Fine-grained Recognition of Ships Under Complex Sea Conditions

Jiaojiao Ma

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: m2578516632@163.com

Hong Jiang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: 249479898@qq.com

Jun Yu

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: yujun@163.com

Wei Li

Technological Center
Xi'an Technological University
Xi'an, 710021, China
E-mail: liwei@163.com

Haoqi Yang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, China
E-mail: yanghaoqi@163.com

Abstract—**For the traditional deep learning cannot solve the fog, coastal background interference, and the difficulty of small ships recognition, a multi-scale deep learning training model is proposed in this paper. Based on Faster R-CNN, this paper uses guided filtering to remove fog, as well as combined with negative sample enhancement learning to train the model, thus solving recognition of ship in complex sea conditions. And with multi-scale training strategy, the multi-scale ship samples are produced and sent to the network for training, so as to solve the problem of small target recognition. The experimental results show that compared with the Faster R-CNN, the precision and recall of our method increase by 6.43% and by 4.68% respectively. It solves the difficulty of ships recognition under complex sea conditions and small ship recognition that cannot be solved by traditional deep learning methods, the trained model has good generalization ability and robustness.**

Keywords- Ship Recognition; Complex Sea Conditions; Multi-scale; Fine-grained; Deep Learning

I. INTRODUCTION

Maritime target recognition is a key content in the field of computer vision. It has important practical significance for the automatic detection and recognition of ships [1-5], both in the civilian and military fields. With its broad application prospects, ship recognition has attracted significant attention all over the world. However, due to the problem of small size and changeable direction of ships, automatic recognition of ship in aerial images still faces many challenges. In addition, the complex sea condition also adds difficulty to ship recognition.

In recent years, with the rapid development of deep learning [6], the target recognition methods [7-11] based on deep learning has been widely used. Chen [12] et al used deep learning for the first time in the detection and recognition of ship targets. The image is searched for areas where ships may exist, and then analyzes candidate areas

to confirm whether or not they contain targets. The recognition rate is 91%.Bousetouane [13] uses the convolutional neural network to extract the features of the ship, and matches the ship with template in the template library. Then, the classification and recognition of the ship is performed, the precision is 89%.Shi [14] used the RBM which is important branch of deep learning to recognize ships, the precision is 95%.Wang [15] mainly analyzes the application of three kinds of mainstream methods (RCNN, Fast RCNN, Yolo) of deep learning about ship recognition, and compares the three methods to obtain the applicable scope of different methods. However, regardless of the above three methods or other existing methods, all of them cannot solve the following three problems: (1) it is impossible to solve the fog obstruction in practice; (2) the small ship cannot be recognized; (3) the target is smaller than the background area that results in the difficultly of training convergence.

II.   RELATED WORK

This article carries on the detection and recognition of ship on the basis of Faster R-CNN, the purpose is to achieve the precise position and recognition of ship in the aerial image. The overall algorithm flow chart for ship recognition is shown in Figure 1.
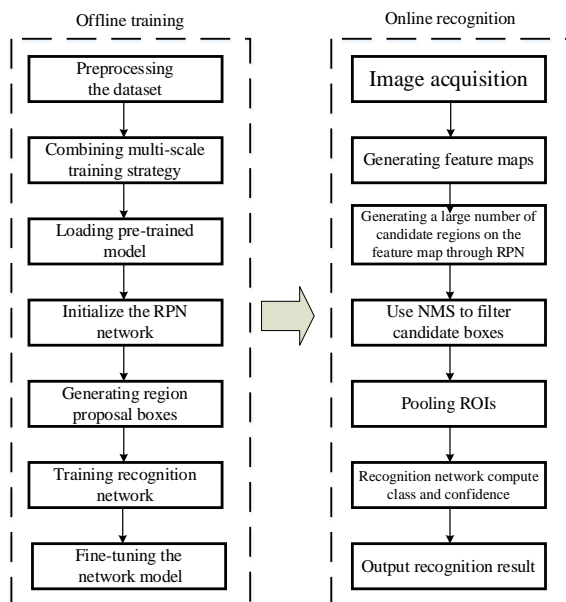

Figure 1.   Algorithm flow chart for the ship recognition.

