

Overview of Deep Convolutional Neural Network Approaches for Satellite Remote Sensing Ship Monitoring Technology

Xiunan Li, Peng Chen and Kaiguo Fan

State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, Zhejiang Province, China.

*Email: chenpeng@sio.org.cn

Abstract. Remote sensing (RS) monitoring of ships has important significance in both military and civilian fields. The RS ship detection aims to locate the position of the ship in the remote sensing image and extract its characteristics. Traditional ship target detection algorithms cannot meet the demands for speed and precision of SAR remote sensing and optical remote sensing data. With the development of artificial intelligence technology, the target detection technology such as deep learning algorithms has made significant progress in RS ship detection. Deep learning has become a heated topic in research. This paper has analyzed and summarized previous researches on the application of deep learning algorithms in ship detection technologies based on SAR and optical remote sensing images in recent years and has provided suggestions for future studies. In the future, deep learning-based technologies for RS ship detection will use more data, such as data from multiple sensors in multiple channels. Deep neural networks will also include more rules and specialized knowledge. Its structure will become more complicated and eventually develop into a neural network like the human brain.

1. Introduction

With the rapid development of remote sensing (RS) information science, RS technology has been widely used in military and civilian fields as a comprehensive technology and has played an important role in terrestrial resource survey, ocean exploration, military investigation, strike analysis, and assessment [1]. Satellite RS ship monitoring can be applied to many areas including management of marine fisheries, water traffic management and supervision, and surveillance of sewage from ships. Synthetic Aperture Radar (SAR) is not affected by clouds or rain, so it can be used 24 hours per day[2]. It is a main research direction of RS ship detection. While a visible light (VL) image is susceptible to factors such as light and clouds, it has high resolution and clear texture, and it can well distinguish a target in the image. On clear and sunny days, the image has rich information and shows clear features of the target structure. VL images have incomparable advantages in the field of maritime reconnaissance, especially ship identification. VL images complement SAR images for marine target monitoring [3, 4].

Both optical and SAR remote sensing images have seen development bottlenecks using traditional ship detection algorithms. For example, the terrain effect in SAR images and effects of thin clouds in optical images still exist. Thanks to the development of artificial intelligence, the Convolutional Neural Networks (CNN) technology has made significant progress in the fields of image classification, detection, and segmentation. The progress of deep learning-based CNNs on ordinary images has drawn attention to many scholars on introducing deep learning into RS ship detection. This paper will summarize, analyze, and discuss recent years' researches on the use of deep learning technology in RS ship detection, and will discuss the future development of RS ship detection.



2. Development of CNN in Target Detection

This section describes the normal structure of a manuscript and how each part should be handled. LeCun et al. have established LeNet5[5], a modern neural network architecture of CNN, mainly for handwritten character recognition (Figure 1). The structure of a CNN has three features: local connections, weight sharing, and spatial or temporal sub-sampling. These features allow a CNN to be invariant to a certain degree of translations, scaling, and distortions [6]. AlexNet[7] proposed by Hinton et al. in 2012 was the winner of the ImageNet competition. Following AlexNet, a larger number of deeper CNN structures have been established, such as VGG[8], GoogleNet[9] and ResNet[10]. The record of the ImageNet competition has been constantly refreshed.

