



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

中国图象图形学报

2020

09

VOL.25

ISSN1006-8961
CN11-3758/TB



遥感边缘智能技术 P1719



第25卷第9期 (总第293期)

2020年9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 顾行发

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@radi.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 Gu Xingfa

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@radi.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

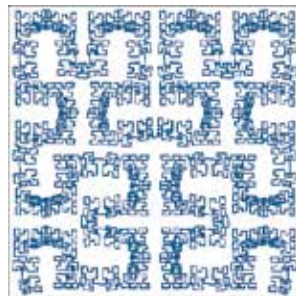
国外发行代号 M1406

国内邮发代号 82-831

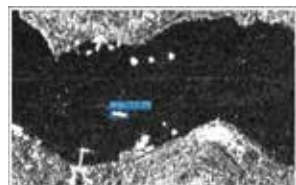
国内定价 60.00元



第一人称视角下的社会力优化多行人跟踪(第1869页)



分形曲线生成的频域方法(第1904页)



双向特征融合的数据自适应SAR图像舰船目标检测模型(第1943页)

学者观点

遥感边缘智能技术研究进展及挑战

孙显, 梁伟, 刁文辉, 曹志颖, 冯瑛超, 王冰, 付琨 1719

综述

红外弱小目标检测算法综述

李俊宏, 张萍, 王晓玮, 黄世泽 1739

目标检测中的尺度变换应用综述

申奉臻, 张萍, 罗金, 刘松阳, 冯世杰 1754

图像处理和编码

分层特征融合注意力网络图像超分辨率重建

雷鹏程, 刘丛, 唐坚刚, 彭敦陆 1773

适用小样本的无参考水下视频质量评价方法

宋巍, 刘诗梦, 黄冬梅, 王文娟, 王建 1787

图像分析和识别

结合面料属性和触觉感测的织物识别

邢寅初, 刘骊, 付晓东, 刘利军, 黄青松 1800

结合GAN的轻量级模糊车牌识别算法

段宾, 符祥, 江毅, 曾接贤 1813

非贪婪的鲁棒性度量学习算法

曾凡霞, 张文生 1825

不规则像素簇显著性检测算法

李明旭, 翟东海 1837

全局区域相异度阈值驱动构建的稀疏尺度集模型

王浩宇, 沈占锋, 柯映明, 许泽宇, 李硕, 焦淑慧 1848

图像理解和计算机视觉

分裂合并运动分割的多运动视觉里程计方法

王晨捷, 张云, 赵青, 王伟, 尹露, 罗斌, 张良培 1859

第一人称视角下的社会力优化多行人跟踪

杨廷召, 刘骊, 付晓东, 刘利军, 黄青松 1869

引入辅助损失的多场景车道线检测

陈立潮, 徐秀芝, 曹建芳, 潘理虎 1882

计算机图形学

旋转骨架法在二值图像分形维数计算中的应用

纪佑军, 何杰, 韩海水, 程忠钊, 曾保全 1894

分形曲线生成的频域方法

陈伟, 乔洁雯, 周晨 1904

医学图像处理

融合残差注意力机制的UNet视盘分割

侯向丹, 赵一浩, 刘洪普, 郭鸿湧, 于习欣, 丁梦园 1915

深层聚合残差密集网络的超声图像左心室分割

吴宣言, 缙新科, 朱子重, 魏域林, 王凯 1930

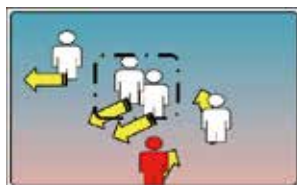
遥感图像处理

双向特征融合的数据自适应SAR图像舰船目标检测模型

张筱晗, 姚力波, 吕亚飞, 简涛, 赵志伟, 藏洁 1943

CONTENTS

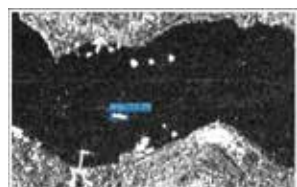
JOURNAL OF IMAGE AND GRAPHICS



Multi-pedestrian tracking optimized by social force model under first-person perspective(P1869)



Fractal curve generation method based on the frequency domain (P1904)



Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images (P1943)

Scholar View

Progress and challenges of remote sensing edge intelligence technology

Sun Xian, Liang Wei, Diao Wenhui, Cao Zhiying, Feng Yingchao, Wang Bing, Fu Kun 1719

Review

Infrared small-target detection algorithms: a survey

Li Junhong, Zhang Ping, Wang Xiaowei, Huang Shize 1739

Scale changing in general object detection: a survey

Shen Fengcan, Zhang Ping, Luo Jin, Liu Songyang, Feng Shijie 1754

Image Processing and Coding

Hierarchical feature fusion attention network for image super-resolution reconstruction

Lei Pengcheng, Liu Cong, Tang Jiangang, Peng Dunlu 1773

Non-reference underwater video quality assessment method for small size samples

Song Wei, Liu Shimeng, Huang Dongmei, Wang Wenjuan, Wang Jian 1787

Image Analysis and Recognition

Cloth recognition based on fabric properties and tactile sensing

Xing Yinchu, Liu Li, Fu Xiaodong, Liu Lijun, Huang Qingsong 1800

Lightweight blurred car plate recognition method combined with generated images

Duan Bin, Fu Xiang, Jiang Yi, Zeng Jiexian 1813

Robust distance metric learning algorithm based on nongreedy solution

Zeng Fanxia, Zhang Wensheng 1825

Significance detection method with irregular pixel clusters

Li Mingxu, Zhai Donghai 1837

Sparse scale set model based on global regional dissimilarity threshold

Wang Haoyu, Shen Zhanfeng, Ke Yingming, Xu Zeyu, Li Shuo, Jiao Shuhui 1848

Image Understanding and Computer Vision

Multi-motion visual odometry based on split-merged motion segmentation

Wang Chenjie, Zhang Yun, Zhao Qing, Wang Wei, Yin Lu, Luo Bin, Zhang Liangpei 1859

Multi-pedestrian tracking optimized by social force model under first-person perspective

Yang Tingzhao, Liu Li, Fu Xiaodong, Liu Lijun, Huang Qingsong 1869

Multi-scenario lane line detection with auxiliary loss

Chen Lichao, Xu Xiuzhi, Cao Jianfang, Pan Lihu 1882

Computer Graphics

Application of skeleton rotating method in fractal dimension calculate of binary image

Ji Youjun, He Jie, Han Haishui, Cheng Zhongzhao, Zeng Baoquan 1894

Fractal curve generation method based on the frequency domain

Chen Wei, Qiao Jiewen, Zhou Chen 1904

Medical Image Processing

Optic disk segmentation by combining UNet and residual attention mechanism

Hou Xiangdan, Zhao Yihao, Liu Hongpu, Guo Hongyong, Yu Xixin, Ding Mengyuan 1915

Left ventricular segmentation on ultrasound images using deep layer aggregation for residual dense networks