The whole process can be divided into two parts: offline training and online recognition .In the offline training process, there are mainly four steps: the first step is to preprocess the dataset ,to remove the fog in the image ;the second step is to join the multi-scale training strategy, the collected images are scaled to three scales training; the third step is to send the feature maps to the region proposal network(RPN) to generate the ship candidate regions; the fourth step, a multi task classifier is applied to target location regression and classification for ship candidate regions. During the online recognition, or testing process, we extracted the trained network model, tested it with real-time images, and output the location and class of ship to be recognized in the image, and analyzed the performance of the network.

A. Removal of Fog

As aerial image is often affected by complex weather, such as fog, this factor will cause blind to the ships and affect the recognition in subsequent experiments. The atmospheric scattering model [16] was proposed by McCartney in 1976 based on the atmosphere. The scattering model is an important theoretical basis for the removal of fog. The model is shown in equation (1).

$$I(x) = J(x)t(x) + A(1 - t(x)) \qquad (1)$$

In equation (1), $I(x)$ is an image with fog, $J(x)$ is an image without fog, $t(x)$ is atmospheric scattering rate, which reflects the ability of ray to penetrate the fog. A denotes atmospheric light intensity, and it is a constant vector.

In this paper, the method of guided filtering [17] is used to defog. The guided filtering obtains the result image based on the equation (1), including a guide image I, an input image P, and an output image q. The guide image I needs to be set in advance according to specific application, it can also be directly taken as the input image P. For the ith pixel in the input image, the calculation method can be expressed as:

$$q_i = a_k p_i + b_k, \quad \forall i \in \omega_k \qquad (2)$$

In equation (2), i is i^{th} pixel label, a_k and b_k are linear coefficients, and ω_k is local neighborhood. By calculating the gradients on both sides of equation (2), $\nabla q = a \nabla p$ can be obtained. When the input image p has gradients, the output q also has similar gradients, so that the feature of the ship edge can be preserved while defogging. The minimum cost function between the input image with fog and the output image is shown in equation (3).

$$E\left(a_k, b_k\right) = \sum_{i \in \omega_k} \left(\left(a_k p_i + b_k - p_i\right)^2 + \varepsilon a_k^2 \right) \quad (3)$$

In the above formula, we introduce ε constants to make a_k converge. We can know from equation (3):

$$a_k = \frac{\frac{1}{|\omega|} \sum p_i^2 - \mu_k \bar{p}_k}{\sigma_k^2 + \varepsilon}, b_k = \bar{P}_k - a_k \mu_k$$

Where μ_k and σ_k^2 are the mean and variance of the guide image p in the ω_k window, $|\omega|$ is the number of pixels in the $\bar{P}_k$ window, $\bar{p}_k$ is the mean of input image p in window ω_k.

For each pixel, the final relationship between the output image q and the input image p is shown as equation (4).

$$q_i = \frac{1}{|\omega|} \sum_{k, i \in \omega_i} \left(a_k p_i + b_k\right) = \bar{a}_i p_i + \bar{b}_i \qquad (4)$$

In this paper, the fog image is used as the guide image to guide the filtering, so as to obtain the result image q after defogging, as shown in Figure 2.
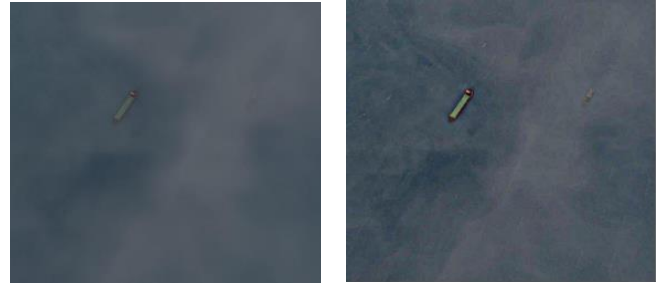
(a) Original image　　　(b) Result image

Figure 2.　Result image after defogging.

B. Multi-scale Training

As aerial images are shoot at high altitude, the ship in images is relatively small so that affect the recognition accuracy. In addition, as samples are limited and the feature of ship cannot be fully extracted during training, which can result in false or missed detection. We use multi-scale samples, that is, set each image to multiple scales. The specific idea is to reduce the training samples to three scales (1024*1024, 512*512,256*256), as shown in Figure 3, it is equivalent to add large numbers of small ship samples for training in the training set. As the number of small target samples increase, that can ensure the network extracts the features of small target effectively, thus avoiding false or missed detections. The experiment shows that the multi-scale training can make the distribution of the target sizes of various classes of training more uniform, so that the trained network model can be more robust to multi-scale targets.

