IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 23, 2018, accepted November 4, 2018, date of publication November 15, 2018,
date of current version December 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2881479

Ship Detection in Optical Satellite Images Using Haar-like Features and Periphery-Cropped Neural Networks

YE YU[1,2], HUA AI[1], XIAOJUN HE[1,3], SHUHAI YU[3], XING ZHONG[1,4], AND MU LU[3]

[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2]University of Chinese Academy of Sciences, Beijing 100039, China
[3]Chang Guang Satellite Technology Co., Ltd., Changchun 130102, China
[4]Key Laboratory of Satellite Remote Sensing Application Technology of Jilin Province, Chang Guang Satellite Technology Co., Ltd., Changchun 130039, China

Corresponding author: Xiaojun He (hexiaojun6@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0502600, in part by the Outstanding Youth Talent Foundation of Jilin province under Grant 20170520166JH, and in part by the Key Science and Technology Development in Jilin Province under Grant 20180201107GX.

ABSTRACT The ship detection field faces many challenges due to the large-scale and high complexity of optical remote sensing images. Therefore, an innovative ship detection method that is simple, accurate, and stable is proposed in this paper. The algorithm consists of the following two steps: 1) the AdaBoost classifier, combined with Haar-like features, is used to rapidly extract candidate area slices, and 2) according to the characteristics of ships, a periphery-cropped network is designed for ship verification. Furthermore, we analyze the characteristics of ocean images to improve the contrast between the target and the background. Thus, an RGB spectrum-stretching method is proposed. Finally, we evaluate our method using spaceborne optical images from the Jilin-1 satellite, Google satellites, and the public dataset NWPU VHR-10. Our experimental results indicate that the proposed algorithm achieves a high detection rate.

INDEX TERMS Object detection, feature extraction, CNN, AdaBoost classifier.

I. INTRODUCTION

Ship detection has attracted increasing amounts of research attention because of its growing significance in the remote sensing field. In early research, many ship detection methods based on synthetic aperture radar (SAR) images were developed [1]–[3]. In recent years, with the rapid development of optical spaceborne imaging technology, some researchers have begun to focus on detecting ships with optical images due to their high resolution and more detailed spatial content compared to those of SAR images [4].

In early research, the proposed detection methods used the physical parameters of ship targets by manually setting parameter constraints and then selecting the target using testing methods, such as the methods in [5] and [6]. These studies identified the ship wake, but this approach is poorly resistant to interference. Thus, with this type of method, it is difficult to meet the demands of actual use. For decades, researchers have chosen a coarse-to-fine detection strategy that involved 1) a ship candidate extraction stage and 2) a ship verification stage. In the ship candidate extraction stage, the detection algorithm searches for all candidate regions that possibly contain the ship targets [4]. In this stage, frequently used methods include image segmentation methods [7]–[10], saliency detection methods [11], wavelet transformation methods [12], and anomaly detection methods [13], all of which are unsupervised methods. In the ship verification stage, each individual ship candidate region is further verified to determine whether it contains a real ship target. Previous studies of this stage mainly applied machine learning methods by converting the related process into feature extraction and binary classification operations (targets vs. backgrounds). The commonly used methods in this stage include support vector machines (SVMs), extreme learning machines [12], AdaBoost [13], sparse representation [14] and neural networks [10]. In addition to these supervised learning methods, some unsupervised methods are also used in this stage, such as the contour analysis method [7] and the shape analysis method [15]. However, the background of remote sensing images is complex, and ship candidate extraction based on unsupervised methods often lacks robustness. In the ship verification stage, precise classification is required, but hand-selected

features cannot be simply adapted to variable satellite imaging conditions.

In recent years, deep learning has made significant progress in the fields of object detection and recognition [16]–[18]. The convolutional neural network structures of Alexnet [19], VGG [20] and Resnet [21] have an advantage over traditional artificial feature engineering structures for feature extraction. For example, [22] used a deep learning target detection method for astrocyte detection that promoted the intelligent development of the medical field, and [23] used deep learning for transportation analysis. Therefore, it is of great significance to study the application of deep learning algorithms in the field of remote sensing.

Detection algorithms based on convolutional neural networks (CNNs) can be mainly divided into one-stage methods and two-stage methods. The earliest two-stage object detection algorithm was R-CNN [24], which selected candidate regions through a selective search algorithm and then used a CNN for classification. Later, a number of improved two-stage algorithms, such as Faster-RCNN [25], R-FCN [26] MSCNN [27] and RON [28], were proposed. The most common one-stage algorithms include YOLOv2 [29], SSD [30], FPN [31] and Retinanet [32], which treat object detection as a regression problem to simultaneously obtain the target position and perform classification. However, these networks are mainly designed for the goals of daily life, and using them in the remote sensing field can cause two common problems. 1) If the remote sensing image is extremely complex and the target scale varies, applying a general neural network algorithm directly to remote sensing images can be ineffective. 2) The size of conventional images rarely exceeds 1000×1000 pixels, but remote sensing images often range in size from thousands to tens of thousands of pixels. Due to hardware resource limitations, large remote sensing images must be split into small slices. After the small slices are detected, they must be recombined. During image splitting and merging, caution must be used so that ships located in the edge region are not split into two parts when splitting the original image, and redundant splitting is required. In addition, to prevent multiple detections of the same target due to redundant cropping, a screening processing is implemented after image merging. Moreover, because the size of a ship is typically large, the area of redundant cutting is also large, which can lead to high consumption of time and resources.

Therefore, in this paper, we combine the traditional machine learning algorithm and a convolutional neural network and apply a coarse-to-fine detection strategy. In the ship candidate extraction stage, unlike in [7]–[13], we abandon the unsupervised method and use the available data and statistical information. Haar-like features are used in an AdaBoost algorithm to quickly search for candidate targets. Next, in the ship verification stage, according to the characteristics of a ship, a neural network called the periphery-cropped network (PCNet) is designed to verify whether candidate regions contain real ship targets. This approach will not limit the size of the detected image and fully utilizes the ability of the CNN in object detection and recognition.