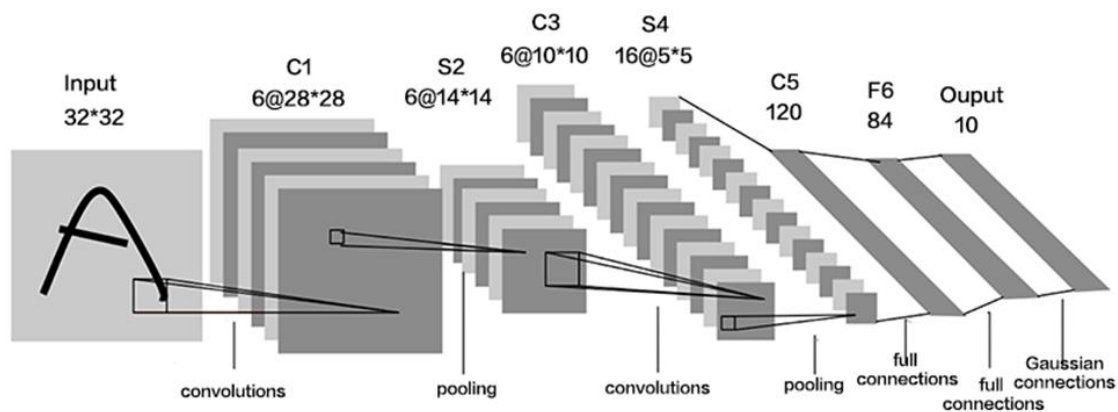


Figure 1. Schematic Diagram of CNN Structure [5]

Image classification determines the type of subject in an image, for example, whether the subject is a cat or a person in an image, while target detection determines the type of object in the image and also locates the subject, combining image classification and positioning. Given the great success of CNN in image classification, researchers have begun to improve it and introduce it into the field of target detection, in which it has also made significant progress. Several target detection frameworks have been born and are commonly used nowadays, such as the region proposal-based Faster R-CNN[11], regression-based YOLO[12], and SSD[13] algorithm.

3. Application of CNN to Ship Detection in SAR Imagery

Ship detection in SAR imagery has the following two problems: missed detections of small targets and the false alarms occurred with noises in blurred land, small islands, and azimuthal directions. CNN can be used for classification and it be used as a network for feature extraction of the target detection framework. There are two common approaches to solve false alarms and missed detections with CNNs. One is to combine traditional detection algorithms and deep learning algorithms; the other is to use the target detection framework with deep learning for ship detection.

3.1. Ship Detection Based on CFAR and Deep Learning Algorithm

Constant false alarm rate (CFAR)[14,15] is the most in-depth and widely-used algorithm among traditional target detection algorithms. Its key concept is modelling the distribution of sea clutter with regional sea clutter statistics. The threshold for dividing target pixels is obtained based on the density curves of the model and the false alarm rate to obtain, which is then used for detecting targets with high grayscale values in an SAR image[16]. A large number of studies and experiments have shown that even at complicated sea state, CFAR detectors can still produce quite satisfactory detection results [17]. It is generally believed that the sea clutter has a Rayleigh distribution with a calm sea and a K distribution at most sea states. Considering that, most of the SAR algorithms for ship detection use the CFAR detection with K distribution[18, 19]. In the face of various sea states, if a single distribution model is used, the detection accuracy may not be satisfactory in some cases. In 2000, Jiang et al.[20]

modelled a probability distribution of the radar backscattering from a sea surface using the probabilistic neural network (PNN). The algorithm can select the sea clutter model adaptively, but it has low accuracy in detecting images with high sea surface background noises. Chen et al.[21] have improved the algorithm which maintains the accuracy with high background noises and is faster than Jiang et al.'s algorithm.

CFAR detects the false alarms occurred with noises in blurred land, small islands, and azimuthal directions. An et al.[22] have found that CNN classification techniques can be used to remove false alarms and improve detection accuracy. Its technical flow chart is shown in Figure 2. Firstly, separate the sea and land by the fully convolutional network (FCN) with 18 layers (ten convolutional layers, five maxpool layers, three deconvolutional layers). Secondly, calculate the distribution of sea clutter and compare it with the K distribution, Gamma distribution and Rayleigh distribution. The results show that in terms of high scores, the Rayleigh distribution is close to the K distribution and is better than the Gamma distribution, and that the Rayleigh distribution is the most efficient in calculation. Thirdly, use the CFAR algorithm for ship detection; finally, classify the ship slices generated from the CFAR detection with a 7-layer deep CNN (two convolutional layers, two maxpool layers, three fully-connected layers) and remove false alarms. The experimental results of UFS images suggest that this method has effectively reduced the false alarm rate.

