IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 9, 2019, accepted April 21, 2019, date of publication April 25, 2019, date of current version May 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2913191

A Novel Data Augmentation Method for Detection of Specific Aircraft in Remote Sensing RGB Images

YIMING YAN, YUMO ZHANG, AND NAN SU

Department of Information Engineering, Harbin Engineering University, Harbin 150001, China

Corresponding author: Nan Su (sunan08@hrbeu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61801142, Grant 61601135, and Grant 61675051, in part by the Natural Science Foundation of Heilongjiang Province of China under Grant QC201706802, in part by the China Postdoctoral Science Foundation under Grant 2018M631912, and in part by the Postdoctoral Science Foundation of Heilongjiang Province under Grant LBH-Z18060.

ABSTRACT In this paper, we propose a novel data augmentation method for the detection of specific aircraft in remote sensing RGB images. For object detection, the number of training data has great influence on the training of deep learning network. Researchers suggest that the extensive training samples in deep learning are indispensable. The extensive training samples can guarantee the accuracy and robustness of object detection. We refer to military aircraft and helicopter as specific aircraft. Due to the number of remote sensing images of specific aircraft is far less than that of civil aircraft, it is difficult to train an ideal detection model only by using the available specific aircraft images. Deep learning networks have the excellent ability on fault tolerance and generalization and can extract features from simulated aircraft samples. This means that the simulated aircraft samples can partly replace real images to some extent and reduce the need for detection models. Inspired by this, true remote sensing images are combined with specific aircraft three-dimensional models to form simulated images. Compared with previous data augmentation method, such as flipping and rotating, our method brings in new sample information. The experiments based on remote sensing images show the feasibility and effectiveness of our method. Meanwhile, the proposed method has compatibility with other data augmentation methods.

INDEX TERMS Aircraft detection, data augmentation, remote sensing images, convolutional neural networks.

I. INTRODUCTION

Aircraft detection is one of the significant areas in remote sensing image analysis and it has attracted increasing attention in the field of remote sensing image analysis [1]–[3]. Numerous experiments have shown that deep learning is the best method in the field of object detection. Recently, object detection methods using deep learning networks [4]–[11] such as Faster-RCNN [4] and SSD [7] have been proposed. Xu *et al.* [12] proposed a unified framework which detected aircraft with the help of the fully convolutional network (FCN). Yu *et al.* [13] presented a rotation-invariant method for detecting geospatial objects from high-resolution satellite images. Han *et al.* [14] proposed a robust pre-trained efficient multi-class geospatial object detection framework for HSR imagery. Deep learning methods can get excellent detection results with sufficient training data.

On the other hand, detection networks based on deep learning generally cannot obtain ideal results without sufficient training data. When the training data is insufficient, low recall rate and over-fitting often occur. Military aircraft and helicopter are referred as specific aircraft in this paper. Available remote sensing images of specific aircraft are insufficient. There are many datasets from remote sensing images, but the aircraft in today's datasets [15]–[17] like DOTA [17] dataset are almost civil aircraft. In addition, due to the great difference between civil aircraft and specific aircraft, we cannot use the network trained by civil aircraft to detect specific aircraft.

Due to the lack of remote sensing specific aircraft samples, training data need to be augmented. Deep learning networks deal with the problem of object characteristics [18], [19].

The associate editor coordinating the review of this manuscript and approving it for publication was Geng-Ming Jiang.

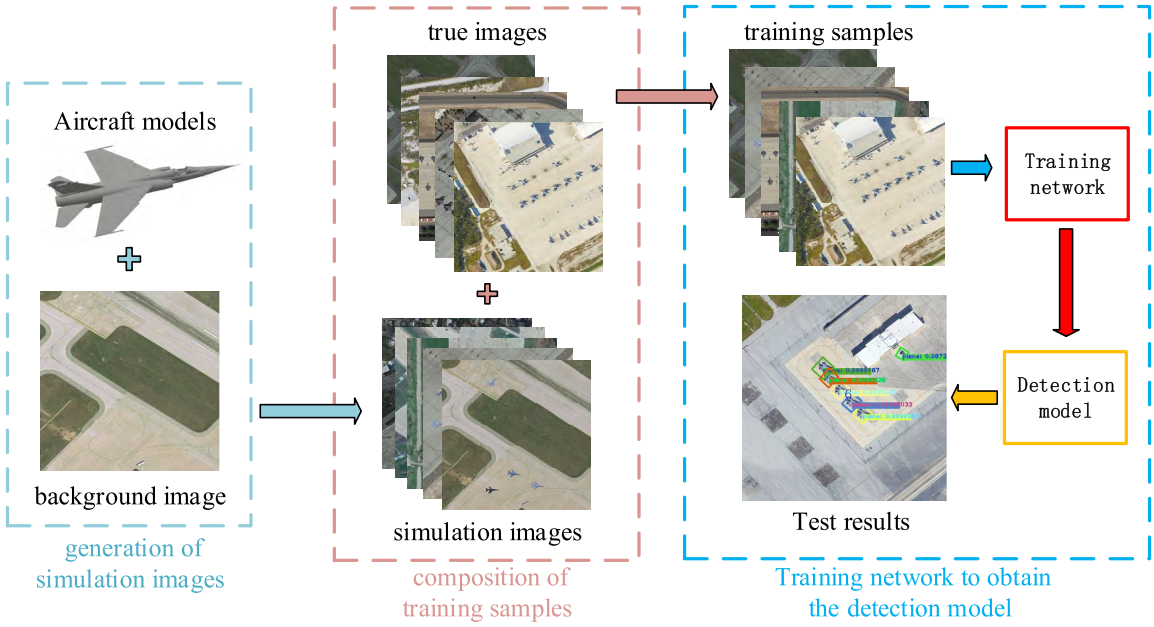


FIGURE 1. The main framework of specific aircraft detection.