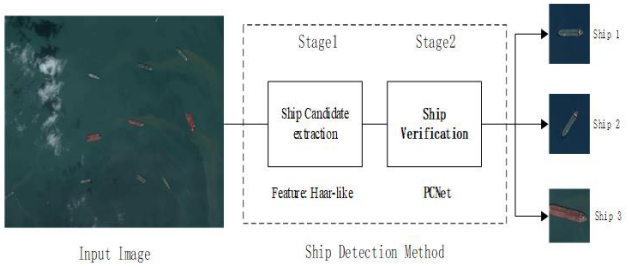


FIGURE 1. Workflow of the proposed method.

PCNet integrates prior ship knowledge and background information by training based on large numbers of images. Instead of using hand-selected features, the features in our method are adaptively learned from spaceborne optical image data. Then, we propose a new residual module used by PCNet for convolution feature extraction. Next, the bow and stern of a ship are used to design a local feature extraction method. We combine the global and local features of ships to improve the ability of the network to identify ships. We also propose an image enhancement preprocessing method, called the RGB spectrum-stretching method, to address the characteristics of a remotely sensed sea area. This approach can improve the accuracy of the ship candidate extraction stage.

Fig. 1. shows the workflow of the proposed method, and this paper is organized as follows. In Section II, we briefly introduce the ship candidate extraction stage and the RGB spectrum-stretching method, which is used to enhance images. Section III presents a detailed introduction to the feature learning algorithms in PCNet. We present the experimental details and analyze the results in Section IV, and Section V concludes this letter.

II. SHIP CANDIDATE EXTRACTION

In this section, we introduce the RGB spectrum-stretching method and discuss why we use Haar-like features as characteristic descriptors. Finally, we provide the details of the algorithm.

A. RGB SPECTRUM-STRETCHING METHOD FOR IMAGE ENHANCEMENT

The goal of object detection algorithms is to find the difference between the target and the background. In the ship candidate extraction stage, we use Haar-like features based on the statistics of the pixel values to separate the ship image from the background image. Therefore, the larger the difference between the pixel value of the surface and that of the ship, the higher the contrast and accuracy of ship detection. When calculating the eigenvalues of the Haar-like features, the grayscale version of the original image is first processed. The conventional grayscale processing method involves calculating the average of the three red, green and blue (RGB) bands. As shown in Fig. 2, after grayscale processing, the

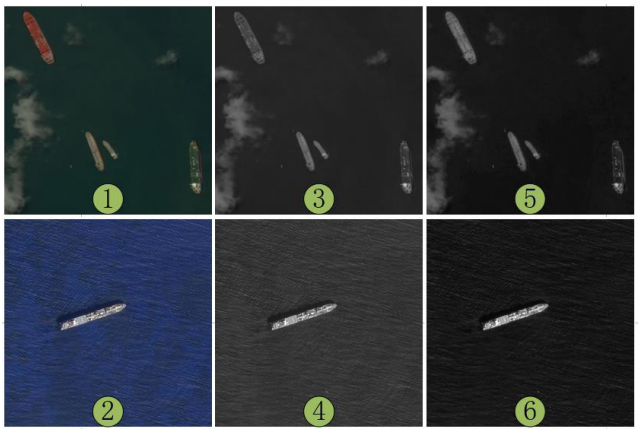


FIGURE 2. Comparison of two image preprocessing methods: (1)–(2) original images; (3)–(4) the results of grayscale transformation; (5)–(6) the results of the proposed RGB spectrum-stretching algorithm.

contrast between the ship and the sea surface is not obvious, which may lead to misdetection. We know that the sea surface is blue in most cases, and it is green when there is vegetation present. Because most commercial and civil vessels do not deliberately disguise the hull, the blue or green spectral information reflected by the ship target is generally weaker than that of the marine area. According to this property, an image contrast enhancement method is designed. Our goal is to improve the contrast between the foreground and the background by suppressing the ocean area and obtaining the ship target, as shown in formula (1), where $R(x, y)$, $B(x, y)$ and $G(x, y)$ represent the pixel information contained in the red, blue and green channels of the remote sensing images, respectively. To prevent the special case where the denominator is 0, eps is introduced. Specifically, eps is the floating point relative precision. $P_{val}(x, y)$ is a pixel value after spectrum processing. α is the gain factor, and in this paper, through the actual test, we got the best result when α is equal to 1.2. Here, the range of image pixels is (0–255), and when the right part of the formula exceeds 255, $P_{val}(x, y)$ is equal to 255.

$$P_{val}(x,y) = \begin{cases} \dfrac{\alpha \times R(x,y) \times B(x,y)}{G(x,y) + eps} & if\ B(x,y) \le G(x,y) \\ \dfrac{\alpha \times R(x,y) \times G(x,y)}{B(x,y) + eps} & if\ B(x,y) > G(x,y) \end{cases} \tag{1}$$

Fig. 2. shows that the contrast between a ship and the sea surface has been significantly improved. Figs. 2 (1)–(2) are the original images, (3)–(4) are images after conventional grayscale processing, and (5)–(6) are images after the proposed RGB spectrum-stretching method is applied. The pixel value of the red hull obviously increases, and that of the ocean area is suppressed. Although the green hull causes the pixel value to decrease based on the formula, the suppression effect is small compared with that of the ocean area, and the contrast is indirectly improved.

TABLE 1. The results of two image preprocessing methods.

Method	N_{tp}	N_{fp}	Precision (%)	Recall (%)
Grayscale	255	31	89.2	82.8
RGB method	270	27	91.0	87.7

Table 1 gives a comparison of the results of the original grayscale processing and the RGB spectrum-stretching method (the meaning of N_{tp}, N_{fp}, etc. are in section 4.3). Notably, the leak detection rate decreases after the enhancement process.

