

# 基于改进 YOLOv5 的船舶目标检测算法

宋策, 尹勇, 王鹏

(大连海事大学航海动态仿真和控制交通行业重点实验室, 辽宁大连, 116026)

**[摘要]** 针对船载视频感知系统存在的船舶目标检测精度不高和检测速度慢的问题, 提出一种改进的 YOLOv5 船舶目标检测方法。利用 SENet 以及 Confluence 边界框选取抑制技术对 Backbone 和边界框的选取方法进行改进, 以提高目标检测精度。通过 SeaShips 公开数据集对改进的算法进行训练及测试。结果显示: 算法的召回率为 98.3%; 精确率可达 88.5%; 检测速度达到 0.019 s/image。表明改进算法具有较高的检测精度, 检测效率可以达到实时。

**[关键词]** 船舶目标检测; 视频感知; 卷积神经网络; YOLO; 边界框选取

**[中图分类号]** U 675.92

## Ship Target Detection Algorithm Based on Improved YOLOv5

SONG Ce, YIN Yong, WANG Peng

(Key Laboratory of Marine Simulation & Control for Transportation Industry, Dalian Maritime University, Dalian 116026, China)

**Abstract:** Aiming at the problems of low detection accuracy and slow detection speed in shipborne video sensing systems, an improved YOLOv5 ship target detection method is proposed in this paper. In this method, SENet and Confluence bounding box selection suppression technology is used to improve the Backbone and bounding box selection method, so as to improve the target detection accuracy. By using Seaships public datasets to train and test the improved algorithm, the recall rate of the proposed algorithm is 98.3%, the precision is up to 88.5%, and the detection speed is up to 0.019 s/image. Experimental results show that the improved algorithm proposed in this paper has high detection accuracy and real-time detection efficiency.

**Keywords:** ship target detection; video sensing; convolutional neural network; YOLO; bounding box selection

## 0 引言

近年来, 视频场景中的目标检测成为计算机视觉领域的研究热点。基于卷积神经网络的目标检测模型能够检测一幅图像中的多个目标, 并判断目标的类别和位置。而在智能船技术发展过程中, 将目标检测技术应用于对其他船舶信息的感知一直是研究者关注的问题之一。随着船舶工业的发展, 海上目标形式更加多样, 在智能船视频系统中需要对各类船舶目标实施更为迅速、精准、直观的检测, 由此获取其他船舶的有关信息。根据不同的图像信息国内外研究者们提出了多种算法。杨一平等<sup>[1]</sup>通

**[收稿日期]** 2021-11-07

**[基金项目]** 国家重点研发计划项目(2019YFB1600602); 中央高校基本科研业务项目(3132019312); 工信部高技术船舶专项(工装函[2018]473号)

**[作者简介]** 宋策(1997—), 男, 硕士生, 从事船载视频感知方向研究。通信作者: 尹勇(1969—), 男, 教授, 博士生导师, 从事航海仿真、航海智能方向研究。E-mail: bushyin@163.com

过传统的多视觉融合方法, 使用局部熵以及连通域对船舶目标进行检测; 石文豪等<sup>[2]</sup>利用不同的结构元素提取船舶边缘特征, 从而提取船舶目标的细节; 杨雪锋<sup>[3]</sup>提出了一种基于 FAST 角点检测、分类、以及区域生长的船舶目标检测方法, 利用目标角点与海浪角点在强度及密集度上的不同对两者进行区分; Arshad 等<sup>[4]</sup>对船舶背景图像首先使用形态学操作进行处理, 然后对船舶的边缘使用 Sobel 算子进行检测, 进而与其背景进行辨别, 但对复杂纹理的情况处理效果不好, 并且会产生较多噪点。