Wu Xuanyan, Gou Xinke, Zhu Zizhong, Wei Yulin, Wang Kai 1930

Remote Sensing Image Processing

Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images

Zhang Xiaohan, Yao Libo, Lyu Yafei, Jian Tao, Zhao Zhiwei, Zang Jie 1943

中图法分类号: TP751 文献标识码: A 文章编号: 1006-8961(2020)09-1943-10

论文引用格式: 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(09): 1943-1952 (张筱晗, 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(09): 1943-1952) [DOI: 10. 11834/jig. 190558]

双向特征融合的数据自适应 SAR 图像舰船目标检测模型

张筱晗¹, 姚力波¹, 吕亚飞¹, 简涛¹, 赵志伟², 藏洁²

1. 海军航空大学信息融合研究所, 烟台 264001; 2. 北京空间飞行器总体设计部, 北京 100089

摘要: 目的 利用合成孔径雷达(synthetic aperture radar, SAR)图像进行舰船目标检测是实施海洋监视的重要手段。基于深度学习的目标检测模型在自然图像目标检测任务中取得了巨大成功,但由于自然图像与 SAR 图像的差异,不能将其直接迁移到 SAR 图像目标检测中。针对 SAR 图像目标检测实际应用中对速度和精度的需求,借鉴经典的单阶段目标检测模型(single shot detector, SSD)框架,提出一种基于特征优化的轻量化 SAR 图像舰船目标检测网络。**方法** 改进模型并精简网络结构,提出一种数据驱动的目标分布聚类算法,学习 SAR 数据集的目标尺度、长宽比分布特性,用于网络参数设定;对卷积神经网络(convolutional neural network, CNN)提取的特征进行优化,提出一种双向高低层特征融合机制,将高层特征的语义信息通过语义聚合模块加成到低层特征中,在低层特征中提取特征平均图,处理后作为高层特征的注意力权重图对高层特征进行逐像素加权,将低层特征丰富的空间信息融入到高层特征中。**结果** 利用公开的 SAR 舰船目标检测数据集(SAR ship detection dataset, SSDD)进行实验,与原始的 SSD 模型相比,轻量化结构设计在不损失检测精度的前提下,样本测试时间仅为 SSD 的 65%;双向特征融合机制将平均精确度(average precision, AP)值由 77.93% 提升至 80.13%,训练和测试时间分别为 SSD 的 64.1% 和 72.6%;与公开的基于深度学习的 SAR 舰船目标检测方法相比,本文方法在速度和精度上都取得了最佳性能,AP 值较精度次优模型提升了 1.23%,训练和测试时间较精度次优模型分别提升了 559.34 ms 和 175.35 ms。**结论** 实验充分验证了本文所提模型的有效性,本文模型兼具检测速度与精度优势,具有很强的实用性。

关键词: 合成孔径雷达图像;舰船目标检测;聚类;特征融合;注意力机制

Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images

Zhang Xiaohan¹, Yao Libo¹, Lyu Yafei¹, Jian Tao¹, Zhao Zhiwei², Zang Jie²

1. Information Fusion Institute, Naval Aviation University, Yantai 264001, China;

2. Beijing Space Vehicle General Design Department, Beijing 100089, China

Abstract: **Objective** Ship detection plays an important role in civil and military fields, including marine object identification, maritime transportation, rescue operation, marine security, and disaster relief. As a basic means of marine monitoring, ship detection in synthetic aperture radar (SAR) images has been studied for years. With the development of sensor

收稿日期: 2019-11-14; 修回日期: 2020-03-21; 预印本日期: 2020-03-28

基金项目: 国家自然科学基金项目(91538201, 61790554, 61971432, 61531020); 山东省泰山学者工程资助项目(tsqn201909156); 山东省高等学校青创科技支持计划项目(2019KJN031)

Supported by: National Natural Science Foundation of China (91538201, 61790554, 61971432, 61531020)

and platform technologies, SAR big data are achieved, making it possible to perform automatic data-driven detection algorithms. Deep learning-based detection models have been proven to be a great success in common object detection tasks for natural scene images; moreover, it outperforms many traditional artificial feature based methods. However, when transferring them to SAR ship detection directly, many challenges emerge, and the results are not satisfying because natural and SAR images have several differences. Ship in SAR images usually appear as some bright parts and lack detail information in comparison with natural images because of the coherent imaging mechanism. The swath of SAR remote sensing images is large, but targets are distributed densely or sparsely; thus, the processing of SAR images is usually more complex than that of natural ones. In addition, the size and shape of ship targets vary, ranging from several pixels to hundreds. All these factors complicate ship detection in SAR images. Aiming to solve these challenges and considering the actual demands in practice, this study proposes a lightweight data-adaptive detector with a feature-optimizing mechanism on the basis of the famous single-shot detector (SSD) to improve detection precision and speed.

Method In this study, the original SSD is modified by having the number of channels halved and the last two convolution blocks removed. The settings of the network parameters follow the outputs of proposed data-driven target distribution clustering algorithm, which learns the distributions of targets in the SAR dataset, including the size of ships and the aspect ratio of ships. The algorithm is free from human experience and can make the detector adapt to the SAR dataset. Trunked visual geometry group 16-layer net (VGG16) is utilized to extract features from input SAR images. Given that the features extracted by convolutional neural networks are hierarchical, low-level features with high spatial resolution usually contain extra local and spatial detail information, whereas more semantic and global information are involved in high-level features with low resolution. For object detection tasks, spatial and semantic information are important. Thus, information must be aggregated through a fusion strategy. A new bidirectional feature fusion mechanism, which contains a semantic aggregation and a novel attention guidance module, is proposed. In feature pyramid networks, the higher features are added to the lower features after an upsampling operation. On this basis, the up-sampled higher features in our model are concatenated with lower features in the channel dimension, and the channel numbers are adjusted through a 1×1 convolution operation. Instead of simply adding lower features to higher features, an inverse fusion from down to top and an attention mechanism are applied. A spatial attention map of each convolution block is generated, and the attention map that contains the most spatial information is selected as a weighted map. In the weighted map, target pixels with higher value are usually more noticeable, whereas the value of background pixels are suppressed. After down sample to weight map, element-wise multiplication is performed between the weighted map and the higher features. The features of the targets are strengthened; thus, spatial information is passed to higher level features. The optimized features are then entered into detector heads to predict the locations and types of targets; the low-level features mainly detect the small ships, whereas the high-level features are responsible for the large ones. The entire network is trained by a weighted sum of location and classification losses. In inference, nonmaximum suppression is used for removing repeated bounding boxes.

