文章编号：1001－9014(2019)03－0290－06

DOI:10.11972/j.issn.1001－9014.2019.03.006

CCNet: A high-speed cascaded convolutional neural network for ship detection with multispectral images

ZHANG Zhong-Xing[1,3], LI Hong-Long[1,3], ZHANG Guang-Qian[1,3], ZHU Wen-Ping[1,3], LIU Li-Yuan[1,3], LIU Jian[1,3], WU Nan-Jian[1,2,3*]

(1. State Key Laboratory of Superlattices and Microstructures, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China;
2. Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100083, China;
3. Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: A novel ship detection method using cascaded convolutional neural network (CCNet) with multispectral image is proposed to achieve high-speed detection. The CCNet employs two cascaded convolutional neural networks (CNN) for extracting regions of interest (ROIs), locating and segmenting ship objects sequentially. Benefit from the abundant details of the multispectral image, CCNet can extract more robust feature for achieving more accurate detection. The efficiency of CCNet has been validated by the experiments on the SPOT 6 satellite multispectral images. Compared with the state-of-the-art deep-learning-based ship detection algorithms, the proposed ship detection algorithm accelerates the processing by more than 5 times with a higher detection accuracy.

Key words: ship detection, remote image processing, convolutional neural network, multispectral image
PACS: 84.40.Xb

CCNet:面向多光谱图像的高速船只检测级联卷积神经网络

张忠星[1,3], 李鸿龙[1,3], 张广乾[1,3], 朱文平[1,3], 刘力源[1,3], 刘 剑[1,3], 吴南健[1,2,3*]

(1. 中国科学院半导体研究所 超晶格国家重点实验室,北京 100083;
2. 中国科学院脑科学与智能技术卓越创新中心,北京 100083;
3. 中国科学院大学 材料与光电研究中心,北京 100049)

摘要:针对实现遥感图像中船只目标的快速检测,提出了一个采用多光谱图像、基于级联的卷积神经网络(CNN)船只检测方法CCNet.该方法所采用两级级联的CNN依次实现感兴趣区域(ROI)的快速搜索、基于感兴趣区域的船只目标定位和分割.同时,采用含有更多细节信息的多光谱图像作为CCNet的输入,能够提升网络提取特征鲁棒性,从而使得检测更加精确.基于SPOT 6卫星多光谱图像的实验表明,与当前主流的深度学习船只检测方法相比,该方法能够在实现高检测精准度的基础上将检测速度提高5倍以上.

关 键 词:船只检测;遥感图像处理;卷积神经网络;多光谱图像

中图分类号:TP751 **文献标识码**:A

Introduction

The image processing and analysis are playing a more and more important role in remote sensing field. As one of the critical applications, ship detection has been widely used to fishery management, ship rescue, and marine traffic security[1-10]. Earlier studies with hand-

Received date: 2018-12-01, **revised date**: 2019-02-11 **收稿日期**:2018-12-01, **修回日期**:2019-02-11

Foundation items:This work is supported by The National Key Research and Development Program of China (Grant No. 2016YFA0202200), National Natural Science Foundation of China (Grant Nos. 61434004, 61234003), National Natural Science Foundation for the Youth of China (61504141, 61704167), National Key R&D Program of Beijing (Z181100008918009) and Youth Innovation Promotion Association Program, Chinese Academy of Sciences (No. 2016107)

Biography:ZHANG Zhong-Xing (1990-), male, Liaocheng, doctor. Research area involves Image processing and digital integrated circuits. E-mail:zhangzhongxing@semi.ac.cn

* **Corresponding author**:E-mail: nanjian@red.semi.ac.cn

crafted feature-based methods are mainly implemented on synthetic aperture radar (SAR) images which can reduce the impact of background objects and noise or low-resolution optical remote sensing images. With the development of the optical image satellite which can provide more detailed spatial contents with the sub-meter spatial resolution, the deep-learning-based ship detection methods have been utilized on optical remote sensing images to acquire a higher detection performance.

Nowadays, satellites always deploy multispectral cameras in order to capture high-resolution visible, near infrared (NIR), short wavelength infrared, panchromatic, and thermal infrared (TIR) images. The method coupled with multispectral band images is emerging to achieve a higher detection accuracy. Zhou et al. proposed a convolutional neural network (CNN) based ship detection algorithm with Landsat8 images as input, which combines NIR, short wavelength infrared, panchromatic, and TIR band images[11]. Jörg Brauchle et al. used independent low-resolution TIR images and high-resolution RGB and NIR images to capture the ship candidates respectively and combined the two branches in classification stage[12].