These characteristics are not necessarily extracted from the real object. The simulated object can also provide characteristic information just like the real object. So, three-dimensional models of specific aircraft could be used to make simulated aircraft. True remote sensing images are combined with specific aircraft three-dimensional models to form simulated images.

Data augmentation aims at artificially enlarging the training dataset from existing data using various methods. It is widely used in the training of deep learning [10], [20]–[23]. Traditional data augmentation methods such as cropping [10], flipping [21] and rotating [23] do not essentially increase the positive sample information of the dataset. Li et al. proposes a weakly supervised deep learning (WSDL) method and they use rotating to augment data. Recently, Wang et al. [24] earned an adversary with Fast-RCNN [5] detection to create hard examples on the fly by blocking some feature maps spatially. Instead of generating occlusion examples in feature space, our method generates simulation images from the original images with very little computation, and does not require any extra parameters learning. Zhong *et al.* [25] randomly selected a rectangle region in an image, and erase its pixels with random values. Random erasure is still the processing of the original image, while our method adds additional sample information.

In this paper, we present a data augmentation method to solve the insufficient of training samples. It combines three-dimensional models of aircraft with remote sensing images to form the simulation images. Compared with other data augmentation methods, our method increases the positive sample information of the dataset, and it can be used together with other augmentation methods to improve the detection results. The major contributions of this paper are as follows.

1) Aiming at the problem of insufficient training samples, a new training strategy based on simulation data is proposed.
2) A simulation data generation method is proposed based on equal scale 3D model.

II. PROPOSED METHOD

In this section, we introduce the main framework of our method and the generation principle of simulation images.

A. MAIN FRAMEWORK

The main framework of specific aircraft detection is shown in Figure 1. The framework mainly is composed of three parts: the generation of simulation images, the composition of the training samples and training network to obtain the detection model.

In the first place, the specific aircraft model is combined with the remote sensing images to form the simulation images. Then, the simulated images are added to the real images to form the training data. Finally, we use the data to train the network, and the obtained detection model is used to detect the specific aircraft in the testing images. In this paper, the R-DFPN (Rotation Dense Feature Pyramid Network) [26] is selected as the training network.

B. SPECIFIC AIRCRAFT SIMULATION IMAGE GENERATION

As mentioned in the introduction, we can obtain a large number of samples of civil aircraft in remote sensing images. However, there is a big difference between specific aircraft and civil aircraft. We cannot use the latter training model to

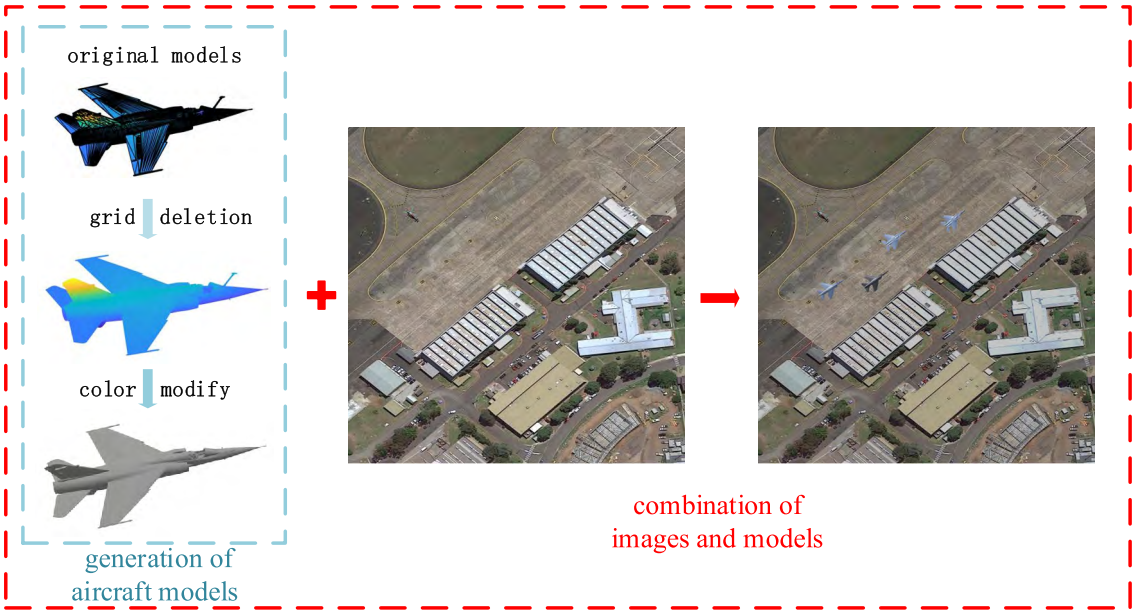


FIGURE 2. The framework of the generated simulation image.

detect the former. At the same time, sufficient samples of specific aircraft cannot be obtained. So, the specific aircraft dataset is augmented by aircraft three-dimensional models in this paper.

The framework of the generated simulation image is shown in Figure 2.

As showed in Figure 2, the generation of simulation images mainly consists of two parts: the generation of specific aircraft models and the combination of true remote sensing images and models. The overall approach is shown below:

$$I = \{I_T, I_S\} \tag{1}$$

where I denote the training images, I_T are true remote sensing images, I_S are simulation images. Simulation images and true remote sensing images constitute training set.