随着卷积神经网络在目标检测领域的逐步完善, 许多研究者采用基于深度学习的算法来实现船舶目标的检测。Chang 等<sup>[5]</sup>基于 SAR 图像评估了 YOLOv2 深度网络架构在各种场景下检测船只的性能, 得出 YOLOv2 方法在接近实时的情况下, 特别是在复杂的情况下, 在精度和性能方面都优于现有技术, 但检测速度较慢。赵文强等<sup>[6]</sup>通过在 YOLOv3 原型网络基础上进行网络尺度的扩展, 利用 K-means 聚类算法得到更为精细化尺度下的先验框, 在目标检测准确率、目标位置预测、模型检测效率方面有显著的表现。Zou<sup>[7]</sup>通过对 SSD 法进行改进, 采用轻量级的 MobilenetV2 网络结构, 满足了当前船舶监督中对船舶目标实施检测的要求。Tanveer 等<sup>[8]</sup>使用了 Faster R-CNN 在 SAR 图像中对船舶进行了检测及计数, 因为卷积层的简洁性提升了计算速度, 具有较高的检测精确度。对于复杂海况下和近岸船舶目标, 基于图像处理的传统船舶目标检测算法泛化性较差, 检测难度大。而一些流行的卷积神经网络算法在船舶检测领域精确率、召回率等还有待提升, 检测速度也难以达到实时检测的要求。为了建立高精度船载视频感知系统, 本文提出一种基于改进 YOLOv5 的船舶目标检测算法, 通过引入主干网络 (Backbone) 增加注意力机制 SENet<sup>[9]</sup>, 对边界框选取进行抑制, 使用 Confluence 替代非极大值抑制 (non-maximum suppression, NMS), 摆脱了边界框对置信度得分以及交并比 (intersection over union, IoU) 的依赖等技术对 YOLOv5 算法进行改进, 以期提高检测精度。

## 1 基于 YOLOv5 的船舶目标检测算法

YOLOv5 船舶目标检测算法首先通过主干神经网络提取输入图像在三个不同尺度上的特征, 通过上采样层将深层特征与浅层特征融合, 最后在三个尺度上进行检测, 得到目标的分类结果以及位置信息。

### 1.1 YOLOv5 网络结构

YOLOv5 的网络结构分为 Backbone, 颈部提取特征 (neck) 和输出三部分, 如图 1 所示。

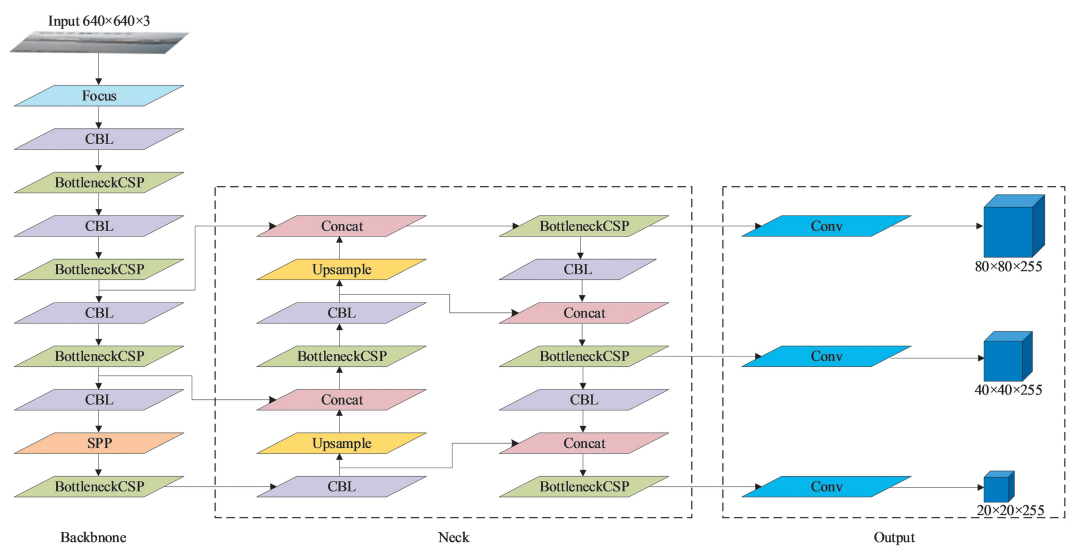


图 1 YOLOv5 网络结构  
Fig.1 YOLOv5 network architecture

在 Backbone 中，分辨率为  $640 \times 640 \times 3$  的输入图像通过 Focus 结构，经过切片运算，首先得到  $320 \times 320 \times 12$  特征图，然后经过 32 个卷积核的卷积运算，最后得到  $320 \times 320 \times 32$  特征图。CBL 模块是一个基本的卷积模块，CBL 模块表示卷积层（Conv2D）+ 归一化层（BatchNormal）+ LeakyReLU。瓶颈层（BottleneckCSP）模块主要对特征图进行特征提取，从图像中提取丰富的信息，与其他大型卷积神经网络相比，BottleneckCSP 结构可以减少卷积神经网络优化过程中的梯度信息重复，其参数量占整个网络参数量的大部分，通过调整 BottleneckCSP 模块的宽度和深度，可以得到 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 四种不同参数的型号。空间金字塔（spatial pyramid pooling, SPP）模块主要增加网络的感受野，获取不同尺度的特征。