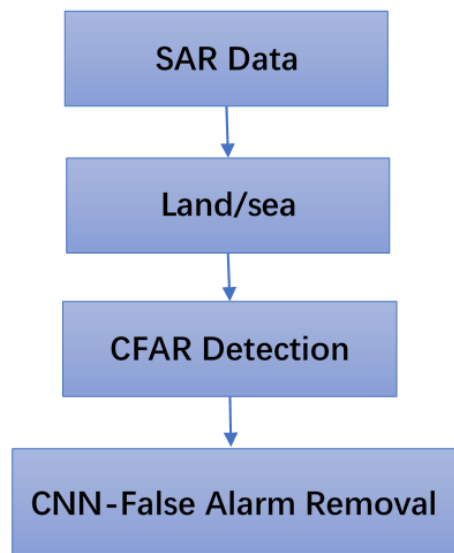


Figure 2. CFAR+CNN Flow Chart

Faster Region-CNN (R-CNN) is directly applied to ship target detection. Its missed detection rate is very high for ships, especially small targets. Kang et al. [23] believe that the missed detection is due to the low confidence score of the test frame generated from the Faster R-CNN of small ships. The deep learning-based detection algorithms are based on features. Small targets have few features, so the their detection should use the pixel-based CFAR detection. The specific detection process includes two steps (see Figure 3). The first step is to perform the Faster R-CNN detection with VGG16 as the network for feature extraction, and then generate regions through the Region Proposal Network (RPN), and finally obtain classification scores and coordinates in object classification. The second step is to determine targets through the CFAR detection of objects with the classification scores between 0.3 and 0.8. If an object has a score above 0.8, it is directly confirmed as a target. If its score is below 0.3, then it is considered as the background. The results of experiments on the Sentinel-1 data suggest that the detection rate of the experiment using is 16% higher than using Faster R-CNN alone, but the false alarm rate has increased by 3.6%. This is because the CFAR algorithm inevitably introduces a small number of false alarms while increasing the detection rate.

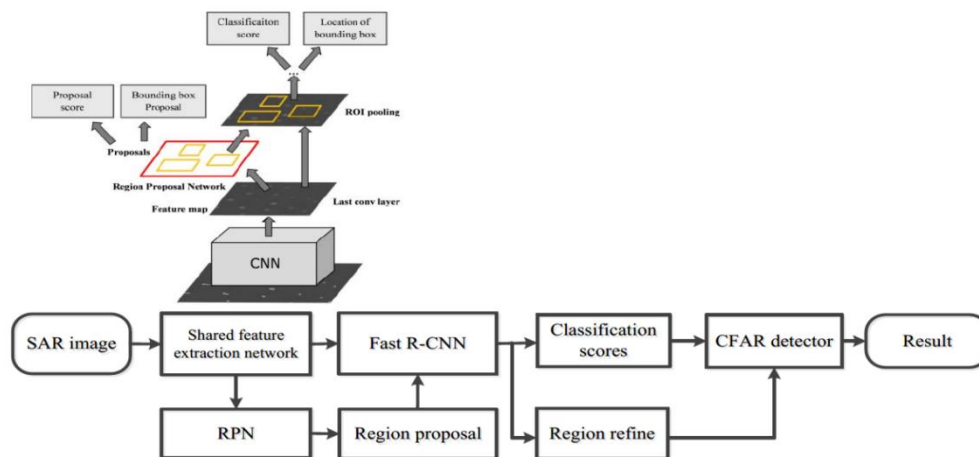


Figure 3. Faster R-CNN+CFAR Detection [23]

The above two schemes both adopt a multi-model strategy, combining the CFAR and CNN for ship target detection. The algorithm is mainly based on the traditional CFAR algorithm, while the CNN is only used for false alarm removal. The combination of Faster R-CNN and CFAR is mainly based on the target detection framework with deep learning, while the CFAR is used for tackling missed detections. Multi-model approaches usually run slower than single-model approaches. The calculation of the parameters of the CFAR algorithm also require a large amount of data.

3.2. Ship Detection Based on Target Detection Framework with Deep Learning

It is not satisfactory to apply deep learning-based CNN models to SAR images. This is because a general CNN is only proposed in the region of the last layer of feature maps. A CNN has higher resolution in shallow layers and higher semantic information in deep layers. Small-scale targets have little information on deep network feature maps. Considering the above, many scholars have improved the feature extraction network and explored the combinations of information at different levels in order to solve this problem.

