Online Exemplar-Based Fully Convolutional Network for Aircraft Detection in Remote Sensing Images

Bowen Cai, Zhiguo Jiang, *Member, IEEE*, Haopeng Zhang, *Member, IEEE*, Yuan Yao, and Shanlan Nie

***Abstract*— Convolutional neural network obtains remarkable achievements on target detection, due to its prominent capability on feature extraction. However, it still needs further study for aircraft detection task, since intraclass variation still restricts the accuracy of aircraft detection in remote sensing images. In this letter, we adopt regularity of aircraft circle response to design our end-to-end fully convolutional network (FCN), and embed online exemplar mining into our network to handle intraclass variation. The mined exemplars are employed to capture different intraclass characteristics, which effectively reduces the burden of network training. Specifically, we first select basic exemplars based on labeled information and initialize the relationships between exemplars and aircraft examples. Then, these relationships will be updated by the similarity of these examples in high-level features space. Finally, aircraft examples will be used to train different exemplar detectors according to updated relationships. Motivated by the geometric shape of aircraft, a circle response map is developed to construct our FCN to achieve more efficient aircraft detection. The comparative experiments indicate that superior performance of our network in accurate and efficient aircraft detection.**

***Index Terms*— Aircraft detection, exemplar mining, fully convolutional network (FCN).**

I. Introduction

AIRCRAFT detection is a representative task in remote sensing images (RSIs) and has attracted increasing attention [1]–[9]. For a cross-shaped geometric structure of aircrafts, a great amount of methods design handcrafted features to accomplish aircraft detection. An *et al.* [2] utilized circle frequency filter to locate the region of interest (RoI) and the extracted histogram of gradients (HoGs) features to classify the regions containing an aircraft. To regulate the dominant orientation of a region and capture more detailed information, Zhang *et al.* [3] designed a new rotation-invariant appearance feature called histogram of oriented gradients normalized by polar angle. Zhao *et al.* [7] adopted aggregate channel features (ACFs) to describe aircrafts in RSIs, which offered richer representations and speeds up computations. However, the handcrafted features always need to adjust the parameters carefully and are not able to accurately handle target detection task within various scales and rotations.

Fig. 1. Aircraft examples with intraclass variation. Examples at right columns show that intraclass variation is derived from aircraft type, rotation, and size.

Recently, researchers introduce convolutional neural network (CNN) into target detection in RSIs, because of its powerful capability on features representation [10]. CNN is mainly used in two strategies: 1) feature extractor [5], [6], [8] and 2) an end-to-end unified network including feature extraction and target detection [4], [9], [11]–[13]. Compared with the latter strategy, only using CNN as a feature extractor is inefficient and cannot meet the requirement for rapid target detection in RSIs. Moreover, the separation and asynchrony between the feature extractor and the classifier will reduce the effectiveness of the training process [14]. Therefore, much more studies focus on the construction of an end-to-end unified network.

The essential issue for constructing detection network is how to address the aircraft intraclass variation that has adverse effect on aircraft detection. Here, intraclass variation is defined as differences among the same category targets and is harmful for detection accuracy. These differences might be caused by different imaging conditions or diversity of target itself and obviously does not form coherent visual category, as shown in Fig. 1.

To tackle the above-mentioned problems and promote the network efficiency, we propose an online exemplar-based fully

Manuscript received February 10, 2018; revised April 11, 2018; accepted April 18, 2018. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0501300 and Grant 2016YFB0501302, in part by the National Natural Science Foundation of China under Grant 61501009, Grant 61771031, and Grant 61371134, and in part by the Aerospace Science and Technology Innovation Fund of China Aerospace Science and Technology Corporation. *(Corresponding author: Haopeng Zhang.)*

The authors are with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and also with the Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China (e-mail: bowencai@buaa.edu.cn; jiangzg@buaa.edu.cn; zhanghaopeng@buaa.edu.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2018.2829147

