

PAPER • OPEN ACCESS

## An On-orbit Target Detection Method in Remote Sensing Images for Micro Satellite

To cite this article: Qihui Zhou *et al* 2021 *J. Phys.: Conf. Ser.* **2006** 012030

View the [article online](#) for updates and enhancements.

You may also like

- [Gravity Probe B gyroscope readout system](#)  
B Muhlfelder, J Lockhart, H Aljabreen et al.
- [THE HERschel INVENTORY OF THE AGENTS OF GALAXY EVOLUTION IN THE MAGELLANIC CLOUDS. A HERSCHEL OPEN TIME KEY PROGRAM](#)  
M. Meixner, P. Panuzzo, J. Roman-Duval et al.
- [Calibration chain design based on integrating sphere transfer radiometer for SI-traceable on-orbit spectral radiometric calibration and its uncertainty analysis](#)  
Wei-Ning Zhao, , Wei Fang et al.



**ECS**  
The  
Electrochemical  
Society  
Advancing solid state &  
electrochemical science & technology

**DISCOVER**  
how sustainability  
intersects with  
electrochemistry & solid  
state science research

# An On-orbit Target Detection Method in Remote Sensing Images for Micro Satellite

Qihui Zhou<sup>1\*</sup>, He Cui<sup>1</sup>, Shengming Liang<sup>1</sup> and Huiqin Li<sup>1</sup>

<sup>1</sup> Space Star technology CO., Ltd, City, Beijing, 100089, China

\*Corresponding author's e-mail: zhouqh@spacestar.com.cn

**Abstract.** Aiming at addressing the contradiction between the information timeliness and on-board computing capacity and data transmission capacity in on-orbit image processing system of small commercial remote sensing satellites, a convolutional neural network based target detection algorithm is proposed. The calculation amount of the model is optimized to 25MFLOPS by using transfer learning, sparse training and weight quantification, while the weight file data is reduced from tens of Mb to 0.5Mb, which makes it possible for small commercial remote sensing satellite to detect target on-orbit quickly, and meets the requirements of the timeliness of remote sensing information.

## 1. Introduction

In recent years, due to the rapid development of aerospace remote sensing earth observation technology, commercial microsatellite technology has achieved rapid breakthroughs [1-4]. Nowadays, commercial remote sensing satellite imaging has the advantages of high resolution, wide coverage, and rich details. At the same time, it has also led to an exponential increase in the data volume of on-orbit optical remote sensing images[5]. For micro-satellites, in the case of limited satellite data transmission bandwidth, it is impossible to solve the contradiction between the short imaging time, the large amount of on-orbit image data, the low efficiency of downloading images and the user's demand for timeliness of remote sensing information [6-7]. With the development of artificial intelligence, deep learning and real-time target detection based embedded system have been potentially used in ground applications [8-13]. Real-time detection function of on-board targets have been achieved by using on-orbit real-time processing of remote sensing images [14], remove invalid information such as cloud cover or calm seas, and extract effective military target information such as ships or aircraft [15-16]. The amount of redundant information have been reduced significantly, and have improved the efficiency of data use, reduced the pressure of data transmission, so as to achieve real-time reconnaissance of large military targets, tracking and monitoring their activities, and grasp the deployment and target trends of large maritime military forces in sensitive areas. Therefore, real-time on-orbit target detection processing has important practical application value in the field of military or commercial satellites.

This paper focuses on the essence of optical remote sensing image and the problem of target detection. A complete set of convolutional neural network based on-orbit real-time target detection and processing method is proposed, and the technical principles involved in the algorithm process of on-orbit real-time target detection and processing are described in detail, including the basic composition and training methods of convolutional neural network. Combined with the characteristics of the actual micro-satellite's weak on-board processing capacity. Through the network compression method based on transfer learning, sparse training and weight quantification, the calculation amount of the model is optimized to 25MFLOPS, and the processing efficiency of the algorithm is improved. The weight file



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

data volume is reduced from tens of Mb to 0.5Mb, which reduces the on-board storage pressure and lays a foundation for the detection of in-orbit targets in small commercial remote sensing satellites.

## 2. Target detection method based on neural network

Artificial Neural Networks (ANN) system has appeared since the 1940s. It is connected by a large number of neurons with adjustable connection weights. Its large-scale parallel processing, distributed information storage, and good self-organization and self-learning capabilities makes it becomes one of the most widely used neural network learning algorithms. It contains an input layer, an output layer, and an intermediate layer between the input and output layers. There are single layer or multiple layers in the middle layer. Because they have no direct connection beyond the network, they are also called hidden layers, similarly, the neurons in the hidden layer are also called hidden units, and their state affects the relationship between input and output. This also means that changing the weight coefficient of the hidden layer can change the performance of the entire multilayer neural network, as shown in Figure 1.

