



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象学报 中国图形学报

2022
07
VOL.27

ISSN1006-8961
CN11-3758/TB



舰船目标检测 P2078



第27卷第7期（总第315期）
2022年7月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



深度学习背景下视觉显著性
物体检测综述(第2112页)



提取全局语义信息的场景图
生成算法(第2214页)



融合弱监督目标定位的细粒
度小样本学习(第2226页)

学者观点

网络监督数据下的细粒度图像识别综述

魏秀参, 许玉燕, 杨健 2057

综述

海事监控视频舰船目标检测研究现状与展望

叶晨, 谔天洋, 肖漓, 陆海, 杨群慧 2078

深度学习行人检测方法综述

罗艳, 张重阳, 田永鸿, 郭捷, 孙军 2094

深度学习背景下视觉显著性物体检测综述

王自全, 张永生, 于英, 闵杰, 田浩 2112

图像分析和识别

图像增强对显著性目标检测的影响研究

郭继昌, 岳惠惠, 张怡, 刘迪, 刘晓雯, 郑司达 2129

混合高斯变分自编码器的聚类网络

陈华华, 陈哲, 郭春生, 应娜, 叶学义 2148

基于半监督对抗学习的图像语义分割

李志欣, 张佳, 吴璟莉, 马慧芳 2157

面向大姿态人脸识别的正面化形变场学习

胡蓝青, 阚美娜, 山世光, 陈熙霖 2171

融合时空域特征的人脸表情识别

陈拓, 邢帅, 杨文武, 金剑秋 2185

非局部注意力双分支网络的跨模态赤足足迹检索

鲍文霞, 茅丽丽, 王年, 唐俊, 杨先军, 张艳 2199

提取全局语义信息的场景图生成算法

段静雯, 闵卫东, 杨子元, 张煜, 陈鑫浩, 杨升宝 2214

融合弱监督目标定位的细粒度小样本学习

贺小箭, 林金福 2226

结合双注意力机制的道路裂缝检测

张志华, 温亚楠, 慕号伟, 杜小平 2240

融合神经网络的布料碰撞检测算法

靳雁霞, 马博, 贾瑶, 陈治旭, 芦烨 2251

双分支特征融合网络的步态识别算法

徐硕, 郑锋, 唐俊, 鲍文霞 2263

图像理解和计算机视觉

问题引导的空间关系图推理视觉问答模型

兰红, 张蒲芬 2274

结合扰动约束的低感知性对抗样本生成方法

王杨, 曹铁勇, 杨吉斌, 郑云飞, 方正, 邓小桐 2287

计算机图形学

动态书法墨迹的可回溯感评测

律睿, 梅莉琳, 吴跃峰, 晏涛 2300

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Review of deep learning based salient object detection(P2112)



Global semantic information extraction based scene graph generation algorithm(P2214)



Weakly-supervised object localization based fine-grained few-shot learning(P2226)

Scholar View

Review of webly-supervised fine-grained image recognition

Wei Xiushen, Xu Yuyan, Yang Jian 2057

Review

Maritime surveillance videos based ships detection algorithms: a survey

Ye Chen, Lu Tianyang, Xiao Yuhao, Lu Hai, Yang Qunhui 2078

An overview of deep learning based pedestrian detection algorithms

Luo Yan, Zhang Chongyang, Tian Yonghong, Guo Jie, Sun Jun 2094

Review of deep learning based salient object detection

Wang Ziquan, Zhang Yongsheng, Yu Ying, Min Jie, Tian Hao 2112

Image Analysis and Recognition

The analysis of image enhancement on salient object detection

Guo Jichang, Yue Huihui, Zhang Yi, Liu Di, Liu Xiaowen, Zheng Sida 2129

A Gaussian mixture variational autoencoder based clustering network

Chen Huahua, Chen Zhe, Guo Chunsheng, Ying Na, Ye Xueyi 2148

Semi-supervised adversarial learning based semantic image segmentation

Li Zhixin, Zhang Jia, Wu Jingli, Ma Huifang 2157

Large pose face recognition with morphing field learning

Hu Lanqing, Kan Meina, Shan Shiguang, Chen Xilin 2171

Spatio-temporal features based human facial expression recognition

Chen Tuo, Xing Shuai, Yang Wenwu, Jin Jianqiu 2185

Non-local attention dual-branch network based cross-modal barefoot footprint retrieval

Bao Wenxia, Mao Lili, Wang Nian, Tang Jun, Yang Xianjun, Zhang Yan 2199

Global semantic information extraction based scene graph generation algorithm

Duan Jingwen, Min Weidong, Yang Ziyuan, Zhang Yu, Chen Xinhao, Yang Shengbao 2214

Weakly-supervised object localization based fine-grained few-shot learning

He Xiaojian, Lin Jinfu 2226

Dual attention mechanism based pavement crack detection

Zhang Zhihua, Wen Yanan, Mu Haowei, Du Xiaoping 2240

Neural network fusion based cloth collision detection algorithm

Jin Yanxia, Ma Bo, Jia Yao, Chen Zhixu, Lu Ye 2251

Dual branch feature fusion network based gait recognition algorithm

Xu Shuo, Zheng Feng, Tang Jun, Bao Wenxia 2263

Image Understanding and Computer Vision

Question-guided spatial relation graph reasoning model for visual question answering

Lan Hong, Zhang Pufen 2274

A perturbation constraint related weak perceptual adversarial example generation method

Wang Yang, Cao Tieyong, Yang Jibin, Zheng Yunfei, Fang Zheng, Deng Xiaotong 2287

Computer Graphics

Traceability evaluation of flowing calligraphy strokes

Lyu Ruimin, Mei Lilin, Ze Yuefeng, Yan Tao 2300

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2022)07-2078-16

论文引用格式: Ye C, Lu T Y, Xiao Y H, Lu H and Yang Q H. 2022. Maritime surveillance videos based ships detection algorithms: a survey. Journal of Image and Graphics, 27(07): 2078-2093 (叶晨, 逯天洋, 肖涌灏, 陆海, 杨群慧. 2022. 海事监控视频舰船目标检测研究现状与展望. 中国图象图形学报, 27(07): 2078-2093) [DOI: 10.11834/jig.200674]

海事监控视频舰船目标检测研究现状与展望

叶晨^{1,2}, 逯天洋¹, 肖涌灏¹, 陆海^{3*}, 杨群慧^{3,4}

1. 同济大学电子与信息工程学院, 上海 201804; 2. 同济大学嵌入式系统与服务计算教育部重点实验室, 上海 201804;
3. 同济大学国家海底科学观测系统项目办公室, 上海 200092; 4. 同济大学海洋地质国家重点实验室, 上海 200092

摘要: 舰船目标检测是海域监控、港口流量统计、舰船身份识别以及行为分析与取证等智能海事应用的基石。随着我国海洋强国建设的推进, 智慧航运和智慧海洋工程迅速发展, 对通过海事监控视频开展有效的舰船目标检测识别以确保航运和海洋工程安全的需求日益紧迫。本文针对基于海事监控视频的舰船目标检测任务, 回顾了舰船目标检测数据集及性能评价指标、基于传统机器学习和基于卷积神经网络的深度学习的目标检测方法等方面的国内外研究现状, 分析了海洋环境中舰船目标检测任务面临的舰船目标尺度的多样性、舰船类别的多样性、海洋气象的复杂性、水面的动态性、相机的运动性和图像的低质量等技术难点, 并通过实验验证, 在多尺度特征融合、数据增广和能耗降低等方面提出了舰船目标检测的优化方法; 同时, 结合前人研究指出舰船目标检测数据集的发展应关注分类粒度的适宜性、标注的一致性和数据集的易扩充性, 应加强对多尺度目标(尤其是小型目标)检测的模型结构的研究, 为进一步提升舰船目标检测任务的综合性能, 促进舰船目标检测技术的应用提供了新的思路。

关键词: 舰船目标检测; 海事监控数据集; 小目标检测; 数据增广; 卷积神经网络性能优化

Maritime surveillance videos based ships detection algorithms: a survey

Ye Chen^{1,2}, Lu Tianyang¹, Xiao Yuhao¹, Lu Hai^{3*}, Yang Qunhui^{3,4}

1. College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China;
2. The Key Laboratory of Embedded System and Service Computing (Ministry of Education), Tongji University, Shanghai 201804, China;
3. China National Scientific Seafloor Observatory, Tongji University, Shanghai 200092, China;
4. State Key Laboratory of Marine Geology, Tongji University, Shanghai 200092, China

Abstract: Maritime surveillance videos based object detection methods aims to meet the quick response requirements through an effective ship detection and recognition system against the backdrop of smart ocean technology. Our research has focused on this aspects as mentioned below: 1) to summarize current approaches and datasets and discuss the challenging issues of them; 2) to analyze the features and the challenges of maritime surveillance ship detection; 3) to clarify the credibility of their accuracy and efficiency and demonstrate our research potential further. In the first phase, we summarize existing ship detection algorithms based on maritime surveillance videos, introduce the common ship detection datasets and

收稿日期: 2020-11-24; 修回日期: 2021-04-22; 预印本日期: 2021-04-29

* 通信作者: 陆海 lh@tongji.edu.cn

基金项目: 国家自然科学基金项目(U1764261); 国家重点研发计划资助(2018YFC1405803); 中央高校基本科研业务费专项资金资助(22120190200); 上海市科技创新资助项目(2019-jmrh1-kj15)