For the first part, we collect three-dimensional model data of specific aircraft. The format of the three-dimensional model is:

$$M = \{P_{N\times 3}, T\} \tag{2}$$

where M is three-dimensional model, $P_{N\times 3}$ is a point cloud with N points, T is an array in which three points are a group, and connects points in the point cloud into triangles. Then, change the color of the model as:

$$M_C = C_m(M) \tag{3}$$

where M_C is the model after texture mapping, C_m is to use the color map to modify the color of the model. The aircraft is mostly gray and black, so the aircraft models were modified to gray and black in order to increase the similarity between the simulated aircraft and the actual aircraft. Generated specific aircraft models are shown in Figure 3.

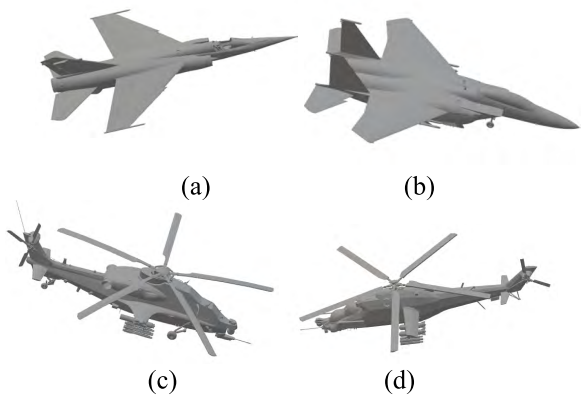


FIGURE 3. Specific aircraft models.

For the second part, the simulation specific aircraft models and remote sensing images are used to generate simulation images. The composition of simulation image as follows:

$$I_S = I_B \cup P_M\{u, v, r, \alpha\} \tag{4}$$

where I_B is the background image, P_M is the projection of three-dimensional model onto the background image. $u$and v control the position of the projection in the background image, r and α controls the size and angle of the projection.

The simulation sample is generated as:

$$P_M = T_v T_p \begin{Bmatrix} x \\ y \\ z \\ 1 \end{Bmatrix} = TN_C \tag{5}$$

where T_v is observing the transformation matrix, T_p is perspective transformation matrix. A three-dimensional

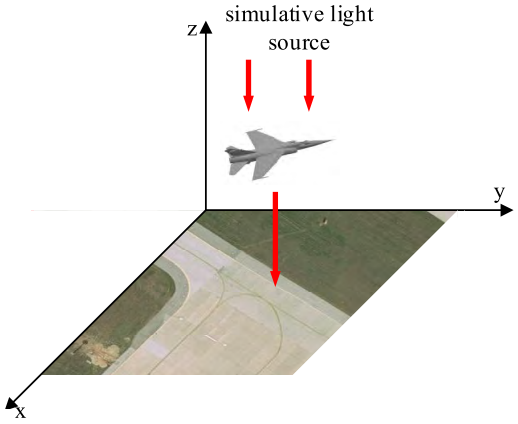


FIGURE 4. The principle of simulation image generation.

coordinate system is constructed, which takes the remote sensing image as x-y plane and takes the upper left corner of the plane as the origin of coordinates. Select the appropriate proportion for the specific aircraft model, make the specific aircraft model size to adapt to the image resolution. The principle of simulation image generation is shown in Figure 4. Parts of the simulated images are shown in the Figure 5

C. TRAINING NETWORK

In this section, we will introduce the main framework of R-DFPN network [26]. The framework mainly consists of two parts: A Dense Feature Pyramid Network (DFPN) for feature fusion and a Rotation Region Detection Network (RDN) for prediction. Specifically, DFPN can generate feature maps that are fused by multiple features for each input image. Then, rotational proposals are obtained from the RPN to provide high-quality regional proposals or the next stage. Finally, location regression and class prediction of proposals are processed in the Fast-RCNN stage.

The feature pyramid is an effective multiscale method to fuse multilevel information. Feature Pyramid Networks (FPN) achieved very good results in small object detection tasks. It uses the feature pyramid, which is connected via a top-down pathway and lateral connection. And DFPN can

FIGURE 5. Simulated images.

significantly improve the detection performance due to the smooth feature propagation and feature reuse.

Similar to the traditional detection framework, Rotation Region Detection Network (RDN) also contains two stages: RPN and Fast-RCNN. In the RPN stage, the representation of the rectangle need to be redefined to get the "Rotation Bounding Box" at first. After that, rotation proposals are generated by regressing the rotation anchors to reduce the impact of non-maximum suppression and improve the recall. Finally, the final result is obtained by non-maximum suppression.

III. EXPERIMENTS

In this section, we evaluate our method on the data which are collected from Google Earth with a spatial resolution of 0.5 m. A total of 150 images are collected, all of which are 512×512 pixels in size, including many international cities. In addition, in order to verify that the training model of civil aircraft cannot accurately detect specific aircraft, we selected 170 images of civil aircraft from the DOTA data set.

In the first part of the experiment, we discuss the necessity and feasibility of using simulation image training model to detect specific aircraft. Then, we design multiple experiments and prove the validity of our method. Finally, we demonstrate that our data augmentation method can be used together with other methods and achieve better detection results. We use the method of Rotation Dense Feature Pyramid Networks (R-DFPN) to train the network. The whole project is written in TensorFlow and python based on CPU (Intel Xeon Gold 6130) and GPU (GTX TITANV).

A. EVALUATION INDICATORS

In this paper, recall, precision and F-measure are used to evaluate our aircraft detection framework on the test images. The calculation method of recall and precision rate is:

$$recall = \frac{TP}{TP + FN} \quad precision = \frac{TP}{TP + FP} \tag{6}$$

Here TP represents the number of true positives, FP represents the number of false positives, and FN represents the number of false negatives.