Figure 3.　Multi-scale training sample images.

C. Feature Extraction

To achieve high recognition accuracy of ship, the effective features of the ship have to be extracted firstly, and the feature extraction of the ship is closely related to the structure of the convolutional neural network for feature extraction. The Convolutional Neural Network [18-20] (CNN) is a

feature extractor composed of convolutional layers and sampling layers. The advantage is that it does not require complex preprocessing. Feature extraction and pattern classification are completely put into a black box. Through continuous optimization to obtain the required parameters, CNN gives the desired classification at the output layer. The combination of convolutional layer and pooling layer will directly affect the effective feature extraction of ships. After a large number of experiments, the pre-trained ZF [21-22] network is chosen. Compared with AlexNet [23-24], GoogleNet [25-26], VGG [27-28], ResNet [29-30] and other networks, ZF has a low degree of structural complexity, and can also reduce the time complexity of feature extraction. The network structure of feature extraction is shown in Figure 4.
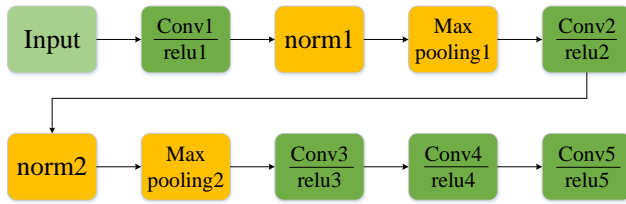

Figure 4.   Feature extraction network structure.

D. Region Proposal Network

The accuracy of the ship recognition is directly related to the quality of the candidate region which is taken in the target detection algorithm. The algorithm combines the candidate regions with CNN. If we can find a way to extract only a few hundred of high quality ship candidate regions and have a high recall, we can not only promote the effect of target recognition, but also upgrade the processing speed of ship recognition. Region Proposal Network (RPN) can be a good solution to the above problem. RPN uses information such as textures, edges, and colors in the image to find out where the ship may exist in image. It can ensure that a higher recall can be maintained when fewer regions (thousands or even hundreds) are selected. This greatly reduces the time complexity of subsequent operations, and the candidate regions obtained are of higher quality than the target of the traditional sliding window method.

The structure of regional proposal network is shown in Figure 5.
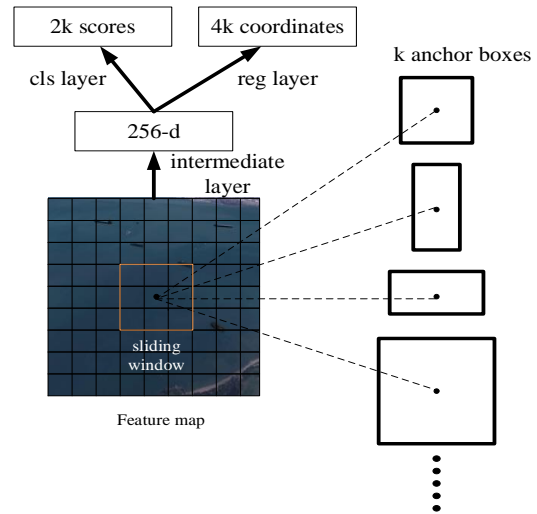
Figure 5.   Region proposal network structure.

A fixed-size window slide is used above the feature map extracted by the feature extraction network. The center point of each window corresponds to k anchor points, and each anchor point corresponds to multiple sizes and aspect ratios of sliding windows (see Fig. 5). On the right side, RPN uses 3 sizes and aspect ratios, so each slide window has k = 9 anchor points. Correspondingly, in the position of each sliding window, nine regions are simultaneously proposed, so the classification layer outputs 2×9 parameters that reflects the probability of the region where ship is. The bounding box regression layer has 4×9 parameters, representing the vertex coordinate of the nine proposal regions.

E. Model Training

After using RPN, Faster R-CNN uses Fast RCNN to recognize and classify. Fast RCNN was proposed by Ross in 2015. This method solves the problems of RCNN detection slowly and training time consuming highly, and achieves end-to-end joint training. The RPN and Fast RCNN share the convolutional features in Faster R-CNN. Fast RCNN uses the high quality proposal regions provided by RPN, which greatly increases the speed of ship recognition. The schematic diagram of the Faster R-CNN is shown in Figure 6.

