High resolution remote sensing image ship target detection technology based on deep learning*

WANG Min (王敏)[1,2,3]**, **CHEN Jin-yong** (陈金勇)[1,2], **WANG Gang** (王港)[1,2], **GAO Feng** (高峰)[1,2], **SUN Kang** (孙康)[1], **and XU Miao-zhong** (许妙忠)[3]

1. The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang 050081, China

2. CETC Key Laboratory of Aerospace Information Applications, Shijiazhuang 050081, China

3. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

(Received 7 January 2019; Revised 9 April 2019)

©Tianjin University of Technology and Springer-Verlag GmbH Germany, part of Springer Nature 2019

With the development of China's high-resolution special projects and the rapid development of commercial satellite, the resolution of the mainstream satellite remote sensing images has reached the sub-meter level. Ship target detection in high-resolution remote sensing images has always been the focus and hotspot in image understanding. Real-time and effective detection of ships play an extremely important role in marine transportation, military operations and so on. Firstly, the full-factor ship target sample library of high-resolution image is synthetically prepared. Then, based on the Faster R-CNN framework and Resnet model, optimize the parameters of the model to achieve accurate results. The simulation results show that the detection model trained in this paper has the highest recall rate of 98.01% and false alarm rate of 0.83%. It can be applied to the practical application of ship detection in remote sensing images.

Document code: A **Article ID:** 1673-1905(2019)05-0391-5

DOI https://doi.org/10.1007/s11801-019-9003-7

With the development of China's high-resolution special projects and the rapid development of commercial satellite companies, the resolution of the mainstream satellite remote sensing images in China has reached the sub-meter level. Target detection has always been the focus research in image target recognition and image understanding[1,2]. Real-time and effective detection and recognition of targets play an important role in both military and civilian applications. High resolution remote sensing images provide important data sources for obtaining detailed information of targets[3]. Traditional target detection methods usually aim at a specific kind of target and strongly depend on the characteristics of professional knowledge and data itself, usually have more stringent requirements for input images and are difficult to solve the impact of image resolution, shooting angle, illumination and noise on detection accuracy.

In recent years, with the research and development of deep learning method and its success in natural scene image interpretation, large data analysis and other applications, deep learning has gradually attracted the attention of remote sensing image interpretation researchers[4-6]. However, although the technology of deep convolution neural network has been developed and made considerable progress, the latest deep learning network model and accelerated algorithm are not applied to the field of remote sensing image processing, and the integrated system design of sample collection, sample production, network design and detection and recognition is lacking. In-depth learning algorithm needs a lot of sample data to optimize the network parameters[7-10], but there is no open large-scale military target sample library at present. Therefore, this paper firstly uses a small number of existing military-civil-commercial satellite manual labeling samples, combined with semi-automatic interactive labeling technology, to establish large-scale targets; The latest residual network model under the framework of deep learning is applied to ship target detection, and a general remote sensing target detection system is designed of sample production, database management, network design, training optimization, detection. We will study how to use deep learning method for feature learning and model building to improve the accuracy and robustness of ship target extraction and detection, which has great theoretical significance for the current research of remote sensing image target detection technology.

The number of training samples does not play a decisive role in the traditional ship target detection algorithm[11-13]. However, in the deep learning algorithm, a large number of training samples are the guarantee of the optimal detection model, and the training samples need to contain various scenarios, so that the detection model

* This work has been supported by the 63rd Batch of China Postdoctoral Science Foundation (No.2018M631767).

** E-mail: wmin0805@163.com

can be good enough in robustness[14-17]. However, on the one hand, the amount of remote sensing image data to meet the requirements is limited, on the other hand, the efficiency of manual interpretation and interpretation is often low, so the preparation of large target sample bank is difficult.

Combining with the characteristics of multi-scale high-resolution image ship target, this paper synthetically applies artificial interactive interpretation and imaging process simulation to study all-factor ship target samples preparation, so as to provide a good foundation for the model training. Firstly, collect the measured images through commercial and shared channels and extract the sample slices. Then, combined with the measured remote sensing image slices, the target slices are interpolated and estimated in the parameter space by image processing algorithm with the support of optical physics/empirical radiation transfer model. Finally, by establishing a unified classification and coding system, appropriate logic and physical storage mechanism, a set of high-resolution image ship total factor samples is formed, which supports the optimization and application of depth neural network model. The technical process of ship total factor sample preparation is shown in Fig.1.