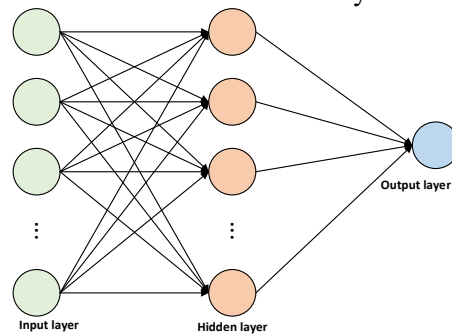


Figure 1. Structure of artificial neural network

In the mathematical model, neurons simulate nerve cells in biology. Biological neuron cells are composed of synapses, dendrites, cell bodies, axons, etc., and axons can output signals to other neurons<sup>[6]</sup>, completing information transfer between neurons. The information processing between neurons is a non-linear process, and an abstract neuron mathematical model of the BP neural network can be obtained, as shown in Figure 2.

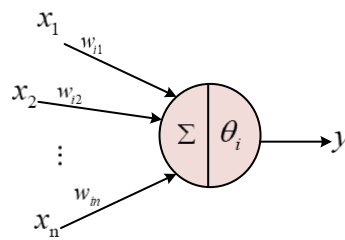


Figure 2. Neuronal structure

For the  $i^{\text{th}}$  neuron,  $x_1, x_2, \dots, x_n$  are the input of the neuron, and the input is often the independent variable that has a key influence on the system model,  $w_{i1}, w_{i2}, \dots, w_{in}$  adjust the weighting ratio of each input quantity for the connection weight. There are many ways to combine and input signals to neurons. Choose the most convenient linear weighted summation to get the neuron  $Net_{in}$  input as:

$$Net_{in} = \sum_{i=1}^n w_i * x_i \quad (1)$$

$\theta_i$  represents the threshold of the neuron, according to the knowledge of biology, it will be activated only when the information received by the neuron reaches the threshold  $\theta_i$ , it will compare  $Net_{in}$  with  $\theta_i$ , then the activation function is used to generate the input of neurons. The transfer function of BP neural network is usually SIGMOD type differentiable function, which can transform the input signal

from negative infinity to positive infinity into output between 0 and 1, so it can realize any nonlinear mapping between input and output, which makes it in such fields as signal processing, computer network, process control, speech recognition, function approximation. It has also been successfully applied in the fields of pattern recognition and data compression.

### 2.1. Convolutional neural network

To recognize an image using traditional neural network, every pixel of the image needs to be connected to the hidden layer node. For a  $1000 \times 1000$  pixel image, if there is 1M hidden layer unit, there are  $10^{12}$  parameters, which is obviously unacceptable, as shown in the Figure 3.

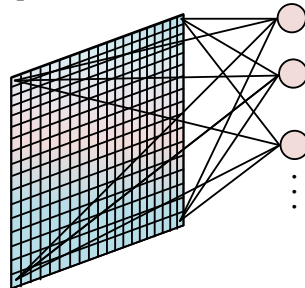


Figure 3.  $1000 \times 1000$  picture connection in traditional neural network

Convolutional neural network (CNN) is a variant of artificial neural network, which comes from Hubel and Wiesel's research on cat's primary visual cortex. The primary visual cortex consists of simple cells and complex cells. Simple cells mainly perceive specific edge stimuli in their local receptive fields, while complex cells respond to edge stimuli with larger local receptive fields as inputs. CNN is good at dealing with large image related machine learning problems. It assumes that the bottom features of images are local and that the features of different small segments and small segments of different images are similar, that is, it can use the same set of classifiers to describe a variety of different images. Different from the conventional neural network, the neurons in each layer of convolutional neural network are arranged in three dimensions: width, height and depth. The width and height correspond to the convolution itself (two-dimensional template), but the depth in convolutional neural network refers to the third dimension of the active data body, not the depth of the whole network, as shown in Figure 4.