Although these achievements improve the performance in ship detection field, there are unfortunately some drawbacks to be addressed. The accuracy of the handcrafted feature-based methods depends heavily on illumination, scale, rotation, and drift. The deep-learning-based ship detection method inherits the CNN defects, such as the huge network model and plenty of training samples. Secondly, it detects objects with a whole image where the size of small objects is different sharply with large objects. The small object is hard to extract efficient features to achieve highly accurate detection, such as the Fast R-CNN based method[8] whose detection performance for small objects reduces 40% compared with the performance for large objects. Moreover, it gives no consideration on the sparsity of ships in remote sensing images, where the area rate is less than 1% rather than 31.69% in the natural image dataset COCO 2017[13]. It generates plenty of unnecessary operations on the background of an image when adopting the state-of-the-art deep learning detection method[8-9] directly. The detection speed is therefore constrained severely to meet the real-time processing requirement. The non-real-time detection restricts its application in scenarios such as marine rescue, fishery management, and marine traffic security.

In this paper, we proposed a novel cascaded CNN model (CCNet) fed with multispectral images, aiming at achieving high-speed and high-accuracy ship detection for marine rescue and marine traffic management task. The main contributions of this paper are as follows:

1) A high-speed ship detection model is proposed by adopting cascaded CNN classification model and CNN detection model to fast acquire the regions of interest (ROIs) and locate ships and segment ship instants respectively. The lightweight CNN classification model improves the detection speed dramatically via filtering out a variety of background regions before implement the CNN detection model. In addition, the size of the input of detection model shrinks a lot, which can reduce the computations furtherly.

2) Compared with the Mask R-CNN method which resizes the different scale objects with a unified proportion for the input image and up-samples the feature maps in a very deep layer for small objects, a self-adapting scale adjustment is inserted between the classification model and detection model of the CCNet for rescaling the size of each objects region independently. It reduces the complexity of the detection model and the computation in CCNet. Furthermore, it makes the CCNet holding abundant detail features for small size objects which is beneficial to achieve high detection performance.

3) A novel combination method of multispectral images is introduced as the input of the CCNet to improve the detection performance. The RGB images and the IR images are concatenated in the classification model for explicating more spatial details and stronger robustness to noise respectively.

This paper is organized as follows. The high efficient ship detection model CCNet is introduced in section 1. The experimental details and results are displayed in section 2. Finally, the conclusion is provided in section 3.

1 The proposed ship detection algorithm

Since the ship objects are very sparse in high-resolution remote sensing images, the traditional deep-learning-based method is hard to achieve a high ship detection speed. We proposed a cascaded CNN model CCNet to achieve high-speed ship detection with the high-accuracy.

1.1 Framework of the CCNet

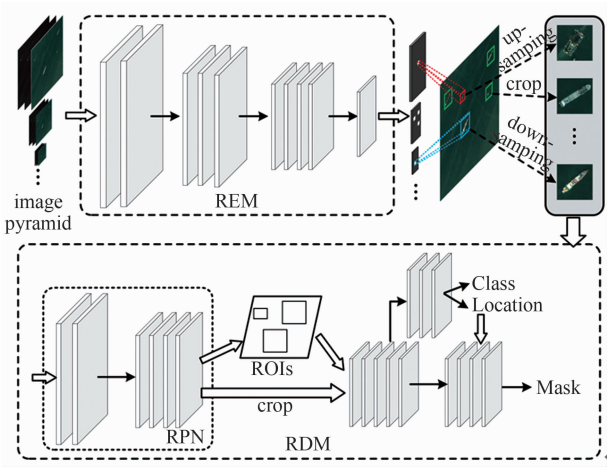

Fig. 1　Framework of the CCNet
图 1　CCNet 框架示意图

As illustrated in Fig. 1, the CCNet framework contains two cascaded CNN models, an ROI extraction model (REM) and an ROI detection model (RDM), which are capable of searching ROIs in a large image and locating objects in ROIs respectively. The REM is a shallow classification network, which takes with the multispectral remote sensing image pyramid as input and generates the location of ROIs. Due to the use of image pyramid, we can rescale ship objects with different size into a similar size, the REM can be designed to extract feature and