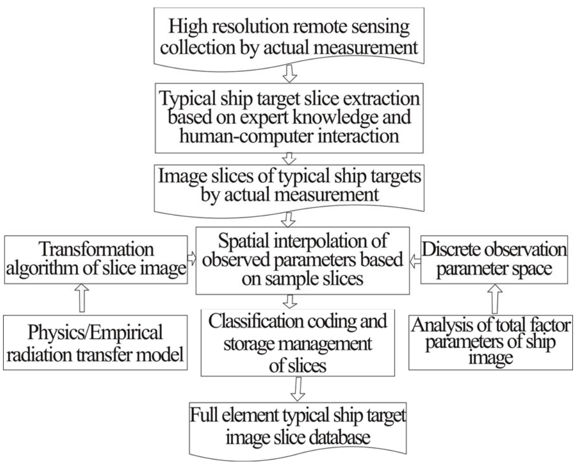


Fig.1 Flow chart of sample making

In this paper, a series of naval base websites of major countries in the world are searched and sorted out manually, and the names and coordinates of military ports are automatically obtained from the websites by using crawler technology according to keywords. Then, according to the longitude and latitude coordinates of ports, the corresponding remote sensing images are downloaded on Google Earth and domestic high-resolution satellites, including warships, destroyers and frigates with the resolution between 0.3 m and 2 m. According to the characteristics of ship targets in high-resolution remote sensing images, five methods of increasing samples are proposed, including image rotation, image noise, brightness contrast transformation, image shearing, and adding different cloud occlusion. Experiments are carried out to verify the validity of the samples for training results.

This paper chooses Faster R-CNN framework model based on depth residual network to detect ship targets in high-resolution remote sensing images[18,19]. Faster R-CNN model consists of three network structures: convolutional neural network (CNN), region Proposal network (RPN) and region-CNN (R-CNN)[8]. The simplified structure of the model is shown in Fig.2. RPN is a fully convolutional neural network. Its purpose is to provide a candidate frame for suspected targets. R-CNN fine-tunes the suspected candidate frame and detects and identifies the targets in the candidate frame. RPN and R-CNN share the deep image features of CNN. CNN network is a deep convolution neural network used to extract the deep features of images. In this paper, ResNet network is selected as the deep residual network.

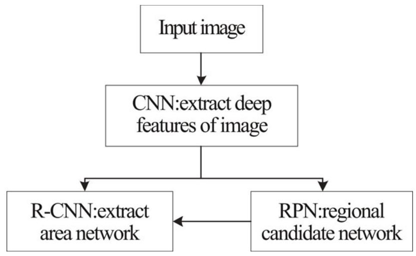


Fig.2 Structure chart of faster R-CNN

The deep residual convolution neural network is mainly used to extract the deep features of images in the whole network framework. The ResNet network has a deep network structure which can extract different levels of image information. At the same time, the residual structure is used to accelerate the training of the ultra-deep neural network. The residual structure can effectively eliminate the upper error increases of training set caused by the deepening of the network, which has a good model detection accuracy as shown in Fig.3. In each residual network, residual learning is used every few layers, which is defined as follows: $y=F(x)+x$, where x and y are the input and output of the unit, $F(x)$ represents the mapping that the residual unit needs to learn, and add X after learning $F(x)$.

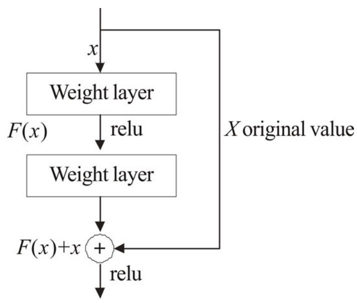


Fig.3 Residual unit

The structure of the ship target detection network model is shown in Fig.4. The convolution network part (CNN) is the part that can be optimized to extract the deep features of the image. The ResNet-50 network is selected in this paper.

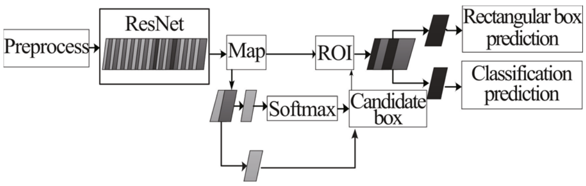


Fig.4 Structure chart of ship detection network model

The software environment in this paper is: Windows 7 system, Python language, Tensorflow + keras, the hardware environment is: CPU Intel Xeon E5-2620 v4-8 Cores, GPU NVIDIA P5000, 64G memory. In order to verify the validity of each amplification algorithm, 462 ship target slices with resolution between 0.2 m and 3 m are selected for sample amplification processing and model training respectively. the sample data amplified is shown in Tab.1.