B. HAAR-LIKE FEATURES FOR THE APPROXIMATE POSITIONING OF SHIPS

The AdaBoost algorithm is an integrated learning method that can integrate multiple weak classifiers into a strong classifier. Therefore, we use the AdaBoost algorithm in the candidate extraction stage. Here, the method is briefly introduced.

Given a set of training samples, the initial weight of the samples is defined as $D_t(i)$, where $D_t(i)$ represents the weight of the i^{th} sample and N is the number of samples. A weak classifier consists of eigenvalue $f(x)$, threshold f_θ and direction vector p_j.

$$\mathrm{h_j}(x) = \begin{cases} 1, & p_j f_j(x) < p_j f_\theta \\ 0, & otherwise \end{cases} \tag{2}$$

Then, the weight values of the samples are updated as follows:

$$D_{t+1}(i) = \frac{D_t(x)}{Z_t} \times \begin{cases} \exp(-a_t), & r \\ \exp(a_t), & w \end{cases} \tag{3}$$

$$a_t = \log_{10}[(1 - \varepsilon_t)/\varepsilon_t] \tag{4}$$

$$\varepsilon_t = P_x(h_t(x) \neq y_i) \tag{5}$$

where "r" represents a correct classification in the previous iteration and "w" represents a classification error in the previous iteration. The final classifier is defined as follows.

$$H(x) = \begin{cases} 1, & \sum_{t=1}^{T} a_t h(t) \ge \frac{1}{2} \sum_{t=1}^{T} a_t \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

The above method can effectively improve the classifier performance, but with a single classifier, it is still difficult to accurately identify the target. Therefore, multiple base classifiers must be connected in an intelligent way based on their complexity. In image detection, many subwindows with invalid background areas are eliminated by the initial base classifiers, and only targets and similar interference areas are used by the next base classifier. This approach can quickly and accurately find the target location.

In terms of feature selection, Haar-like features reflect changes in the grayscale of an image, and the eigenvalues correspond to the difference between the white rectangular pixels and the black rectangular pixels. In a remote sensing image, the gray value of a ship is higher than that of the

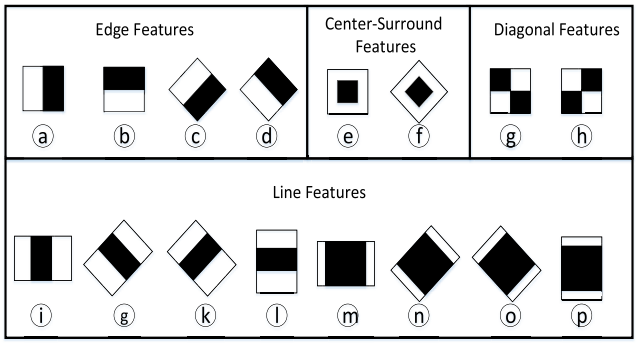


FIGURE 3. The Haar-like features used in this paper.

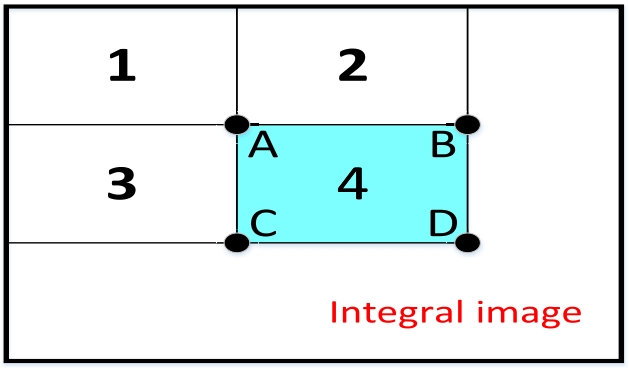


FIGURE 4. An intuitive explanation of an integral image.

surrounding surface. Therefore, Haar-like features are suitable for ship detection based on remote sensing images. Fig. 3 shows the Haar-like feature templates used in this paper. Additionally, to accelerate the calculations, an integral graph is designed for the Haar-like features. The value of I (x, y) in the integral image is the integral result of the summation of all pixels above and to the left of I (x, y), which is calculated in the original image. In this manner, we can quickly calculate the sum of all pixels in area 4 using the values of A, B, C, and D, as shown in Fig. 4:

$$P_{sum\_4} = P_{val\_D} - P_{val\_B} - P_{val\_C} + P_{val\_A} \tag{7}$$

where P_{sum_4} is the summation of all pixels in area 4 and P_{val_A} is the value of point A in an integral image.

Furthermore, in the classifier training phase, there is a tradeoff between a high false alarm rate and the guaranteed extraction of all ships by the classifier. The specific parameter values used in this experiment are given in Section IV. After image enhancement by the RGB spectrum-stretching method, an AdaBoost classifier was trained to determine the approximate positions of ships. Fig. 5 shows a remote sensing image that was obtained by the Jilin-1 satellite.

III. SHIP VERIFICATION

After a set of ship candidate windows is obtained, each window is analyzed using discriminative features to verify whether it truly contains a ship instance.

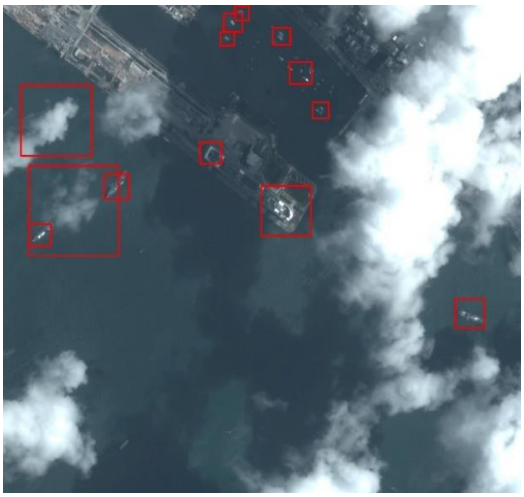


FIGURE 5. Result of ship candidate extraction, including real ships and background regions.