Supported by: National Natural Science Foundation of China (U1764261); National Key R&D Program of China (2018YFC1405803); Fundamental Research Funds for the Central Universities (22120190200); Shanghai Science and Technology Innovation Funding Project (2019-jmrh1-kj15)

ship detection methods available, and some evaluation metrics followed for ship detection tasks. Customized attention is yielded to the interconnected results between traditional computer vision algorithms for ship identification, which mainly consist of modules such as horizon detection, background subtraction and foreground extraction, and some deep learning methods based on fast region convolutional neural network (Fast R-CNN), single shot multibox detector (SSD) and you only look once (YOLO). It can be sorted out that although mean average precision (mAP) metric remains recognized index to measure the performance of models, its effectiveness issue is still discussed in terms of ship detection tasks and present novel metrics, including bottom edge proximity (BEP), n-multiple object detection precision (N-MODP) and n-multiple object detection accuracy (N-MODA). Current datasets are capable to detect vessels motion via deep learning models. But, the accuracy and robustness of training are required to be improved greatly due to extreme weather condition and light variation or inconsistent labels. In the second phase, we evaluate the features and challenges for ship detection. The difference lies between ship detection and regular object detection. For example, a coastline platform or a ship sensor have very large visible ranges and leading to a big scale variability. In addition, it is challenged to design a set of models adapt to various image domain scenarios derived from extreme marine weather conditions. Photographic system has to withstand exposure to extremes of temperature, high vibration levels, humidity and chemicals as well. The harsh environment combined with noise pollution and limited network bandwidth can cause the loss of image quality and make uncertainty with information loss for the models. In the third phase, we improve the accuracy and efficiency of ship detection algorithms and evaluate some common methods for ship detection technology on the three aspects as following: 1) multi-scale feature fusion: we carry out convolutional neural network (CNN) models manipulation based on different input scales and backbones. Some of the object detection models are degraded when facing large variations result from large field of view among ship objects during voyage. It is suggested that input scale determines the upper bound of accuracy of CNN models, and CNN models or backbones which are specially designed for multi-scale detection tasks narrow the gap of accuracy between different input scales. 2) Data augmentation: Waves and wind induce pitch and roll rotations on the sea, which is demanding for ship detection. Moreover, weather change and day-night brightness variation mean the image data shall be in multiple domains, which requires the detection models to be robust to images from different domains, or even from domains that have not been included in training samples. In light of the above-mentioned variables and marine based camera motion results, we evaluate the performance improvement when data augmentation is applied to image translation and rotation. We also adjust photo brightness and use Gaussian blur to simulate the blur caused by water condensed on cameras. Almost 5% increase is observed in mAP, verifying the robustness of data augmentation in ship detection. Other effective approaches include domain transfer based on generative adversarial network (GAN) or domain-independent models derived of multi-domain object detection tasks. 3) Light-weighted models and energy optimization: Computing complexity is constrained of semantic constraints like the horizon; Common object detection optimization is used to lower computation load, including light-weighted backbones like MobileNet and ShuffleNet. We calculate the parameter quantity and computing operation quantity of object detection models as well as each accuracy of them. We recommended that further studies should be considered on the following aspects: 1) to develop new datasets or existing datasets improvement based on sufficient coverage of possible conditions, high-quality annotations, precise classification and easier extension, respectively; 2) to decrease arithmetic operations and energy consumption in object detection models; 3) to strengthen multi-scale target detection modeling; 4) to enhance data fusion between object detection in multi-sensors images and the semantic ability of single image or multiple images interpretation.

Key words: ship object detection; maritime surveillance dataset; small object detection; data augmentation; performance optimization for convolutional neural network

0 引 言

人类进行海洋活动离不开舰船作为载体和工

具,在特定环境中对舰船目标进行自动化检测是港口流量统计、碰撞回避等领域的关键技术。随着海洋探测技术的发展,人类对海洋的开发利用已经从海面走向海底,特别是在 21 世纪初,海底长期观测

组网技术出现。海底观测网可将一系列的海洋观测仪器布放到海底,通过海底光电缆与陆基信息处理设备互联而成为开放式的海洋综合观测系统,具备水下大功率远程供电、大规模数据采集和信息传输能力,可实现对海底地壳深部、海底界面到海水水体及海面的大范围、全天候、综合性、长期、连续和实时的高分辨率和高精度的观测。海底观测网是继地面与海面观测、空中遥测遥感之后,人类建立的第3种地球科学观测平台,将成为未来海洋探测与研究的主要方式。然而,大量仪器设备布放到海底面临的首要问题是人类渔业活动对海底设施的破坏,尤其是在我国渔业活动频繁的东海海域。据报道,我国国际海光缆故障的90%均是受到人类渔业活动破坏的影响(叶银灿等,2015)。海底光电缆、海底设备等一旦受损或断裂,会造成系统整个工程电力、通信中断,且维修周期长、维修难度大,会带来巨大的损失。为避免舰船在海底设施附近进行抛锚、捕捞等活动,目前主要采取对海面舰船目标进行舰载自动识别系统(automatic identification system, AIS)位置监控,但是这种监控对那些刻意关闭AIS进行走私和偷捕的船舶无法起到监控作用,并且难以对发生破坏活动的舰船进行取证和执法。因此,在海底设施附近布设安装视频传感摄像头的浮标,并对视频图像开展及时的舰船目标识别,成为监控周边船舶、保护海底工程设施的一项非常有效的手段。

随着我国海洋强国建设的推进,智慧航运和智慧海洋工程迅速发展,对通过海事监控视频开展有效的舰船目标检测识别以确保航运和海洋工程安全的需求日益紧迫。本文针对基于海事监控视频的舰船目标检测,回顾了检测方法的国内外研究现状,阐

明了舰船目标检测任务中的技术难点,通过实验证明了针对性的预处理和数据增广的必要性,从适应目标尺度的多样性和提高模型的性能并降低模型对设备运算量的要求等方面展开调研,并对相应模型进行了实验评估,为促进舰船目标检测技术的应用提供了思路。

1 海事监控视频舰船目标检测数据集与性能评价指标

1.1 舰船目标检测公开数据集

数据集在目标检测等计算机视觉研究中扮演着重要角色,用于模型训练、过程评估和性能评价等环节。在基于数据驱动的深度学习方法中,质量高或者相关性高的数据集是训练高性能模型的先决条件。当前,海面舰船目标图像主要有两种,一是基于机载、星载传感器获取的遥感图像;二是基于陆基、船基和浮标安装的传感器获取的监控图像。遥感图像通常采用雷达和红外线传感器采集,其中合成孔径雷达(synthetic aperture radar, SAR)在各种遥感系统中广泛使用。监控图像多来源于可见光(visible spectrum, VI 或 VIS)或红外线传感器。红外线传感器又分为红外线传感器(infrared, IR)和近红外传感器(near-infrared, NI)。图1为陆基可见光传感器、船基近红外传感器、浮筒可见光传感器和星载SAR传感器采集的图像,分别来自SeaShips数据集(Shao等,2018)、新加坡海事数据集(Singapore-marine-dataset, SMD)(Prasad等,2017)、Buoy数据集(Fefilat'yev等,2006;Kristan等,2014;Patino等,2016;Ribeiro等,2019)和Sentinel-1数据集(Kang等,2017)。



图1 不同数据集中的海面舰船图像

Fig. 1 Maritime images of different datasets((a) SeaShips; (b) SMD; (c) Buoy; (d) Sentinel-1)

海事目标检测、分类与跟踪数据集(maritime detection, classification and tracking data set, MarDCT)(Bloisi等,2015)是2017年前舰船目标检测领域的

少数海事监控视频数据集之一,该数据集在包含目标检测和跟踪标注的MarDT数据集的基础上补充了船只分类标签,含有在威尼斯采集的可见光

与远红外视频,包括固定摄像头和云台全方位变焦镜头(pan/tilt/zoom cameras, PTZ Cameras)两种视角的数据。MarDCT数据集在舰船目标检测任务的研究初期起到了关键作用,其局限性在于缺少船只重叠、海岸线和陆地相互交叉的样例。

为了提升对舰船目标检测方法的性能评估,Prasad等人(2017)公开了SMD数据集。Moosbauer等人(2019)分析了SMD数据集的数据分布情况,并提出了一套较完善的训练集、验证集与测试集划分方法。SeaShips数据集(Shao等,2018)由珠海横琴的海岸陆基摄像头采集,包含6种船只类别的图像数据,在尺度、视角、光照、背景动态和目标相互遮挡等方面进行了分析和针对性的优化。近年来,由应用威胁检测与启发式对策技术的智能海盗回避项目(intelligent piracy avoidance using threat detection and countermeasure heuristics, IPATCH)收集的IPATCH(Patino等,2016)等数据集也应用于面

向海面目标的目标检测、目标跟踪和行为分析等领域研究。

基于海面浮动摄像头的公开数据集有Buoy(Fefilatyeve等,2006; Kristan等,2014; Patino等,2016; Ribeiro等,2019)、多模海洋环境障碍物检测数据集(multi-modal marine obstacle detection dataset, MODD)(Kristan等,2014,2016)和MODD2(Bovcon等,2018)等。其中,Buoy由固定在浮筒上的低成本可见光摄像头采集,在图像稳定性、图像质量等方面对算法提出了挑战;MODD与MODD2由固定在无人艇(unmanned surface vehicle, USV)上的摄像头采集,标注分为大型障碍和小型障碍两类,可用于检测效果的定性分析,也可对其进行进一步的分类标注作为其他数据集的补充。此外,SMD等数据集也有部分舰载摄像头数据可供定量的模型评估使用。表1汇总了舰船检测领域中部分公开数据集。

表1 舰船目标检测领域部分公开数据集
Table 1 Public datasets for ship target detection

数据集	数据类型	传感器	标注类别	数据量
MarDCT	视频/陆基	VI/IR	DCT	12
SMD	视频/混合	VI/NI	DT/7 + 地平线	36/12 604(Prasad等,2016a)
SeaShips	图像/陆基	VI	D/6	168/31 455(Shao等,2018)*
Buoy	视频/浮筒	VI	地平线	10/998(Prasad等,2016a)
MODD	视频/舰载(USV)	VI	D/2**	12/4 544(Kristan等,2014)
MODD2	视频/舰载(USV)	VI	D/2**	28/11 675(Bovcon等,2018)
IPATCH(Patino等,2016)	视频/混合	VI/IR/NI	D/未知,以及行为标注	未知
SEAGULL(Ribeiro等,2019)	视频/机载(UAV)	VI/IR/NI/超光谱	DT/5	19/151 753(Ribeiro等,2019)
Annapolis	图像/陆基	VI	D/9	13 958(Morris等,2012)