Tab.1 Results of sample amplification

	Original	Rotation	Noise	Brightness	Shear	Cloud occlusion
Magnitude	No	Rotating anti-clockwise every 45°	Adding 4 different pepper and salt noises	2 types of brightness, 2 of types contrast	Shearing in 4 directions randomly	Occlusion of 4 different clouds at different locations
Number	462	3 696	2 310	2 310	2 310	2 310

It is mainly divided into two processes: training and testing, shown in Fig.5. In the training stage, this paper transfers and learns from the weight of ImageNet data set, inputs the training set and verification set data. First normalizes and standardizes the data, then inputs the standard data into the network, and through layer-by-layer calculation, propagates it forward to the network, the final classification results are obtained, and compared with the labels. When the *IOU* (intersection-parallelism ratio = intersection/union) is not less than 0.7, it is considered that the detection is accurate and labeled as positive label, otherwise the detection is error. Then calculate the loss function of the positive label, adjust the parameters of each layer of the network by the SGD method. When the value of loss function does not change, the iteration is stopped after many iterations. Relu is selected as Excitation function in this paper. At this time, the network serves as the network model for ship detection. In the testing stage, the test sample set is input into the trained network, then the classification score and location parameters are obtained. The accuracy and false alarm rate of the model are further calculated as the evaluation criteria of the model.

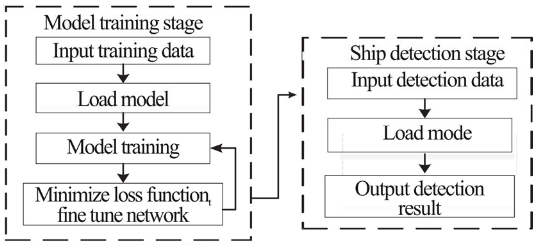


Fig.5 Flow chart of the detection and training

In this paper, three evaluation criteria are used to evaluate the test results, which are expressed as follows: accuracy (probability of correct target in all targets detected), recall rate (probability of correct target detected), false alarm rate (probability of incorrect target in all targets detected):

$$P=\frac{m_1}{m_1+n}, \tag{1}$$

$$R=\frac{m_1}{m}, \tag{2}$$

$$F=\frac{n}{m_1+n}, \tag{3}$$

where m denotes the total number of correct targets, m_1 denotes the number of correct targets detected, and n denotes the number of false alarms that the detected targets are not correct targets.

The model is trained for the above sample amplification data, and 540 additional images are selected for ship detection experiments. The total number of targets is 855 by statistics. The training time of each experiment is about 30 h, which ensures that the experiment achieves its best convergence result. At last, we choose all sample slice with 9 240 images (except shear) to train model, then detect with same sample set. The last detect results and false alarm results are shown in Figs.6 and 7. The test results are shown in Tab.2 below.

From Fig.7, it can be seen that the network model trained in this paper can detect several side-by-side ships and coastal ships which are difficult to detect. It can be seen that the false alarm targets originally generated, including waves, islands were not made mistake except Fig.7(d). Compared with the model trained from the original sample, the recall rate of target detection is increased from 85.49% to 96.03%, and the false alarm rate is reduced from 2.79% to 0.83%. Through experiments, we find that image rotation has the greatest impact on the detection recall rate of ship detection in remote sensing images, which can reach 98.01% because of the symmetry would not change the information of ship, but the network needs to learn. At last, we test on a wide range sensing image, getting the results with 9 right and 1 false of 10 targets which is shown in Fig.8.

Tab.2 Experimental results

Data set	Test number	Test target number	Correct target number	False alarm number	False alarm rate	Recall rate
Original sample			731	21	2.79%	85.49%
Rotation			838	31	3.57%	98.01%
Noise	540	855	755	24	3.08%	88.30%
Brightness			754	19	2.46%	88.19%
Shear			740	21	2.76%	86.55%
Cloud occlusion			757	29	3.69%	88.54%
All (except shear)			824	7	0.83%	96.03%