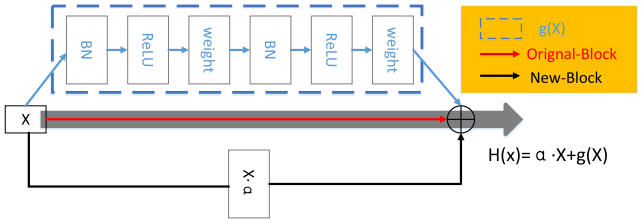


FIGURE 6. The structures of the two residual models.

CNNs have the ability to automatically extract features. Compared with traditional artificial design features, a CNN can obtain hidden abstract features of the target and provides high robustness. However, remote sensing image processing is very difficult because of the highly complex backgrounds. In addition, unpredictable imaging conditions add to the difficulty of effective information extraction. Therefore, according to the characteristics of remote sensing images, we designed PCNet for ship verification. The PCNet has two main contributions, which are discussed below in part A and part B.

A. PART A

Many excellent classification networks exist, such as VGG and Inception. However, these networks have a common problem. When the network depth increases, the performance decreases because the gradient disappears. In 2015, [21] proposed a residual network called Resnet to solve this problem by introducing a shortcut. Resnet can be illustrated by the dotted line and red arrow in Fig. 6, where X represents an n-dimensional input and $g(x)$ represents the conventional convolution process. To prevent vanishing gradient issues, we choose Relu as the activation function. The batch normalization method used in the model is a feature-scaling method that has many advantages, such as preventing the gradient from disappearing, improving the convergence speed of

TABLE 2. The detection results of the two residual methods.

Method	N_{tp}	N_{fp}	Precision (%)	Recall (%)	F (%)
Original Resnet	268	34	88.7	87.0	87.8
Our Resnet	270	27	91.0	87.7	89.3

the model, and partially preventing overfitting. $H(x)$ is the final output. The Resnet formula is as follows.

$$H(x) = X + g(x) \tag{8}$$

In this paper, we made some changes to the shortcut function in Resnet [21] to form a new residual module, which is shown in Fig. 6. This approach follows the black arrow instead of the red arrow, and the formula is as follows:

$$H(x) = X \cdot \partial + g(x) \tag{9}$$

where ∂ is an n-dimensional vector $\{\partial_1, \partial_2, \ldots \partial_n\}$, n represents the dimension of the input X, and $X \cdot \partial$ denotes that each component of the vector ∂ operates on each dimension of X. In this paper, we modified the shortcut connection method in Resnet [21]. In this modification, the input X in Resnet [21] is directly added to $g(x)$ based on the shortcut connection. Because each dimension of the input X produces a different effect in the shortcut connection, each channel of X is given a trainable coefficient according to the vector $\{\partial_1, \partial_2, \ldots \partial_n\}$, and unlike the $1 \times 1 \times$ n convolution, each coefficient in $\{\partial_1, \partial_2, \ldots \partial_n\}$ does not affect other coefficients. The network can determine the role of each channel of input X in the shortcut connection by statistical learning based on relevant samples. In addition, in the parameter initialization phase, the initial value of ∂ is $\{1, 1, 1 \ldots, 1\}$. Thus, the initial residual network is the same as that of Resnet [21]. Therefore, although the change is small, the new residual structure of this approach can be regarded as an extended structure of Resnet [21]. In addition, in the training process, referring to a strategy that is similar to warm-up, we used the regular resnet structure in the first 3 to 5 epochs, after which α is used. Table 2 shows a comparison of the two residual methods.

B. PART B

Remote sensing images can have many characteristics, such as a complex background and a low image resolution. It is difficult for conventional classification networks to accurately identify objects in these images due to the presence of complicated interference. Currently, improving the effectiveness of network feature description requires the use of multiscale structures by combining the characteristics of different features. The conventional multiscale structures connect feature maps from different layers to fully utilize fine-scale features. In this paper, we propose a novel multiscale structure for ship verification. First, we extract the local ship information by cropping the original image. Then, we simultaneously input the original image and the local image to the network

TABLE 3. The detail parameters of the proposed PCNet model.

Layer Name	Layer	Output Size
Conv_1	3×3×32	112×112
Pool_1	Averagepool Stride=2	56×56
Res-block-1	{1×1,3×3,1×1}×3 64	28×28
Res-block-2	{1×1,3×3,1×1}×4 128	14×14
Res-block-3	{1×1,3×3,1×1}×6 256	7×7
Res-block-4	{1×1,3×3,1×1}×3 512	7×7
GlobalPool	GlobalAveragePool	-
Concatenate	Add	-
Dense	Fc	512
Softmax		

TABLE 4. The results of three feature fusion methods.

Crop Method	Precision (%)	Recall (%)	F (%)
Center Crop	88.8	84.7	86.7
Periphery Crop	91.0	87.7	89.3
No Local Features	87.2	81.8	84.4

to improve the classification accuracy. The morphological characteristics of a ship are distinct from other types of maritime interference. The bow/stern of a ship is always at the periphery of the image. Therefore, we designed two types of cropping methods, which are shown in Fig. 7. We can consider the bow and stern of a ship to be local features, or we can consider the middle area of a ship to be a local feature. A comparison of the accuracies of the two designed methods is shown in Table 4. Notably, the local feature fusion method is effective, and the bow/stern fusion method is more efficient for ship identification. Finally, the structure of PCNet that we designed is shown in Fig. 8, and the network parameters are presented in Table 3. Feature extraction uses parameter settings that are similar to those of Resnet, and the main change is related to the shortcut function.

The advantages of the proposed version of PCNet in the ship detection field are as follows: 1. the local characteristics of the bow/stern improve the feature identification ability of the network; 2. the bow/stern features are unique to ships and can effectively be distinguished from other interference types, such as clouds, waves, and reefs; and 3. since PCNet uses both the local features and the global features of the ship, ships with a small area of cloud occlusion can still be detected well.