注:数据类型中,USV表示无人艇,UAV表示无人机。标注类别中,D表示目标检测,C表示图像分类,T表示目标跟踪,斜线后的数字表示船只类别的数量,**表示分为大型障碍和小型障碍两类。数据量中,斜线前/后的数字表示视频/帧数,*表示SeaShips数据集目前仅可取得总帧数7 000帧的部分数据集。

目前,针对舰船目标检测任务收集的数据集能够基本满足常用目标检测方法的数据需求。然而,这一领域的部分数据集仍然存在一些不足。安那波利斯海事监控(Annapolis)数据集(Morris等,2012; Bousetouane和Morris,2016)仅标注了前景中的舰船,但其背景中包含了许多船只,其中不乏与前景舰船十分相似的船只,且相应模型将较多的属于背景的船只检测了出来(Bousetouane和Morris,2016)。

MODD和MODD2数据集用于船只避障时,其目标仅分为小型障碍和大型障碍两类,如果将其用于目标检测任务,可能需要重新进行类别标注。

SMD数据集在舰船目标相互遮挡时存在标注上的不一致,在某些视频中只标注了遮挡舰船的可见部分,而另外一些视频中标注了预测的全部船体,也没有考虑在出现目标遮挡时将可见区域与估计的全部区域一同标注出来(Moosbauer等,2019);同

时,一部分包围框也没有紧贴船体的轮廓线,导致模型的性能受到影响。SeaShips 数据集虽然在舰船目标的尺度多样性上有所考虑,但是其中按照微软富环境常见目标(Microsoft common objects in context, MS COCO)数据集标准统计出的小尺度目标标注仅约占全部标注的 0.13%。

本文实验主要采用 SMD 和 SeaShips 数据集。考虑到 SMD 数据集标注上的不一致性和 SeaShips 数据集小尺度目标偏少的问题,本文使用 SMD 数据集对比各个模型在多尺度目标检测任务中的性能,而使用 SeaShips 数据集对比模型的检测精度和效率。

1.2 舰船目标检测性能评价指标

在目标检测领域,对检测结果精确性的评价指标分为包围框和数据集两个层面。

在包围框层面,衡量检测精确度最常用的指标为交并比(intersection over union, IoU),即结果包围框 D 与标注包围框 G 两者的交集区域的面积与并集区域面积的比值,具体为

$$IoU(D, G) = \frac{|D \cap G|}{|D \cup G|} \quad (1)$$

若为交并比设置一个阈值,高于此阈值的检测结果视为真阳性(true-positive, TP),其集合记为 TP ;低于阈值的检测结果视为假阳性(false-positive, FP),其集合记为 FP ;综合标注包围框数据计算出假阴性(false-negative, FN)结果,记其集合为 FN ,即可计算出模型的精确率 P 、召回率 R 以及 F-score(用 F_β 表示,其中 β 为权重参数),具体为

$$P = \frac{|TP|}{|TP| + |FP|} \quad (2)$$

$$R = \frac{|TP|}{|TP| + |FN|} \quad (3)$$

$$F_\beta = \frac{(1 + \beta^2)PR}{\beta^2 P + R} \quad (4)$$

在数据集层面,衡量目标检测模型精确度的常用指标是类别平均精确率 mAP(mean average precision)。mAP 在 2007 年模式分析、统计建模与计算学习视觉目标分类竞赛(pattern analysis, statistical modeling and computational learning, visual object classes, PASCAL VOC)中提出,该方法按照置信度降序排列检测结果,依次遍历每个检测结果,统计精确率和召回率,并基于固定的 11 个召回率采样点对模

型的精确率进行采样(VOC 2007)(Everingham 等, 2010)或使用所有的精确率和召回率计算值(VOC 2010)(Everingham 等, 2015),绘制精确率—召回率曲线,计算该曲线与精确率和召回率两条坐标轴围成的面积,即平均精确率(average precision, AP)。mAP 是不同目标类别的 AP 值的平均值。AP 与 mAP 避免了不同模型中置信度的不等价性对评估产生的影响,可用于目标检测领域绝大多数模型。

Prasad 等人(2020)从舰船目标检测领域的避撞等实际应用需求出发,认为在纵向位置和横向跨度上,可以容忍低估船只的距离(即包围框底边偏低)、高估船只的横向跨度(即可能将船航行产生的航迹也认为是船的一部分),即为检测结果不同方向上的误差赋予了一种权重,构造了底边与边缘相似度(bottom edge proximity, BEP)评价指标。若设检测结果包围框与标注包围框的高度分别为 y_{DO} 与 y_{GT} ,在水平方向上的重叠长度与左右两端非重叠长度为 x_b 和 x_a 、 x_c ,底边的纵坐标差的绝对值为 Δy_{BE} ,定义 $BEP_1 = X_1 Y_1$, $BEP_2 = X_2 Y_2$ 。其中

$$X_1 = \frac{x_b}{x_a + x_b + x_c}, Y_1 = 1 - \frac{\Delta y_{BE}}{\min(y_{GT}, y_{DO})} \quad (5)$$

$$X_2 = \frac{x_b}{x_a + x_b}, Y_2 = 1 - \frac{\Delta y_{BE}}{y_{GT}} \quad (6)$$

BEP_1 较 BEP_2 更为严格(Prasad 等, 2020)。

基于视频的目标检测任务与基于图像的目标检测任务的最大区别在于视频中的帧与帧之间在时域上的关联性要比单幅图像丰富。针对视频的性能评价,如何从视频片段中抽样出参与评价指标计算的帧是一个重要问题,不同的抽样方法将导致不同的性能偏好。Stiefelhagen 等人(2007)在第一届国际事件、活动和关系分类评价标准工作坊(first international evaluation workshop on classification of events, activities and relationships, CLEAR 2006)中提出多帧图像多目标检测精确率(n -multiple object detection precision, N-MODP)与多帧图像多目标检测正确率(n -multiple object detection accuracy, N-MODA)两个指标。其中,多目标检测精确率 MODP 是单帧图像中每个检测结果框 IoU 的平均值, N-MODP 是一段序列中全部图像的 MODP 的均值,即

$$MODP(f) = \frac{\sum_{k=1}^{N_{map}^f} IoU(D_k^f, G_k^f)}{N_{map}^f}$$

$$N-MODP = \frac{\sum_{f=1}^{N_{\text{frame}}} MODP(f)}{N_{\text{frame}}} \quad (7)$$

式中, N_{frame} 表示该段序列的总帧数, f 为帧的序号, N_{map}^f 表示第 f 帧中所有能够与某个标注包围框对应的检测结果包围框的数量, D_k^f 与 G_k^f 分别表示第 f 帧中的第 k 个目标的标注包围框与检测包围框。多目标检测正确率 MODA 则由 1 减去某帧图像中模型的加权错误率(误检与漏检的数量之和除以标记总数)计算得到, N-MODA 则由 1 减去整个序列中模型的错误率计算得到。具体为

$$MODA(f) = 1 - \frac{c_M(m_f) + c_{FP}(fp_f)}{N_G^f}$$

$$N-MODA = 1 - \frac{\sum_{f=1}^{N_{\text{frame}}} (c_M(m_f) + c_{FP}(fp_f))}{\sum_{i=1}^{N_{\text{frame}}} N_G^f} \quad (8)$$

式中, m_f 与 fp_f 分别表示第 f 帧中漏检与误检的数量, c_M 与 c_{FP} 为其对应的权重或损失函数; N_G^f 表示第 f 帧中标注包围框的数量。

N-MODA 与 N-MODP 体现了模型在实际运行时的精确率和正确率, 但与 mAP 相比, 它依赖于算法中置信度阈值的选择。

另外, 基于实时视频的目标检测结果常用于距离估计和运动估计, 因此视频目标检测在时域上的稳定性也影响到决策和规划的稳定性。Zhang 和 Wang(2017) 关注到视频目标检测任务中的相应需求, 提出了一种由分段误差(在同一个目标出现的全程中, 目标由检测到漏检, 或重新检出的状态变化的次数与全程总帧数的比值)、中心点偏移和尺度与长宽比例误差组合而成的对视频目标检测结果稳定性的度量。总误差计算为

$$\Phi = E_F + E_C + E_R \quad (9)$$

式中, E_F 、 E_C 和 E_R 分别表示分段误差、中心点偏移损失和尺度与长宽比例误差。

分段误差 E_F 计算为

$$E_F = \frac{1}{N} \sum_{k=1}^N \frac{s_k}{t_k - 1} \quad (10)$$

式中, N 为一段视频中目标的总数, k 为目标的序号, s_k 与 t_k 分别为编号为 k 的目标状态改变的次数与其出现全程的总帧数。

记第 k 个目标在第 f 帧中的检测包围框为

$(x_D^{kf}, y_D^{kf}, w_D^{kf}, h_D^{kf})$, 标注包围框为 $(x_G^{kf}, y_G^{kf}, w_G^{kf}, h_G^{kf})$, 其中 x 、 y 、 w 与 h 分别表示对应包围框的中心点横坐标、纵坐标以及包围框的宽和高, 则中心点偏移损失 E_C 计算为

$$E_C = \frac{1}{N} \sum_{k=1}^N (\sigma_X^k + \sigma_Y^k) \quad (11)$$

$$\sigma_X^k = sd(e_X^k), e_X^{kf} = \frac{x_D^{kf} - x_G^{kf}}{w_G^{kf}}$$

$$\sigma_Y^k = sd(e_Y^k), e_Y^{kf} = \frac{y_D^{kf} - y_G^{kf}}{h_G^{kf}}$$

式中, $sd(\cdot)$ 为标准差函数。 e_X^k 为第 k 个目标在其出现的所有帧上的中心点横坐标误差项 e_X^{kf} 的集合, 同理, e_Y^k 为第 k 个目标中心点纵坐标误差项 e_Y^{kf} 的集合。

尺度与长宽比例误差 E_R 计算为

$$E_R = \frac{1}{N} \sum_{k=1}^N (\sigma_S^k + \sigma_R^k) \quad (12)$$

$$\sigma_S^k = sd(e_S^k), e_S^{kf} = \sqrt{\frac{w_D^{kf} h_D^{kf}}{w_G^{kf} h_G^{kf}}}$$

$$\sigma_R^k = sd(e_R^k), e_R^{kf} = \left(\frac{w_D^{kf}}{h_D^{kf}} \right) / \left(\frac{w_G^{kf}}{h_G^{kf}} \right)$$