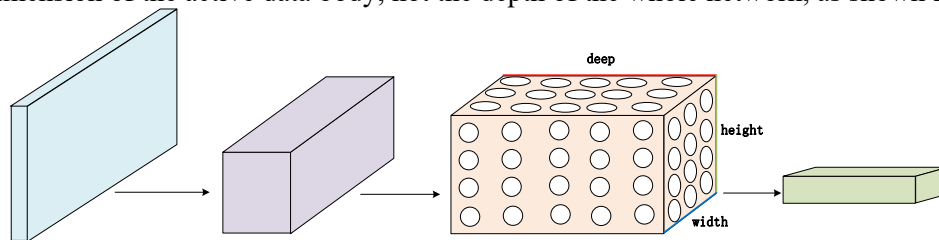


Figure 4. The infrastructure of CNN

Through a series of methods, the dimension of the image recognition problem with huge amount of data is continuously reduced, and finally it can be trained. The main ideas of CNN are as follows:

- Local connection and weight sharing

Convolutional neural network uses the way of local connection and weight sharing to reduce the number of parameters, accelerate the training speed of the model and play a role of regularization. The local receptive field can greatly reduce the model parameters. The so-called local receptive field refers to the network is not fully connected, each neuron is connected to the previous layer of local small area. By this way, the parameters between layers are greatly reduced.

In order to further reduce the number of weights, CNN adopts the network mode of weight sharing. The weight sharing makes the convolution kernel of each local receptive field detect the same feature in different positions, forming a two-dimensional response output: feature map. Using weight sharing can greatly reduce the number of training parameters, especially when the input image resolution is large. From the perspective of feature extraction, the local receptive field can extract primary visual features

(such as corner, edge contour, etc.) from the two-dimensional image, and the subsequent layers combine these primary features to form higher-level and more abstract features.

- Pooling operation

Another important idea of CNN is pooling. After obtaining the convolution features of the image, the dimension of the feature map of the convolution layer should be reduced by pooling operation. Pooling operation is to map features of continuous sub regions with fixed size in the horizontal and vertical directions with a certain step size. Pooling operation is usually maximum pooling or average pooling, that is, selecting the maximum value or calculating the mean value in the fixed size sub region as the mapping value. Through pooling operation, the number of neurons is reduced and the complexity of subsequent network is simplified, which makes CNN invariant to the local changes of input and effectively simulates the complex cells of primate visual cortex.

## 2.2. The basic structure of convolutional neural network

The typical structure of convolutional neural network is composed of stacked convolution layer, pooling layer and multi-layer perceptron in series. The convolution layer and pooling layer contain multiple feature maps. The feature map is a two-dimensional response output obtained by performing the same operation (convolution or pooling) on the local receptive field of the previous layer.

- Convolutional layer

Convolution operation extracts rich features through a variety of convolution kernels to form feature maps with different responses. Assume that the input of the previous layer of the convolution layer is expressed as  $X \in R^{W \times H \times C}$ , which is a three-dimensional matrix, W and H are represented respectively the width and height of the feature graph, C represents the number of channels, that is, the number of characteristic graphs. X is convoluted by K groups of convolution kernels  $\{W_k\}_{k=1}^K$  ( $X \in R^{W \times H \times C}$ ). The convolution layer is obtained by adding bias and nonlinear transformation  $Y \in R^{W_0 \times H_0 \times C_0}$ . The relationship between each feature graph in convolution layer  $Y^k \in R^{W_0 \times H_0 \times C_0}$  and the previous layer as show

$$Y^k = g(\sum_{c=1}^C X^c * W_k^c + b_k) \quad (2)$$

where \* represents the convolution operation of two-dimensional data,  $b_k$  indicates offset, g represents the activation function. Activation function is a kind of nonlinear transformation, which can increase the expression ability of network. Due to the characteristics of local connection and weight sharing, CNN greatly reduces the network parameters and the risk of over fitting during model training.

- Pooling Layer

Pool means down-sampling, in order to reduce the number of feature maps. Pooling operation is independent for each depth slice, and the scale is generally, convolution operation is performed relative to the convolution layer. The operations performed by pooling layer generally include the following: Max pooling, mean pooling, Gaussian pooling and trainable pooling. The maximum pooling is shown in the Figure 5, where each operation is performed on four numbers.

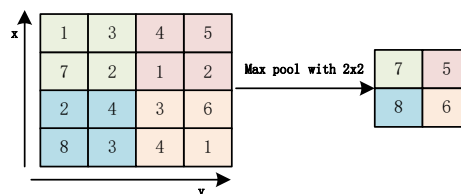


Figure 5. Maximum pooling

- Fully-connected layer

The full connection layer and convolution layer can be converted to each other. For any convolution layer, to turn it into a full connection layer, we only need to change the weight into a matrix that is

mostly 0 except for some specific blocks. On the contrary, any fully connected layer can also become a convolution layer.