Result The public SAR ship detection dataset widely used in SAR ship detection references is adopted in experiments. All the experiments are implemented using Python language under the TensorFlow framework on a 64-bit computer with Ubuntu 16.06, CPU Intel (R) Core (TM) i7-6770K @ 4.00 GHz $\times$ 8, and NVIDIA GTX 1080Ti with CUDA9.0 and cuDNN7.0 for acceleration. The training iteration, initial learning rate, and batch size are set as 120 k, 0.000 1, and 24, respectively. A momentum optimizer is used, with weight decay, gamma, and momentum values of 0.000 5, 0.1, and 0.9, respectively. An ablation study is operated to verify the effectiveness of each proposed module, and the model is compared with five published state-of-art methods. Precision rate, recall rate, average precision (AP), and the average training and testing time on a single image, are taken as evaluation indicators. In the original SSD, a model with parameters from the proposed data-driven target distribution clustering algorithm improves the AP by 1.08% in comparison with the model with original parameters. The lightweight design of the network significantly improves the detection speed; compared with that of the SSD, the training and testing time of the proposed model decrease from 20.79 ms to 12.74 ms and from 14.02 ms to 9.17 ms, respectively. The semantic aggregation and attention fusing modules can improve detection precision, whereas when the two modules are used together, the optimum performance in detection precision is achieved. The AP increased from 77.93% to 80.13%, and the precision and recall rates increased from 89.54% to 96.68% and from 88.60% to 89.60%, respectively. However, speed is not considerably affected, and the model still runs faster than SSD.

The proposed model outperforms other models in terms of precision and speed; moreover, it improves AP by 6.9%, 1.23%, 9.09%, and 2.9% in comparison with other four methods. **Conclusion** In this study, we proposed a lightweight data adaptive single shot detector with feature optimizing mechanism. Experiment results show that our model have remarkable advantages over other published state-of-the-art detection approaches in terms of precision and speed.

Key words: synthetic aperture radar (SAR) image; ship detection; cluster; feature fusion; attention mechanism

0 引言

舰船目标检测是实施海洋监视的重要环节,在军事侦察、海洋运输管理、海上犯罪打击等领域都有着重要应用(Allard等,2008)。得益于全天时、全天候的优势,合成孔径雷达(synthetic aperture radar, SAR)图像成为一种有效的舰船目标遥感数据源, SAR 舰船目标检测得到了广泛研究。传统的基于海杂波建模思想的 SAR 舰船检测方法通常针对某一场景进行建模(Dalal 和 Triggs, 2005; Agrawal 等, 2015; Pappas 等, 2018),无法同时兼顾多类应用场景。而基于数据驱动的深度学习目标检测方法则不受场景影响,只要提供充足多样的数据,即可训练深度网络自动提取图像特征,实现目标检测。

得益于计算机技术与人工智能技术的发展,在计算机视觉领域,基于深度学习的目标检测方法取得了很多检测任务的最佳性能。经典的目标检测模型,包括以 Faster RCNN(region convolutional neural network)(Ren 等,2017)为代表的双阶段检测模型、以 YOLO(you only look once)(Redmon 等,2015)和 SSD(single shot multi-box detector)(Liu 等,2016)为代表的单阶段检测模型都被引入到 SAR 舰船检测中。与传统方法相比,这些模型多是端到端的,并且在精度和泛化能力上有了很大提升。但是,将这些模型从自然图像目标检测迁移到 SAR 舰船目标检测时仍面临着一些问题:1)这些模型参数,如网络层数和锚框设置多是针对 Pascal VOC(visual object classes)(Everingham 等,2015)和 MS COCO(Lin 等, 2014)等自然图像数据集设定的,而这些设置未必适合 SAR 舰船数据;2)数据量相对较小的检测数据集往往不足以从头训练规模较大的检测模型,通常做法是在检测模型初始化时,将特征提取部分直接载入大型自然图像分类数据集(如 ImageNet 等)预训练好的模型参数,然后使用检测数据集进行微调;3)分类任务与检测任务、自然分类数据集与 SAR 目

标检测数据集存在偏差;4)SAR 图像自身的一些特点给检测增加了难度,如受空间分辨率与成像机理影响,图像中舰船目标细节信息少,各类舰船尺度不一、分布各异等。在设计 SAR 舰船目标检测模型时,要考虑这些问题。Kang 等人(2017)和 Yang 等人(2018)根据使用的遥感舰船数据重新设置了检测模型参数。Deng 等人(2019)利用密集连接网络作为特征提取网络,并采用了特征重用等策略,可以从头训练。为了提升 SAR 舰船目标检测的精度,学者利用多种方式对网络提取的特征进行优化,其中,特征融合是一项常用策略(Jiao 等,2018; Zhang 等, 2019)。自从 Lin 等人(2017)提出特征金字塔网络(feature pyramid network, FPN),其高低特征层融合的思想大量用于多种模型。卷积神经网络(convolutional neural network, CNN)在提取特征过程中,随着网络层数的加深与池化层的应用,深特征层有着更大的感受野,能够提取出更多抽象的全局语义特征,而浅层特征空间信息更加丰富。为弥补浅层特征语义信息的不足,将高层特征处理后加入到浅层特征。Wang 等人(2018)将 FPN 中的语义聚合策略应用在 SAR 舰船检测中,与原始的 SSD 模型相比,显著提升了舰船检测精度。Kang 等人(2017)将浅层特征、中层特征与高层特征都融合起来,输入到 Faster RCNN 模型。特征融合的策略能够提升模型性能,但也增加了运算量,此外,大部分融合模型的融合通道是由上至下的,没有考虑高层信息中空间信息的不足。

针对上述问题,本文对 SAR 图像舰船目标检测进行研究,针对 SAR 图像的特点改进了经典检测模型 SSD,同时提升了目标检测的速度和精度。本文工作包括两部分:1)根据 SAR 数据的特点,精简了 SSD 模型结构,提出了一种数据集目标特性聚类算法,用于指导网络参数的设置;2)借鉴特征融合思想,提出一种双向特征融合机制,除了利用语义聚合增强浅层特征的语义信息,同时设计了一种新的基于注意力机制的融合方法,将浅层特征丰富的空间信息加成到高层特征。前者主要改进模型适应

SAR 舰船检测任务,后者进一步优化模型,提升检测精度。在公开 SAR 舰船目标检测数据集 SSDD (SAR ship detection dataset) (Li 等,2017) 上的实验充分验证了本文方法的有效性。

1 LSSD 检测模型

1.1 模型结构

本文模型是在经典的 SSD 模型基础上改进的,并以 VGG-16(visual geometry group 16)作为特征提取网络。标准的 SSD-300 模型(输入图像大小为 $300 \times 300 \times 3$)保留了 VGG-16 的前 5 个卷积单元,去掉全连接层,并继续在后面连接了 6 个卷积单元,最终卷积单元输出的特征图空间维变成 1×1 。但是,对 SAR 图像来说,某种极化模式下获取的图像是单通道的。在使用通用检测模型时,通常做法是将单通道 SAR 图像扩展为 3 通道作为输入,这种