式中, e_S^k 为第 k 个目标在其出现的所有帧上的尺度误差项 e_S^{kf} 的集合, 同理, e_R^k 为第 k 个目标长宽比例误差项 e_R^{kf} 的集合。

2 海事监控视频舰船目标检测方法

计算机视觉领域的目标检测方法主要分为传统的目标检测算法和基于卷积神经网络(convolutional neural network, CNN)的深度学习两类。深度学习是一种数据驱动的方法, 为了发挥 CNN 的潜力, 需要高质量的数据集和运算量足够的计算设备。由于深度学习相对于传统方法无需进行人为的特征工程设计, 且模型表现能力更强, 近年来, 在诸多领域中的性能表现都超过了传统模型。但是, 目前舰船目标检测领域多使用传统方法, 深度学习的方法和模型尚处于初步适配的阶段。

2.1 传统的舰船目标检测方法

基于传统机器学习和计算机视觉的舰船目标检测方法大多采用地平线检测—背景减除—前景分割的通用技术框架。其中, 地平线检测方法分为基于

直线特征的检测方法 (Fefilatyevev, 2012; Fefilatyevev 等, 2012)、基于区域建模的方法 (Chen 等, 2018b) 以及上述两种方法的混合方法 (Prasad 等, 2016a)。背景减除的方法可分为 3 类: 1) 基于单幅图像的统计信息判断每个像素是否属于前景的方法 (Fefilatyevev, 2012; Fefilatyevev 等, 2012; Wang 等, 2017); 2) 基于高斯混合模型 (Gaussian mixture model, GMM) 等概率模型计算像素属于前景和背景的概率的方法 (Chen 等, 2018b); 3) 提取每个像素或小区域的图像特征, 计算它属于前景或背景概率的方法 (Zhang 等, 2017)。前景分割采用基于背景减除的结果利用形状学方法生成检测结果 (Westall 等, 2008; Zhang 等, 2017)。

近年来, 学者们对既有模型的缺陷开展了进一步优化。为解决海浪等背景动态问题, Zhang 等人 (2017) 提出一种基于离散余弦变换 (discrete cosine transform, DCT) 的监控视频舰船目标检测方法, 使用离散余弦变换分析每个 8×8 的图像块, 使用变换后该图像块交流系数 A 的平均值 $\bar{A}$ 度量图像块中纹理信息的丰富程度, 并将其与图像中所有图像块 $\bar{A}$ 值中最大值 $\bar{A}_{\max}$ 的比值 a 作为判断图像块属于天空或者海面的依据 (阈值 $T = 0.1$, 超过则判定为海面)。对前景和背景进行分类时, 对于地平线上的图像块, 若有 $a > T$, 则作为前景; 对于地平线下的图像块, 则将 DCT 系数矩阵中的固定区域作为竖直方向、水平方向和对角线方向上图像亮度变化剧烈程度的量度, 作为特征使用 GMM 计算其为前景和背景的概率, 最后使用形状学方法生成检测结果。

为了提高含有较多海浪等动态元素的背景建模效果, Chen 等人 (2018b) 提出了一种使用自适应调整的 GMM 进行舰船检测的方法, 在一个刻画图像背景的 GMM 中维护固定数量的子高斯分布, 在检测过程中对各个子高斯分布的参数和权重进行调整, 若发现有不能符合上述 GMM 中任意一个子高斯分布的像素点, 则构造一个新的分布替换掉所有子高斯分布中权重最低的分布。

本文使用支持向量机 (support vector machine, SVM) 方法, 基于方向梯度直方图 (histogram of oriented gradients, HOG) 特征, 并且结合多尺度交叉模态线性特征 (multi-scale cross modal linear feature, MSCM-LiFe) 地平线检测算法 (Prasad 等, 2016b) 对检测结果进行过滤, 考虑到 SeaShips 数据集中没有

海天分界线, 在 SMD 数据集中对上述方法进行了实验测试。将与任一标注框相交面积超过自身面积 30% 的检测结果作为真阳性, 反之作为假阳性, 并将与任一检测结果框相交面积均小于自身面积 30% 的标注框作为假阴性, 计算得到的准确率为 74.1%, 召回率为 44.7%。

2.2 卷积神经网络舰船目标检测方法

Girshick 等人 (2014) 第一次实现了基于 CNN 的目标检测, 所提出的基于区域的卷积神经网络 (region convolutional neural network, R-CNN) 采用选择搜索的方法进行候选区域提议, 再对这些候选区域逐个使用 CNN 进行分类, 计算出属于各个类别的置信度。此后, 目标检测领域中基于 CNN 的方法迅速发展, 许多学者将其应用到舰船目标检测领域, 评估了各类网络模型在舰船目标检测任务中的性能表现, 分析了它们在舰船目标检测任务中的优势和缺点, 并初步做出了一些针对性的改进。

Bousetouane 和 Morris (2015) 将 CNN 用于舰船细粒度分类任务, 使用卷积神经网络提取图像特征, 再使用 SVM 对 CNN 提取的特征进行分类, 在 Annapolis 数据集上对 AlexNet、OverFeat、GoogLeNet 和牛津大学视觉几何小组 (Visual Geometry Group, VGG) 所提出的 VGG 等基础网络进行评估, OverFeat 网络的 mAP 成绩最好, 为 70.21%。

Bousetouane 和 Morris (2016) 基于 Fast R-CNN 的目标检测网络架构, 对比了通过滑动窗口、选择搜索、候选区域提议网络 (region-proposal network, RPN) 以及使用 HOG 等传统计算机视觉特征进行预检测获得候选区域的方法, 分析结果认为, R-CNN 与 Fast R-CNN 的候选区域提议方法并不完全适合于在宽幅图像中检测尺度较小的舰船目标, 主要原因是选择性搜索算法作为一种自底向上的区域提议算法, 它的结果数量会随着图像尺度的变大呈指数级增长, 在目标数量多且重叠严重的情况下表现不佳。

Cruz 和 Bernardino (2016) 评估了滑动窗口和基于图像显著性进行候选区域提议再使用 CNN 对候选区域进行分类的目标检测方法, 证明了将滑动窗口替换为其他区域提议算法在减少运算量上的有效性。

Cane 和 Ferryman (2018) 受到基于图像语义分割的目标检测模型的启发, 解决了基于锚框的 CNN

目标检测方法的过拟合和实时性等问题。作者基于全景分割数据集 ADE20k(Adela 20000)中包含海事目标的图像子集构建了训练集,并使用 MODD、SMD、IPATCH 和海鸥数据集(SEAGULL)中的一部分视频片段作为测试集,分别在接受所有非背景检测结果和只接受感兴趣类别的检测结果的设置下,使用 SegNet 获得了 -1.520 和 0.207 的 N-MODA 以及 0.240 和 0.080 的 N-MODP;使用高效率空间金字塔网络(efficient spatial pyramid net, ESPNet)获得了 -2.743 和 -0.624 的 N-MODA 以及 0.208 和 0.138 的 N-MODP;使用高效率神经网络(efficient neural network, ENet)获得了 -0.762 和 -0.115 的 N-MODA 以及 0.272 和 0.024 的 N-MODP。然而,语义分割模型引入了相互重叠的目标难以有效分开的问题。

Zou 等人(2019)使用对舰船目标进行重新分类的 SMD 数据集,分别对 Faster R-CNN 和单阶段多框预测目标检测模型(single shot multibox detector, SSD)进行测试,获得了 84.35% 和 80.23% 的 mAP,并且使用 ResNet 替换了 Faster R-CNN 中的基础网络 VGG,获得了 88.08% 的 mAP。Moosbauer 等人(2019)使用 Faster R-CNN 与 Mask R-CNN 模型,在其提出的对 SMD 数据集的划分上进行测试,其中带特征金字塔网络(feature pyramid network, FPN)的 Faster R-CNN 模型获得了 0.773 (IoU 阈值为 0.5) 和 0.854 (IoU 阈值为 0.3) 的 F-score。Schöller 等人(2019)基于长波长红外图像数据集,对 Faster R-CNN、RetinaNet 和 YOLO(you only look once)系列网络中的 YOLOv3 网络进行评估,分别获得了 81%、86% 和 90% 的 mAP,并通过图像上采样的方

法,使 RetinaNet 的 mAP 提升到 90%。

Shao 等人(2020)结合海岸线检测、注意力机制与 YOLOv2 网络,将检测范围限制在海岸线向上平移 30 像素形成的分界线以下区域,并将 YOLOv2 网络检测的结果结合显著性信息和海岸线特征进行了优化。作者在 SeaShips 数据集上进行测试,结果较 YOLOv2 有了 4.4% 的 mAP 提升(提升至 87.4%),减小了与 Faster R-CNN 的差距。

上述研究表明,基于海事监控图像和视频的舰船目标检测与通用目标检测相似,两阶段目标检测网络的检测精确率高于一阶段检测网络,但检测速度较慢。CNN 模型表现出对目标尺度的敏感性,将区域提议网络更换为其他尺度不敏感的区域提议方法后,性能有了较大提升。

表 2 是本文基于 SMD 数据集,使用 GeForce GTX 1080Ti GPU 对各种目标检测经典方法的评估结果。本文选择了在目标检测领域最具代表性的若干种 CNN 模型,包括两阶段目标检测模型 Faster R-CNN(Ren 等,2015)及其通过 FPN(Lin 等,2017)在多个尺度的特征图上分别计算区域提议的优化模型,以及一阶段目标检测模型 SSD(Liu 等,2016)。同时,实验选取了一阶段目标检测模型 YOLOv4(Bochkovskiy 等,2020)和无锚框目标检测模型 CenterNet(Zhou 等,2019)作为最新目标检测模型的代表。各模型均使用基于 ImageNet(Faster R-CNN 与 SSD)或 MS COCO(YOLOv4 与 CenterNet)预训练的模型进行 fine-tune 训练。