It is generally believed that the deeper the CNN is, the better the performance is. However, problems such as gradient disappearance or gradient explosion arise with the increase of network depth, making the network difficult to be trained. ResNet can make the training easier while increasing the depth of the network. However, when it comes to small targets, it is difficult to retain information about small targets in feature maps generated by deep networks. Therefore, Jiao J. et al.[24] merged the feature maps generated from four residual blocks of ResNet101. The specific combining process is: first, make the feature map in the last three stages of resnet the same as the one in the first stage by sampling grid sizes of its nearest neighbor. Second, make the number of channels in the last three stages the same as that in the first stage through a 1x1 convolution. Third, add them up for combining. Finally, obtain the feature map we need through a 3x3 convolution (see Figure 4). According the tests on RadarSat2, TerraSAR-X, and Sentinel-1, the accuracy rate was up to 92.8% and the F1 score was 0.879.

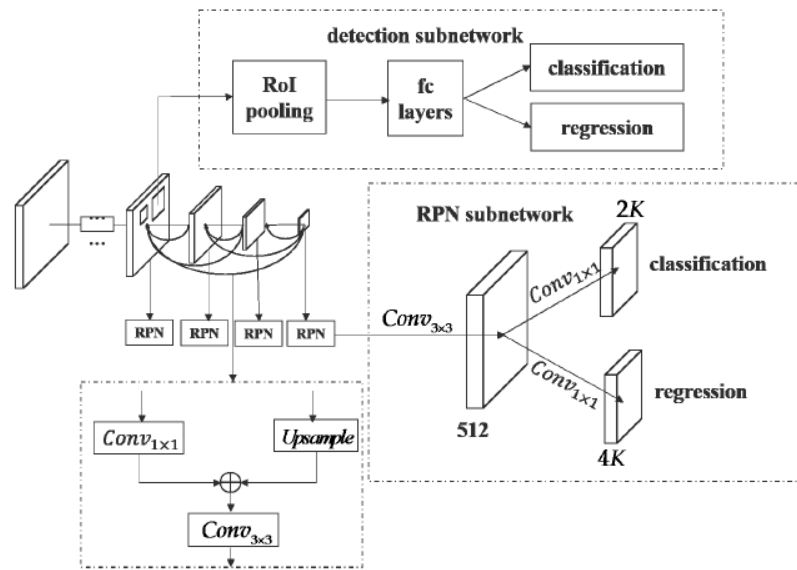


Figure 4. Target Detection Process [24]

Contextual information helps to increase confidence in decision making. Kang et al.[25] use not only multilayer fusion but also contextual information to improve the accuracy of detection. They have used VGG16 as the feature extraction network. In VGG16, conv1_2, conv2_2, conv3_3, conv4_3, and conv5_3 are called conv1, conv2, conv3, conv4, and conv5, respectively. Conv1 was reduced by max pooling; conv3 was made in the size as conv2 through sampling; and the three feature maps were merged by L2 regularization. The borders of three layers were obtained from ROI pooling. After extracting context features, three ROI features and the three context features were combined and input into two fully-connected layers. The classification and the borders were obtained by fusing the fully-connected layers (see Figure 5). The F1 score was 0.873 with the Sentinuel-1 data.