做法增加了网络的冗余信息。此外,SAR 遥感图像中的舰船目标尺度变化很大,所占像素从几个到几百不等,但受空间分辨率影响,单个舰船目标通常不会占满图像(300×300 像素标准输入大小)。因此,对模型进行更改,直接以单通道图像作为输入,去掉 SSD 模型的后 3 个卷积模块,并将全部卷积层的通道维减半。这样,与 SSD 相比,网络需要训练的参数数量少于原来的 $1/4$,模型显著变轻量化,将其命名为 LSSD (lightweight SSD)。经验证,即使是较小的检测数据集如 SSDD 也可以从头训练 LSSD,使其不必依赖于分类任务预训练模型。借鉴 SSD 模型的设置,同时考虑 SAR 图像中的小舰船目标,选取 conv3_3、conv4_3、conv7、conv8 输出的特征图输入到检测器中进行后续检测。

本文模型整体框架如图 1 所示,图中右侧为 LSSD 网络的具体设置,图中 S 代表语义聚合,A 代表注意力引导的融合。

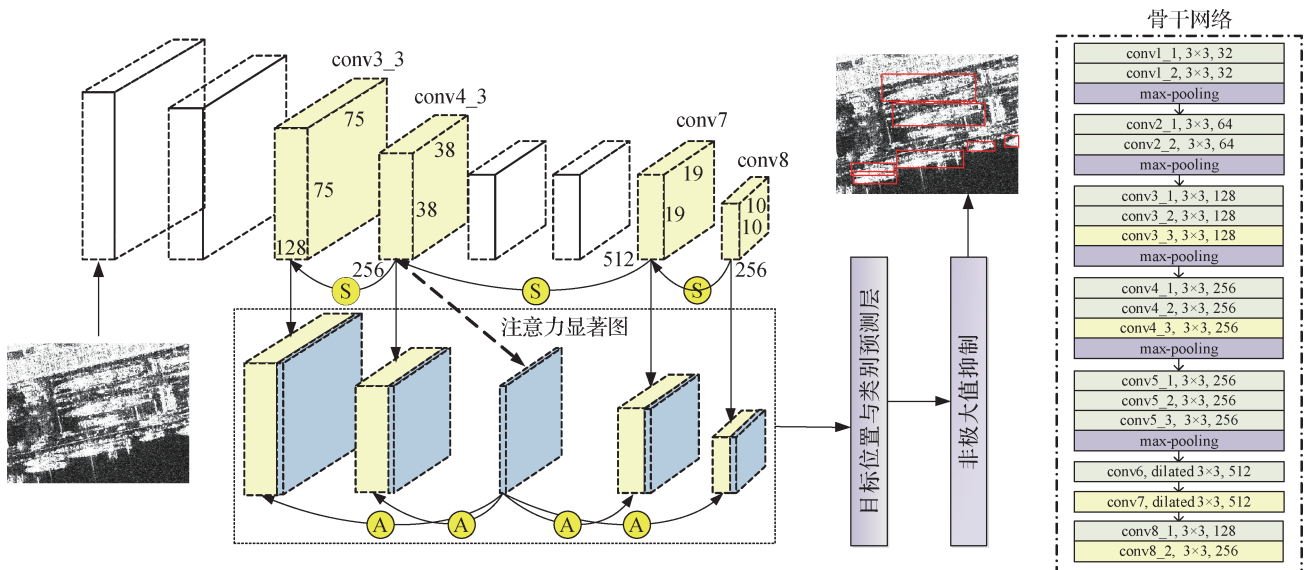


图 1 本文模型框架图

Fig. 1 Framework of proposed model

1.2 基于聚类算法的数据集目标特性统计

与 Faster RCNN 和 SSD 模型相同,LSSD 也是基于锚框(anchor)实施检测。锚框的设置包括尺度和比例,能够反映数据集中目标的尺度和比例分布特性,对网络检测性能具有重要影响。受 YOLOv2 (Redmon 和 Farhadi,2017) 中的维度聚类算法启发,本文设计了一种基于聚类的目标特性分布自动学习算法,具体内容以 SSDD 数据集为例进行说明。

假设图像大小为 $M \times N$,舰船目标检测框大小

为 $w \times h$ 。目标尺度 s 和长宽比 r 的定义为

$$s = \frac{w \times h}{M \times N} \quad (1)$$

$$r = \frac{w}{h} \quad (2)$$

1) 学习数据集中舰船目标的尺度分布。假设数据集中所有舰船目标的尺度集合为 $S: \{s_1, s_2, \dots, s_k\}$, 设定 k 个尺度聚类中心 $U: \{\mu_1, \mu_2, \dots, \mu_k\}$, 其中 μ_i 初值为 $[0, 1]$ 区间的随机值, μ_1 到 μ_k 从小

到大排列。通过最小化目标各尺度与聚类中心欧氏距离之和 d_s , 以更新 U 的参数, 即

$$\min d_s = \sum_{i=1}^n \|s_i - \mu_j\|^2 \quad (3)$$

式中, $j \in \{1, 2, \dots, k\}$, μ_j 为最接近 s_i 的聚类中心。每次迭代中, μ_j 更新为以其为聚类中心的所有目标尺度的均值, 不断迭代直到 d_s 小于设定阈值。迭代完成后, 将学习到的尺度聚类中心 U 分配到合适的特征层中, 较小的尺度分配给低的特征层, 大的锚框尺度则分配给高特征层。

2) 学习数据集中舰船目标的长宽比分布特性。设定特征层 l_x 的长宽比聚类中心个数为 g , 尺度接近该层各尺度聚类中心的舰船目标长宽比构成集合 $R_x: \{r_{x1}, r_{x2}, \dots\}$, g 个长宽比聚类中心构成集合 $V_x: \{\tau_1, \tau_2, \dots, \tau_g\}$, τ_i 初值设为随机正数, 通过最小化目标长宽比到各自聚类中心欧氏距离之和 d_{rs} 进行优化, 即

$$\min d_{rs} = \sum_{xi=1}^{n_x} \|r_{xi} - \tau_j\|^2 \quad (4)$$

式中, $j \in \{1, 2, \dots, g\}$, τ_j 为最接近 r_{xi} 的长宽比聚类中心。每次迭代中, τ_j 更新为聚类中心目标长宽

比的均值。不断进行迭代, 直到 d_{rs} 小于设定阈值。最后根据长宽比聚类中心的值设置该层锚框的长宽比参数。

2 基于双向融合的特征优化

2.1 原理分析

卷积神经网络提取的分层特征有着不同的特点。与 SSD 一样, 为检测多尺度目标, LSSD 也使用了多层特征。其中浅层特征通常包含更多目标细节信息, 缺乏高级语义信息; 深层特征信息更加抽象, 但缺乏空间位置信息。以 FPN 网络为代表的从上到下的语义聚合的特征融合方式已经得到了广泛应用, 针对如何将浅层丰富的空间信息传递到深层特征, 本文提出了基于注意力机制的融合方式。