由于 SMD 数据集中的标注不一致问题,本文参考 Moosbauer 等人(2019)方法,将 IoU 阈值分别设为 0.3 和 0.5,并且仅使用舰船和背景两个类别进

表 2 CNN 方法舰船目标检测评估结果
Table 2 Evaluation results of CNN ship detection methods

模型	基础网络	输入分辨率 /像素	AP _{0.3} /%	AP _{0.5} /%	AP ^S /%	AP ^M /%	AP ^L /%	帧率/(帧/s)
Faster R-CNN(Girshick 等,2020)	ResNet-50	500 × 500	80.1	65.7	4.6	35.9	49.7	2.4
Faster R-CNN + FPN(Girshick 等,2020)	ResNet-50	500 × 500	84.7	74.6	6.7	41.3	53.0	11.0
SSD	ResNet-101	512 × 512	65.0	40.0	0.6	20.4	30.4	12.9
YOLOv4	Darknet53	512 × 512	86.2	77.4	1.8	40.7	55.4	49.7
CenterNet	ResDCN-101	512 × 512	73.7	56.2	0.4	26.9	50.3	41.7

注:加粗字体表示各列最优结果;ResDCN 全称为 residual dense convolutional network;AP^S、AP^M 和 AP^L 分别表示小型目标、中型目标和大型目标(MS COCO 标准)在不同 IoU 阈值下的平均值。

行测试。 AP^S 、 AP^M 和 AP^L 分别表示小型目标、中型目标和大型目标 (MS COCO 标准) 在不同 IoU 阈值下的平均值,其 IoU 阈值取值范围为 0.3 ~ 0.95,每隔 0.05 采样一次。帧率 (frame per second, FPS) 表示网络在单张 GPU 上每秒可以处理的图像数量。从结果来看,针对多尺度优化的 Faster R-CNN 模型对多尺度目标的适应能力较好,而基于各种优化改进的 YOLOv4 模型在平均精确度上位居前列。需要注意的是,由于不同模型对 SMD 数据集中相关标注不一致的敏感性不同,表 2 中的评估结果不能完全代表相关模型对舰船目标检测任务的适合程度。

3 舰船目标检测任务中的技术难点

舰船目标检测任务的应用场景决定了一些通用目标检测技术面临的技术难点会放大。视野开阔、气象情况较为复杂是海洋环境的典型特征。海面波浪、反射阳光,以及舰船航行留下的航迹等为水面带来了更丰富的动态因素,从而对各种模型的表现力提出了更高要求。具体来讲,舰船目标检测面临以下 6 个技术难点:

1) 舰船目标尺度多样性。在海事监控图像数据中,由近到远,开阔的视野使得帆船、皮划艇等小型船舶和油轮、集装箱船等大型舰船均可以出现在同一帧图像中,导致图像数据中的目标具有极大的尺度多样性。并且,远处的大型舰船和近处的小型船舶可能具有相似的视觉尺度。在 CNN 模型中,较小尺度的目标会在卷积和池化过程中在特征图上对应越来越小的区域,甚至消失;在最深层,网络的感受野如果大幅超过目标的尺度,目标将很难体现在特征图上。在当前的一些研究中,研究者注意到一部分 CNN 模型 (如 R-CNN) 对占画面比例较小的目标的检测效果比起足够大的目标来说有明显的下降 (Boussetouane 和 Morris, 2016; Shao 等, 2020)。

2) 舰船类别的多样性。舰船目标有很多细分的类别,例如,在 SeaShips 数据集中,将舰船分为油轮、客船、集装箱船、散装货船、普通货船和渔船等 6 类 (Shao 等, 2018)。不同细分类别目标的形状、纹理等图像特征有着较大的差异。舰船类别的多样性要求在设计模型时仔细评估模型是否有足够的能力刻画相同目标类别内部的多样性,在构建数据集

时仔细考虑类别划分的粒度;在 CNN 模型中,依据模型的特点提供分类粒度合适的标签,有利于模型训练过程平稳进行;分类标签的粒度还在一定程度上决定了各个类别的样本数量,从而影响了各个类别间的平衡性。

3) 海洋气象的复杂性。海洋气象环境相较于陆地更加复杂,也更容易发生极端天气。面对海事监控系统持续性、高可靠性监测的要求,数据集需要对各种天气和光照条件进行充分覆盖。同时,目标检测算法中建立的背景模型需要具有对不同的天气和光照条件的描述能力。基于时域上连续的视频数据,使用自适应的背景建模算法 (如 Chen 等人 (2018b) 的方法) 不断调整背景和前景的模型参数,是无需事先精确确定 GMM 等由多个子分布或子模型构成的一类模型中子模型的数量也可以对变化的背景进行描述的一种思路。

4) 水面的动态特性。船只的尾流、水面的浪花和反射光斑等背景中的动态元素对背景建模和背景减除提出了挑战。在背景减除阶段,一些算法无法减除尾流、浪花等动态元素,致使误检产生。针对水面动态元素造成的误检,可以考虑对它们进行建模,以及对检测结果进行事后的分类过滤。

5) 相机的视角和运动。在同一或不同位置以不同角度固定的摄像头拍摄同一个物体时,照片上物体的大小、位置和姿态都会有相应的差别。在海洋环境中,浮标等平台常常随着波浪晃动,固定在可动平台上的摄像头会随着平台的颠簸、晃动出现物体位置的抖动,甚至倾斜;有时甚至会出现摄像头淹没的情况。因此,应用于可动平台上的舰船目标检测模型应具有一定的旋转和平移的健壮性。

6) 低质量图像。安装在浮筒等位置上的摄像头,由于成本限制,其图像分辨率和质量一般较受限;同时,拍摄和传输过程中的噪声干扰也会使图像质量下降。在通信手段缺乏的远洋海域,通信带宽的限制造成了数据采集和传输的困难,致使数据集难以构造,使用服务器部署集中算力的方法也难以实现。

4 舰船目标检测性能优化方法

针对第 3 节所述舰船目标检测的技术难点,以往的研究亦做出过很多改进。对传统方法影响较大

的技术难点是水面动态特性、气象情况和昼夜变化以及舰船类别的多样性3类,对基于传统方法的舰船目标检测技术框架中地平线检测与背景减除两个阶段的影响最大,近年来的研究也大多基于提升背景减除健壮性和检测效率(Zhang等,2017;Chen等,2018)的思路;CNN方法则开始关注舰船目标的尺度多样性问题(Bousetouane和Morris,2016;Shao等,2020),同时尝试了使用语义分割网络(Cane和Ferryman,2018)、结合注意力机制和融合先验信息(Shao等,2020)等改进。

本文关注到舰船目标检测领域数据集数量和质量较欠缺、小型目标检测效果较差以及缺少CNN方法在嵌入式设备上的应用研究的问题,在预处理与数据增广、尺度多样性两个方面进行了实验研究,针对实际应用中嵌入式平台的功率和算力受限等问题,开展了运算量优化的模型研究,并评估了模型在较低运算量和功耗下的性能表现。

4.1 预处理与数据增广

舰船目标检测面临图像质量低、传感器运动以及天气和光照条件的多样性等技术难点,可以通过预处理和数据增广的方式加以改善。但是,目前大多数基于CNN的舰船目标检测仍缺少针对这些技术难点在数据预处理和数据增广等方面的研究。

远洋监测平台常常受到传感器成本和传输带宽的限制,难以获得高质量的图像数据,较低的分辨率和较大比例的噪声是影响图像质量的主要问题。有代表性的图像滤波方法包括均值滤波、中值滤波及高斯滤波等,其依据自身的特点适用于不同类型的噪声,如中值滤波适合过滤脉冲噪声,均值滤波适合过滤高斯噪声。解决图像分辨率低的问题则是图像

超分辨率技术的用武之地。图像超分辨率方法(苏衡等,2013)分为基于插值的方法(如双线性插值、最近邻插值)、基于重建的方法(Stark和Oskoui,1989;Katsaggelos等,1993)和基于学习的方法(Stark和Oskoui,1989;Katsaggelos等,1993;沈明玉等,2019)。

在舰船目标检测任务中,固定在舰船和浮筒上的传感器会随着其附着平台的俯仰、偏航、滚转和平动以及PTZ摄像头自身的视角和缩放的变化,在拍摄所得的图像数据中产生视角和视距的变化。在数据增广环节,可以尝试通过平移、倾斜和缩放等方式模拟这些变化,提高模型的健壮性。

本文基于SeaShips数据集尝试了一些具有针对性的数据增广方法,并对比了数据增广前后模型的检测性能。为了模拟舰船和浮筒的沉浮和摆动,为数据集增加 $[-15^{\circ}, 15^{\circ}]$ 的随机旋转;为了模拟海面上时常出现的风浪以及各种恶劣天气导致的图像模糊,添加部分高斯模糊后的图像(噪声系数范围为 $[20, 60]$);为了模拟海面天气变化造成的图像亮度不一,在数据集中添加经过随机亮度处理后的图像(亮度范围为 $[0.5, 1.25]$)。此外,为了消除舰船航行方向的影响,以及进一步扩充数据集的样本量,本文添加了水平翻转后的图像。

表3为使用数据增广前后的SeaShips数据集在1080Ti GPU上训练的Faster R-CNN模型(基于ResNet-50基础网络)的性能表现。可以看出,使用增广后的数据集训练的模型在各类别上的AP值均比使用未增广数据集训练的模型要高,mAP值则高出了近5%,表明数据增广能够有效提高模型的健壮性。

表3 SeaShips数据集数据增广前后性能比较

Table 3 Performance comparison before and after data augmentation

数据集	各类别 AP						mAP
	油轮	客船	集装箱船	散装货船	普通货船	渔船	
未增广	0.813	0.898	0.836	0.878	0.808	0.906	0.857
增广后	0.908	0.905	0.909	0.904	0.906	0.909	0.907

注:加粗字体表示各列最优结果。

目前处于研究热点的生成对抗网络(generative adversarial network,GAN)在图像去噪、超分辨率和数据增广任务中显现出了潜力。将白天拍摄的图像转换为夜晚拍摄的图像,或转换为不同天气条件下