Figure 6.   The architecture of proposed multi-scale Faster R-CNN for ship recognition. The simplified CNN model is surrounded by green boxes.

This article uses the RPN and Fast RCNN two-part network for ship recognition. Therefore, the back propagation algorithm cannot be used directly during the training process. Therefore, alternate training methods are used during training.

In the first stage, train RPN. The ImageNet pre-trained model (M0) is used to initialize the RPN network, and then the dataset is used to train the RPN network. After the training is completed, the model M1 is obtained.

In the second stage, train Fast RCNN. The ImageNet pre-trained model (M0) is also used to initialize the Fast RCNN network. Then the trained RPN network in first stage is used to obtain the proposal area P1, and the P1 is used to train the Fast RCNN network. After the training is completed, the model M2 is obtained.

In the third stage, train the RPN network again. Using M2 to initialize the RPN network and get the M3 network, this stage only fine-tunes the parameters of the RPN and sets the network parameters of ZF.

In the fourth stage, Fast RCNN is trained again .Use the M3 network to initialize the Fast RCNN, then use the third stage of the trained RPN network to obtain the proposal regions P2, and use P2 to train the Fast RCNN. This stage only fine tunes the parameters of the full connecting layer of Fast RCNN. In this way, both networks share the convolutional layer and form a unified network. The number of iterations for each stage is shown in Table 1.

TABLE I.   FASTER R-CNN TRAINING PROCESS

Training stage	Network	Number of iterations
1	RPN	40000
2	Fast RCNN	40000
3	RPN	80000
4	Fast RCNN	40000

The algorithm adopted in this paper uses an end-to-end network for ship recognition, which avoids the trouble of buffering transfer data in the previous multi-stage training. It has greatly improved both in speed and accuracy, and achieved robust, rapid and accurate of ship recognition.

III.   EXPERIMENT RESULTS AND ANALYSIS

A. Dataset

The dataset is provided by DataFountain and is composed of large numbers of ships photoed in bad weather, including four classes of ship: cargo ship, cruise ship, fishing ship, yacht. In order to show the recognition results better in the experiments, we call these four kinds of ships as huochuan, youlun, yuchuan and youting. Parts of the sample image are shown in Figure 7:

Figure 7.   Part of the sample images (huochuan is cargo ship, youlun is cruise ship, yuchuan is fishing ship, youting is yacht).

In the experiment, the dataset consisted of 33,397 images, of which the number of huochuan is 33756, the number of youlun is 8028, the number of youting is 13,608, and the number of yuchuan is 10,120.In addition, 5000 images with fog and coastal background were selected as nega-

tive samples of another two classes .As in deep learning, the background is also a class in the target recognition, so this experiment is set to 7 classes.

Since the recognition model adopted in this experiment is Faster R-CNN, the location of the ship must be marked in the training set firstly. That is, finding the region of interest (ROI), and setting the ground-truth of ships in the image, and mark the specific class of ship. The ROIs of some training samples are shown in Figure 8.
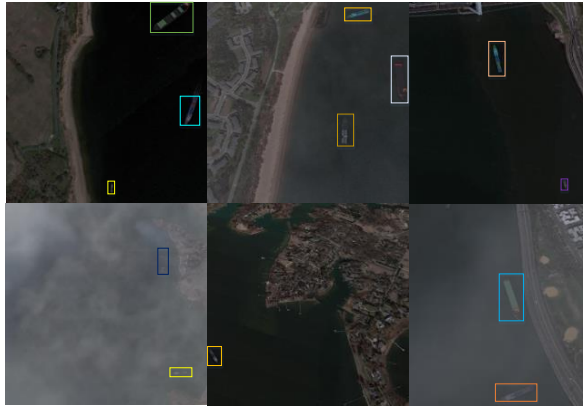
Figure 8.   The ROIs of some training samples.

In order to evaluate the effectiveness of the algorithm, this paper uses precision and recall to measure the performance of the model. Both the recall and the precision range are between [0,1]. The calculation equations are as shown in (5) and (6):

$$precision = \frac{TP}{TP + FP} \qquad (5)$$

$$recall = \frac{TP}{TP + FN} \qquad (6)$$

In the above formula, TP is the number of samples which is recognized correctly, FP is the number of false recognition, and FN is the number of missed recognition.