为直观理解 CNN 提取的特征, 本文将训练好的 LSSD 模型中不同层的特征进行可视化。图 2 是训练好的 LSSD 模型提取的两幅 SAR 图像 conv3_3、conv4_3、conv7、conv8 的第 1 通道和第 64 通道的特征图及各通道平均热力图。可以看出, 针对输入的特定图像, 不同卷积层在不同通道提取了不同特征,

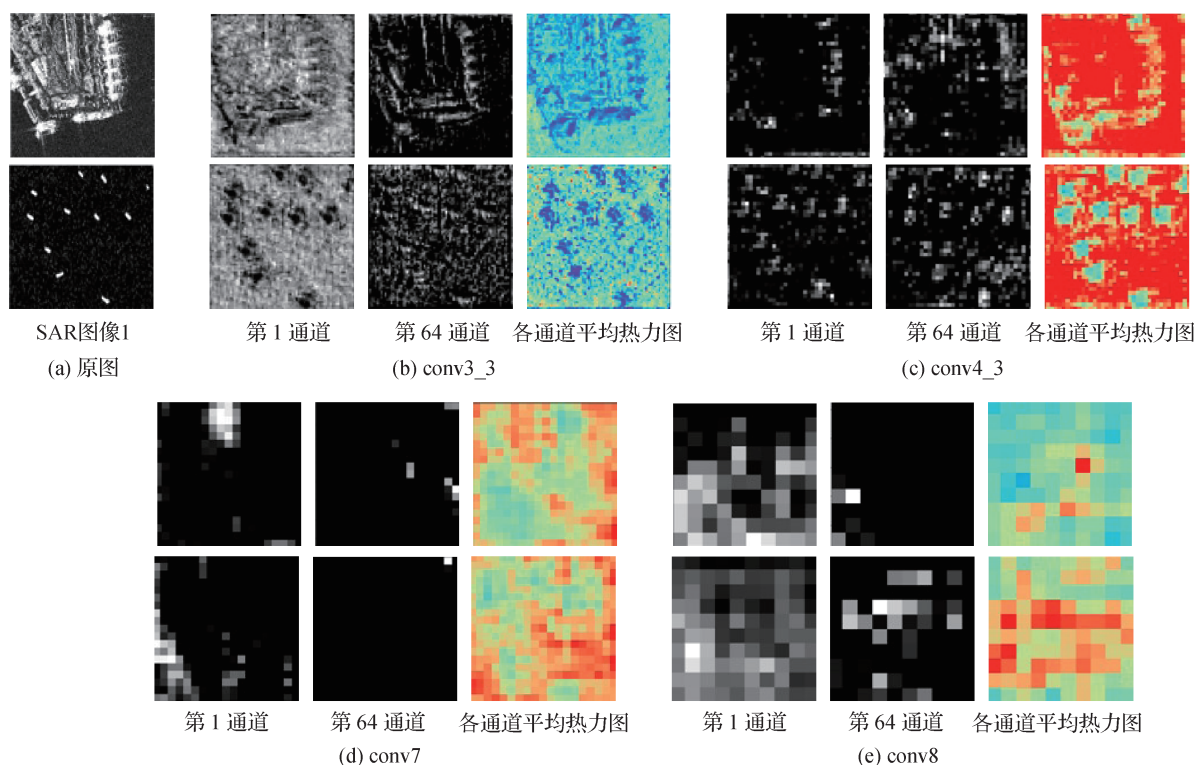


图2 训练好的 LSSD 模型从 SAR 图像提取的特征图

Fig. 2 Feature maps extracted by our trained model from SAR image

((a) original images; (b) conv3_3; (c) conv4_3; (d) conv7; (e) conv8)

但很难解释特定通道特征的含义。为展示每个卷积层提取的整体特征,在特征层通道维进行全局平均池化,得到特征的空间维平均图,并以热图的形式展示。从平均图可以看出,随着特征层的增深,平均特征中关于目标的表征越来越抽象。其中,在 conv4_3 的特征平均图中,无论背景是较为简单的水面,还是更复杂的陆地,舰船目标与背景分离最明显。受计算机视觉中注意力机制(Woo 等,2018)启发,将 conv4_3 特征平均图视为空间注意力显著图,指导网络更关注的位置。可以直接利用 conv4_3 特征平均图作为权重图,去对其余更深的特征层进行加权,以这种方式将低特征层的目标位置信息传递到更深的特征层中,实现特征自下而上的融合。

2.2 特征双向融合实施流程

采用的双向特征融合操作流程如图3所示。首先进行自上而下的语义聚合,然后实施注意力机制引导下的自下而上的特征融合。语义聚合模型(图3(a))主要包括深层特征上采样、特征图通道维串联以及 1×1 卷积修正通道数等过程,最终融合特征与浅层特征在空间维和通道维的尺寸相同。基于注意力引导的特征融合模块如图3(b)所示,来自 conv4_3 的特征图在通道维进行平均池化,得到大小为 p 的平均特征图 F ,作为融合中的注意力图。 F 通过 sigmoid 函数进行归一化,得到注意力权重图,具体计算为

$$F(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

式中, x 代表 F 中各像素值, $F(x) \in (0,1)$ 。该权重图被用来对其余特征层进行加权。融合前,需要通过下采样等方式将权重图的大小改为高层特征图的空间大小,然后对高层特征的每个通道的特征图进行逐元素相乘的操作,得到融合特征图。注意该操作只自下而上进行,conv4_3 特征图得到的注意力权重图不再影响 conv3_3。

3 实验与结果分析

对本文方法利用公开的 SSDD 数据集进行实验,首先利用基于聚类的数据集目标特性统计方法指导的网络参数设置与原 SSD 网络参数设置进行对比,然后逐步对比本文方法每一模块的效果,最后将本文方法与其他 SAR 舰船目标检测的方法进行对比。