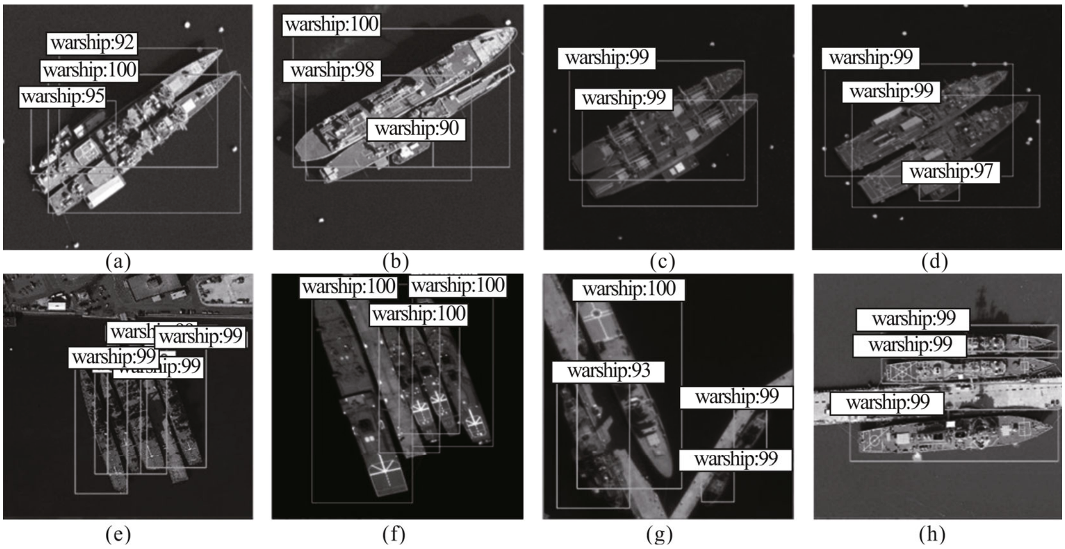
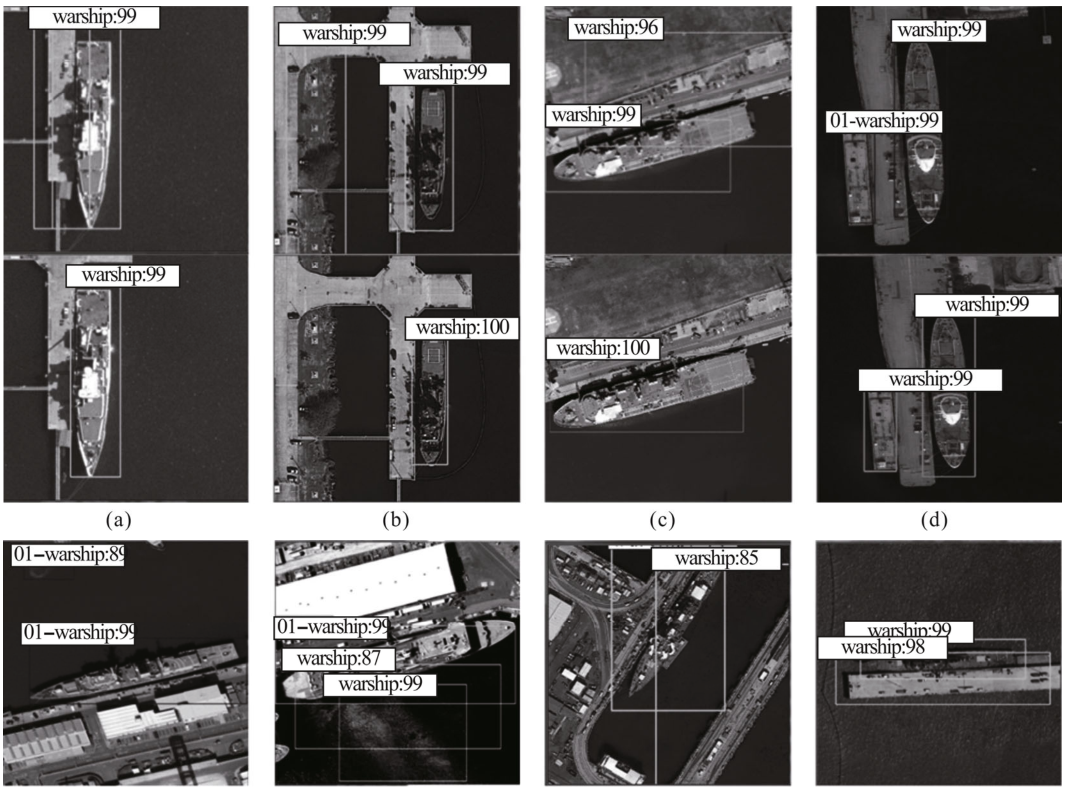
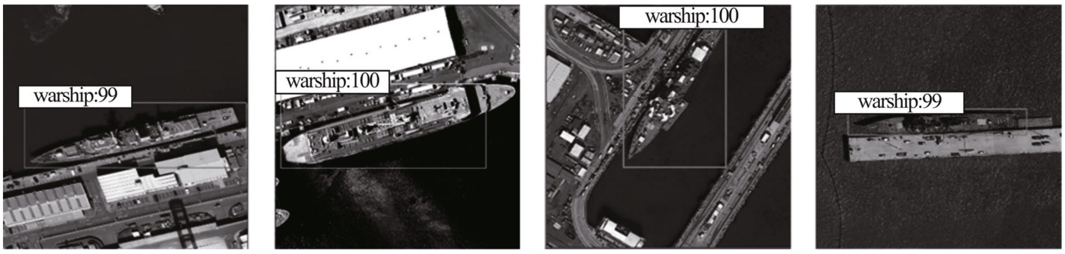