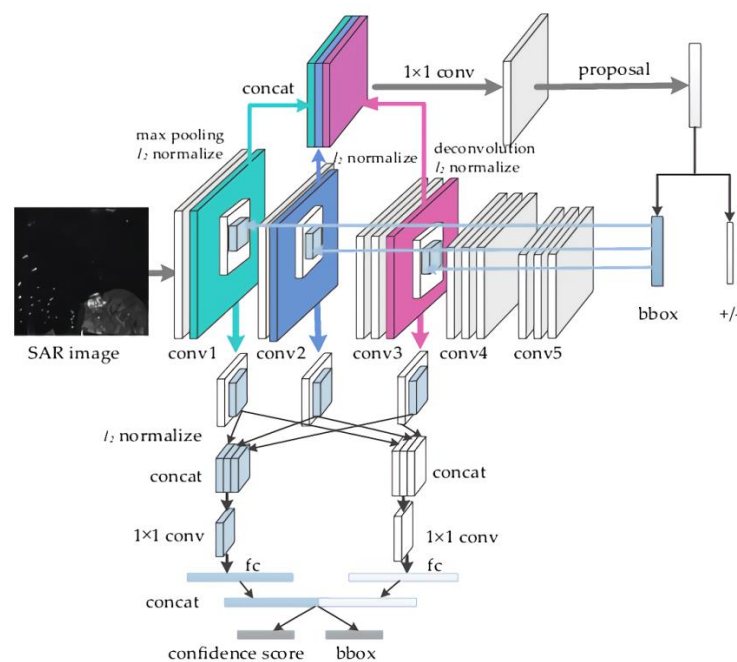


Figure 5. Fused-layer CNN with Contextual Information [25]

Single deep learning detection frameworks can meet the requirements of ship detection accuracy. The key point is the fusion of different feature layers. Feature fusion aims to provide solutions to multi-scale ship detection problems, especially the information loss in deep CNNs for small targets. There are no answers to the problems of false alarms with some land and islands and blurred azimuthal directions.

4. Optical Image for Ship Inspection

The difficult issues in the field of target detection in optical remote sensing imagery remain to be the detection of small targets, the impact of complicated near-shore sea states, and the false alarms caused by cloud coverage and broken clouds. Traditional optical RS target detection algorithms, such as the algorithm based on image gray values, are suitable for cases where the sea is relatively calm with well-distributed texture and low gray-level waterbody[26, 27, 28]. Edge-based algorithms are influenced by a lot of interference from the edges of large wave and light wave blocks[29]. In terms of algorithms based on fractal theory and fuzzy logic, if an image contains disturbance of clouds, the self-similarity of background will decrease, and the fitting error of fractal model will be quite large [30]. All the above traditional algorithms will have a lower detection accuracy at complicated sea states.

Deep learning applications for ship detection in optical imagery are mainly based on three major detection frameworks, namely Faster R-CNN, SSD, and YOLO, together with the improved versions of these frameworks.

Unlike SAR images, optical remote sensing images are more similar to natural images and match human vision more. As deep CNNs have slow convergence speed and small data set, overfitting is easy to occur in training. Therefore, people often train deep CNNs using migration learning. Nie et al. [31] have used the SSD detection framework, VGG-16 for feature extraction and migration learning to train the network, as shown in Figure 6. They tested on the Google Earth data and discovered the average precision (AP) of the SSD framework can reach 87.9% using migration learning. This has proven the effectiveness of migration learning for detecting ship targets in optical remote sensing images.

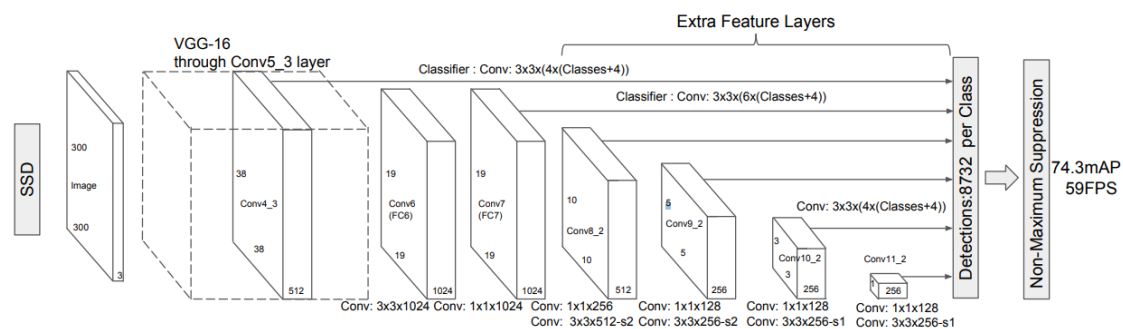


Figure 6. Flow Chart of SSD Detection [13]