In general, precision rate and recall rate indicators are contradictory, so they need to be considered comprehensively. The most common method is f-measure (also known as f-score). F-measure is the weighted harmonic average of Precision and Recall:

$$\mathrm{F} = \frac{(\alpha^2 + 1)\mathrm{P} \times \mathrm{R}}{\alpha^2(\mathrm{P+R})} \tag{7}$$

When the parameter $\alpha = 1$, this is the most common F1-measure. F1- measure combines the results of precision and recall, and when F1- measure is high, the test method is more effective.

B. AIRCRAFT DETECTION WITH CIVIL ARICRAFT MODEL

In this section, we discuss the necessity of using simulation image training model to detect specific aircraft. 100 images with 500 civil aircraft samples are used as the set of training images. Then, use these images to train the network and the trained models are tested both 70 civil aircraft images with 350 samples, 50 specific aircraft images of 210 samples and 50 helicopter images of 190 samples. The results of the detection are shown in Figure 6 and Table 1.

TABLE 1. Detection results of civil aircraft training model.

Aircraft category	recall	precision	F1-measure
Civil aircraft	87.5%	90.6%	89.1%
Specific aircraft	19.9%	60.7%	30.1%
Helicopter	23.2%	65.0%	34.2%

As can be seen from the (c), (d) of Figure 6, although the background of civil aircraft image is complex, almost all the civil aircraft can be detected. On the contrary, the (e), (f), (g), (h) of Figure 6 show that the background of specific aircraft and helicopter image is simple but very few aircraft samples can be detected. This is because the shape of civil aircraft and the other two types of aircraft are quite different. In the detection of two types of aircraft, the convolution network cannot match the corresponding features in the test images. As a result, two types of aircraft and civil aircraft are considered as different objects and the test results are poor.

Table 1 shows the recall, precision and F1-measure of the civil aircraft model in detecting each types of aircraft. The results show that the training model of civil aircraft is very poor in detecting specific aircraft. The recall rate of specific aircraft is only 19.9% and the recall rate of the helicopter is only 23.2%. This means that when testing specific aircraft and helicopter, we cannot use civil aircraft as training data.

C. AIRCRAFT DETECTION WITH SIMULATED SAMPLES

In this section, we discuss the feasibility of using simulated samples training model to detect specific aircraft. For testing whether the model trained by simulation image can detect specific aircraft, 80 empty scene images which include no aircraft are selected for training. Then, we construct simulation images through our method and every image is combined with 4 specific aircraft models. After getting 80 images with about 320 samples, we use these images to train the network and the trained models are tested with 50 real remote sensing specific aircraft images with 210 samples. The results of the test are shown in Figure 7 and Table 2.

As can be seen from the (a), (b) of Figure 7, because the simulation samples are generated by three-dimensional models of real airplanes, the simulated aircraft samples are very similar to the real aircraft in shape and material. Meanwhile, (c), (d) of Figure 7 show that the model trained by simulation samples can detect partial specific aircraft in real remote sensing specific aircraft images.

Table 2 shows the recall, precision and F1-measure of the model trained by simulation images. The recall rate of specific aircraft is 38.1% and the precision rate is 88.6%.

