通过重新分配损失占比,改善了模型训练时损失分布不均的情况,提高了模型对难样本的检测性能;在 MS-SOD-4 中,移除了 Adapt-Anchor 锚框聚类,其算法精度下降

了 1.6%, 各尺度目标检测精度均有所下降,表明 Adapt-Anchor 提高了预设锚框与数据集的匹配度,改善了模型对各尺度目标的回归能力。

因此,消融实验结果表明,MS-SOD 算法模型中相关的设计策略对于提升多尺度水上船舶目标的检测性能是有效的。

3.3.4 结果可视化分析

为进一步验证 MS-SOD 模型的检测性能,采用梯度加权类激活映射(Grad-CAM)方法^[30],可视化分析模型对船舶目标的检测情况。其中,Grad-CAM 方法是利用预测结果中的类别分数,计算特定网络层中该类别的分类权重,并将该分类权重通过热力图形式显示到原始输入图像上,能较好地反馈模型对目标的检测情况,可作为模型分类与定位能力的解释依据。可视化结果见图 4。

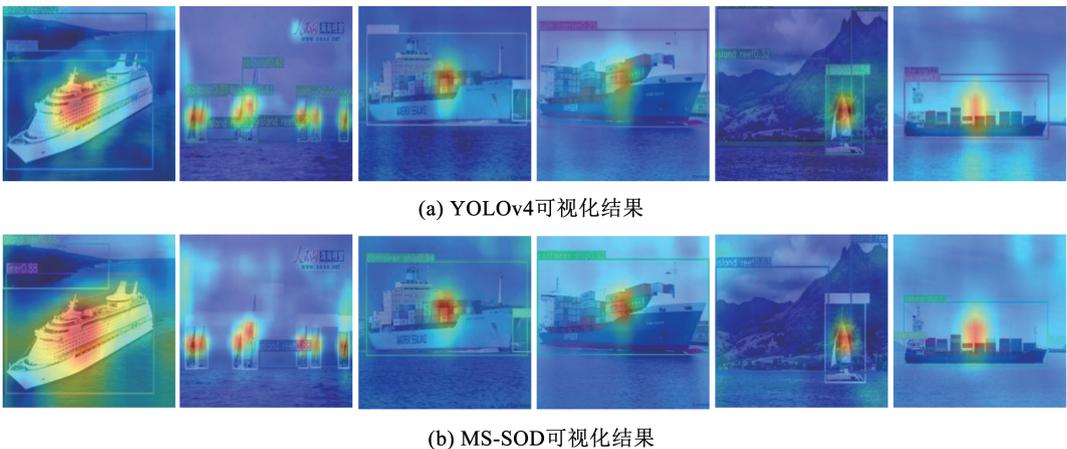


图 4 不同检测模型基于 Grad-CAM 的可视化结果比较

Fig.4 Comparison of visual analysis results of different models based on Grad-CAM

图 4 展示了 YOLOv4 和 MS-SOD 针对水上船舶目标样例的 Grad-CAM 可视化检测结果。对比结果显示,图 4(b) 中待检测目标中心区域的热力图值明显高于图 4(a) 中对应的区域结果,并且图 4(b) 中的响应范围较大,说明 MS-SOD 对于目标区域的特征响应明显强于 YOLOv4。结合图中的检测结果可以看到,MS-SOD 对于目标的类别预测更加准确,结

果置信度更高。由此可见,MS-SOD 的主干网络模型在 CSP-ResBlock 模块的残差单元嵌入 CBAM 和优化残差单元对输入特征的处理,增强了模型对有效特征的提取能力,提高了船舶目标检测性能。