### 2.3. Training of CNN

Convolutional network is a kind of mapping from input to output in essence. It can learn a large number of mapping relations between input and output without any precise mathematical expression between input and output. As long as convolutional network is trained with known patterns, the network has the mapping ability between input and output pairs. Convolution network performs supervised training, so its sample set is composed of vector pairs such as (input vector, ideal output vector). All these vector pairs should come from the actual "running" results of the system to be simulated by the network. They can be collected from the actual running system.

Convolutional neural network has unique advantages in image processing because of its special structure of local weight sharing. Its layout is closer to the actual biological neural network. Weight sharing reduces the complexity of the network, especially the multi-dimensional input vector image can be directly input into the network, which avoids the complexity of data reconstruction in the process of feature extraction and classification. In addition, it avoids the explicit feature sampling, but implicitly learns from the training data, which makes the convolutional neural network significantly different from other neural network-based classifiers. The feature extraction function is integrated into the multi-layer perceptron by restructuring the structure and reducing the weight. In the negative sample set, the false alarm rate can be greatly reduced by adding the samples which are easy to cause false alarm to the ship detection, such as islands, reefs, strong scattering clutter, etc., and adding the ship target samples which are interfered by noise, sea clutter, coast and other information in the easily confused samples.

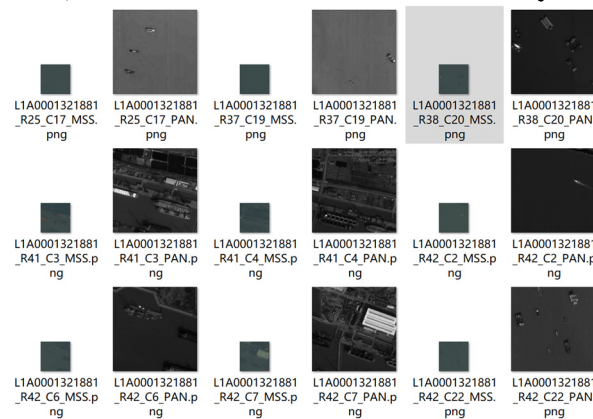


Figure 6. Positive sample(17523 in total)

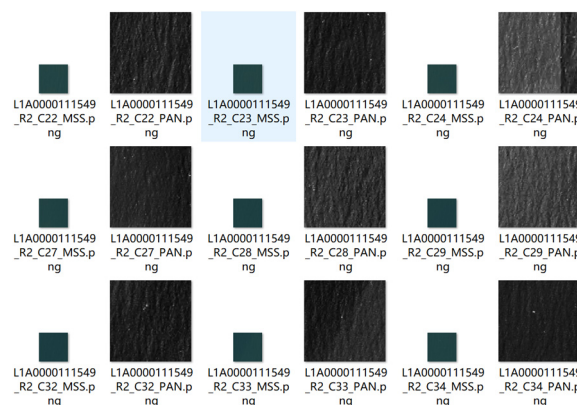


Figure 7. High brightness sea clutter samples(91523 in total)

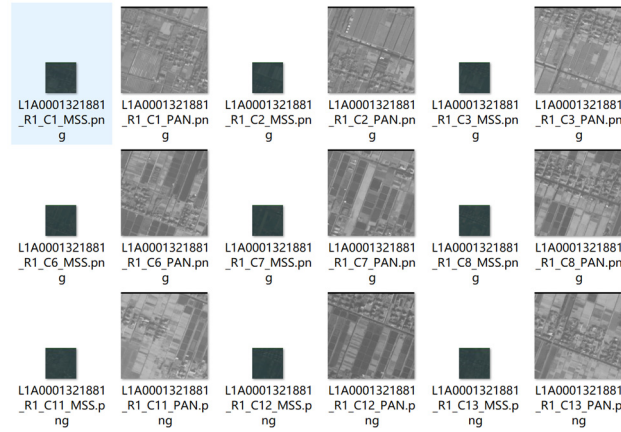


Figure 8. Confusing sample (14728 in total)

By extracting the target features from the three kinds of samples, the network can distinguish the detected targets, thus reducing the false alarm and improving the detection accuracy.