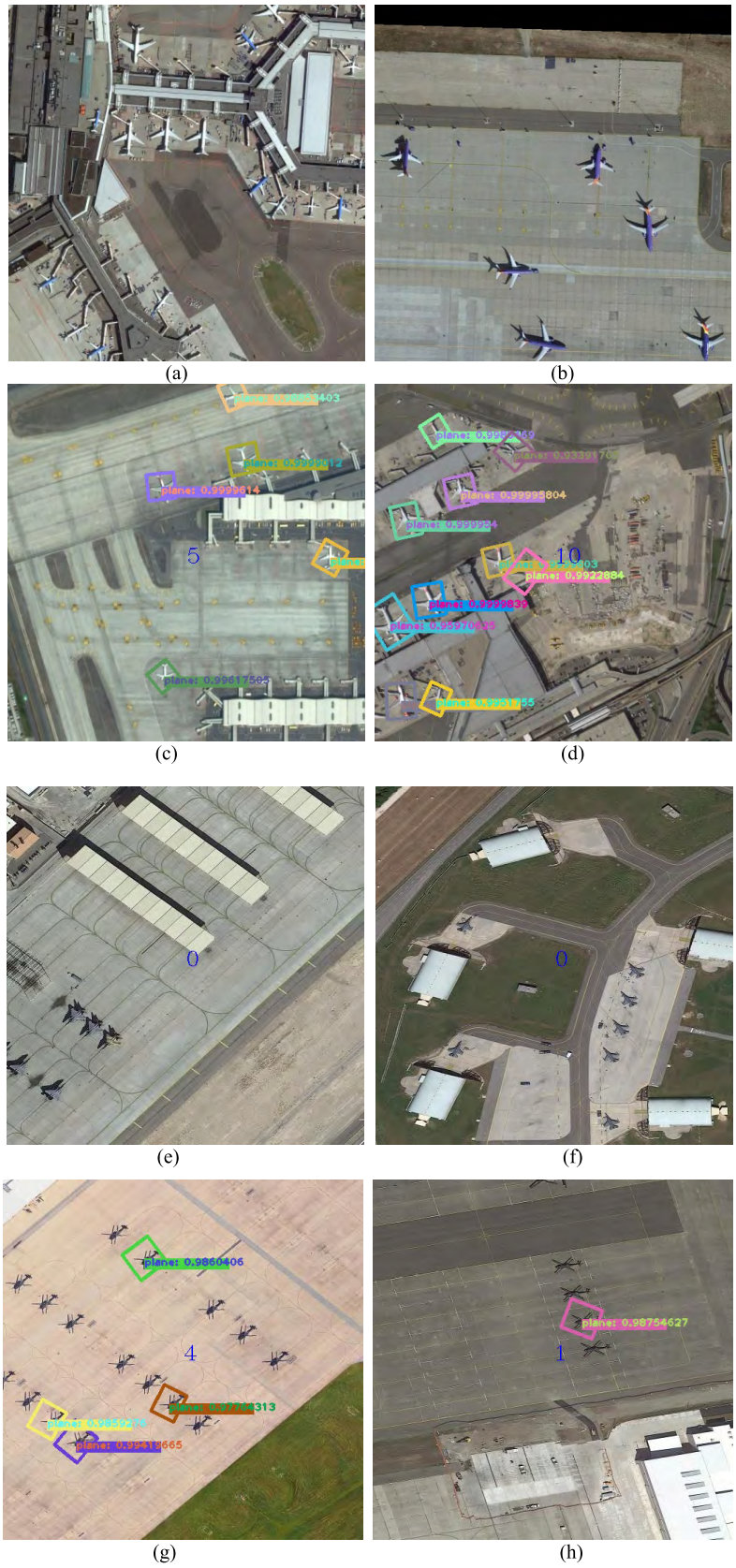


FIGURE 6. (a), (b) are part of training images. (c), (d) are part of civil aircraft test results, (e), (f) are part of specific aircraft test results, (g), (h) are part of helicopter test results.

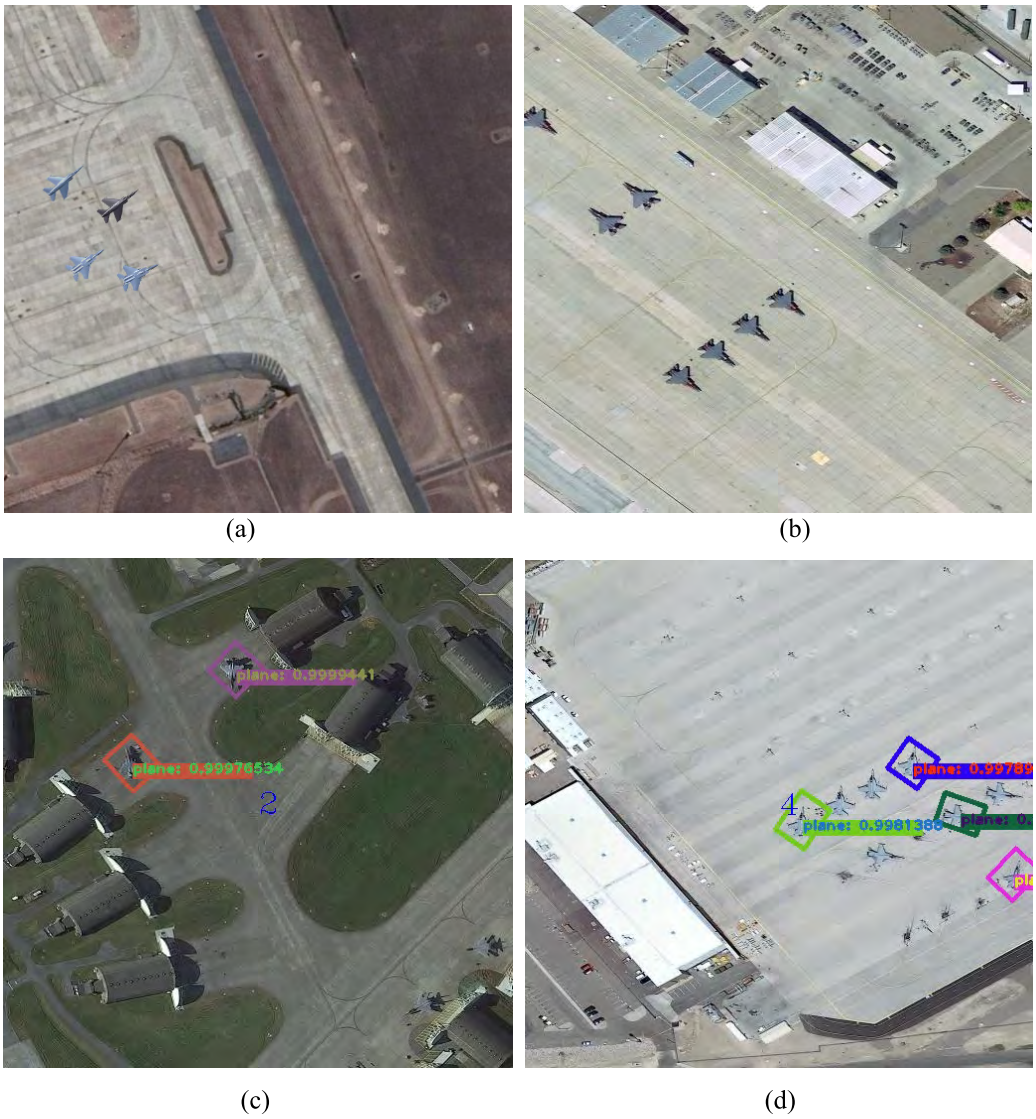


FIGURE 7. (a) Simulation aircraft samples (b) true aircraft samples (c), (d) results trained by simulation samples.

TABLE 2. Detection result of simulation image.

Aircraft category	recall	precision	F1-measure
Specific aircraft	38.1%	88.6%	52.3%

The color, Angle and other factors of the simulation sample can be arbitrarily set. This makes the simulation sample very similar to the real aircraft. Moreover, convolutional neural network has certain fault tolerance and generalization ability. So, simulated images constructed by our method can partly replace the real image to some extent and reduce the need for the detection model.