YOLOv5 还在特征金字塔网络（feature pyramid networks, FPN）结构的基础上增加了一个自底向上的特征金字塔结构。通过这种组合操作，FPN 结构自上而下传递鲁棒语义特征，特征金字塔自下而上传递鲁棒定位特征<sup>[10]</sup>。聚合不同特征层的输出，提高网络的检测能力，以应对不同尺度目标。在图的最后，输出分类结果和对象坐标<sup>[11]</sup>。

1.2 SENet 网络结构

本文的网络结构在 YOLOv5 的基础上进行了一定改进，调整了部分参数，并在 Backbone 部分的 BottleneckCSP 模块之后引入了压缩激励结构 SENet。注意力机制的引入，对于输入的每条特征通道有着不同的重要性，通过全连接层（FC）进行学习，从而得到来自不同特征通道空间维度为  $H \times W \times C$  的权重。SENet 增加了模型的复杂度和计算负担，但这样关键特征的提取能力得以提升，并能有效地抑制无用的特征。图 2 为修改后的 Backbone。

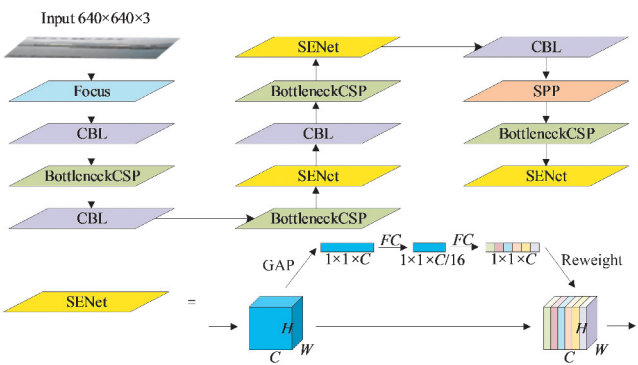


图 2 修改后的 Backbone  
Fig.2 The modified Backbone

2 基于 Confluence 的输出端

在检测过程中，由于需要检测不同种类的船舶，且船舶之间存在结构及大小的差异（如散货船、集装箱船、渔船、邮轮，等），因此需要使用 YOLOv5 从多个尺度进行检测。Backbone 中深卷积层在分类任务中能够获取图像深层次的细节，而浅卷积层在对象位置回归任务中更优，对外观细节明显的对象进行检测时较为适用，因此在预测时，会先将主干网络产生的三个尺度的特征图通过 FPN 和像素聚合网络（pixel aggregation network, PAN）结构通过拼接（concat）层操作融合起来，这样能更好地获得目标的位置信息、分类信息和不同尺度物体的特征。

YOLOv5 在输入图像的尺寸为  $640 \times 640$  时，输入检测网络特征图的大小为  $20 \times 20$ ， $40 \times 40$ ， $80 \times 80$ 。较深的  $20 \times 20$  特征图负责预测外观特征较为明显的目标，而较浅  $80 \times 80$  的特征图负责预测外观特征较为模糊的目标。随后采用了 Confluence 方法<sup>[12]</sup>，替代 YOLOv5 中原本所采用的加权 NMS，该方法不依赖于置信度得分来选择最佳边界框，也不依赖 IoU 来消除误检。其通过使用加权曼哈顿距离来选择最接近群集中其他所有边界框的边界框，并删除具有高度融合的近邻框。

Confluence 是一个两阶段的算法，它保留了最优边界框，并消除了假阳性。第一阶段使用置信加权曼哈顿距离对相邻边界框进行计算，来评估边界框的一致性。第二阶段涉及移除所有与保留的边界框在同一边界框簇中的其他框，首先通过加权曼哈顿距离  $M$  衡量两个边界框的重叠程度，得到  $O$  值，如图 3 所示，其计算公式为

$$O_{(u,v,m,n)} = M_{(u,v)} + M_{(m,n)} = |x_1 - p_1| + |x_2 - p_2| + |y_1 - p_1| + |y_2 - p_2|。$$
 (1)