Ships are multi-scale on remote sensing images. Large ships may have hundreds of thousands of pixels, while small ships may have only a dozen pixels. Li et al.[32] believe that information about shapes may be more important than local details regarding the identification of multi-scale ships. Therefore, they have proposed the Hierarchical Selective Filtering-Net (HSF-Net) based on the Faster R-CNN target detection algorithm (see Figure 7). With respects to HSF-Net, three different sizes of kernels were used for different proportions of anchor frames. The receptive fields from the smallest to the largest are 1×1 , 3×3 , and 5×5 . 1×1 kernel was used for detecting small-scale targets, while 5×5 kernel was for detecting large-scale targets. Tests were conducted with Google Earth, GaoFen-2, and UAV data, and the AP achieved 89.32%.

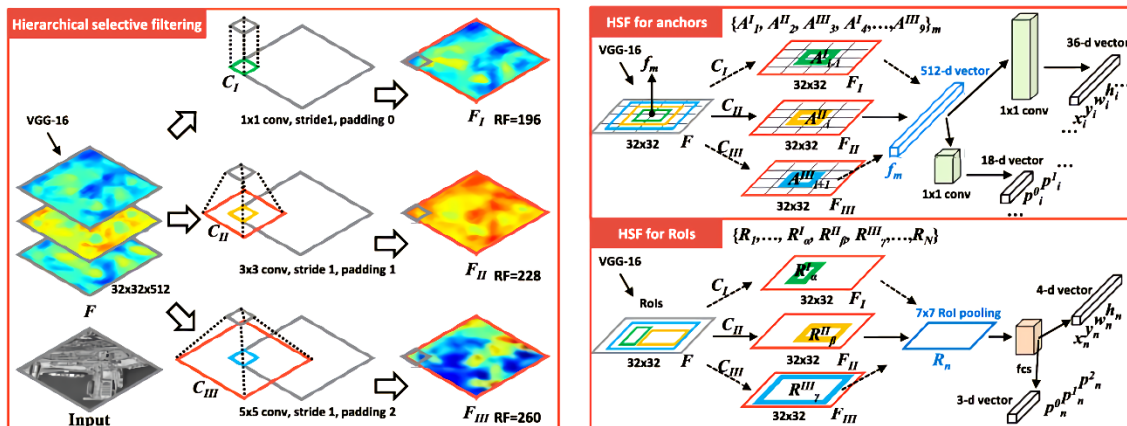


Figure 7. Schematic diagram of HSF-Net [32]

The detection of congested small-scale ships has always been a daunting task in the field of ship detection. Congested ships confuse the algorithm, so the algorithm takes multiple ships as one ship. Then, ships are often marked by only one test frame, causing missed detection of ships. Van Etten et al. have proposed a You Only Look Twice (YOLT) [33] framework based on the YOLO detection framework to tackle the detection of congested small-scale ships. In an original YOLO network, a 416×416 output picture is converted into a 13×13 feature map through feature extraction by 32 time-down sampling. If the target is less than 32 pixels, it will be difficult to be detected. Therefore, the author used the 16-time down sampling to obtain a 26×26 feature map. This method has enriched the information of small-scale ships and has also connected the information in shallow and deep layers with the passthrough layer. These strategies all contribute to the detection of small targets. The F1 score of the Digital Globe data for congested small-scale ships was 0.82.

Optical remote sensing images can accelerate the training of the network through migration learning. On the last layer of feature maps, using larger feature maps and kernels in different sizes can help solve the problems related to multi-scale ships and congested ships. However, the above studies did not consider the multi-channel characteristics of optical remote sensing images, and they did not analyze the remote sensing images with complicated sea states and cloudy sky.