IV. EXPERIMENTS

A. EXPERIMENTAL DATA

Supervised learning algorithms require a large number of samples, and the richness of the samples will affect the final detection results. Therefore, to allow the model to rapidly converge, a large number of remote sensing images are used

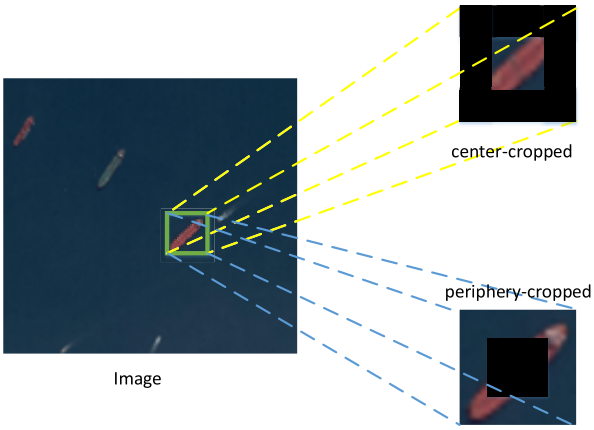


FIGURE 7. The structures of two local features.

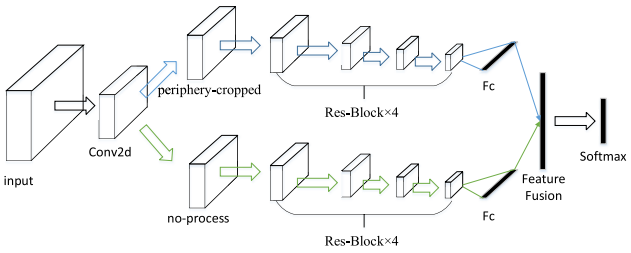


FIGURE 8. The structure of PCNet.

as training samples. In addition, to improve the robustness of the algorithm, we collect multisource remote sensing samples. Furthermore, we use public data to test and accurately analyze the performance of the algorithm.

The experimental data are from the commercial optical remote sensing satellite Jilin-1 and Google satellites. The Jilin-1 satellite has a resolution of 0.72 meters, and Google satellite data are from Google Maps with a resolution of 1 meter. After cropping, approximately 1500 ship samples were obtained for model training. In addition, to improve the versatility of the model, we use the public data set NWPU VHR-10 [33]–[35] for testing. NWPU VHR-10 contains 61 images with ships, with a total of 308 ships.

B. DATA AUGMENTATION

A statistical learning algorithm depends not only on the design of the network structure and parameters but also on the quality of the dataset, which can affect the final ship detection accuracy. A remote sensing image is different from a conventional image, and such images are difficult to obtain in large quantities. Therefore, an artificial data augmentation method is designed as follows.

1)Translation Augmentation

The translation is the position of the moving sample in the original image. Although the CNN provides translation invariance in feature extraction, translation augmentation can avoid overfitting when the network learns the positioning parameters.

2) Rotation Augmentation

In remote sensing images, even if the same ship target has different locations at different times, because the amount of data collected is limited, it is impossible to cover all possible target angles, and the neural network algorithm only provides partial rotation invariance. Therefore, the samples are subjected to rotation augmentation processing to improve the generalizability of the model.

3) Color Transform Augmentation

Because satellite imaging is affected by the weather, ships may exhibit different colors at different times. Thus, color transformation can improve the generalizability of the model.

During the training process, data enhancement is performed for each batch of training samples, and the data are then transmitted to the network for training.

C. ALGORITHM EVALUATION CRITERIA

To objectively evaluate the performance of the target detection algorithm, we use precision–recall (P–R) curves, F values, and the Ap metric.

Precision reflects the false detection rate of an algorithm and is formulated as follows:

$$Precision = \frac{N_{tp}}{N_{tp} + N_{fp}} \tag{10}$$

where N_{tp} is the number of correctly detected ships and N_{fp} is the number of falsely detected ships.

Recall represents the effectiveness of detection and can be expressed as follows:

$$Recall = \frac{N_{tp}}{N_{tp} + N_{fn}} \tag{11}$$

where N_{fn} is the number of nondetected ships.

F is the harmonic mean, which directly reflects the performance of the algorithm.

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{12}$$

Furthermore, since the image used in this paper is the meter-level resolution, we only detect ships that are larger than 100 pixels.

D. EXPERIMENTS

In this paper, three sets of experiments were designed. 1. In the ship candidate extraction stage, the influence of Haar-like features and other features, such as HOG and LBP, on the final test results is investigated. 2. In the ship verification stage, the influence of local feature fusion on ship detection is verified, and the two feature fusion modes discussed above are compared. 3. Through comparison experiments, the proposed algorithm is compared with several typical ship detection methods to verify the feasibility of the algorithm.

1) THE FIRST EXPERIMENT

The AdaBoost classifier has the advantage of fast detection; therefore, we use it for initial approximate positioning. In this

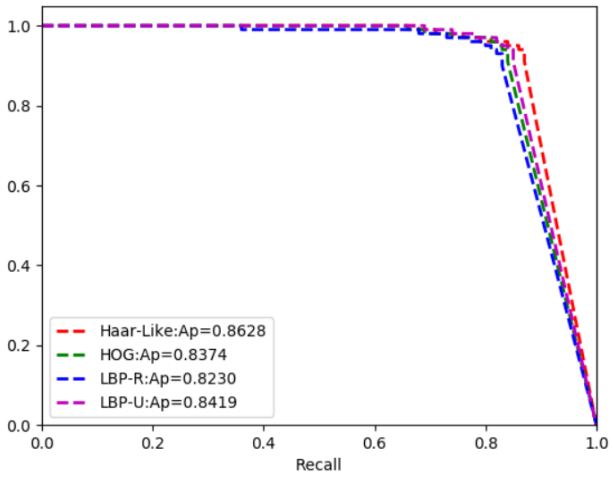


FIGURE 9. P–R curves of four artificial features.

TABLE 5. The parameters of the AdaBoost classifier.