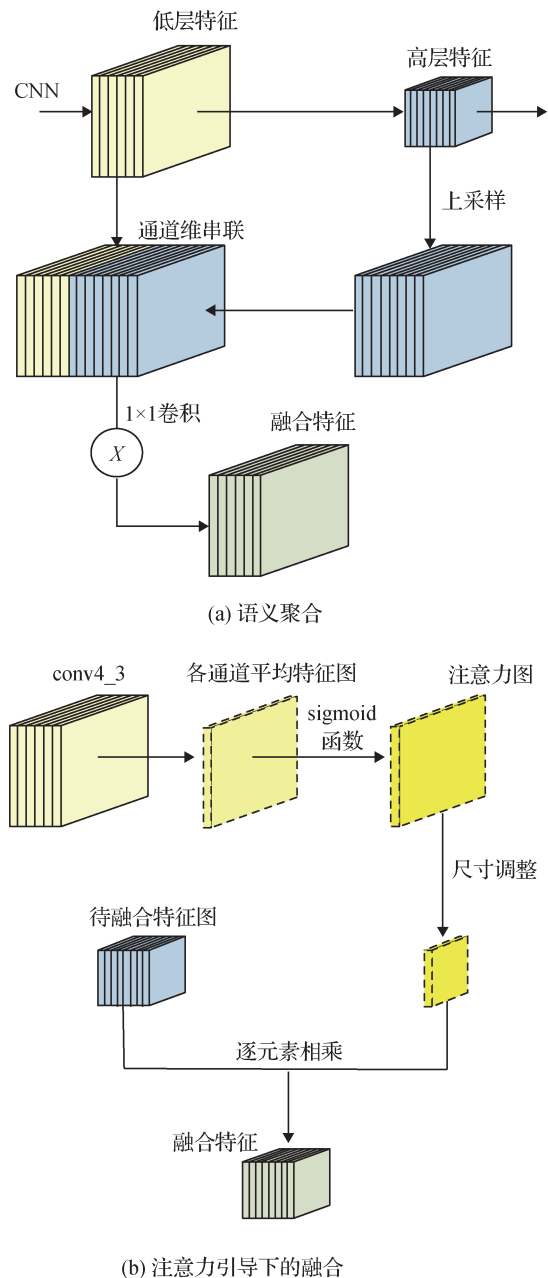


图3 特征融合算法流程图

Fig. 3 Framework of our feature fusion module

- ((a) semantic aggregation block;
(b) attention guided fusion block)

3.1 数据集说明与实验设置

SSDD 数据集是国内第1个公开的 SAR 舰船目标检测数据集,包含1160幅图像与2456个舰船实例。实验平台操作系统为64位单 GTX 1080Ti GPU 计算机,操作系统为 Ubuntu 16.06,编程语言为 Python,使用的深度学习框架为 TensorFlow,并使用 CUDA9.0 与 cuDNN7.0 加速。训练次数设为12万次,前2万次初始学习率为0.0001,接下来6万次学习率

设为 0.000 01, 最后 4 万次学习率设为 0.000 001。训练中, 批次容量设为 24, 使用 momentum 训练优化器。

为评价测试结果, 采用精确率 (precision)、召回率 (recall) 以及平均精确率 (average precision, AP) 作为客观评价指标, 指标的具体定义与 Kang 等人 (2017) 采用的定义相同。精确率和召回率是在交并比 (intersection-over-union, IoU) 阈值设为 0.5、分类得分阈值设为 0.2 的情况下计算的。这些指标值越高, 说明算法检测性能越好。此外, 利用单幅图像的平均训练与测试时间衡量算法的运行速度。

3.2 实验结果

3.2.1 网络参数设置对检测结果的影响

实验采用两组参数设置在 SSD 模型上进行。一组为 SSD 原始参数, 锚框采用 SSD 模型中的 6 层特征, 如表 1 所示; 另一组为使用本文提出的基于聚类的数据集目标特性学习方法为指导的网络参数, 锚框采用 LSSD 模型中的 4 层特征, 如表 2 所示。

表 1 采用 SSD 模型中 6 层特征的锚框参数

Table 1 Anchor settings of SSD
in the 6 feature layers

卷积层	尺度	长宽比	额外尺度
conv4_3	0.1	1:1, 1:2, 2:1	0.141 4
conv7	0.2	1:1, 1:2, 2:1, 1:3, 3:1	0.273 9
conv8_2	0.375	1:1, 1:2, 2:1, 1:3, 3:1	0.454 1
conv9	0.55	1:1, 1:2, 2:1, 1:3, 3:1	0.631 5
conv10	0.725	1:1, 1:2, 2:1	0.807 8
conv11	0.9	1:1, 1:2, 2:1	0.983 6

表 2 采用 LSSD 模型中 4 层特征的锚框参数

Table 2 Anchor settings of LSSD
in the 4 feature layers

卷积层	尺度	长宽比	额外尺度
conv3_3	0.005	1:1, 1:2, 2:1	0.01
conv4_3	0.020	1:1, 4:7, 20:11	0.04
conv7	0.060	1:1, 4:7, 3:1, 1:4, 20:11	0.12
conv8_2	0.250	1:1, 4:7, 3:1, 1:4, 20:11	0.40

两组实验除锚框设置不同外, 其余实验条件均相同, 实验结果如表 3 所示。可以看出, 锚框的参数设置对检测结果有很大影响。在其他条件不变的前

提下, 适合检测数据集的网络参数设置对挖掘模型的性能潜力非常重要。本文方法不受人为经验影响, 完全基于数据集特性优化网络参数设置。

表 3 SSD 模型采用不同锚框参数的检测结果

Table 3 Detection results of SSD with
different anchors parameters

参数	平均 精确率	精确率	召回率
SSD 中 6 层特征的锚框参数	76.95	86.35	87.00
LSSD 中 4 层特征的锚框参数	78.03	89.15	88.80

注: 加粗字体为每列最优结果。

3.2.2 模型简化实验

通过模型简化实验的方法验证所提轻量化设置、双向特征融合模块对检测结果的影响。将采用 LSSD 模型中 4 层特征的锚框参数设置的 SSD 模型作为比较的基准模型, 与 LSSD 模型、结合了语义聚合的 LSSD 模型、结合了注意力引导融合的 LSSD 模型以及结合了双向特征融合方法的 LSSD 模型分别进行性能指标与速度指标的对比, 结果如表 4 所示。可以看出, 与 SSD 模型相比, 本文提出的 LSSD 模型在速度方面具有显著优势, 测试时间仅为 SSD 模型的 65%, 而检测精度与 SSD 模型基本一致。语义聚合和注意力引导特征融合都可以提高检测的精度, 而在训练和测试中, 只增加了很少的时间。两个模块的组合在精度方面实现了最佳性能, 平均精确率与 SSD 相比提高了 2.1%。

为进一步展示这些模型的检测结果, 图 4 给出了部分典型检测结果样例。可以看出, 结合了双向特征融合模块的模型在提升召回率、减少虚警率等方面表现最佳, 说明本文的融合方法是一种有效的特征优化方法。