随后通过考虑置信分数  $c$  和与之竞争的边界框的  $O$  值来评估给定边界框的最优性，利用公式

$W_{(u,v,m,n)} = O_{(u,v,m,n)} / c$  计算其置信度得分, 可以得到加权接近性  $W$ 。

### 3 实验结果及分析

#### 3.1 实验数据集

随着计算机视觉的迅速发展, 目前在可见光图像的船舶目标检测领域中公开的并被广泛使用的数据集逐渐增多, 其中包括新加坡海事数据集、VAIS 数据集以及 Seaships 数据集等。本文使用 Seaships 数据集<sup>[13]</sup>公开的部分数据集, 其中 Seaships 数据集包括 7000 张图片, 分辨率为  $1920 \times 1080$ ; 使用 VOC 数据集的格式, 其中包括散货船、矿砂船、集装箱船、渔船等不同船型。将 Seaships 中 5040 张图像用于训练, 700 张图像用于测试。

#### 3.2 实验平台训练设置

实验平台在 Windows10 系统上进行, 开源深度学习框架 Pytorch1.6 通过 CUDA 对 GPU 加速进行训练, 实验平台使用了一块 NVIDIA RTX2060 显卡, 显存为 6 GB。为了减少过拟合的风险, 使用广义交并比 (generalized intersection over union, GIoU) 作为损失函数, 将网络的 Backbone 部分作为分类网络, 在 VOC 训练集上进行预训练, 得到初始的权重, 并使用 Adam 优化算法, 通过动量和自适应学习率来加快收敛速度。此外, 训练集采用 Mosaic 数据增强的策略, 随机使用 4 张图片, 随机缩放, 再随机分布进行拼接, 大大丰富了检测数据集, 特别是随机缩放增加了很多小目标, 让网络的鲁棒性更好。

#### 3.3 实验结果评价指标

为了对检测结果进行分析评价, 需要引入如下指标:  $N_p$  为被正确检测的各类船舶数量;  $N_D$  为被检测为各类船舶的数量;  $N_T$  为被标记为各类船舶的数量<sup>[14]</sup>。当本文所使用的 Confluence 中检测框与正确目标的曼哈顿距离阈值超参数设置为 0.45, 当超过阈值时认为正确检测出了目标。

预测的精确率  $P = N_p / N_D$ ; 召回率  $R = N_p / N_D$ ; 考虑  $P$  以及  $R$  的综合指标  $F_1 = 2 \times P \times R / (P + R)$ 。

#### 3.4 实验结果分析

##### 3.4.1 不同网络参数下 YOLOv5 的检测性能比较

比较 YOLOv5 在网络宽度与深度不同时网络的检测精度、模型参数的大小, 结果如表 1 所示。

表 1 不同网络模型大小的 YOLOv5 网络性能						
Tab.1 YOLOv5 network performance with different network model sizes						
模型大小 Model size	$N_T$	$N_D$	$N_p$	精确率 Precision/%	召回率 Recall/%	模型参数大小 Parameter/MB
YOLOv5s	922	1051	901	85.7	97.8	14.1
YOLOv5m	922	1028	904	87.9	98.1	41.3
YOLOv5l	922	1035	907	87.6	98.4	90.8
YOLOv5x	922	1085	906	83.5	98.3	169.0

从表 1 中可以看出, 增加宽度与深度在一定程度上对网络的性能带来了部分提升, 但提升结果并

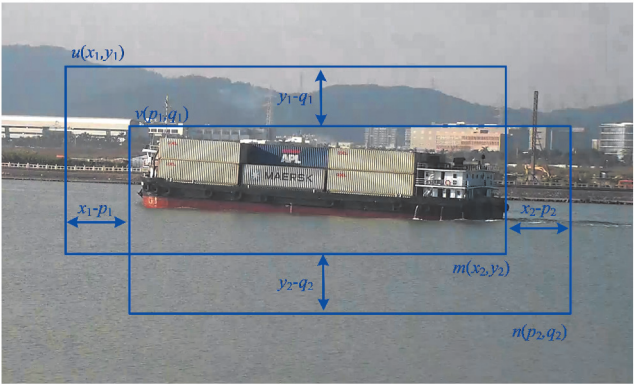


图 3 坐标为  $x_1, y_1, x_2, y_2$  的检测框与另一个坐标为  $p_1, q_1, p_2, q_2$  的检测框的重叠度计算

Fig.3 Overlap calculation of a detection with coordinates  $x_1, y_1, x_2, y_2$  with another detection represented by  $p_1, q_1, p_2, q_2$