classification for objects with the similar size, which can compress the depth of networks and reduce the computation. The ROIs of different size ship objects can be captured in different layer of the image pyramid. As the ROIs are captured in different layers of the image pyramid, multiple ROIs with different scale are derived. A self-adapting scale adjustment can be used to rescale each ROI independently. If the scale of the ROI is too small, the up-sampling operation will be implemented for achieving high detection accuracy. Also, the down-sampling operation will be implemented when the size of the ROI is very large to reduce the computations. The RDM is a two-stage CNN detection network, which is fed with the suspected objects and gets a precise location and instant segmentation mask. The first stage is a region proposal network (RPN) which is responsible for extracting the hierarchical features of the image and generating the location of an object with a more accurate range than REM. The second stage can generate a more precise location and instant segmentation based on the results of the first stage.

The CCNet uses a full convolutional model, which is more robust to diverse conditions of illumination, scale, rotation, and drift. Moreover, compared with the state-of-the-art deep learning detection method such as Mask R-CNN[14], SSD[15], and YOLO[16], the introduction of the REM rejects large part of background reduces the computation dramatically. As the output ROI of the REM has the scale information of an object, the followed self-adaptive scale adjustment can resize the ROI to a similar level, which makes the RDM designed simpler and is benefit to detect small objects with a high performance.

1.2 ROI extraction model REM

Table 1 Network structure of REM
表 1 REM 网络结构

Layer	Kernels	Size	Stride	RF
*conv*1	8	3×3	1	3×3
*max-pooling*1	–	3×3	2	5×5
*conv*2	16	3×3	1	9×9
*max-pooling*2	–	3×3	2	13×13
*conv*3	32	3×3	1	21×21
*max-pooling*3	–	3×3	2	29×29
avg-pooling	–	4×4	3	53×53
*conv*4	4	1×1	1	53×53
softmax	–	1×1	1	–

The REM is a CNN classifier which is adopted to search ROIs in a whole remote sensing image. Instead of the previous network employed a very deep network to classify the objects whose scale vary in a very large range[8-9, 11, 19], the REM adopts the shallow network with pyramid image input. This strategy is mainly based on two considerations. The first is that the landscape of the remote sensing image is very simple except the scale variety. The simple landscape can be easily divided into 4 categories: ship, sea, cloud, and land. The second is that the very deep network is unfriendly for small objects through the classification accuracy improvements with the depth of networks for large objects, whereas the classification accuracy does not lose as much as rescaling the large objects to a small size. So the most efficient method is to design a shallow network with the image pyramid as input which can rescale the different scale object into an approximate size. A 5-level image pyramid with a factor of 2 is lead up in the REM to tackle different scale ships.

The detailed network configuration and receptive field (RF) of each layer in the REM are listed in Table 1. Each convolution layer is followed by a ReLU active function. The feature map size of the output of *conv*4 is 1/24 of the size of the original REM input image. And the receptive field of *conv*4 feature maps can reach 53×53. Each point of the *conv*4 represents a 53×53 image patch with stride 24. We use the central 32×32 pixels of the REM input image as the mapping of a point in *conv*4 feature maps. Then the softmax is used to each point in *conv*4 feature map. The result of *softmax* represents the category of the corresponding 32×32 region in a layer of the image pyramid. The region of the original image which is represented by the 32×32 region in a layer of the image pyramid is captured as the ROI, such as a 32×32 region in the top of the 5-level image pyramid represents a 512×512 image patch of the original image. The ROIs generated in different layers of the image pyramid will be screened by using non-maximum suppression (NMS) method. If the adjacent image patches are classified as ships, they will be merged into an entire region as the final ROI.

1.3 Self-adapting scale adjustment

As the input image of REM is in a pyramid fashion and the classification window is 32×32, the ROIs whose sizes vary from 32×32 to 512×512 contain roughly the size information of the suspected object. To extract effective features and improve the detection accuracy for small ship objects, a novel image size sampling method is introduced. The calculation of the sample rate is:

$$S = \begin{cases} 4, & if\ L_{\mathrm{ROI}} <= 64 \\ 256/L_{\mathrm{ROI}}, & \text{others} \end{cases}, \quad (1)$$

where the L_{ROI} represents the maximum side length of a ROI. If the ROI is very small, the up-sampling operation will be employed to improve the resolution of ROI. For a ship whose length is 10 pixels, the Self-adapting scale adjustment can rescale it to 40 pixels. Besides, the down-sampling operation will be implemented for large ship objects with rare detect performance decline, whereas it can reduce a large amount of computation.