Parameter	Value
μ_1	0.995
μ_2	0.5
σ	14

experiment, we compared the effects of four different types of features, including HOG, rotation-invariant LPB, uniform LBP and Haar-like features, on ship candidate extraction. The P–R curves are shown in Fig. 9. The results indicate that the Haar-like features are most suitable for ship detection.

The main AdaBoost classifier parameters used in the proposed method are listed in Table 5. The total recall rate of the classifier is equal to the product of the values. In Table 5, μ_1 is the minimum recall rate threshold of each classifier, μ_2 represents the virtual alarm rate threshold of each classifier, and σ represents the total number of weak classifiers. To force the final recall of the detector to be close to 1, μ_1 must be sufficiently large so that the positive sample pass rate of each classifier is very high. Similarly, μ_2 must be sufficiently low. The selection of μ_2 greatly influences the precision. If the value of μ_2 is too large, there will be too many false regions in the ship candidate extraction stage, which will not only decrease the error detection rate but also seriously affect the speed of the algorithm.

2) THE SECOND EXPERIMENT

The main CNN classifier hyperparameters are listed in Table 6.

Traversing one side of the entire sample is called an epoch. The learning rate reflects the parameter update speed during backpropagation. If the learning rate is too large, the optimal parameter cannot be calculated. If the learning rate is too small, the model will converge slowly. The batch

TABLE 6. The hyperparameters of the proposed PCNet model.

Name	Value
Epoch	20
Learning rate	0.001
Batch size	32
Loss	Cross-entropy
Optimization	Adam

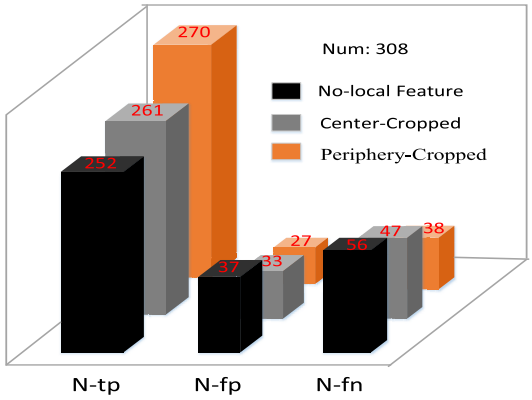


FIGURE 10. Statistical results of the three structures.

size is defined as the number of samples that are read into memory at a given time. If the batch size is too small, the model will obtain local optimal parameters, but the batch size is limited by the memory size. Cross-entropy is the loss function used in our method. The Adam algorithm is utilized as the optimization method. This algorithm provides faster convergence for the proposed model than does the stochastic gradient descent (SGD) algorithm.

Fig. 10 illustrates the results of the periphery-cropped, center-cropped, and noncropped network structures in the NWPU VHR-10 test set. It can be seen that the fusion of local features and global features can improve the network classification performance, and periphery-cropped local feature fusion is more suitable for ship detection.

3) THE THIRD EXPERIMENT

In this experiment, we compare several mainstream object detection methods, including [11], YOLOv2 [29], Faster-RCNN [25], FPN [31], MSCNN [27], and SSD [30], which correspond to methods [1]–[6] in Table 7. Fig. 11 presents the P–R curves of these methods. Method 1 involves unsupervised learning based on the test set; therefore, it is not shown in Fig. 11. In Table 7, the harmonic mean 'F' of the proposed method is higher than that of all other algorithms except Methods 3 and 6.

Although Methods 3 and 6 yielded better test results than the proposed algorithm, as the size of the remote sensing image increases, the resource consumption of these methods becomes extremely high, as noted in Section I. Therefore, they are not conducive to the rapid processing of remote

TABLE 7. The detection results of seven methods.

Method	Precision (%)	Recall (%)	F (%)
PCNet	91.0	87.7	89.3
Method [1]	84.4	47.6	60.8
Method [2]	82.6	76.9	79.7
Method [3]	91.3	88.9	90.1
Method [4]	88.0	86.0	87.0
Method [5]	91.7	89.9	90.8
Method [6]	87.2	84.1	85.6

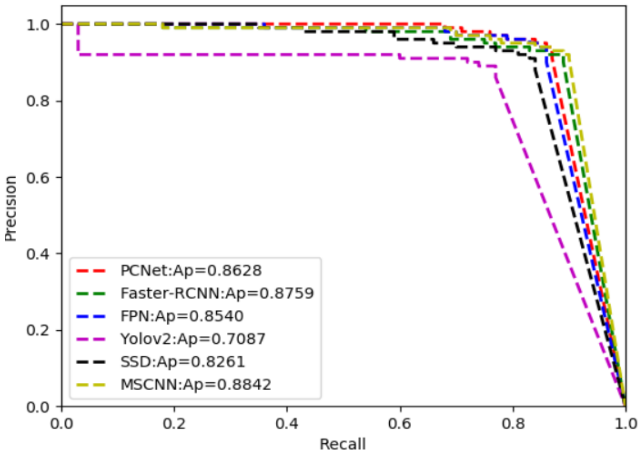


FIGURE 11. P–R curves of six methods.

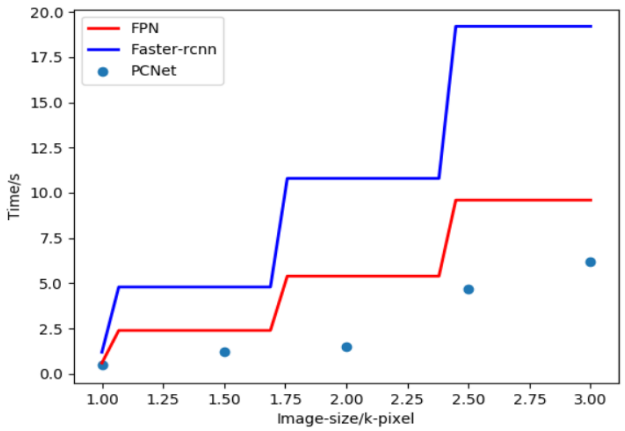


FIGURE 12. Time consumption results of three methods.