D. AIRCRAFT DETECTION WITH SIMULATED IMAGES

In this section, we compare the detection results of the original specific aircraft image training model and the simulation image training model. Firstly, two sets of remote sensing images are selected for training, the first set containing 10 images and the second set containing 20 images. Then, we combine each image with 4 specific aircraft models through our method and use these simulation images to train the network. Finally, testing data included 50 specific aircraft images are tested with the models obtained from these two sets. The helicopter images are processed and tested in the same way. ResNet-101 network and Faster-RCNN ZF network are used to initialize the network. The results are showed in Figure 8 and Table 3,Table 4.

As can be seen from Figure 8, the simulation image detection results of the aircraft model are better than those of original remote sensing image. Due to insufficient training data, three aircraft in figure (a) four and four aircraft in figure (b) were not detected. Each simulation image has 4 aircraft samples more than the original image, which makes

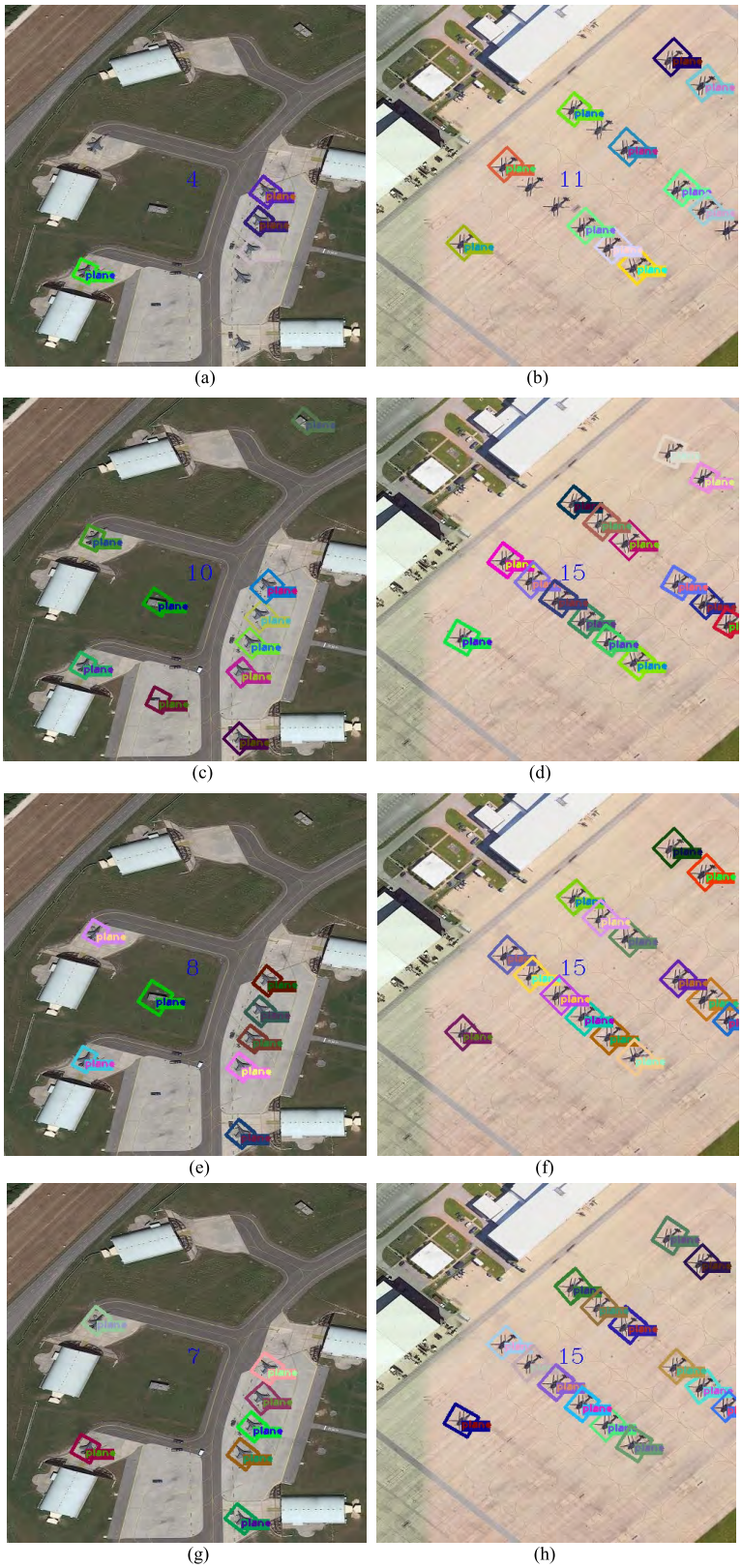


FIGURE 8. (a), (b) are part of the results trained by 10 true images (c), (d) are part of the results trained by 10 images with 40 simulated samples (e), (f) are part of the results trained by 20 true images (g), (h) are part of the results trained by 20 images 80 simulated samples.