Before training, all weights should be initialized with different small random numbers "Small random number" is used to ensure that the network will not enter the saturation state due to too large weights, which will lead to training failure "Different" is used to ensure that the network can learn normally. In fact, if the weight matrix is initialized with the same number, the network will not be able to learn. The training algorithm is similar to the traditional BP algorithm. It mainly includes four steps, which are divided into two stages, also referring to the back propagation algorithm (BP algorithm).

- Dissemination of positive

The input samples are processed from the input layer through the hidden cells layer by layer, and then passed through all the hidden layers to the output layer; In the process of layer by layer processing, the state of each layer of neurons only affects the state of the next layer of neurons. In the output layer, the current output is compared with the expected output. If the current output is not equal to the expected output, the back-propagation process is entered. Set  $m$  as the number of layers of neural network,  $X$  as input sample; The sum of the inputs of the neurons  $i$  in the  $k$  layer is expressed as  $U_i^k$ ,  $X_i^k$  as output. The weight from the neuron  $j$  of the layer  $k-1$  to the neuron  $i$  of the layer  $k$  is  $W_{ij}$ . The activation function of each neuron is  $f$ , Then the relationship of each variable can be expressed by the following mathematical formula:

$$X_i^k = f(U_i^k) \quad (3)$$

where,

$$U_i^k = \sum_j W_{ij} X_j^{k-1} \quad (4)$$

- Back-propagation

In back propagation, the error signal is transmitted back according to the original forward propagation path, and the weight coefficient of each neuron in each hidden layer is modified to minimize the error signal.

### 3. Experiment on orbit

The above CNN model is verified in orbit by using the navigation anti-interference enhancement hardware platform of a scientific experimental satellite, as shown in the Figure 9. Because the payload has no image processing function in the early stage, the image processing algorithm and ground training model are injected into the payload memory through on orbit reconstruction technology, and the remote sensing image captured by the satellite compound eye camera is received through CAN bus for on orbit real-time processing. The ship image slice data and target traceability information are transmitted back.

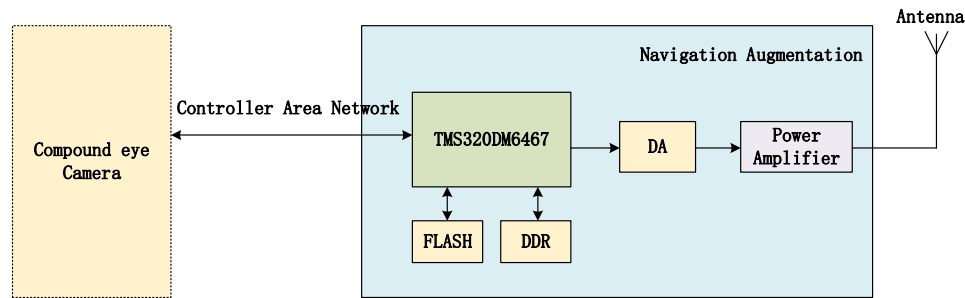


Figure 9. On orbit image processing data stream

The traditional deep learning remote sensing image target detection algorithm has the characteristics of computation intensive and storage intensive (model weight file size 36.9mb; 7 tflops), while the computing performance and storage space of embedded processing units on small commercial satellites are limited, which puts forward strict requirements for model design training and hardware platform transplantation.

Using more than 400000 sample data sets, the deep learning optimization of the deep neural network model is carried out, and 37 layers of network processing levels are constructed. By keeping the performance of the deep convolution neural network for ship target recognition unchanged, the network compression method based on transfer learning, sparse training and weight quantization is realized, and the initial weights of the network are optimized. Compared with the mainstream Resnet, Inception, VGG and other deep learning models, the optimized algorithm model optimizes the calculation amount of the model to 25MFlops, and reduces the weight file data amount from tens of MB to 0.5Mb, so that it can be transplanted into the load hardware platform smoothly, and several key parameters such as error convergence and generalization performance can be reconstructed later.

After on orbit image intelligent processing and image slice return, nine ship target slices in the original image are obtained, and the slice data is restored and reversed to obtain slices as shown in the Figure 10. From the figure, it can be seen that ship targets can be detected in each slice.