不明显，而且需要以牺牲模型大小为代价，因此最终选择了参数量较少，更容易在船载设备部署的 YOLOv5s 轻量化版本，并在此基础上对模型进行了改进。

3.4.2 与不同深度学习目标检测算法比较

为了验证本文算法的性能，在相同数据集上与其他检测算法进行了比较，结果如表 2 所示。从表 2 可以看出，本文改进的 YOLOv5s 网络采用了 SENet 作为注意力机制添加到 Backbone 中，提升了原版 YOLOv5s 目标检测精确率，随后使用了 Confluence 检测框筛选方法，使网络不依赖 IoU 的情况下从三个尺度进行检测，同时小幅度提升了检测精确率，模型参数大小也基本不变，具有轻量化优点，硬件实现较为容易。检测能力已达到较高水平，且相比 SSD 网络<sup>[16]</sup>拥有更好的检测性能。SSD300 因为输入图像分辨率并不高，并且多尺度检测的方法并不占据优势，所以检测精度并没有很好的表现。但由于主干网络使用了 VGG16，所以模型参数量也较大。本文相比原版 YOLOv5s，虽然牺牲了一部分速度，但提高了检测精度，达到与两阶段的目标检测算法 Faster R - CNN<sup>[17]</sup> 相接近的水准，基本能实现目标实时检测。图 4 为不同检测算法的部分检测结果。其中：1) 表示识别标签为 Ore carrier；2) 表示识别标签为 General cargo ship 和 Bulk cargo carrier；3) 识别标签为 General cargo ship 和 Fishing ship。从图 4 可看出，改进后的 YOLOv5s 网络在船舶之间距离较近时，边界框对船舶轮廓的识别更为准确，没有同其他检测模型一样出现边界框交叉的情况，同时在检测渔船等目标时，没有出现误检或漏检的现象。

表 2 不同深度学习检测算法比较  
Tab.2 Comparison of different deep learning detection algorithms

网络 Network	$N_T$	$N_D$	$N_P$	精确率 Precision /%	召回率 Recall /%	$F_1$	模型参数 大小 Parameter /MB	检测速度 Detection speed/ (s · image <sup>-1</sup> )
YOLOv5s	922	1051	901	85.7	97.8	0.914	14.1	0.015
YOLOv5s + SENet	922	1040	906	87.1	98.3	0.924	14.1	0.019
YOLOv5sS + SENet + Confluence	922	1024	906	88.5	98.3	0.931	14.4	0.021
YOLOv4 <sup>[15]</sup>	922	1036	899	86.8	97.5	0.918	244.0	0.095
SSD300 <sup>[16]</sup>	922	1066	904	84.8	98.1	0.909	93.1	0.086
Faster R - CNN <sup>[17]</sup>	922	1035	907	87.6	98.4	0.927	108.0	0.230