5. Conclusion and Discussion

Remote sensing ship detection with deep CNNs has made great progress so far. A few studies have provided solutions to the problems occurred in traditional SAR and optical RS algorithms for ship detection. For example, concerning the difficulty small targets posed on deep learning target detection, it is generally believed that using deep learning directly does not have satisfactory detection results. Different strategies and methods are required to integrate shallow and deep layers to highlight the information of small targets. The combination of CFAR and CNN has borrowed the classification capacity of CNN. The algorithm is mainly based on the CFAR algorithm, while the CNN plays a role of reducing false alarms. The combination of Faster R-CNN and CFAR is mainly based on the deep learning-based target detection framework, while the CFAR is used for tackling missed detections. Both the above two methods adopt the multi-model strategy. As two models are used instead of one, a multi-model algorithm usually run more slowly than a single-model algorithm, and it requires a large amount of data to estimate some parameters of the CFAR algorithm. In terms of the influence of cloud on optical remote sensing, deep learning algorithms have good performance if the data is large enough. Our own experiments also show that its performance surpasses traditional algorithms, which is a very big progress. All the studies mentioned previously have failed to fully analyze and discuss the false alarms caused by islands, land margins, and other factors, which will also be a difficult task in follow-up studies.

In the future, deep learning-based RS ship detection will develop in the direction of business application. Based on the analysis of current issues, we believe that the technology can be further improved in the following aspects:

(1) With respects to date, multi-channel remote sensing data should be used to their full potential. SAR data have multiple polarization modes, and optical data have multiple channels. The above studies did not consider the influence of SAR data polarization on ship detection. In terms of optical data, only traditional true color images were used for ship detection. Future research should use more data from channels, such as four channels of high-resolution optical data, in order to improve detection performance. Where the training samples are small, data enhancement strategies such as rotation, flipping, and random addition of noise can be used to improve the generalization ability of the model. Training samples should be large enough to cover as many types of false alarms as possible including small islands.

(2) With regard to the network, future research should focus on how to effectively add some priori or specialized domain rules or knowledge to make it reflected in the network. An example of that would be using multiple classification networks. Regarding convolution implementation of sliding windows, if the scale of the ship changes greatly, different window sizes will be used. This will increase the calculation amount. If the ships are congested, and the results can be quite poor, hierarchical classification can be introduced to the structure design.

In summary, deep learning convolutional neural networks have significantly promoted the development of remote sensing technology for ship detection. Deep learning algorithms are essentially data-driven. Remote sensing data will see an explosive growth due to the development of sensing technology. In the future, the remote sensing technology for ship detection may no longer be required to distinguish optical data and SAR data, or to consider different imaging modes with different resolutions. Building and training an end-to-end model could suffice. Let us wait and see.

6. Acknowledgement

This work was supported by the National Natural Science Foundation of China (No.41476088).

7. References

- [1] Chen L, Wang Z R, Han Z, Wang G Q, Zhou H T, Shi H, Hu C and Long T 2017 A review of ship detection and recognition based on optical remote sensing image. *science and technology review*, **35**(20): 77-85.
- [2] Nieto-Hidalgo M, Gallego A J, Gil P and Pertusa A 2018 Two-Stage convolutional neural network for ship and spill detection using SLAR images. *IEEE Transactions on Geoscience and Remote Sensing*, **56**(9):5217-30.
- [3] Wang Y Q, Ma L and Tian Y 2011 State-of-the-art of ship detection and recognition in optical remotely sensed imagery. *Acta Automatica Sinica*, **37**(9):1029-39.
- [4] Liu F, Shen T S and Ma X X 2017 Convolutional neural network based multi-band ship target recognition with feature fusion. *Acta Optica Sinica*, **37**(10): 1015002.
- [5] Lecun Y, Bottou L, Bengio Y, and Haffner P 1998 Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**(11): 2278-2324.
- [6] Zhou X Y, Wang K, Li and L Y 2017 Review of object detection based on deep learning. *Electronic Measurement Technology*, **40**(11):89-93.
- [7] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc.*:1097–1105.
- [8] Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition. (preprint arXiv:1409.1556).
- [9] Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed Scott, Anguelov D, Erhan D, Vanhoucke V and Rabinovick A 2015 Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1-9.
- [10] He K M, Zhang X Y, Ren S Q and Sun J 2016 Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.