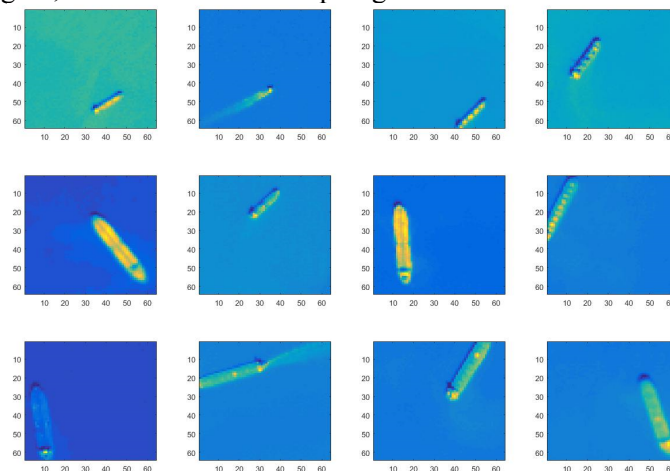


Figure 10. Targets in each slice

#### 4. Conclusions

In this paper, based upon the advantages in special structure of local weight sharing, actual biological neural network like layout, a CNN based on-orbit real-time target detection method for remote sensing images was proposed. By employing transfer learning, sparse training and weight quantification, the calculation amount of the model was optimized to 25MFLOPS, and the processing efficiency of the algorithm was improved. The proposed method made possible for on-orbit target detection in small commercial remote sensing satellites.

## References

- [1] Xuejun Zhang, Yanchao Fan, He Bao. Applications and development of ultra large aperture space optical remote sensors[J]. Optics and Precision Engineering, 2016, 24(11): 2613-2626.
- [2] Dun Gong. Review on mapping space remote sensor optical system [J]. Chinese Optics, 2015(5): 714-724.
- [3] Kuchеров B, Pribyl O, Artyushenko V. Increasing Efficiency of Getting Results of Satellite Remote Sensing for Smart Cities[C], Prague, Czech Republic, 2017: 1-6.
- [4] Qingjun Zhang. Optical Remote Sensing Technology Progress Promoted by ZY-1 Series Satellites [J]. Spacecraft Recovery & Remote Sensing, 2018, 39(4): 45-54.
- [5] Hongtao Lu, Qinchuan Zhang. Applications of Deep Convolutional Neural Network in Computer Vision [J]. Journal of Data Acquisition & Processing, 2016, 31(1): 1-17.
- [6] YUAN W, Lining X. Remote Sensing Satellite Networking Technology and Remote Sensing System: A Survey[C], Qingdao, China, 2015. 3: 1251-1256.
- [7] Zhongguo Wang, Dabao Wang. Research on Transmission Effectiveness of Remote Sensing Data from LEO Satellite-to-Earth at Ka-band [J]. Spacecraft Engineering, 2013, 22(1): 72-77.
- [8] Jun Gao, Kai Wang, Xiaoyu Tian, et. A BP-NN based Cloud Detection Method for FY-4 Remote Sensing Images [J]. Journal of Infrared and Millimeter Waves, 2018, 37(4): 477-485.
- [9] Xiaobin Huang, Jun Shen, Guoxin Li, et. Remote sensing image classification based on BP neural network and optimized with genetic algorithm [J]. Modern Electronic Technology, 2020, 43(12): 47-49, 54.
- [10] Huiyuan Chen, Zeyu Liu, Weiwei Guo, et al. Fast detection of ship targets for large-scale remote sensing image based on a cascade convolutional neural network [J]. Journal of Radars, 2019, 8(3): 413-424.
- [11] Donghang Yu, Baoming Zhang, Haitao Guo, et al. Joint salient feature and convolutional neural network for ship detection in remote sensing images [J]. Journal of Image and Graphics, 2018, 23(12): 1947-1958.
- [12] Gang Wang, Jinyong Chen, Gao Feng, et al. Remote sensing image target detection based on deep neural network and feature extraction [J]. Radio Engineering, 2018, 48(9): 760-766.
- [13] Chunhui Zhao, Yao Zhou. Ship target detection and recognition based on improved faster RCNN algorithm [J]. Journal of Shenyang University (Natural Science), 2018, 30(5): 366-371, 380.
- [14] Tao Chu, Chuanzhao Han. Algorithm of Ship Detection for Optical Remote Sensing Satellite on Orbit [J]. Spacecraft Engineering, 2021, 30(2): 31-39.
- [15] Fang Xu, Jinghong Liu, Hui Sun, et al. Research progress on vessel detection using optical remote sensing image [J]. Optics and Precision Engineering, 2021, 29(4): 916-931.
- [16] Yuwen Chen, Jianlin Xiong, Bin Liu. Survey of ship detection and recognition technology in optical remote sensing images [J]. Journal of Telemetry, Tracking and Command, 2021, 42(3): 64-73.