sensing images. To verify the above analysis, we selected two typical models (i.e., FPN (Method 4) and Faster-RCNN (Method 3)), including a one-stage algorithm and a two-stage algorithm, for time consumption comparisons. An NVIDIA Tx2SOC in maximum power mode was used for this analysis. When the FPN and Faster-RCNN algorithms are used, the pixel size of the detection image should not exceed 1000×1000 to ensure high accuracy. For large-scale remote sensing images, redundant image splitting is required, and in the case of image merging, some postprocessing algorithms

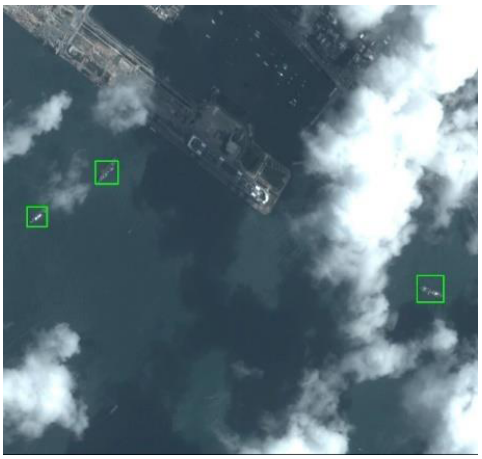


FIGURE 13. The result obtained with the proposed algorithm (the image is from the Jilin-1 satellite).

are needed to prevent multiple detections of the same target. Since the image resolution is 1 meter and commercial and civilian ships are generally less than 300 meters, the redundancy is 300 pixels. This experiment roughly calculates the corresponding time consumption statistics. We do not consider the image splitting, merging and partial postprocessing steps. The time consumption calculation for the FPN and Faster-RCNN methods is as follows:

$$F(x) = T_0 * g(\frac{x - 0.3}{1 - 0.3})^2 \tag{13}$$

where x represents the size of the image and the unit is per thousand pixels (here, the image to be tested is cropped to a square for convenience). $g(x)$ indicates that the input x is rounded to the right; for example, $g(1.2) = 2$ and $g(1.8) = 2$. For a single 1K image, the FPN and Faster-RCNN methods take 0.6 s and 1.2 s, respectively. Therefore, the T_0 values of these methods are 0.6 and 1.2. The time consumption of PCNet mainly depends on the AdaBoost classifier used for suspected region extraction. Therefore, when training the AdaBoost classifier, it is necessary to control the number of suspected regions through hyperparameters. Because there is no specific formula for the detection time of PCNet, we test remote sensing images of five sizes. Fig. 12 gives a comparison of the results of the three algorithms.

Fig. 13 shows the final image result obtained using the proposed algorithm, and the results of ship candidate extraction are shown in Fig. 5.

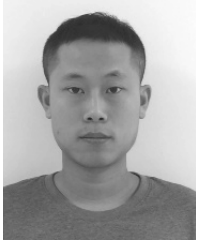
V. CONCLUSION

In this paper, we proposed a novel ship detection method. In the ship candidate extraction stage, Haar-like features are used to determine the initial positions of ships. Based on the characteristics of remote sensing images, we proposed an image enhancement method based on the RGB spectrum-stretching method. In the ship verification stage, we designed PCNet, which integrated the local and global features of ships to improve the accuracy of the proposed method.

Although there is room to improve the proposed algorithm, we consider the method to provide a novel and promising ship detection framework that represents a new potential direction for research in the ship detection field. Our future work will focus on the following two tasks. The first task is to improve the accuracy of the ship candidate extraction process. PCNet has a high accuracy for ship verification, but there is some room for improvement in the accuracy of the ship detection method. The second task is to reduce the cost of computational resources.