1.4 ROI detection model RDM

The RDM is a two-stage CNN based detection model with fixed input size 256×256 which is inspired by the Mask R-CNN[14]. The first stage is a region proposal network (RPN) which is responsible for extracting the hierarchical features of the image and generating the location of an object with a more accurate range than REM. The second stage can generate a more precise location and instant segmentation based on the results of the first stage.

The RPN of the RDM contains a backbone network and several convolutional layers for extracting hierarchical features and outputting the locations respectively. The structure of the backbone is listed in Table 2. As the object scales of the RDM inputs are similar, we can just extract location information on a narrow range: *conv*3, *conv*4, and *conv*5 feature maps. The FPN method[17] is

implemented on *conv*3, *conv*4, and *conv*5 layers to merge more high-level features to form the *F*3, *F*4, and *F*5 feature maps. Then the *F*3, *F*4, and *F*5 feature maps are convoluted by 512 filters. Then the softmax and convolution operation is employed to generate the category and location respectively. The NMS operation is followed to suppress the number of ROIs. The ROIs cropped on *F*3, *F*4, and *F*5 are the input of the second stage of RDM. A 2-layer convolution operation with 512 filters is used to generate the final location information and category. Also, the *F*3, *F*4, and *F*5 are cropped based on the final location information as the input to generate the instant segmentation information. An 8-layer convolution operation with 512 filters respectively is employed to generate the segmentation.

Compared with the state-of-the-art Mask R-CNN[14], SSD[15], and YOLO[16] methods which have to detect objects whose scales vary in a very large range, the size fluctuation of the object detected in the RDM is confined in a small range. The RDM can reduce the network's depth and extract ROIs in a less number of feature maps compared with the Mask R-CNN and other method.

Table 2 Backbone network structure of the RDM
表 2 RDM 中主干网络配置

Layer	Kernels	Size	Stride	RF
conv1	16	7×7	2	7×7
max-pooling1	–	3×3	2	11×11
conv2	32	3×3	1	19×19
max-pooling2	–	3×3	2	27×27
conv3	64	3×3	1	43×43
max-pooling3	–	3×3	2	59×59
conv4_a	128	3×3	1	91×91
conv4_b	128	3×3	1	123×123
max-pooling4	–	3×3	2	155×155
conv5	256	3×3	1	219×219

1.5 Details of training and inference

As the CCNet contains two independent cascaded CNN models which are designed for different usages, the datasets for the two models are different. The training of the two models are independent. The REM is trained on a dataset which contains ship, sea, cloud, and land four categories. The island is categorized as land or sea depending on the coverage of the island due to the number of island samples is much less than other categories. It is hard to use data augmentation method for expending the island samples to achieve a comparable number with the remaining three categories. In the training phase, the input of REM is fixed to 32×32 which is sufficient for achieving a high classification performance of 4 categories.

The training of the RDM is similar to other end-to-end deep learning detection models which select the objects and backgrounds from a single image. As our RDM method focuses on detecting objects in a suspected region for reducing the unnecessary computation, we need to build an image dataset which only consists of the ship object samples cropped from the original remote sensing images. Since the ship object samples carry very little backgrounds which dissatisfy to extract abundant background information, the training of the RDM is implemented based on a more broad-scenario image dataset which crop size of the object image patch is expended out to N times of the ground object region. Here N is a random number between 1.2 and 2. Then the self-adopted scale adjustment is used to rescale the size of these image patches. As the ROI whose size is less than 64×64, the zero padding is used for adapting the input of RDM.

As the number of samples we can obtain is hardly satisfied the training of the CCNet start from scratch. The CCNet is trained on the Airbus image dataset[18] previously, which contains more than 10 thousand RGB training samples. The images in Airbus dataset are rescaled to the same resolution of the Sentinel images. Owing to the Airbus images only contain RGB channel, the gray image which is converted by the RGB image is doubled as the NIR input channels and pan. Channel. Then the model of CCNet is fine-tuned in our SPOT 6 images.