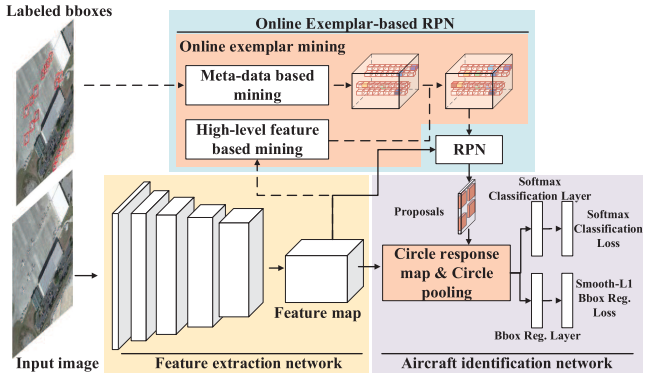

Fig. 2. Flowchart of proposed method. It consists of feature extraction network, online exemplar-based region proposal network, and aircraft identification network. The dotted line indicates that the data are only delivered in training process.

convolutional network (OE-FCN) in this letter. The proposed method introduces an online exemplar-mining mechanism into CNN and utilizes the exemplars to represent different intraclass characteristics of aircrafts. In detail, metadata, including scales and aspect ratios of labeled bounding boxes in each minibatch, are separated into different groups. Each group corresponds to a basic exemplar, and, thus, the relationships between exemplars and aircraft examples are initialized by these groups. To construct the final division of aircraft examples, high-level features of these examples in the same minibatch are also divided into different clusters. These clusters will be used to update mappings from aircraft examples to exemplars. The detectors in region proposal network will be trained independently for each exemplar using its corresponding examples. Moreover, based on the cross shape of the aircrafts, we design a circle response map (CRM) to construct our FCN.

The contributions of proposed network are three folds. 1) Our network removes the requirement for priors including the size and aspect ratio of aircrafts. 2) Compared with traditional exemplar mining, it is much more reliable, robust, and extensible by combining high-level features and labeled information. 3) With the design of the CRM, our network achieves rapid and accurate aircraft detection in RSIs.

II. Proposed Method

A. Overview

As shown in Fig. 2, the proposed network consists of a feature extraction network, an online exemplar-based region proposal network (OE-RPN), and an aircraft identification network. Feature extraction network is constructed based on Resnet-50 [15]. We keep all convolutional layers and remove the average pooling layer and fully connected layer. In addition, an extra convolutional layer is attached to reduce feature dimension from 2048 to 1024. The feature map obtained from the feature extraction network will be fed into the OE-RPN and the aircraft identification network.

In the OE-RPN, we first augment labeled bounding boxes in each minibatch for the diversity of training data and compute the scales and aspect ratios of aircrafts based on these bounding boxes. Then, the scale and aspect ratio of each aircraft example will be concatenated into a 2-D labeled feature. We will cluster all labeled features in a minibatch and use the cluster centers to represent the aircraft exemplars. Thus, all aircraft examples are separated into different exemplars based on metadata. Besides, the OE-RPN also uses the high-level features to group all aircraft examples and updates the mapping between these positive examples and typical exemplars. The losses of aircraft examples will be computed by their assigned exemplar detectors and backward propagated to update the network parameters in corresponding exemplars. The OE-RPN finally collects all candidate regions extracted by different exemplar detectors and outputs top N regions with the highest scores to the aircraft identification network.

Aircraft identification network employs the high-level features to construct a CRM, as shown in Fig. 3. The candidate regions generated from the OE-RPN are projected to the CRM and evaluated their scores. The region with higher score leads a much greater probability of containing an aircraft target, and vice versa. In this letter, a fully convolutional strategy is adopted to design the OE-RPN and aircraft identification network, and it improves the computational efficiency of our network.

B. Online Exemplar Mining

To achieve exemplar mining, conventional methods group positive examples into clusters based on the metadata [1], [11]. These clusters are used to represent mined exemplars. The prerequisite for conventional approaches is that the intraclass variation could be completely reflected on the metadata, and all the difference of metadata should be paid attention to. However, the intraclass variation really needed to be focused is that leads great different features which cannot be correctly recognized by aircraft detectors. In this letter, both metadata and high-level features are utilized to mine the aircraft exemplars more effectively. A purely online form of exemplar mining is embedded into our end-to-end FCN. Thus, the exemplar mining is exactly as frequently as with the network forward propagation; therefore, the CNN features are made full use of without learning delay. Our online exemplar mining (OEM) consists of two parts: metadata-based mining and high-level feature-based mining.