REFERENCES

[1] C. P. Schwegmann, W. Kleynhans, and B. P. Salmon, "Synthetic aperture radar ship detection using Haar-like features," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 154–158, Feb. 2017.
[2] G. Gao, G. Shi, G. Li, and J. Cheng, "Performance comparison between reflection symmetry metric and product of multilook amplitudes for ship detection in dual-polarization SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 5026–5038, Nov. 2017.
[3] G. Gao and G. Shi, "Ship detection in dual-channel ATI-SAR based on the notch filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4795–4810, Aug. 2017.
[4] Z. Zou and Z. Shi, "Ship detection in spaceborne optical image with SVD networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5832–5845, Oct. 2016.
[5] F. Y. M. Lure and Y.-C. Rau, "Detection of ship tracks in AVHRR cloud imagery with neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Surface Atmos. Remote Sens., Technol., Data Anal. Interpretation*, vol. 3, Aug. 1994, pp. 1401–1403.
[6] J. M. Weiss, R. Luo, and R. M. Welch, "Automatic detection of ship tracks in satellite imagery," in *Proc. IEEE Int. Geosci. Remote Sens. (IGARSS), Remote Sens. Sci. Vis. Sustain. Develop.*, vol. 1, Aug. 1997, pp. 160–162.
[7] G. Jubelin and A. Khenchaf, "Multiscale algorithm for ship detection in mid, high and very high resolution optical imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2014, pp. 2289–2292.
[8] F. Bi, B. Zhu, L. Gao, and M. Bian, "A visual search inspired computational model for ship detection in optical satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 749–753, Jul. 2012.
[9] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3446–3456, Sep. 2010.
[10] C. Corbane, F. Marre, and M. Petit, "Using SPOT-5 HRG data in panchromatic mode for operational detection of small ships in tropical area," *Sensors*, vol. 8, no. 5, pp. 2959–2973, May 2008.
[11] Y. Yin, N. Liu, C. Li, W. Wan, and T. Fang, "Coarse-to-fine ship detection using visual saliency fusion and feature encoding for optical satellite images," in *Proc. IEEE Int. Conf. Audio, Lang. Image Process.*, Jul. 2017, pp. 705–710.
[12] J. Tang, C. Deng, G.-B. Huang, and B. Zhao, "Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1174–1185, Mar. 2014.
[13] Z. Shi, X. Yu, Z. Jiang, and B. Li, "Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4511–4523, Aug. 2014.
[14] N. Yokoya and A. Iwasaki, "Object localization based on sparse representation for remote sensing imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2014, pp. 2293–2296.
[15] G. Yang, B. Li, S. Ji, F. Gao, and Q. Xu, "Ship detection from optical satellite images based on sea surface analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 641–645, Mar. 2014.
[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[17] Y. LeCun, "Generalization and network design strategies," in *Connectionism in Perspective*, R. Pfeifer, Z. Schreter, F. Fogelman, and L. Steels, Eds. Zürich, Switzerland: Elseveir, 1989, pp. 143–155.
[18] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.* 2017, pp. 3856–3866.
[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[20] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2016, pp. 770–778.
[22] I. Suleymanova *et al.*, "A deep convolutional neural network approach for astrocyte detection," *Sci. Rep.*, vol. 8, Aug. 2018, Art. no. 12878.
[23] X.-L. Zhang and G.-G. He, "Forecasting approach for short-term traffic flow based on principal component analysis and combined neural network," *Syst. Eng.-Theory Pract.*, vol. 27, no. 8, pp. 167–171, 2007.
[24] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentatio," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Jun. 2014, pp. 580–587.
[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[26] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 379–387.
[27] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 354–370.
[28] T. Kong, F. Sun, A. Yao, H. Liu, M. Lu, and Y. Chen, "RON: Reverse connection with objectness prior networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2017, pp. 5936–5944.
[29] J. Redmon and A. Farhadi. (Dec. 2017). "YOLO9000: Better, faster, stronger." [Online]. Available: https://arxiv.org/abs/1612.08242
[30] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.
[31] T-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature Pyramid Networks for Object Detection," in *Proc. CVPR*, 2017, vol. 1, no. 2, pp. 1–9.
[32] T.-Y. Lin, P. Goyal, K. He, P. Dollár, and R. Girshick. (Aug. 2017). "Focal loss for dense object detection." [Online]. Available: https://arxiv.org/abs/1708.02002
[33] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS J. Photogramm. Remote Sens.*, vol. 98, pp. 119–132, Dec. 2014.
[34] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, Jul. 2016.
[35] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.



YE YU was born in Siping, Jilin, China, in 1992. He received the B.S. degree in optical engineering from Jilin University, Jilin, in 2015, and the M.S. degree in optical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin, in 2018, where he is currently pursuing the Ph.D. degree in optical engineering with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences.

His research interests include image processing and artificial intelligence, and he is currently involved on remote sensing image processing and satellite on-orbit information processing technology. Until now, he has published three articles (including non-first authors) in scientific journals and holds two patents

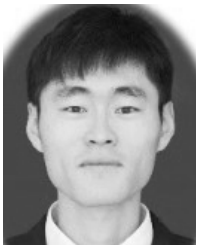
Mr. Yu was a recipient of several scholarships from Jilin University, Academic Scholarship from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 2015–2017.

HUA AI was born in Changchun, Jilin, China, in 1961. He has been with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, since 1984. His research direction is photoelectric displacement measuring technology and instruments.

XIAOJUN HE was born in Guang'an, Sichuan, China, in 1983. He received the B.S. degree from Jilin University in 2011 and the Ph.D. degree from the Chinese Academy of Sciences in 2016.

He is currently the Director of the Photoelectric Imaging Room at Chang Guang Satellite Technology Co., Ltd., and the Chief of the Spectrum Series Satellite. He is responsible for the overall development of the load electronics system and the spectrum series satellites, and has long been committed to the research of new imaging technologies. The relevant research results are widely used in the Jilin-1 Satellite. As a major member, he has participated in over 10 aerospace projects, such as 863, natural fund, and final assembly research. He has published over 10 papers and granted two patents.

Dr. He received the first prize of the Army Science and Technology Progress Award (ranked ninth) and was a member of the National Nuclear High Base 01 Special Expert Group. He was selected as the first batch of research and development talent team of major science and technology projects in Jilin Province.



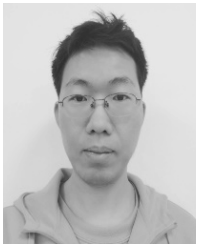
SHUHAI YU was born in China in 1985. He received the B.S. degree from Jilin University in 2009 and the Ph.D. degree from the Changchun Institute of Optics and Mechanics, Chinese Academy of Sciences, in 2014.

He was an Assistant Researcher at the Changchun Institute of Optics and Mechanics from 2014 to 2016. In 2016, he joined the Optical Design Research Office, Chang Guang Satellite Technology Co., Ltd., working on the on-orbit intelligent processing of satellites.

XING ZHONG was born in 1982. He received the master's degree and the Ph.D. degree in optical engineering from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences (CAS). He is currently the Vice President and the Chief Engineer at Chang Guang Satellite Technology Co., Ltd., and also a Full Professor at the University of CAS. His research interests are satellite's overall design, especially the payload and platform integration technologies. Until now, he has published over 60 papers in scientific journals and holds a number of patents.

During his institute career, as a Principal Investigator, he has held a series of space optical instruments developing and played an important technical role in many national space remote-sensing projects. He currently puts emphasis on the construction of Jilin-1 satellite's constellation, for high resolution and quick-revisit remote-sensing applications.

Dr. Zhong received the CAS's special prize of President Scholarship in 2009, the special prize of the 9th Changbai Youth Science and Technology Award in 2013, the CAS's LuJiaxi Young Scholar's Prize in 2016, and so on.



MU LU was born in Changchun, China, in 1989. He received the B.E. and M.E. degrees in communication engineering from Jilin University in 2012 and the Ph.D. degree in mechatronic engineering from the University of Chinese Academy of Sciences in 2017.

He has authored three articles. His research interests include digital imaging processing and moving target detection and multitarget tracking, and moving target recognition.