拍摄的图像,这类任务属于图像的跨域转换任务。Arruda等人(2019)使用CycleGAN网络利用包含标注的源域图像和不包含标注的目标域图像生成包含标注的伪目标域图像,实现了源域与目标域的转换,

并使用有标注的源域和伪目标域的数据训练目标检测模型,检测性能提高了 10%。Katayama 等人(2019)使用多项式损失函数训练 CycleGAN,将水下摄像机拍摄的图像色彩校正为常规图像,并使用色彩校正后的图像训练 YOLO9000 目标检测模型,mAP 为 41.8%,相比原图像提升了 7.0%,相比使用原版 CycleGAN (38.8%) 进行色彩校正提高了 3.0%。

在其他领域,GAN 也用于跨域数据增广,表现出缓解数据集类别不平衡问题的潜力(Frid-Adar 等,2018;Mikołajczyk 和 Grochowski,2018)。

除了基于 GAN 的跨域数据增广方式,通过对抗训练的方式也可得到域无关的目标检测模型,从而提升单个模型在多个域上的检测性能。

为了使目标检测模型具有自适应不同域的能力,Chen 等人(2018a)在 Faster R-CNN 模型的尾部加入两个域分类器网络,其中图像级别的域分类器添加到最深层的特征图之后,另一个实例级别的域分类器以基于感兴趣区域(region of interest, ROI)的特征向量作为输入,两个域分类器之间通过一致性损失进行正则化。在训练 Faster R-CNN 模型时,使用一种对抗的方式学习域不变特征,使得域分类器无法通过网络提取的特征判断图像来自哪一个域。一方面,对于域分类网络,需要最小化其分类损失,提升其分类性能;另一方面,Faster R-CNN 网络则需要使得域分类器的分类损失最大化。使用了一种梯度翻转层(gradient reversal layer, GRL)(Ganin 和 Lempitsky,2014)实现对抗学习。

4.2 针对尺度多样性的优化

传统计算机视觉方法是否受到尺度多样性的影响取决于所用方法对目标尺度的敏感程度。CNN 模型的不同特征层级具有不同的感受野,其适宜检测的目标尺度也各不相同。在较深层次中,较大的感受野会使得尺度较小的目标在特征图上对应的区域缩小。在舰船目标检测领域,已经有学者关注到尺度多样性问题,尤其是小尺度目标对于检测性能的影响。Boussetouan 和 Morris(2016)使用传统计算机视觉方法进行区域提议,但这样的网络无法进行端到端的训练。因此,本文提议参考通用目标检测领域针对尺度多样性进行优化的方法对模型进行优化设计。

通用目标检测领域针对目标尺度多样性进行优

化的主要方法是通过构建多尺度的特征图,在不同尺度的特征图上分别对目标尺度与感受野相适合的目标进行检测。图像金字塔是这一思路的最朴素实现。它按照一个尺度比例逐级放大和缩小图像,某一级别的模型只输出尺度落在该级别的检测结果。其最大问题在于需要将每一个尺度的图像都使用目标检测模型进行一次检测,因此运算量很大。同时,多尺度的图像造成了模型在训练阶段与测试阶段尺度上的不一致,易导致检测精确性下降(Lin 等,2017)。

特征金字塔方法则不再通过放缩图像的方式构建多尺度的特征,而是通过对特征图进行逐步卷积、池化等操作产生特征的尺度多样性。Lin 等人(2017)认为浅层特征图中过多的初级特征对目标检测任务不利,然而高分辨率的特征图又对小尺度目标的检测十分重要,因而提出了一种自顶向下构造的特征金字塔网络,即 FPN。该网络由特征提取网络中最深层的特征图开始,逐层使用上采样后的特征图与特征提取网络中对应层级的特征图经过 1×1 的卷积后所得的特征图相加,构造出多个层级的特征,供任务网络使用。

Fu 等人(2017)提出了反卷积单阶段多框预测目标检测模型(deconvolutional single shot multibox detector, DSSD),使用反卷积(转置卷积)对深层次特征进行上采样,并使用逐像素相乘的方法通过侧连接与特征提取网络中对应层级的特征图融合而构造多尺度的特征图。

姜文涛等人(2019)尝试了不同的上采样方法和不同层级之间特征融合的模式,简化了 DSSD 网络中特征融合网络的结构,改用逐像素相加的方式进行特征融合,并在特征融合之后增加了一个 3×3 的卷积层用以减轻混叠效应,较 DSSD 模型降低了运算量,且取得了更高的 mAP。张筱晗等人(2020)设计了先自深层次向浅层次构造融合特征,再将低层特征用做注意力线索,自浅层次向深层次融合的双向特征融合网络,在 SAR 图像舰船检测任务中取得了较 SSD 模型更好的 AP。

但是,上述方法都没有突破特征图分辨率最大为原始图像的 $1/4$ 的限制。根据增高分辨率网络(higher high-resolution network, HigherHRNet)(Cheng 等,2020)的研究, $1/2$ 甚至更高分辨率的特征图在检测小尺度目标时是有较重要作用的。该研

究针对特征金字塔方法在小尺度目标关键点热力图分辨率不高的问题,通过双线性插值上采样各个尺度下的热力图至原始图像大小,并对所有尺度上采样后的热力图进行平均,同时提高了热力图的分辨率和语义信息的载量。目前,还有将高分辨率热力图应用于基于关键点估计的无锚框(anchor-free)方法(如 CenterNet(Zhou 等,2019)),也是值得进一步研究的方向。

本文基于 Detectron 平台(Girshick 等,2020)和 mmdetection 平台(Chen 等,2019),在 Faster R-CNN 架构下,对 ResNet 基础网络以及针对多尺度目标进行优化的高分辨率网络(high-resolution network,

HRNet)基础网络,使用 SMD 数据集将目标分为船和背景两类,在 500×500 像素和 800×800 像素两个输入分辨率上进行了性能测试。实验使用 GeForce GTX 1080Ti GPU,结果如表 4 所示。实验中分别使用 0.3 和 0.5 作为 IoU 阈值统计 mAP,对不同尺度目标的 AP 值统计的 IoU 阈值取值范围为 $[0.30, 0.95]$,每隔 0.05 取样一次。实验数据表明,训练和测试阶段使用的图像分辨率对于多尺度目标的检测效果起到了决定性作用。同时,FPN 方法在减少运算量的情况下缩小了不同分辨率、基础网络之间多尺度目标检测效果的差距,HRNet 基础网络则较大幅度地改善了小尺度目标的检测效果。

表 4 不同输入分辨率及多尺度优化模型的性能比较

Table 4 Performance comparison of different input scales and multiscale-optimized models

模型	基础网络	输入 分辨率/像素	AP _{0.3} /%	AP _{0.5} /%	AP ^S /%	AP ^M /%	AP ^L /%	帧率/ (帧/s)
Faster R-CNN(Girshick 等,2020)	ResNet-50	500 × 500	80.1	65.7	4.6	35.9	49.7	2.4
Faster R-CNN + FPN(Girshick 等,2020)	ResNet-50	500 × 500	84.7	74.6	6.7	41.3	53.0	11.0
Faster R-CNN + FPN(Chen 等,2019)	HRNetv2-W18 (Sun 等,2019)	500 × 500	70.8	57.1	0.6	31.5	46.4	11.2
Faster R-CNN + FPN(Girshick 等,2020)	ResNet-101	500 × 500	83.3	69.9	7.1	40.5	53.4	8.7
Faster R-CNN(Girshick 等,2020)	ResNet-50	800 × 800	84.2	74.5	15.0	42.0	56.5	2.3
Faster R-CNN + FPN(Girshick 等,2020)	ResNet-50	800 × 800	85.5	73.4	8.8	42.6	55.7	8.9
Faster R-CNN + FPN(Chen 等,2019)	HRNetv2-W18 (Sun 等,2019)	800 × 800	82.3	69.5	16.8	39.3	52.1	10.4
Faster R-CNN + FPN(Girshick 等,2020)	ResNet-101	800 × 800	85.9	72.9	10.3	41.4	57.3	7.1

注:加粗字体表示各列最优结果。AP^S、AP^M 和 AP^L 分别表示小型、中型和大型目标在不同 IoU 阈值下的平均值。

4.3 运算量与能耗优化

舰船目标检测方案多部署于船只、浮标等平台上,这些平台搭载的计算设备多数计算能力较弱,存储资源有限,同时供电功率受到限制,因而在很大程度上限制了基于 CNN 的目标检测模型的使用。对舰船目标检测方案进行运算量与能耗方面的优化是促进 CNN 模型实际应用的重要条件之一,也是在海事智能设备上提高舰船目标检测效果的关键。

在舰船目标检测任务中,目标的位置受到地平线等语义条件的约束。因此,有研究通过结合语义约束对检测区域进行限制,提高了模型的推理速度。Shao 等人(2020)基于 YOLOv2 网络建立了舰船目标检测模型,通过过滤海岸线上移 30 像素之后的直线上的所有单元格减小了模型的运算量。Fan 和

Chen(2019)使用一种类注意力机制的方法从图像中提取出可能包含目标的子图像块,再使用 Faster R-CNN 进行目标检测,将推理速度从 13 帧/s 提高到 18 帧/s。

目前,舰船目标检测领域的研究多使用标准的基础网络,然而这些网络存在着较高的计算量冗余。MobileNet 系列基础网络(Howard 等,2017,2019; Sandler 等,2018)通过设计高效的网络层、优化网络结构和神经网络搜索等方法,在保证模型的精确度不出现明显下降的情况下,大幅度地优化了模型所需的运算量和存储空间。本文基于 VGG16 和不同版本 MobileNet 的 SSD 目标检测模型搭建了舰船目标检测网络模型,计算了模型的数量和运算量,并基于 SeaShips 数据集测试了 mAP 值,结果如表 5 所示。其中,浮点运算次数(floating point operations,

表 5 目标检测模型的参数量和运算量
Table 5 Parameter quantity and calculation
quantity of object detection model