B. Comparison of Recognition Result After Defogging

In order to verify the removal of fog algorithm used in this paper is practical and effective, some samples with fog are selected from the dataset for testing. The experiment results are shown in Figure 9.
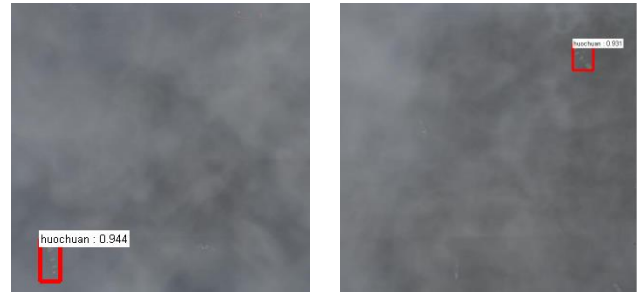
(a) Recognition results without defogging
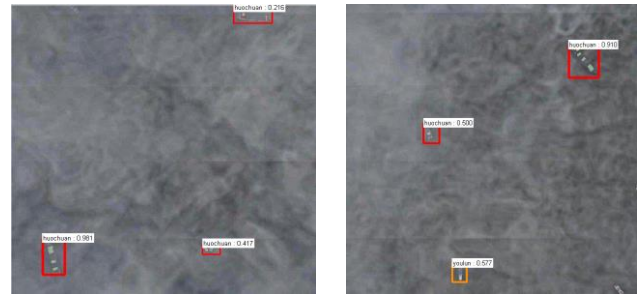
(b) Recognition results after defogging

Figure 9.   Comparison of ship recognition experiment with fog.

From the above experiment results, we can see that the method used in this paper not only can defog effectively, but also can retain the effective features of the ship as much as possible, and eliminate the unfavorable factors for ship recognition.

C. Comparison of Two Algorithms in the Same Sea Condition

To further verify the efficiency of the proposed algorithm, the algorithm is compared with Faster R-CNN, 100 images (including 341 ships) are selected from the training set for testing. The precision and recall of the two algorithms are shown in the table 2.

TABLE II.   COMPARISON OF RECOGNITION EFFICIENCY OF THE TWO ALGORITHMS

Detection Method	TP	FP	TN	precision/%	recall/%
Faster R-CNN	297	40	44	88.13	87.1
Our	313	18	28	94.56	91.78

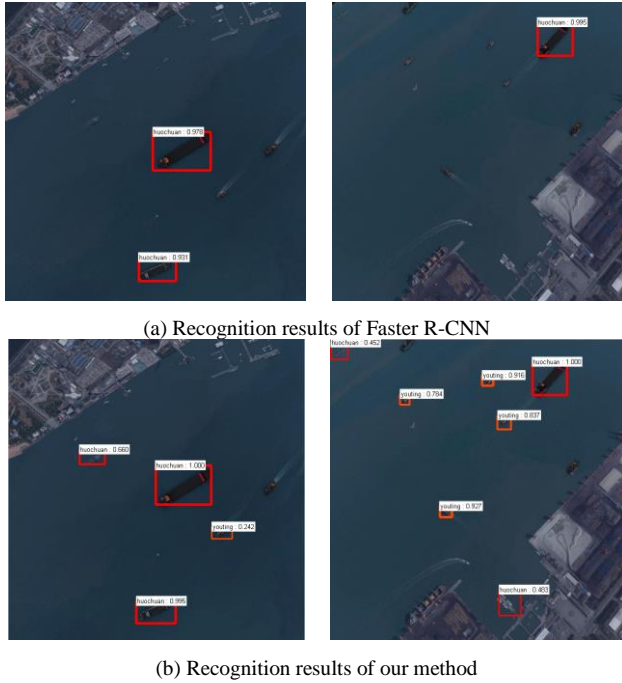(a) Recognition results of Faster R-CNN

(b) Recognition results of our method

Figure 10. Comparison of two algorithms in the same sea state.

From Table 2 above, it can be seen that the model trained by our method is higher than the model trained by Faster R-CNN in terms of precision or recall. Compared with Faster R-CNN, the precision is increased by 6.43%, the recall is increased by 4.68%. It can recognize the various classes of ship in the images. From Figure 10, it can be seen that both the precision and confidence of ship recognition have increased significantly.