- [11] 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):1137-1146.
- [12] Redmon J, Divvala S, Girshick R and Farhadi A 2016 You only look once: unified, real-time object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779-788.
- [13] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y, and Berg A C 2016 SSD:single shot multibox detector. *European conference on computer vision, Springer, Cham*, 21-37.
- [14] Eldhuset K 1996 An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions. *IEEE transactions on Geoscience and Remote Sensing*, **34**(4): 1010-19.
- [15] Wackerman C C, Friedman K S, Pichel W G, Clement-colon P and Li X 2001 Automatic detection of ships in RADARSAT-1 SAR imagery. *Canadian Journal of Remote Sensing*, **27**(5): 568-577.
- [16] Gao G, Liu L, Zhao L J, Shi Gong T and Kuang G Y 2009 An adaptive and fast CFAR algorithm based on automatic censoring for target detection in high-resolution SAR images. *IEEE transactions on geoscience and remote sensing*, **47**(6): 1685-1697.
- [17] Tang M N, Lin T Q and Wen G J 2011 Overview of ship detection methods in remote sensing image. *Application Research of Computers*, **28**(1):29-36.
- [18] Vachon P W, Thomas S J, Cranton J, Edel H R and Henschel M D 2000 Validation of ship detection by the RADARSAT synthetic aperture radar and the ocean monitoring workstation. *Canadian Journal of Remote Sensing*,**26**(3): 200-212.
- [19] Vachon P W, Adlakha P, Edel H, Henschel M, Bruce Ramsay B, Flett D, Rey M, Staples G and Thoms S 2000 Canadian progress toward marine and coastal applications of synthetic aperture radar. *Johns Hopkins APL Technical Digest*, **21**(1): 33-40.
- [20] Jiang Q, Aitnouri E, Wang S and Zhou D 2000 Automatic detection for ship target in SAR imagery using PNN-model. *Canadian Journal of Remote Sensing*, **26**(4): 297-305.
- [21] Chen P, Huang W G, Fu B and Shi A Q 2005 An improved CFAR algorithm for ship detection in SAR imagery. *Journal of Remote Sensing*, **9**(3):260-264.
- [22] An Q Z, Pan Z X, You H J 2018 Ship detection in Gaofen-3 SAR images based on sea clutter distribution analysis and deep convolutional neural network. *Sensors*,**18**(2): 334.
- [23] Kang M, Leng X G, Lin Z and Ji K F 2017 A modified faster R-CNN based on CFAR algorithm for SAR ship detection. *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP). IEEE*,1-4.
- [24] 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 multiscene SAR ship detection. *IEEE Access*, **6**: 20881-20892.
- [25] 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.
- [26] Wang M, Luo J C and Ming D 2005 Extract ship targets from high spatial resolution remote sensed imagery with shape feature. *Geomatics and Information Science of Wuhan University*, **30**(8): 685-688.
- [27] You X J, Xu S S and Hou L 2015 A new method for ship detection based on feature fusion in optical image. *Computer Engineering and Applications*, **41**(19):199-202.
- [28] Xiao L P, Cao J and Gao X Y 2007 Detection for ship targets in complicated background of sea and land. *Opto-Electronic Engineering*, **34**(6):6-10.
- [29] Zhu C G, Zhou H, Wang R S and Guo J 2010 A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Transactions on geoscience and remote sensing*, **48**(9): 3446-3456.
- [30] Zhang D X, He S H and Yang S Q 2009 Ship targets detection method based on multi-scale fractal feature. *Laser and Infrared*, **39**(3):315-318.
- [31] Nie G H, Zhang P, Niu X, Dou Y and Xia F 2017 Ship detection using transfer learned single shot multi box detector. *ITM Web of Conferences. EDP Sciences*, **12**: 01006.

- [32] Li Q P, Mou L C, Liu Q J, Wang Y H and Zhu X X 2018 HSF-Net: Multiscale deep feature embedding for ship detection in optical remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, (99): 1-15.
- [33] Van Etten A 2018 You only look twice: Rapid multi-scale object detection in satellite imagery. (preprint arXiv:1805.09512).