2 Experiment and performance

2.1 Experiment data

The multispectral images of SPOT 6 are employed to demonstrate the effectiveness of the CCNet. The multispectral images cover 5 different bands in the visible, near infrared range. The detailed band information is listed in Table 3. We obtain 1 000 multispectral images with the size of 1 024×1 024 in different illumination and weather conditions, such as weak contrast condition, strong reflective condition, cloud-covered condition, or strong wave condition. The RGB and NIR images are rescaled 4 times to achieve the same resolution of the panchromatic (Pan.) images. These images are divided by 4:1 as the training and test samples. The training and test samples contain 3 132 ships and 756 ships respectively whose lengths vary from 15 pixels to 305 pixels.

Table 3 Detailed information of SPOT 6 satellite images
表 3 SPOT 6 卫星图像详细信息

Type	Bandwidth (μm)	Resolution (m)
Pan.	0.450 ~ 0.745	1.5
Blue	0.450 ~ 0.525	6
Green	0.530 ~ 0.590	6
Red	0.625 ~ 0.695	6
NIR	0.760 ~ 0.890	6

2.2 Performance of the CCNet

The performance of the CCNet is shown in Table 4. We evaluate the performance of the CCNet by using average precision (AP) and average recall (AR) under different intersection over union (IOU). $AP_{50:95}$ and $AR_{50:95}$ are the mean AP and AR values of the IOU 0.5 to IOU 0.95 with the interval 0.05 respectively. The AP_S/AP_S, AP_M/AP_M, and AP_L/AP_L are used to evaluate the AP/AR for objects whose areas are lower than 32×32, between 32×32 and 96×96, and larger than 96×96 respectively. Fig. 2 illustrates the detection performance comparison for objects in different sizes between the CCNet and original Mask R-CNN which is fine-tuned on our dataset. The value of AP and AR are both between 0 and

1. The higher value represents the better performance. Compared with Mask R-CNN, the CCNet achieves comparable performance for big objects and more than 2 times better performance for small objects. Table 5 shows the size of total filters, FLOPS counts and inference time of the two cascaded stage REM and RDM respectively, and compares these indicators with Mask R-CNN. The parameter memories of REM and RDM are only 0.012% and 0.217% of Mask R-CNN respectively. Also, the parameter memory of REM is much smaller than RDM. The FLOPS of REM and RDM are only 4.03% and 3.08% respectively. The inference times are evaluated on NVIDIA GeForce GTX TITAN X (Maxwell). The REM and RDM can achieve 30 *fps* for 1 024 × 1 024 5-level image pyramid and 10 *fps* and 256 × 256 images, respectively. Whereas the Mask R-CNN can only achieve 4 *fps* for 1 024 × 1 024 image. So the model complexity of CCNet is much lower than Mask R-CNN.

Table 5 Model complexity comparison between CCNet and Mask R-CNN

表 5 CCNet 和 Mask R-CNN 模型复杂度比较

Model		Input size	Parameter memory	GigaFLOPS	Inference time (ms)
CCNet	REM	1 024 × 1 024 (5-level pyramid)	0.029 MB	25.95	32.8
	RDM	256 × 256	53 MB	19.86	50.7
Mask R-CNN		1 024 × 1 024	244 MB	644.12	239.6