3.2.3 与其他方法对比

为进一步验证本文方法的有效性与实用性, 与 Faster RCNN (Ren 等, 2017)、基于卷积神经网络的检测模型 (CNN-based model) (李健伟 等, 2018)、基于生成对抗网络 (generative adversarial network, GAN) 和线上难例挖掘 (online hard example mining, OHEM) 的检测模型 (李健伟 等, 2019) 以及密集连接的端到端检测网络 (densely connected end-to-end neural network) (Jiao 等, 2018) 等模型进行对比实

表 4 模型简化实验结果
Table 4 The testing results in ablation study

模型	语义聚合	注意力引导融合	平均精确率/%	精确率/%	召回率/%	训练时间/ms	测试时间/ms
SSD	-	-	78.03	89.15	88.80	20.79	14.02
LSSD	-	-	77.93	89.54	88.60	12.74	9.17
LSSD + 语义聚合	✓	-	78.68	94.55	88.00	13.01	9.72
LSSD + 注意力引导融合	-	✓	72.20	95.26	89.20	12.97	9.57
LSSD + 双向特征融合	✓	✓	80.13	96.68	89.60	13.33	10.18

注:加粗字体为每列最优结果。“-”表示没有采用对应方法,“✓”表示采用了该方法。

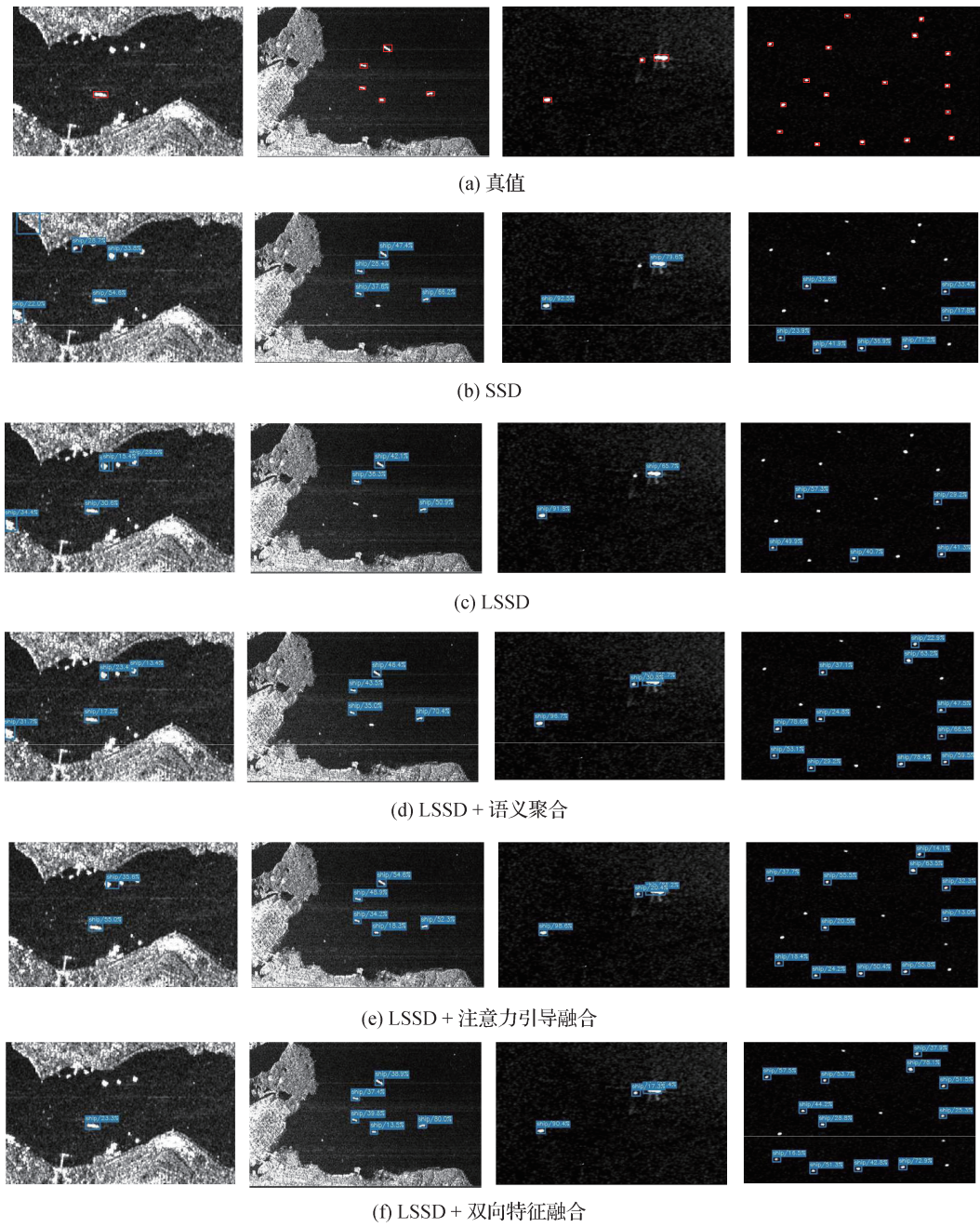


图 4 不同方法部分检测结果样例图

Fig. 4 Some samples of detecting results using different methods((a) ground truths; (b) SSD; (c) LSSD; (d) LSSD with semantic aggregation module; (e) LSSD with attention fusion module; (f) LSSD with bi-directional fusion module)

验,以上对比模型依次称为模型 1、模型 2、模型 3、模型 4,结果如表 5 所示。

表 5 本文方法与其他舰船检测方法对比结果
Table 5 Detection results of different models for comparison

模型	平均精确率/%	精确率/%	召回率/%	训练时间/ms	测试时间/ms
模型 1	73.24	92.31	88.80	540.42	159.11
模型 2	78.90	93.10	86.40	572.67	185.53
模型 3	70.24	82.01	78.40	2 152.09	330.78
模型 4	77.84	95.21	83.60	561.82	180.95
本文	80.13	96.68	89.60	13.33	10.18

注:加粗字体为每列最优结果。

模型 2 和模型 4 都是基于 Faster RCNN 模型提出的,与 Faster RCNN 模型相比,它们提高了检测精度,同时也增加了平均训练和测试的时间。但是,与本文的单阶段检测方法相比,Faster RCNN 在训练精度上并没有表现出太大优势,原因可能由于 Faster RCNN 系列结构更复杂,像 SSDD 这样的小型检测数据集无法很好地对其进行训练。模型 3 利用生成对抗网络和在线难例挖掘技术来扩大训练中的数据,并基于 RCNN 方法进行检测,不仅速度慢,而且精度也是几种方法中最低的。而本文方法在召回率和检测精度等方面都实现了最佳性能,并且在速度方面具有明显优势,反映出本文方法的有效性与实用性。