Fig.6 Typical results of ship detection





(e) (f) (g) (h)

Fig.7 Comparison of false alarm results of ship detection

Fig.8 Ship detection results of wide range sensing image

This paper mainly studies the application of deep learning algorithm in the field of ship target detection in high resolution remote sensing images. Firstly, the multi-scale and multi-type sample database is constructed by means of sample expansion technology, and then the Faster RCNN network model based on ResNet is constructed. Taking the ship target detection of high-resolution remote sensing image as an example, the simulation experiment shows that the maximum ship detection recall rate is 98.01% and the false alarm rate is 0.83%, which can meet the needs of practical engineering applications.

References

[1] C X Gao and N Sang, Bulletin of Surveying and Mapping **S1**, 108 (2014). (in Chinese)

[2] L P Xiao, J Cao and X Y Gao, Opto-Electronic Engineering **34**, 6 (2007). (in Chinese)

[3] Y H Zhao, X Q Wu and L Y Wen, Opto-Electronic Engineering, **35**, 8 (2008).

[4] M Busta,L Neumann and J Matas, An End-to-End Trainable Scene Text Localization and Recognition Framework, IEEE International Conference on Computer Vision, 2204 (2017).

[5] Alex Krizhevsky, Ilya Sutskever and Geoffrey E Hinton, Imagenet Classification with Deep Convolutional Neural Networks, The 25th International Conference on Neural Information Processing Systems, 1097 (2012).

[6] L Eikvil, L Aurdal and H Koren, ISPRS Journal of Photogrammetry and Remote Sensing **64**, 65 (2009).

[7] M Everingham, S M Eslami, L Gool, C K Williams, J Winn and A Zisserman, International Journal of Computer Vision **111**, 98 (2015).

[8] R Girshick, Fast R-CNN, IEEE International Conference on Computer Vision, 1440 (2015).

[9] K. Simonyan and A. Zisserman, Very Deep Convolutional Networks For Large-Scale Image Recognition, arXiv:1409.1556v6, 2015.

[10] C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke and A Rabinovich, Going Deeper with Convolutions, IEEE Conference on Computer Vision and Pattern Recognition, 1 (2015).

[11] R Girshick, J Donahue, T Darrell and J Malik, Rich Feature Hierarchies for Accurate Object Detection and Segmentation, IEEE Conference on Computer Vision and Pattern Recognition, 580 (2014).

[12] D M Gonzalez, M Volpi, N Komodakis and D Tuia, Rotation Equivariant Vector Field Networks, arXiv:1612.09346v3, 2017.

[13] J F Henriques and A Vedaldi, Warped Convolutions: Efficient Invariance to Spatial Transformations, The 34th International Conference on Machine Learning, 1461 (2017).

[14] A Kembhavi, D Harwood and L S Davis IEEE Transactions on Pattern Analysis and Machine Intelligence **33**, 1250 (2011).

[15] J Leitloff, D Rosenbaum, F Kurz, O Meynberg and P Reinatz,. Remote Sensing **6**, 11315 (2014).

[16] K Liu and G Mattyus, IEEE Geosci. Remote Sensing Letters **12**, 1 (2015).

[17] T Y Tang, S L Zhou, Z P Deng, H X Zou and L Lei, Sensors **17**, 336 (2017).

[18] G S Xia, X Bai, J Ding, Z Zhu, S Belongie, J Luo, M Datcu, M Pelillo and L Zhang, DOTA: A Large-Scale Dataset for Object Detection in Aerial Images, arXiv:1711.10398v3, 2019.

[19] Y Zhou and O Tuzel, Voxelnet: End-to-End Learning for Point Cloud Based 3D Object Detection, arXiv:1711.06396, 2017.