模型	参数量/M	FLOPs/G	mAP/%
SSD300 VGG16	20.03	29.27	83.1
SSD300 MobileNet V1	6.94	1.28	84.3
SSD300 MobileNet V2	6.27	0.80	80.3
SSD300 MobileNet V3-Large	6.71	0.57	77.4
SSD300 MobileNet V3-Small	2.24	0.18	74.5

注:加粗字体表示各列最优结果。

FLOPs)表示对于每一个输入样本,模型推理需要执行的浮点操作的数量。

除减少模型的参数量和计算量外,针对不同计算平台的特点进行优化也是进一步提高模型推理速度的可行思路。针对存储资源较充足但计算资源较集中的计算平台,可以通过设计合理的访存方式进行优化。标准的卷积操作访存不规则,不利于 CPU 进行计算,而使用 Im2col 等方法可以利用数据冗余减少卷积过程中不规则的内存访问,起到平衡空间与时间复杂度的作用。

5 结 语

基于海事监控视频的舰船目标检测技术是目标检测的一个细分领域。解决舰船目标检测任务中的技术难点是推动其应用的重要条件。在使用特征提取能力更强的深度神经网络模型时,研发人员主要面临目标尺度多样化、数据集质量较低、样本类型不均衡,以及边缘计算设备的算力和功耗受限等问题。本文尝试了一部分通用目标检测领域的优化方法,初步证明这些方法在舰船目标检测领域同样有效,可以纳入到后续研究的考虑中。该领域可在以下方向展开进一步的研究:

1)数据集的进一步优化。针对海事图像和视频数据集应具有的特征进行优化。首先,以适宜的粒度进行分类。提供较细粒度的标注,可以为模型提供更加充分的信息,使各个类别之间的样本量趋于平衡,降低模型的训练难度。在推理阶段有利于基于细分类别进行差异化的决策;其次,优化标注的一致性、易用性。数据集的构建应该保证标注的一致性,参考目前使用最广泛的目标检测数据集的标

注格式,并保证数据集的标注与之兼容,有利于相互转换;最后,使数据集具有易扩充性能。由于数据收集的时间有限,单独的数据集难以涵盖所有的天气变化和光照变化情况。对原有的数据集进行增补,或者与其他数据集集成是提升模型覆盖范围的方法之一。这就要求数据集遵循一定的规范,例如规范的舰船目标分类、易于读取的标注结构等。对已有数据集进行扩展标注和分析也有利于数据集质量的进一步提升。研究者可以通过划分数据集、细化类别标注和增加图像分割标注等方法扩充可用的数据集,增加数据集的适用领域。MarDCT 数据集的发展以及 Moosbauer 等人(2019)对 SMD 数据集的分析和划分就是数据集扩展和分析的代表。

2)模型设计的进一步优化。针对海事监控视频舰船目标检测任务的难点,首先,应进一步研究多尺度目标(尤其是小型目标)检测的模型结构。其次,应考虑图像信息与语义信息的融合方式。海事监控数据的一个特点是地平线检测较容易,同时所有待检测的目标都集中在与地平线共面的同一个水平面上,因此地平线信息可以作为参考,起到限制检测区域、过滤检测结果的辅助作用;一部分数据增广方法(如小尺度目标的随机复制)也受到语义信息的约束。最后,考虑到海事监控设备实际的计算能力和功耗限制,探索性能—计算量比以及性能—功率比更高的模型,是推动舰船目标检测技术更大规模应用部署的重要方向。

致 谢 本文多项实验使用新加坡南洋理工大学和劳斯莱斯—南洋理工大学合作实验室的研究团队采集的 SMD 数据集以及武汉大学研究团队采集的 SeaShips 数据集等公开数据集,在此表示感谢。

参考文献(References)

Arruda V F, Paixão T M, Berriel R F, De Souza A F, Badue C, Sebe N and Oliveira-Santos T. 2019. Cross-domain car detection using unsupervised image-to-image translation: from day to night//Proceedings of 2019 International Joint Conference on Neural Networks (IJCNN). Budapest, Hungary: IEEE: 1-8 [DOI: 10.1109/IJCNN.2019.8852008]

Bloisi D D, Iocchi L, Pennisi A and Tombolini L. 2015. ARGOS-Venice boat classification//Proceedings of the 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Karlsruhe, Germany: IEEE: 1-6 [DOI: 10.1109/AVSS.2015.

- 7301727]
- Bochkovskiy A, Wang C Y and Liao H Y M. 2020. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. [2020-09-25]. <https://arxiv.org/pdf/2004.10934.pdf>
- Bousetouane F and Morris B. 2015. Off-the-shelf CNN features for fine-grained classification of vessels in a maritime environment//The 11th International Symposium on Visual Computing (ISVC). Las Vegas, USA: Springer: 379-388 [DOI: 10.1007/978-3-319-27863-6_35]
- Bousetouane F and Morris B. 2016. Fast CNN surveillance pipeline for fine-grained vessel classification and detection in maritime scenarios//Proceedings of the 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Colorado Springs, USA: IEEE: 242-248 [DOI: 10.1109/AVSS.2016.7738076]
- Bovcon B, Mandeljc R, Perš J and Kristan M. 2018. Stereo obstacle detection for unmanned surface vehicles by IMU-assisted semantic segmentation. *Robotics and Autonomous Systems*, 104: 1-13 [DOI: 10.1016/j.robot.2018.02.017]
- Cane T and Ferryman J. 2018. Evaluating deep semantic segmentation networks for object detection in maritime surveillance//Proceedings of the 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Auckland, New Zealand: IEEE: 1-6 [DOI: 10.1109/AVSS.2018.8639077]
- Chen K, Wang J Q, Pang J M, Cao Y H, Xiong Y, Li X X, Sun S Y, Feng W S, Liu Z W, Xu J R, Zhang Z, Cheng D Z, Zhu C C, Cheng T H, Zhao Q J, Li B Y, Lu X, Zhu R, Wu Y, Dai J F, Wang J D, Shi J P, Ouyang W L, Loy C C and Lin D H. 2019. Mmdetection: open mmlab detection toolbox and benchmark [EB/OL]. [2020-09-25]. <https://arxiv.org/pdf/1906.07155.pdf>
- Chen Y H, Li W, Sakaridis C, Dai D X and Van Gool L. 2018a. Domain adaptive faster R-CNN for object detection in the wild//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 3339-3348 [DOI: 10.1109/CVPR.2018.00352]
- Chen Z H, Yang J X and Kang Z. 2018b. Moving ship detection algorithm based on gaussian mixture model//Proceedings of the 3rd International Conference on Modelling, Simulation and Applied Mathematics (MSAM 2018). Shanghai, China: Atlantis Press: 197-201 [DOI: 10.2991/msam-18.2018.42]
- Cheng B W, Xiao B, Wang J D, Shi H H, Huang T S and Zhang L. 2020. Bottom-up higher-resolution networks for multi-person pose estimation [EB/OL]. [2020-09-25]. <https://arxiv.org/pdf/1908.10357.pdf>
- Cruz G and Bernardino A. 2016. Aerial detection in maritime scenarios using convolutional neural networks//Proceedings of the 17th International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS). Lecce, Italy: Springer, 373-384 [DOI: 10.1007/978-3-319-48680-2_33]
- Everingham M, Eslami S M A, Van Gool L, Williams C K I, Winn J and Zisserman A. 2015. The pascal visual object classes challenge: a retrospective. *International Journal of Computer Vision*, 111 (1): 98-136 [DOI: 10.1007/s11263-014-0733-5]
- Everingham M, Van Gool L, Williams C K I, Winn J and Zisserman A. 2010. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88 (2): 303-338 [DOI: 10.1007/s11263-009-0275-4]
- Fan H X and Chen L K. 2019. Ship target detection algorithm based on improved deep learning//Proceedings of the 4th IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). Chengdu, China: IEEE: 286-290 [DOI: 10.1109/IAEAC47372.2019.8997993]
- Fefilatyev S. 2012. Algorithms for Visual Maritime Surveillance with Rapidly Moving Camera. Tampa: University of South Florida, USA: 28-44
- Fefilatyev S, Goldgof D, Shreve M and Lembke C. 2012. Detection and tracking of ships in open sea with rapidly moving buoy-mounted camera system. *Ocean Engineering*, 54: 1-12 [DOI: 10.1016/j.oceaneng.2012.06.028]
- Fefilatyev S, Smarodzinava V, Hall L O and Goldgof D B. 2006. Horizon detection using machine learning techniques//Proceedings of the 5th International Conference on Machine Learning and Applications (ICMLA'06). Orlando, USA: IEEE: 17-21 [DOI: 10.1109/ICMLA.2006.25]
- Frid-Adar M, Klang E, Amitai M, Goldberger J and Greenspan H. 2018. Synthetic data augmentation using GAN for improved liver lesion classification//The 15th IEEE International Symposium on Biomedical Imaging (ISBI 2018). Washington, USA: IEEE: 289-293 [DOI: 10.1109/ISBI.2018.8363576]
- Fu C Y, Liu W, Ranga A, Tyagi A and Berg A C. 2017. DSSD: deconvolutional single shot detector [EB/OL]. [2020-09-25]. <https://arxiv.org/pdf/1701.06659.pdf>
- Ganin Y and Lempitsky V. 2014. Unsupervised domain adaptation by backpropagation [EB/OL]. [2020-09-25]. <https://arxiv.org/pdf/1409.7495.pdf>
- Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, USA: IEEE: 580-587 [DOI: 10.1109/CVPR.2014.81]
- Girshick R, Radosavovic I, Gkioxari G, Dollár P and He K M. 2020. Detectron [CP/OL]. [2020-09-22]. <https://github.com/facebookresearch/detectron>
- Howard A, Sandler M, Chen B, Wang W J, Chen L C, Tan M X, Chu G, Vasudevan V, Zhu Y K, Pang R M, Adam H and Le Q. 2019. Searching for MobileNetV3//Proceedings of 2019 IEEE International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 1314-1324 [DOI: 10.1109/ICCV.2019.00140]

- Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. Mobilenets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2020-09-25]. <https://arxiv.org/pdf/1704.04861.pdf>
- Jiang W T, Zhang C, Zhang S C and Liu W J. 2019. Multiscale feature map fusion algorithm for target detection. *Journal of Image and Graphics*, 24(11): 1918-1931 (姜文涛, 张驰, 张晟翀, 刘万军. 2019. 多尺度特征图融合的目标检测. *中国图象图形学报*, 24(11): 1918-1931) [DOI: 10.11834/jig.190021]
- Kang M, Ji K F, Leng X G and Lin Z. 2017. Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection. *Remote Sensing*, 9(8): # 860 [DOI: 10.3390/rs9080860]
- Katayama T, Song T, Shimamoto T and Jiang X T. 2019. GAN-based color correction for underwater object detection//Proceedings of OCEANS 2019 MTS/IEEE SEATTLE. Seattle, USA; IEEE: 1-4 [DOI: 10.23919/OCEANS40490.2019.8962561]
- Katsaggelos A K, Lay K T and Galatsanos N P. 1993. A general framework for frequency domain multi-channel signal processing. *IEEE Transactions on Image Processing*, 2(3): 417-420 [DOI: 10.1109/83.236528]
- Kristan M, Perš J, Sulič V and Kovačič S. 2014. A graphical model for rapid obstacle image-map estimation from unmanned surface vehicles//Proceedings of the 12th Asian Conference on Computer Vision (ACCV). Singapore, Singapore; Springer: 391-406 [DOI: 10.1007/978-3-319-16808-1_27]
- Kristan M, Sulič Kenk V, Kovačič S and Perš J. 2016. Fast image-based obstacle detection from unmanned surface vehicles. *IEEE Transactions on Cybernetics*, 46(3): 641-654 [DOI: 10.1109/TCYB.2015.2412251]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA; IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot MultiBox detector//Proceedings of the 14th European Conference on Computer Vision (ECCV). Amsterdam, the Netherlands; Springer: 21-37 [10.1007/978-3-319-46448-0_2]
- Mikołajczyk A and Grochowski M. 2018. Data augmentation for improving deep learning in image classification problem//Proceedings of 2018 International Interdisciplinary PhD Workshop (IIPHDW). Świnoujście, Poland; IEEE: 117-122 [DOI: 10.1109/IIPHDW.2018.8388338]
- Moosbauer S, König D, Jäkel J and Teutsch M. 2019. A benchmark for deep learning based object detection in maritime environments//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach, USA; IEEE: 916-925 [DOI: 10.1109/CVPRW.2019.00121]
- Morris B, Aha D W, Auslander B and Gupta K. 2012. Learning and leveraging context for maritime threat analysis: vessel classification using Exemplar-SVM. ADA574666. Naval Research Lab Washington DC Navy Center for Applied Research in Artificial Intelligence
- Patino L, Cane T, Vallee A and Ferryman J. 2016. PETS 2016: dataset and challenge//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Las Vegas, USA; IEEE: 1240-1247 [DOI: 10.1109/CVPRW.2016.157]
- Prasad D K, Dong H X, Rajan D and Quek C. 2020. Are object detection assessment criteria ready for maritime computer vision? *IEEE Transactions on Intelligent Transportation Systems*, 21(12): 5295-5304 [DOI: 10.1109/TITS.2019.2954464]
- Prasad D K, Prasath C K, Rajan D, Rachmawati L, Rajabaly E and Quek C. 2016a. Challenges in video based object detection in maritime scenario using computer vision [EB/OL]. [2020-09-25]. <https://arxiv.org/pdf/1608.01079.pdf>
- Prasad D K, Rajan D, Prasath C K, Rachmawati L, Rajabally E and Quek C. 2016b. MSCM-LiFe: multi-scale cross modal linear feature for horizon detection in maritime images//Proceedings of 2016 IEEE Region 10 Conference (TENCON). Singapore, Singapore; IEEE: 1366-1370 [DOI: 10.1109/TENCON.2016.7848237]
- Prasad D K, Rajan D, Rachmawati L, Rajabally E and Quek C. 2017. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey. *IEEE Transactions on Intelligent Transportation Systems*, 18(8): 1993-2016 [DOI: 10.1109/TITS.2016.2634580]
- Ren S Q, He K M, Girshick R and Sun J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks//Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS). Montreal, Canada; MIT Press: 91-99
- Ribeiro R, Cruz G, Matos J and Bernardino A. 2019. A data set for airborne maritime surveillance environments. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(9): 2720-2732 [DOI: 10.1109/TCSVT.2017.2775524]
- Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetV2: inverted residuals and linear bottlenecks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA; IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]
- Schöller F E T, Plenge-Feidenhans' l M K, Stets J D and Blanke M. 2019. Assessing deep-learning methods for object detection at sea from LWIR images. *IFAC-PapersOnLine*, 52(21): 64-71 [DOI: 10.1016/j.ifacol.2019.12.284]
- Shao Z F, Wang L G, Wang Z Y, Du W and Wu W J. 2020. Saliency-aware convolution neural network for ship detection in surveillance video. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(3): 781-794 [DOI: 10.1109/TCSVT.2019.2897980]

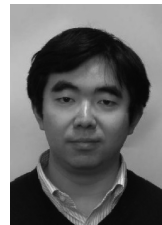
- Shao Z F, Wu W J, Wang Z Y, Du W and Li C Y. 2018. SeaShips: a large-scale precisely annotated dataset for ship detection. *IEEE Transactions on Multimedia*, 20(10): 2593-2604 [DOI: 10.1109/TMM.2018.2865686]
- Shen M Y, Yu P F, Wang R G, Yang J and Xue L X. 2019. Image super-resolution reconstruction via deep network based on multi-staged fusion. *Journal of Image and Graphics*, 24(8): 1258-1269 (沈明玉, 俞鹏飞, 汪荣贵, 杨娟, 薛丽霞. 2019. 多阶段融合网络的图像超分辨率重建. *中国图象图形学报*, 24(8): 1258-1269) [DOI: 10.11834/jig.180619]
- Stark H and Oskoui P. 1989. High-resolution image recovery from image-plane arrays, using convex projections. *Journal of the Optical Society of America A*, 6(11): 1715-1726 [DOI: 10.1364/JOSAA.6.001715]
- Stiefelhagen R, Bernardin K, Bowers R, Garofolo J, Mostefa D and Soundararajan P. 2007. The CLEAR 2006 evaluation//*Proceedings of the 1st International Evaluation Workshop on Classification of Events, Activities and Relationships (CLEAR)*. Southampton, UK: Springer; 1-44 [DOI: 10.1007/978-3-540-69568-4_1]
- Su H, Zhou J and Zhang Z H. 2013. Survey of super-resolution image reconstruction methods. *Acta Automatica Sinica*, 39(8): 1202-1213 (苏衡, 周杰, 张志浩. 2013. 超分辨率图像重建方法综述. *自动化学报*, 39(8): 1202-1213) [DOI: 10.3724/SP.J.1004.2013.01202]
- Sun K, Zhao Y, Jiang B R, Cheng T H, Xiao B, Liu D, Mu Y D, Wang X G, Liu W Y and Wang J D. 2019. High-resolution representations for labeling pixels and regions [EB/OL]. [2020-09-25]. <https://arxiv.org/pdf/1904.04514.pdf>
- Wang B, Dong L L, Zhao M and Xu W H. 2017. Fast infrared maritime target detection: binarization via histogram curve transformation. *Infrared Physics and Technology*, 83: 32-44 [DOI: 10.1016/j.infrared.2017.03.009]
- Westall P, Ford J J, O'Shea P and Hrabar S. 2008. Evaluation of maritime vision techniques for aerial search of humans in maritime environments//*Proceedings of 2008 Digital Image Computing: Techniques and Applications*. Canberra, Australia: IEEE: 176-183 [DOI: 10.1109/DICTA.2008.89]
- Ye Y C, Jiang X M, Pan G F and Jiang W. 2015. *Submarine Fiber Optic Cable Engineering*. Beijing: China Ocean Press: 197-198 (叶银灿, 姜新民, 潘国富, 江伟. 2015. 海底光缆工程. 北京: 海洋出版社: 197-198)
- Zhang H and Wang N Y. 2017. On the stability of video detection and tracking [EB/OL]. [2020-09-25]. <https://arxiv.org/pdf/1611.06467.pdf>
- Zhang X H, Yao L B, Lyu Y F, Jian T, Zhao Z W and Zang J. 2020. Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images. *Journal of Image and Graphics*, 25(9): 1943-1952 (张筱哈, 姚力波, 吕亚飞, 简涛, 赵志伟, 藏洁. 2020. 双向特征融合的数据自适应 SAR 图像舰船目标检测模型. *中国图象图形学报*, 25(9): 1943-1952) [DOI: 10.11834/jig.190558]
- Zhang Y, Li Q Z and Zang F N. 2017. Ship detection for visual maritime surveillance from non-stationary platforms. *Ocean Engineering*, 141: 53-63 [DOI: 10.1016/j.oceaneng.2017.06.022]
- Zhou X Y, Wang D Q and Krähenbühl P. 2019. Objects as points [EB/OL]. [2020-09-24]. <https://arxiv.org/pdf/1904.07850.pdf>
- Zou J J, Yuan W and Yu M H. 2019. Maritime target detection of intelligent ship based on faster R-CNN//*Proceedings of 2019 Chinese Automation Congress (CAC)*. Hangzhou, China: IEEE: 4113-4117 [DOI: 10.1109/CAC48633.2019.8996260]

作者简介



叶晨, 1980 年生, 男, 高级工程师, 主要研究方向为机器学习、图像处理、大数据分析及其在工业智能领域的应用。

E-mail: yechen@tongji.edu.cn



陆海, 通信作者, 男, 工程师, 主要研究方向为深度学习及多源数据融合。

E-mail: lh@tongji.edu.cn

逯天洋, 男, 硕士研究生, 主要研究方向为深度学习目标检测。E-mail: tylu@alumni.tongji.edu.cn

肖潇灏, 男, 硕士研究生, 主要研究方向为深度学习目标检测。E-mail: 2033089@tongji.edu.cn

杨群慧, 女, 研究员, 主要研究方向为海洋观测技术。E-mail: yangqh@tongji.edu.cn