4 结 论

本文主要研究 SAR 图像舰船目标检测,着重解决了检测中速度与精度的平衡问题。为了提高算法的处理速度,设计了轻量化的 LSSD 模型,与 SSD 模型相比,需训练的参数大幅减少,模型可从头开始训练而不必依赖预训练模型。为了提高模型检测的精度,提出了一种用于参数设置的基于聚类的目标分布特性学习算法,并采用语义聚合块和新的注意力引导融合块的双向特征融合模块来优化检测特征,改进了传统方法中单向特征融合不能优化高层特征的不足。利用公开的 SSDD 数据集验证了所提各模块的效果,结果表明,所提模型在速度和精度方面均表现出良好性能,与基准模型以及其他经典检测算法相比,训练和测试速度均有明显提升,检测精度也高于其他检测算法,从检测结果图中也可以看出,所提模型有效降低了目标漏检情况。

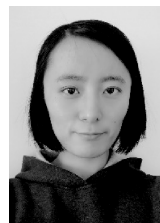
然而,尽管本文模型较其他检测模型表现出更好的速度和精度性能,但测试中仍存在一些失败案例,如无法较好地检测密集排列的舰船目标,这也是实验中各检测模型存在的普遍问题。其原因主要在于包括本文模型在内的这些模型都采用了垂直矩形边框作为目标边界框实施检测,而相邻的密集排列舰船目标的检测框存在大量重叠区域,容易被模型自动合并、删减,造成检测框位置不准确甚至漏检。而采用目标最小外接倾斜矩形框作为目标边界检测框可以有效避免这类问题。因此,在今后工作中,将继续研究基于目标倾斜边界框的检测算法,提高检测效果。此外,本文模型没有利用 SAR 图像的极化信息。如何利用 SAR 图像的极化信息提升检测性能也是接下来一个有意义的研究重点。

参考文献 (References)

- Agrawal A, Mangalraj P and Bisharwal M A. 2015. Target detection in SAR images using SIFT//Proceedings of 2015 IEEE International Symposium on Signal Processing and Information Technology. Abu Dhabi: IEEE: 90-94 [DOI: 10.1109/ISSPIT.2015.7394426]
- Allard Y, Germain M and Bonneau O. 2008. Ship detection and characterization using polarimetric SAR data//Shahbazian E, Rogova G and DeWeert M J, eds. Harbour Protection Through Data Fusion Technologies. Dordrecht: Springer: 243-250 [DOI: 10.1007/978-1-4020-8883-4_29]
- Dalal N and Triggs B. 2005. Histograms of oriented gradients for human detection//Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE: 886-893 [DOI: 10.1109/CVPR.2005.177]
- Deng Z P, Sun H, Zhou S L and Zhao J P. 2019. Learning deep ship detector in SAR images from scratch. IEEE Transactions on Geoscience and Remote Sensing, 57(6): 4021-4039 [DOI: 10.1109/TGRS.2018.2889353]

- Everingham M, Ali Eslami S M, 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]
- Jiao J, Zhang Y, Sun H, Yang X, Gao X, Hong W, Fu K and Sun X. 2018. A densely connected end-to-end neural network for multiscale and multiscale SAR Ship detection. *IEEE Access*, 6: 20881-20892 [DOI: 10.1109/access.2018.2825376]
- 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]
- Li J W, Qu C W, Peng S J and Deng B. 2018. Ship detection in SAR images based on convolutional neural network. *Systems Engineering and Electronics*, 40(9): 1953-1959 (李健伟, 曲长文, 彭书娟, 邓兵. 2018. 基于卷积神经网络的 SAR 图像舰船目标检测. *系统工程与电子技术*, 40(9): 1953-1959) [DOI: 10.3969/j.issn.1001-506X.2018.09.09]
- Li J W, Qu C W and Shao J Q. 2017. Ship detection in SAR images based on an improved faster R-CNN//*Proceedings of 2017 SAR in Big Data Era: Models, Methods and Applications*. Beijing: IEEE: 1-6 [DOI: 10.1109/BIGSDATA.2017.8124934]
- Li J W, Qu C W, Peng S J and Jiang Y. 2019. Ship detection in SAR images based on generative adversarial network and online hard examples mining. *Journal of Electronics and Information Technology*, 41(1): 143-149 (李健伟, 曲长文, 彭书娟, 江源. 2019. 基于生成对抗网络和线上难例挖掘的 SAR 图像舰船目标检测. *电子与信息学报*, 41(1): 143-149) [DOI: 10.11999/JEIT180050]
- 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*. Honolulu: IEEE: 2117-2125 [DOI: 10.1109/CVPR.2017.106]
- Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L. 2014. Microsoft COCO: common objects in context//*Proceedings of the 13th European Conference on Computer Vision*. Zurich: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]
- 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*. Amsterdam: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]
- Pappas O, Achim A and Bull D. 2018. Superpixel-level CFAR detectors for ship detection in SAR imagery. *IEEE Geoscience and Remote Sensing Letters*, 15(9): 1397-1401 [DOI: 10.1109/LGRS.2018.2838263]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2015. You only look once: unified, real-time object detection//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]
- Redmon J and Farhadi A. 2017. YOLO9000: better, faster, stronger//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE: 6517-6525 [DOI: 10.1109/CVPR.2017.690]
- Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]
- Wang J Z, Lu C H and Jiang W W. 2018. Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression. *Sensors*, 18(9): #2851 [DOI: 10.3390/s18092851]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module [EB/OL]. [2019-11-14]. <https://arxiv.org/pdf/1807.06521.pdf>
- Yang X, Hao S, Fu K, Yang J R, Sun X, Yang M L and Guo Z. 2018. Automatic ship detection in remote sensing images from Google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sensing*, 10(1): #132 [DOI: 10.3390/rs10010132]
- Zhang X H, Wang H P, Xu C A, Lyu Y F, Fu C L, Xiao H C and He Y. 2019. A lightweight feature optimizing network for ship detection in SAR image. *IEEE Access*, 7: 141662-141678 [DOI: 10.1109/ACCESS.2019.2943241]

作者简介



张筱晗, 1992 年生, 女, 博士研究生, 主要研究方向为遥感图像目标检测识别、深度学习。

E-mail: xhan_zhang@163.com



姚力波, 通信作者, 男, 讲师, 主要研究方向为天基遥感图像处理、信息融合。

E-mail: ylb_rs@126.com

吕亚飞, 男, 博士研究生, 主要研究方向为遥感图像检索、多源信息处理、深度学习。E-mail: yfei_Lv@163.com

简涛, 男, 教授, 主要研究方向为信号与信息处理、信息融合。E-mail: work_jt@163.com

赵志伟, 男, 高级工程师, 主要研究方向为卫星系统论证、卫星应用。E-mail: waly_zzw@163.com

藏洁, 女, 硕士, 工程师, 主要研究方向为遥感卫星系统论证、卫星应用。E-mail: 964249204@qq.com