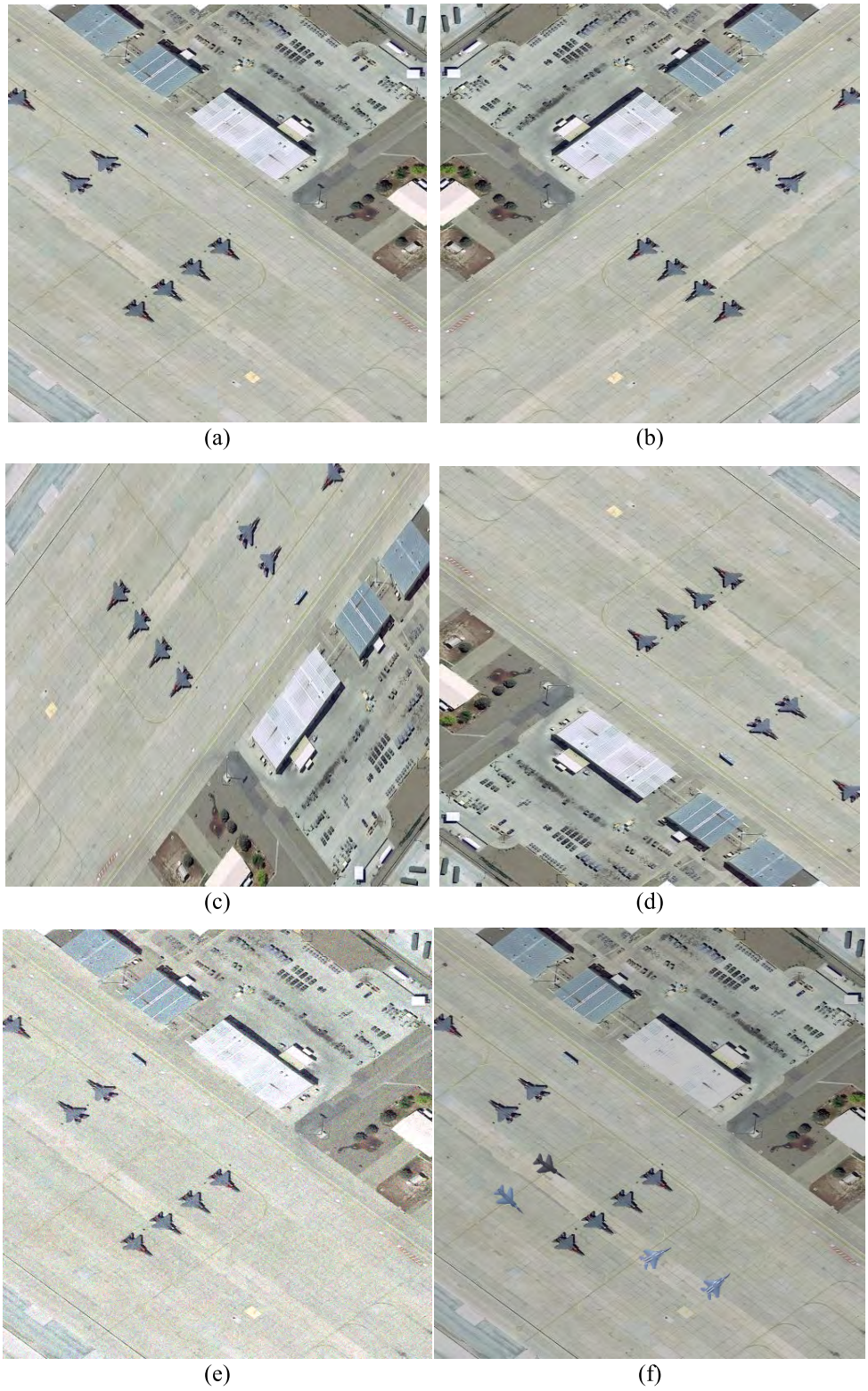


FIGURE 9. Different training images of four data augmentation methods.

the simulation image have more aircraft information. After adding simulation samples, the detection model can detect more aircraft.

Table 3 and Table 4 shows the recall, precision and F1-measure of different image models. As can be seen from Table 3 and Table 4, after the 10 images are further processed

TABLE 3. Specific aircraft detection results of different images.

images	network	recall	precision	F1-measure
10 original images	ResNet-101	64.9%	98.1%	78.1%
	ZF	76.2%	76.2%	76.2%
10 images+40 simulated samples	ResNet-101	90.2%	85.1%	87.6%
	ZF	80.4%	86.9%	83.5%
20 original images	ResNet-101	89.3%	90.1%	89.4%
	ZF	84.2%	80.2%	82.2%
20 images+80 simulated samples	ResNet-101	92.3%	88.9%	90.6%
	ZF	81.7%	87.9%	84.7%

TABLE 4. Helicopter detection results of different images.

images	network	recall	precision	F1-measure
10 original images	ResNet-101	82.3%	96.2%	88.7%
	ZF	88.2%	91.3%	89.7%
10 images+40 simulated samples	ResNet-101	91.5%	94.5%	92.9%
	ZF	90.6%	90.6%	90.6%
20 original images	ResNet-101	87.5%	95.3%	91.2%
	ZF	92.1%	88.2%	90.1%
20 images+80 simulated samples	ResNet-101	93.1%	93.6%	93.3%
	ZF	91.1%	90.7%	90.9%

TABLE 5. Detection results with different data augmentation methods.

Training data	recall	precision	F1-measure
original	43.4%	87.3%	58.0%
original*	66.9%	83.6%	74.4%
flipping	64.9%	98.1%	78.1%
flipping*	90.2%	85.1%	87.6%
noise	61.1%	84.6%	70.9%
noise*	72.3%	82.3%	76.9%
rotating	81.7%	85.8%	83.7%
rotating*	88.9%	83.8%	86.3%

with our method, the detection results of both ResNet-101 network and ZF network are obviously improved. Compared with the original images, the network trained by the simulation image shows improvement in the F1-measure. And as can be seen from the third and fourth rows of Table 3 and Table 4, this rule is also applicable to detection result of 20 images. The experimental results show that we can obtain better detection results by our data augmentation strategy under small sample condition.

E. THE COMPATIBILITY WITH OTHER METHODS

In this section, we discuss the influence of different methods on the detection results. Our experiments use the method of Rotation Dense Feature Pyramid Networks (R-DFPN) and adopted the preordained model ResNet-101 to initialize the network. We chose four different data augmentation methods: flipping [21], adding Gaussian noise [22], rotating [23] and simulation image to carry out the experiment. In order to test the influence of different methods on the detection results, we conduct four comparative experiments.