D. Recognition Results Under Various Sea Conditions

In order to verify the robustness of the proposed algorithm, large numbers of images that contain various ships under different sea states were selected from the dataset for testing. Its recognition result is shown in Figure 11.

(a) Coastal background interference

(b) Dark light conditions

(c) Fog covers

(d) Different size of the ships

(d) Small ship
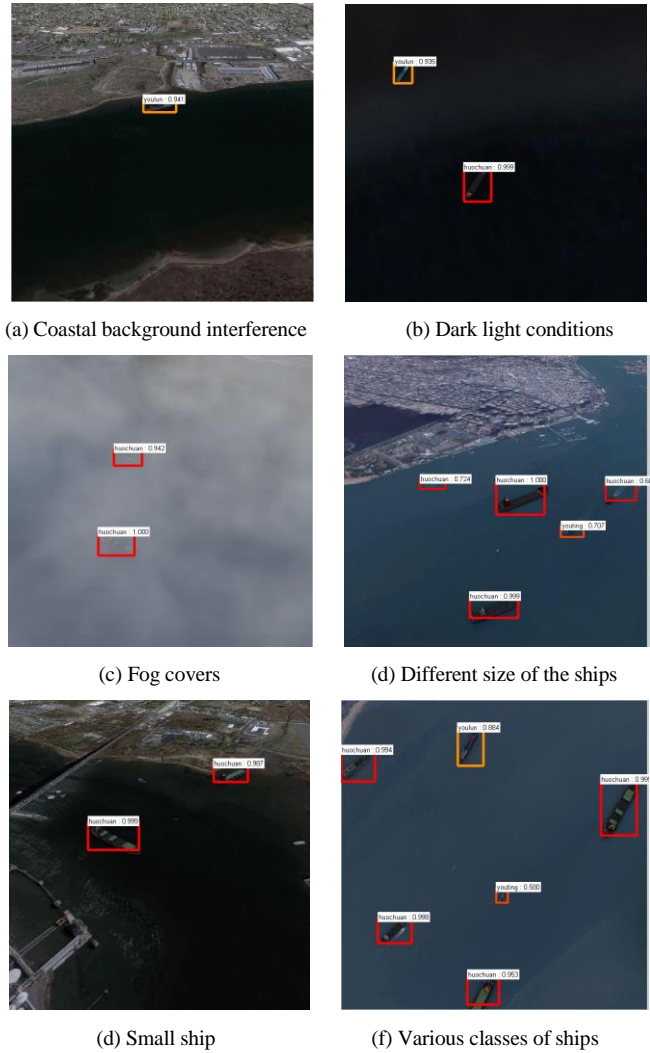

(f) Various classes of ships

Figure 11. Recognition results under various sea states.

It can be seen from Figure 11, our method can deal with the interference of the coastal background as shown in (a), (d), (e), the fog cover as shown in (c), the size of the ship differs greatly as shown in (d), the small ships as shown in (e), and the various classes of ships are as shown in (f), all have better recognition results. The experiment results show that the proposed algorithm is widely applicable to the ship recognition under complex sea states, and the accuracy and real-time performance meet the actual requirements, and it has strong robustness.

IV.   CONCLUSIONS

The purpose of this paper is to solve the difficulties of ship recognition and small target recog-

nition under complex sea conditions. Firstly, through the guidance filtering to remove fog and combined it with the negative sample reinforcement learning method to train model, the problem of the influence of the complex sea situation is solved. Then, combined with multi-scale training strategy, different sizes of ship targets are sent to the network for training, thus solving the problem of small target ship recognition. Experiments show that this algorithm can extract the ship features of different sizes and conditions. The network model has strong robustness to dark light condition, fog cover, different size of ships and interference of coastal background, and the processing time of each image is at millisecond level that can meet the requirement of real time.

REFERENCES

[1] W. Wang, Y. Ji, X. Lin, "A novel fusion-based ship detection method from Pol-SAR images," Sensors, vol. 15, pp. 25072-25089, Sep 2015.

[2] H.-L. Wang, M. Zhu, C.-B. Lin and D.-B. Chen, "Ship detection in optical remote sensing image based on visual saliency and AdaBoost classifier," Optoelectronics Letters, vol. 13, no. 2, pp. 151–155, 2017.