此外,图 5 展示了 YOLOv4 和 MS-SOD 针对可见光样例图片进行检测的对比结果。

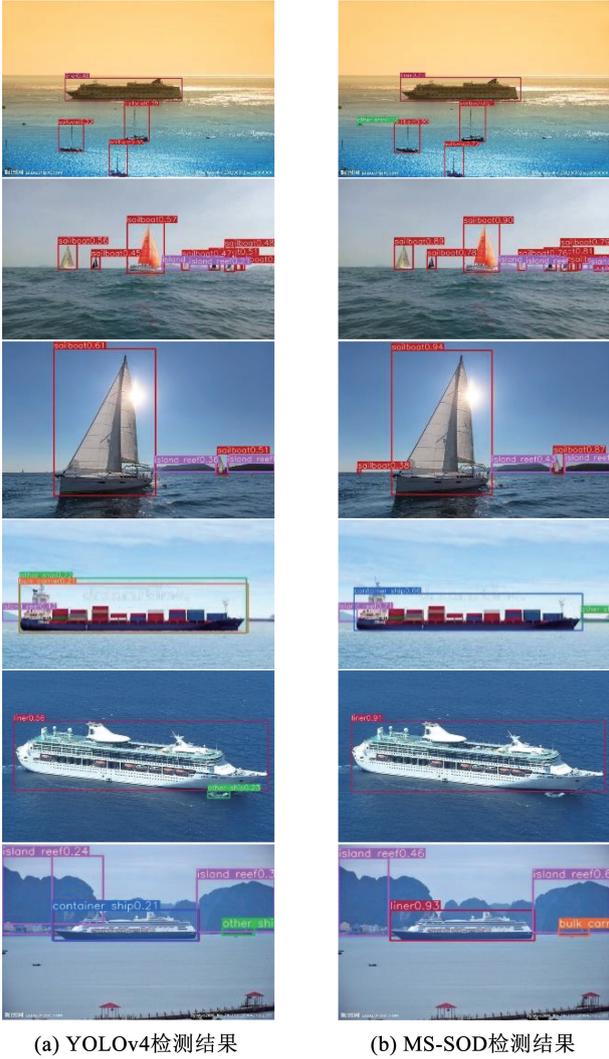


图5 不同检测模型针对可见光图片的检测结果比较

Fig. 5 Comparison of visible light detection results of different models

图5结果表明,在船舶分类上,MS-SOD 准确率要高于YOLOv4;在目标位置框预测上,MS-SOD 通过预设锚框优化改善了目标框的回归能力,NMS 后处理优化消除了冗余的预测框,为每个目标保留了更加精准的预测框;在预测置信度上,MS-SOD 对目标的预测结果置信度更高。

4 应用实例

为了更好地验证 MS-SOD 方法服务于当前智慧海事建设的实际效果,结合长江海事智能监管平台的开发,部署了本文提出的算法模型,并开展了必要的应用实测。

长江海事智能监管平台主要功能之一,是对目标水域通航船舶进行全天候、全时域的海事监管,使

用视觉目标检测算法完成船舶目标的机器识别,同时自动记录下该船舶目标的通行时间、通行画面、船舶类别等重要信息。

图6 简要描述了本文提出的模型在长江海事监管平台中的运行流程。基于该平台,我们对长江武汉段航道进行了海事监管实测。其中,表10 列出了实测当日(能见度较低)下午4 个小时内通行船舶的相关统计数据;图7 展示了部分实测效果。

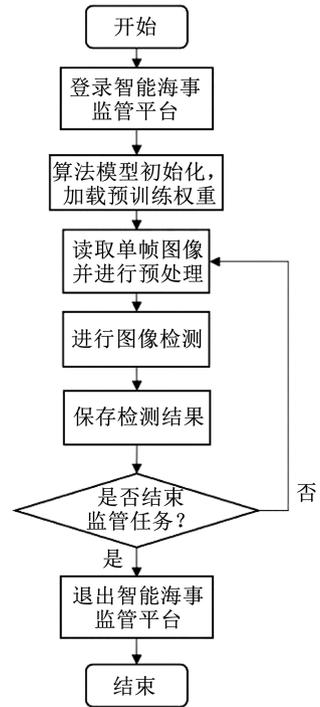


图6 长江海事监管平台中船舶检测流程

Fig. 6 Ship detection process of maritime supervision platform of Yangtze River

表10 实测通行船舶统计数据

Tab. 10 Statistics of measured ships

类别	数量
邮轮	5
集装箱船	24
散货船	47
其他船	19
总计	95

实测结果表明,本文提出的 MS-SOD 方法对于复杂水域中多尺度船舶目标具有较好的检测效果,且能满足海事智能监管任务的实际需求。



图 7 实测结果(长江武汉段航道)

Fig. 7 Experimental results of ship detection (Wuhan reach of Yangtze River)

5 结 论

为更好地服务于各类智慧海事应用,在当前基于深度学习目标检测方法的研究基础上,提出并实现了一个高效、精准的多尺度水上船舶目标视觉检测算法。经广泛的实验验证,该多尺度船舶目标检测算法在构建的较大规模水上船舶目标数据集上,各项性能指标均超过了目前主流的目标检测算法模型。例如,在 V_{mAP} 方面,所提方法相对于经典的 YOLOv4 算法,提高了 11.4%,其中各尺度目标检测精度均有不同幅度提升,对于中、小尺度的目标检测性能提升较为明显,分别为 6.0% 和 10.5%,且对船舶目标的类别判定准确度也有所提高。与此同时,该检测算法已被部署于长江海事智能监管平台,并进行了相关性能实测。此外,基于提出的算法模型,团队在 2020 年首届海洋目标智能感知国际挑战赛中荣获了一等奖。