In Figure 9, (a) shows the original image, (b) shows the flipped image, (c) and (d) show the rotated image, (e) shows the image added noise and (f) shows the simulated image. At the same time, in order to verify the compatibility of our method, the simulation image is used on the basis of the other three methods. The results are showed in Table 5.

Table 5 shows the recall, precision and F1-measure of different detection results. The '$*$' indicates use simulation images to data augmentation and each image is combined with 4 aircraft models. Four groups of training samples are respectively the original image, using three data augmentation methods and using simulation images on the basis. As can be seen from Table 5, detection result trained by the original image is the worst. The results of training by single augmentation method are better than that of the original image. Furthermore, using simulation images on the basis

of other methods will make the detection result better. This shows that our method can be compatible with other data augmentation methods and make the detection results better.

IV. CONCLUSION

In this paper, we propose a novel strategy to augment the training positive samples by using simulated images. Specific aircraft three-dimensional model is projected onto real remote sensing images and generates simulation images. In the case of insufficient images, this method can effectively increase the number of positive samples.

The experiments showed that:

1) When the training samples are insufficient, we cannot obtain a good detection model. And civil aircraft training network cannot be used to detect specific aircraft

2) The proposed method improves the comprehensive ability of detection effectively. The detection performance of the method is improved obviously at each comparative experiments.

3) The proposed method is well compatible with other data augmentation methods.

CONTRIBUTION

Yumo Zhang, Na Su, and Yiming Yan conceived and designed the experiments; Yumo Zhang performed the experiments; Yumo Zhang, Na Su, and Yiming Yan analyzed the experiments; Yumo Zhang wrote the paper.

REFERENCES

[1] H. Wu, H. Zhang, J. Zhang, and F. Xu, "Fast aircraft detection in satellite images based on convolutional neural networks," in *Proc. ICIP*, Sep. 2015, pp. 4210–4214.

[2] L. Liu and Z. Shi, "Airplane detection based on rotation invariant and sparse coding in remote sensing images," *Optik*, vol. 125, no. 18, pp. 5327–5333, Sep. 2014.

[3] Y. Zhang, K. Fu, H. Sun, X. Sun, X. Zheng, and H. Wang, "A multi-model ensemble method based on convolutional neural networks for aircraft detection in large remote sensing images," *Remote Sens. Lett.*, vol. 9, no. 1, pp. 11–20, 2018.

[4] 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.*, Montreal, QC, Canada, Dec. 2015, pp. 91–99.

[5] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, Dec. 2015, pp. 1440–1448.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788.

[7] W. Liu *et al.*, "Ssd: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. ECCV*. Springer, 2016, pp. 630–645.

[9] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Proc. NIPS*, 2016, pp. 379–387.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. ICCV*, Oct. 2017, pp. 2980–2988.

[12] T.-B. Xu, G.-L. Cheng, J. Yang, and C.-L. Liu, "Fast aircraft detection using end-to-end fully convolutional network," in *Proc. IEEE Int. Conf. Digit. Signal Process.*, Beijing, China, Oct. 2016, pp. 139–143.

[13] Y. Yu, H. Guan, and Z. Ji, "Rotation-invariant object detection in high-resolution satellite imagery using superpixel-based deep hough forests," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2183–2187, Nov. 2015.

[14] X. Han, Y. Zhong, R. Feng, and L. Zhang, "Robust geospatial object detection based on pre-trained faster R-CNN framework for high spatial resolution imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 3353–3356.

[15] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (voc) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.

[16] H. Xiao, K. Rasul, and R. Vollgraf. (2017). "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms." [Online]. Available: https://arxiv.org/abs/1708.07747

[17] G.-S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3974–3983.

[18] M. Zhang, "Aircraft recognition via structural features," *Comput. Eng. Appl.*, vol. 49, no. 4, pp. 174–177, 2013.

[19] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015, pp. 1097–1105.

[22] H.-J. Bae *et al.*, "A Perlin noise-based augmentation strategy for deep learning with small data samples of HRCT images," *Sci. Rep.*, vol. 8, Dec. 2018, Art. no. 17687.

[23] Y. Li, Y. Zhang, X. Huang, and A. L. Yuille, "Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 182–196, Dec. 2018.

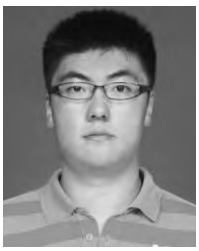
[24] X. Wang, A. Shrivastava, and A. Gupta, "A-fast-RCNN: Hard positive generation via adversary for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3039–3048.

[25] Z. Zhong *et al.*, "Random erasing data augmentation," Tech. Rep., 2017.

[26] X. Yang *et al.*, "Automatic ship detection in remote sensing images from Google earth of complex scenes based on multiscale rotation dense feature pyramid networks," *Remote Sens.*, vol. 10, no. 1, p. 132, 2018.

YIMING YAN was born in 1982. He received the bachelor's, master's, and Ph.D. degrees in information and communication engineering from the Harbin Institute of Technology, China, in 2006, 2008, and 2013, respectively. Since 2016, he has been with Harbin Engineering University. His research interests include remote sensing image processing and pattern recognition.

YUMO ZHANG was born in 1995. He received the bachelor's degree in information and communication engineering from Harbin Engineering University, China, in 2018, where he is currently pursuing the master's degree. His research interests include remote sensing image processing and pattern recognition.



NAN SU was born in 1986. He received the bachelor's, master's, and Ph.D. degrees in information and communication engineering from the Harbin Institute of Technology, China, in 2010, 2012, and 2017, respectively. Since 2017, he has been with Harbin Engineering University. His research interests include remote sensing image processing and pattern recognition.