[3] M. Kang, X. Leng, Z. Lin, K, Ji, "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," Proc. Int. Workshop Remote Sens. Intell. Process. , pp. 1-4, May 2017.

[4] Yao. Y, Jiang. Z, Zhang. H, Zhao. D, Cai. B, "Ship detection in optical remote sensing images based on deep convolutional neural networks," J. Appl. Remote Sense, Nov 2017.

[5] C. Schwegmann, W. Kleynhans, B. Salmon, "Synthetic aperture radar ship detection using Haar-Like features," IEEE Geosci. Remote Sens. Lett., vol. 14, no. 2, pp. 154-158, Feb. 2017.

[6] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, May 2015.

[7] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Proc. IEEE Conf. Comput. Vis. Pattern Recog, pp. 580-587, 2014.

[8] R. Girshick, "Fast R-CNN," Proc. IEEE Conf. Comput. Vis. (ICCV), pp. 1440- 1448, Jun. 2015.

[9] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137-1149, Jun. 2017.

[10] J. Redmon, S. Divvala, R. Girshick, A. Farhadi,"You only look once: Unified real-time object detection," Proc. IEEE Conf. Comput. Vis. Pattern Recogn., pp. 779-788, 2016.

[11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, "SSD: Single shot multibox detector," Proc. Eur. Conf. Comput. Vis., pp. 21-37, 2016.

[12] Chen Liang, Wang Zhiru, Han Zhong, et al, "Ship Target Detection and recognition method based on visible remote sensing image," Science & Technology Review,pp.77-85, 2017.

[13] F.Bousetouane, B.Morris, "Off-the-shelf cnn features for fine-grained classification of vessels in a maritime environment," Advances in Visual Computing, pp. 379-388, 2015.

[14] Shi Danrong, "Ship detection of SAR image based on deep learning," Xidian University, 2015.

[15] Wang Bing, "Research on ship detection Based on deep learning," Xiamen University,2017.

[16] E. J. McCartney, "Optics of the Atmosphere: Scattering by molecules and particles," John Wiley and Sons, 1975.

[17] K. He, J. Sun, X. Tang, "Guided image filtering," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 6, pp. 1397-1409, Jun. 2013.

[18] N. Ketkar, "Convolutional neural networks in Deep Learning with Python,"Springer,pp. 61-76, 2017.

[19] P. Kim, "Convolutional neural network in MATLAB Deep Learning," Springer, pp.121-147, 2017.

[20] E. Maggiori, Y. Tarabalka, G. Charpiat, P. Alliez, "Fully convolutional neural networks for remote sensing image classification," Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), pp. 5071-5074, Jul. 2016.

[21] M.A. Rafique, W. Pedrycz, M. Jeon, "Vehicle license plate detection using region-based convolutional neural networks, "Soft Comput., pp. 1-2, 2017.

[22] M. D. Zeiler, R. Fergus, "Visualizing and understanding convolutional networks," European conference on computer vision, pp. 818-833, 2014.

[23] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Proc. 25th Int. Conf. Neural Inf. Process. Syst., pp. 1097-1105, 2013.

[24] J. Hosang, M. Omran, R. Benenson, B. Schiele, "Taking a deeper look at pedestrians," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 4073-4082, 2015.

[25] C. Szegedy et al., "Going deeper with convolutions," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1-9, 2015.

[26] Z. Zhong, L. Jin, Z. Xie, "High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps," Proc. 13th Int. Conf. Document Anal. Recognit., pp. 846-850, 2015.

[27] A. Vedaldi, K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," Proc. 23rd Annu. ACM Conf. Multimedia Conf., pp. 689-692, 2015.

[28] T. Sercu, C. Puhrsch, B. Kingsbury, Y. LeCun, "Very deep multilingual convolutional neural networks for LVCSR," Proc. IEEE Int. Conf. Acoust. Speech Signal Process, pp. 4955-4959, Mar. 2016.

[29] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE Conference on ComputerVision and Pattern Recognition, pp. 770-778, 2016.

[30] J. B. Zhang, Y. Zheng, D. K. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," Proc. 31st AAAI Conf. Artif. Intell., pp. 1655-1661, 2016.