但是,本文方法也存在着不足。例如,对于小尺度水上船舶目标的检测效果仍有改进空间,可进一步研究论证。

参考文献

- [1] 严新平, 柳晨光. 智能航运系统的发展现状与趋势[J]. 智能系统学报, 2016, 11(6): 807
YAN Xiping, LIU Chenguang. Review and prospect for intelligent waterway transportation system[J]. CAAI Transactions on Intelligent Systems, 2016, 11(6): 807
- [2] GIRSHICK R. Fast R-CNN [C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440
- [3] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904
- [4] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 7263
- [5] REDMON J, FARHADI A. YOLOv3: an incremental improvement[Z]. arXiv:1804.02767, 2018. DOI: 10.48550/arXiv.1804.02767
- [6] LIU Wei, ANGELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//European Conference on Computer Vision. Amsterdam: Springer International Publishing, 2016: 21
- [7] 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. Las Vegas: IEEE, 2016: 779
- [8] 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. Columbus: IEEE, 2014: 580
- [9] 张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报, 2017, 43(8): 1289
ZHANG Hui, WANG Kunfeng, WANG Feiyue. Advances and perspectives on applications of deep learning in visual object detection[J]. Acta Automatica Sinica, 2017, 43(8): 1289. DOI: 10.16383/j. aas. 2017. c160822
- [10] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 91
- [11] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [Z]. arXiv: 2004. 10934, 2020. DOI: 10.48550/arXiv.2004.10934
- [12] CHEN Zhijun, CHEN Depeng, ZHANG Yishi, et al. Deep learning for autonomous ship-oriented small ship detection [J]. Safety Science, 2020, 130: 104812
- [13] 王新立, 江福才, 宁方鑫, 等. 基于改进卷积神经网络的船舶目标检测[J]. 中国航海, 2018, 41(2): 41
WANG Xinli, JIANG Fucui, NING Fangxin, et al. Ship detection with improved convolutional neural network [J]. Navigation of China, 2018, 41(2): 41
- [14] 聂鑫, 刘文, 吴巍. 复杂场景下基于增强 YOLOv3 的船舶目标检测[J]. 计算机应用, 2020, 40(9): 2561

- NIE Xin, LIU Wen, WU Wei. Ship detection based on enhanced YOLOv3 under complex environments [J]. *Journal of Computer Applications*, 2020, 40(9): 2561
- [15] LIU Bo, WANG Shengzheng, XIE Z X, et al. Ship recognition and tracking system for intelligent ship based on deep learning framework [J]. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 2019, 13
- [16] SHAO Zhenfeng, WANG Linggang, WANG Zhongyuan, et al. Saliency-aware convolution neural network for ship detection in surveillance video [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 30(3): 781
- [17] KIM K, HONG S, CHOI B, et al. Probabilistic ship detection and classification using deep learning [J]. *Applied Sciences*, 2018, 8(6): 936
- [18] 潘慧, 段先华, 罗斌强. 多尺度特征 DCA 融合的海上船舶检测算法研究 [J]. *计算机工程与应用*, 2022, 58(4): 177
- PAN Hui, DUAN Xianhua, LUO Binqiang. Research on marine ship detection based on multi feature fusion interaction and DCA [J]. *Computer Engineering and Applications*, 2022, 58(4): 177
- [19] HAN Xu, ZHAO Lining, NING Yue, et al. ShipYOLO: an enhanced model for ship detection [J]. *Journal of Advanced Transportation*, 2021: 1
- [20] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision. Munich: Springer International Publishing, 2018: 3
- [21] WANG Chien-Yao, LIAO Hong-Yuan Mark, WU Yueh-Hua, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition workshops. Seattle: IEEE, 2020: 390
- [22] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2117
- [23] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759
- [24] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] // Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980
- [25] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-YOLOv4: scaling cross stage partial network [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13029
- [26] HU Jie, SHEN Li, SUN Gang. Squeeze-and-Excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Beijing: IEEE, 2018: 7132
- [27] ZHOU Xingyi, WANG Dequan, KRÄHENBÜHL P. Objects as points [Z]. arXiv: 1904. 07850, 2019. DOI: 10.48550/arXiv.1904.07850
- [28] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. FCOS: fully convolutional one-stage object detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9627
- [29] JOCHER G. YOLOv5 [CP/OL]. <https://github.com/ultralytics/yolov5>. 2020
- [30] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization [C] // Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 618

(编辑 苗秀芝)