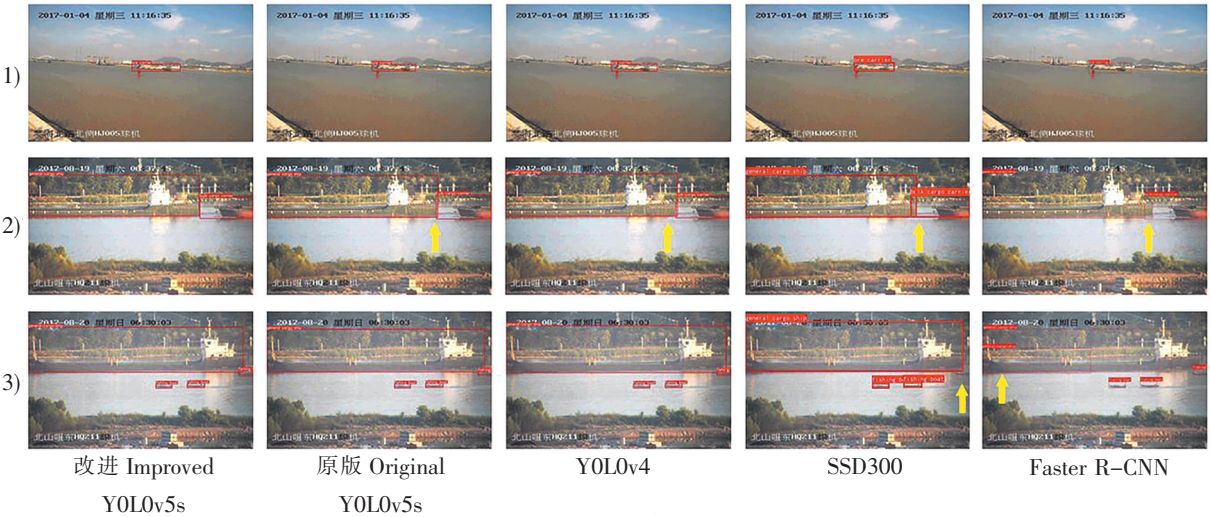


图 4 不同检测算法的部分检测结果  
Fig.4 Part of the detection results of different detection algorithms

## 4 结论

本文提出了一种新的卷积神经网络结构, 用于检测可见光视频影像中的船舶目标。该网络结构主要基于 YOLOv 融合了 SENet 压缩激励结构, 对输出端使用 Confluence 方法进行改进, 使网络能够不依赖 IoU 也可以检测出多尺度的目标特征。在可见光视频影像中的船舶目标检测任务中, 本文算法相比同类检测算法有着较高的准确度, 并且模型参数量较小, 便于在车载视频设备上实现。

## [ 参考文献 ]

- [1] 杨一平, 柳圣军, 周超, 等. 多视觉信息融合的内河航道智能监控系统 [J]. 中国图象图形学报, 2018, 13(8): 1608-1616.
- [2] 石文豪, 安博文. 基于多结构形态学的港口船舶检测方法 [J]. 计算机系统应用, 2016, 25(10): 283-287.
- [3] 杨雪锋. 基于船载红外视频分析的海盗船艇智能辨识方法研究 [D]. 大连: 大连海事大学, 2016.
- [4] ARSHAD N, MOON K S, KIM J N. Multiple ship detection and tracking using background registration and morphological operations [M]. Signal Processing and Multimedia. Berlin: Springer, 2010.
- [5] CHANG Y L, AMARE ANAGAW, LENA CHANG, et al. Ship detection based on YOLOv2 for SAR imagery [J]. Remote Sensing, 2019, 11(7): 1-14.
- [6] 赵文强, 孙巍. 基于 S4-YOLO 的海上目标检测识别方法 [J]. 光学与光电技术, 2020, 18(4): 38-46.
- [7] ZOU Y Y, ZHAO L N, QIN S X, et al. Ship target detection and identification based on SSD-MobilenetV2 [C] // IEEE 5th Technology and Mechatronics Engineering Conference. Chongqing: IEEE, 2020. 1676-1680.
- [8] HASHIR TANVEER, TIMO BALZ, BAHAA MOHAMDI. Using convolutional neural network (CNN) approach for ship detection in sentinel-1 SAR imagery [C] // 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR). Xiamen: [s. n.], 2019.
- [9] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 7132-7141.
- [10] 李宇, 刘孔玲, 黄蒲荪. 基于深度学习的布匹疵点检测方法 [J]. 毛纺科技, 2021, 49(04): 98-103.
- [11] ZHOU F B, ZHAO H L, NIE Z. Safety helmet detection based on YOLOv5 [C] // IEEE International Conference on Power Electronics. Computer Applications (ICPECA). IEEE. 2021: 6-11.
- [12] SHEPLEY A J, FAIZON G, KWAN P. Confluence: a robust non-IoU alternative to non-maxima suppression in object detection. [EB/OL]. (2020-12-01) [2021-07-12]. <https://arxiv.org/pdf/2012.00257.pdf>.
- [13] SHAO Z F, WU W J, WANG Z Y, et al. Sea ships: a large-scale precisely annotated dataset for ship detection [J]. IEEE Transactions on Multimedia, 2018, 20(10): 2593-2604.
- [14] 沈丰毅, 王琴, 蒋剑飞. 基于改进卷积神经网络舰船的实时目标检测算法 [J]. 计算机应用研究, 2020, 37(S1): 316-319.
- [15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [EB/OL] [2020-04-23]. <https://arxiv.org/pdf/2004.10934.pdf>.
- [16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] // Proc of European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [17] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2015, 39(6): 1137-1149.

(责任编辑 陈 敏 英文审校 周云龙)