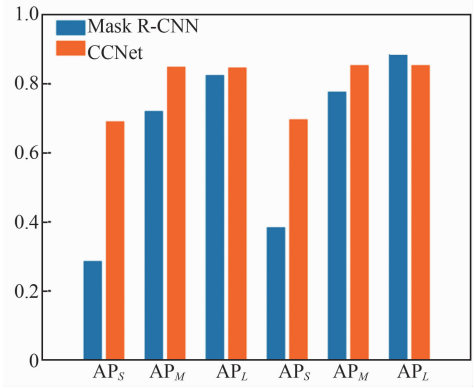

Fig. 2 Performance comparison between Mask R-CNN and CCNet

图 2 Mask R-CNN 与 CCNet 性能对比

The final results of our method are shown in Fig. 3. We select four kinds of representative scenarios to illustrate the performance of CCNet. These scenarios are weak contrast and strong wave condition, the condition of single image with different size ships, cloud-covered condition, and complex background condition. The image samples with these conditions are listed in Fig. 3 (a) from the first row to the fourth row respectively. The results of CCNet are shown in Fig. 3 (b). Also, we give the results of an end-to-end deep learning method Mask R-CNN which is fine-tuned on our image dataset with 256 × 256 in training and 1024 × 1024 in inference stage in Fig. 3 (c), which demonstrates the efficiency of our model to detect small size objects. The CCNet achieves an excellent performance under these complex scenarios. As the Mask R-CNN method takes the entire image as input, the tiny object can only occupy a litter area which is hard to extract effective features by CNN model. It causes the missing detection for tiny objects. Also, some ships are not recalled with Mask R-CNN owing to the weak contrast condition. Our method avoids these issues by using the self-adapting scale adjustment operation and detecting objects in each ROI.

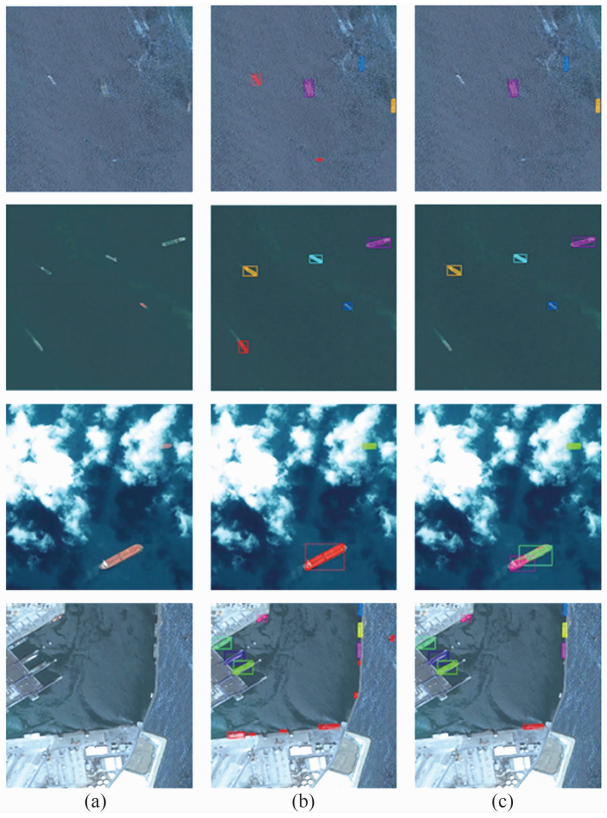

Fig. 3 Detection performance. (a) The original test images. (b) The results of CCNet. (c) The results of Mask R-CNN model fed with the entire image directly

图 3 检测结果. (a) 原始测试图片. (b) CCNet 结果. (c) 采用整幅图像输入的 Mask R-CNN 结果

To demonstrate the efficiency of our CCNet, we compare it with the state-of-the-art deep-learning-based ship detection algorithm. The comparisons of the per-

Table 4 Performance of the CCNet

表 4 CCNet 性能

	$AP_{50:95}$	AP_{50}	AP_S	AP_M	AP_L	$AR_{50:95}$	AR_{50}	AR_S	AR_M	AR_L
Location	0.761	0.938	0.690	0.848	0.846	0.769	0.918	0.696	0.853	0.853
Segmentation	0.621	0.914	0.555	0.715	0.689	0.639	0.894	0.562	0.729	0.725

formances are listed in Table 6. Due to lack of evolution for small objects in previous works, we just compare the detection performance (Precision and recall) and detection speed. The precision and recall are evaluated under IOU 0.5 while Zhang *et al*[19]. only provided under IOU 0.4. The algorithms with multispectral image have higher performance than those algorithms with visible images. Our method owns the best precision compared with previous methods. Though the recall is less the Zhou *et al*., the rate of scale variation of the largest ship and the smallest ship in our method is more than 20 times whereas only 4 times in the work by Zhou *et al*.[11]. Thus, our CCNet is more robust than the method proposed by Zhou. As the ship occupy rate in our image dataset is far above the real remote sensing images, the FLOPS of our method is evaluated under the executive rate of REM and RDM is 5:1, which is also far above the real remote sensing images as the ship only occupies less 1% of the sea. Compared with the other methods, our method reduces the counts of FLOPS more than 5 times because the CCNet uses a cascaded model to extract ROIs and locate objects in ROIs.

Table 6 Comparison with related works
表 6 与相关工作比较

	Ours	Zhou *et al*.[11]	Yao *et al*.[9]	Zhang *et al*.[19]
Image	multispectral image	multispectral image	RGB	RGB
Precision	0.957	0.910	0.733	0.6
Recall	0.918	0.940	0.864	0.9
GigaFLOPS @ 1024×1024	30	172	314	340

3 Conclusion

A new CNN-based ship detection model, CCNet, is proposed to achieve high detection performance with high speed. The CCNet employs cascaded CNN model REM and RDM, where the REM is used to extract ROIs and the RDM is used to locate and segment ship object in the ROIs respectively. Benefited from the lightweight REM model, the CCNet eliminates numerous unnecessary computations. Moreover, the detection performance for small size ships improves as the self-adapting scale adjustment method for each ROI. The efficiency of the CCNet is demonstrated by the experiment on SPOT 6 multispectral images. The proposed ship detection algorithm achieves higher precision compared with the previous deep-learning-based ship detection algorithm. In addition, the CCNet reduces the computations more than 5 times. It is a huge step towards the real-time marine rescue and marine traffic management task.

References

[1] Jubelin G, Khenchaf A. Multiscale algorithm for ship detection in mid, high and very high resolution optical imagery[C]//Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International. IEEE, 2014: 2289-2292.

[2] L Liu G, Zhang Y, Zheng X, *et al*. A new method on inshore ship detection in high-resolution satellite images using shape and context information[J]. *IEEE Geoscience and Remote Sensing Letters*, 2014, **11**(3): 617-621.

[3] Bi F, Zhu B, Gao L, *et al*. A visual search inspired computational model for ship detection in optical satellite images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2012, **9**(4): 749-753.

[4] Zhang R, Yao J, Zhang K, *et al*. S-CNN-BASED ship detection from high-resolution remote sensing images[J]. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2016, 41.

[5] He H, Lin Y, Chen F, *et al*. Inshore ship detection in remote sensing images via weighted pose voting[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, **55**(6): 3091-3107.

[6] Shi Z, Yu X, Jiang Z, et al. Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, **52**(8): 4511-4523.

[7] Zhu C, Zhou H, Wang R, *et al*. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features[J]. *IEEE Transactions on geoscience and remote sensing*, 2010, **48**(9): 3446-3456.

[8] Liu Z, Hu J, Weng L, *et al*. Rotated region based CNN for ship detection[C]//Image Processing (ICIP), 2017 IEEE International Conference on. IEEE, 2017: 900-904.

[9] Yao Y, Jiang Z, Zhang H, *et al*. Ship detection in optical remote sensing images based on deep convolutional neural networks[J]. *Journal of Applied Remote Sensing*, 2017, **11**(4): 042611.

[10] Zou Z, Shi Z. Ship detection in spaceborne optical image with SVD networks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, **54**(10): 5832-5845.

[11] Zhou M, Jing M, Liu D, *et al*. Multi-resolution Networks for Ship Detection in Infrared Remote Sensing Images[J]. *Infrared Physics & Technology*, 2018.

[12] Brauchle J, Bayer S, Berger R. Automatic ship detection on multispectral and thermal infrared aerial images using MACS-Mar remote sensing platform[C]//Pacific-Rim Symposium on Image and Video Technology. Springer, Cham, 2017: 382-395.

[13] Lin T Y, Maire M, Belongie S, *et al*. Microsoft coco: Common objects in context[C]//European conference on computer vision. Springer, Cham, 2014: 740-755.

[14] He K, Gkioxari G, Dollár P, *et al*. Mask r-cnn[C]//Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE, 2017: 2980-2988.

[15] Liu W, Anguelov D, Erhan D, *et al*. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[16] Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[17] Lin T Y, Dollár P, Girshick R B, *et al*. Feature Pyramid Networks for Object Detection[C]//CVPR. 2017, **1**(2): 4.

[18] Airbus dataset [OL]. 2018. https://www.kaggle.com/c/airbus-ship-detection/data.

[19] Zhang Z, Guo W, Zhu S, *et al*. Toward Arbitrary-Oriented Ship Detection With Rotated Region Proposal and Discrimination Networks [J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, (99): 1-5.