1) Metadata-Based Mining: To promote the diversity of exemplars mined from metadata, we first employ data augmentation on the labeled bounding boxes according to the constraint of intersection over union (IoU). Based on the assumption that the IoU between augmented bounding box and ground-truth box are greater than a desired IoU t, the scale and aspect ratio of the bounding box can be adjusted, respectively, to obtain a more diversified bounding box. Given a ground-truth bounding box with scale s_g and aspect ratio ar_g, we can obtain the scale s_a and aspect ratio ar_a of augmented bounding box by

$$\left\{ s_a \mid \sqrt{t} \cdot s_g \le s_a \le \frac{s_g}{\sqrt{t}} \right\} \tag{1}$$

$$\left\{ ar_a \mid \left(\frac{2t}{1+t}\right)^2 \cdot ar_g \le ar_a \le \left(\frac{1+t}{2t}\right)^2 \cdot ar_g \right\}. \tag{2}$$

Then, we concatenate the scale and the aspect ratio of augmented bounding box as a 2-D labeled feature. All these features in each minibatch are allocated into different groups by minibatch K-means algorithm [16]. The metadata of exemplars will be computed by labeled features in the same group. Besides, the dependence between aircraft examples and exemplars is constructed based on their metadata. In detail, given the center coordinate of an aircraft as an example, we utilize metadata of all exemplars to generate multiple candidate bounding boxes. The IoU values between the candidate bounding boxes and the ground truth of this example are calculated. This example will be assigned to the exemplar whose candidate bounding box has the highest IoU with ground truth.

2) High-Level Feature-Based Mining: In order to compensate for the limited representability of metadata, high-level features are also applied for exemplar mining. In our network, the feature maps generated by feature extraction network will be input into OE-RPN. The OE-RPN first selects the positive CNN features from the feature map, then employs clustering algorithm on these features to aggregate similar features into the same group. Based on the groups in metadata mining, the mapping from aircraft examples to exemplars is supplemented by the sibling relationship between the positives in high-level feature space, which constructs final divisions of aircraft examples to achieve OEM.

Algorithm 1 shows the process of our OEM in detail.

Algorithm 1 OEM

Input: In each iteration, ground-truth bounding boxes B_g, high-level features F. Meta-data clusters c_m, high-level feature clusters c_n. M_f is a index set of positive examples in mini-batch.

Output: In each iteration, meta-data clusters c_m, high-level feature clusters c_n. G_e denotes all groups in exemplar space.

compute scales and aspect ratios of B_g to construct the labeled features M_g;
M_g is augmented to obtain M_a using Eq.(1) and Eq.(2);
for $y \in M_a$ **do**
  allocate labeled feature into nearest meta-data center by $K_m[y] \leftarrow \arg\min_i \|y - c_m^i\|^2$;
for $y \in M_a$ **do**
  $k \leftarrow K_m[y]$;
  update meta-data center count $v_m[k] \leftarrow v_m[k] + 1$;
  compute a gradient descent step on meta-data centers $c_m^k \leftarrow (1 - \frac{1}{v_m[k]}) \cdot c_m^k + \frac{y}{v_m[k]}$;
$G_m \leftarrow \emptyset$;
$G_n \leftarrow \emptyset$;
for $x \in M_f$ **do**
  assign positives to meta-data centers by highest IoU $d_m[x] \leftarrow \arg\max_i \frac{area(C_m^i(x) \cap B_g)}{area(C_m^i(x) \cup B_g)}$;
  $d \leftarrow d_m[x]$;
  update meta-data groups $G_m[d] \leftarrow G_m[d] \cup \{x\}$;
  allocate positive CNN features into nearest high-level feature centers $k \leftarrow \arg\min_i \|F[x] - c_n^i\|^2$;
  update high-level feature groups $G_n[k] \leftarrow G_n[k] \cup \{x\}$;
  $K_n[x] \leftarrow k$;
$G_e \leftarrow \emptyset$;
for $x \in M_f$ **do**
  $k \leftarrow K_n[x]$;
  update center count of high-level feature $v_n[k] \leftarrow v_n[k] + 1$;
  compute a gradient descent step on high-level feature center, $c_n^k \leftarrow (1 - \frac{1}{v_n[k]}) \cdot c_n^k + \frac{F[x]}{v_n[k]}$;
  $d \leftarrow d_m[x]$;
  update the clusters G_e in exemplars space $G_e[d] \leftarrow G_m[d] \cup G_n[k]$;

C. Circle Response Map and Circle Pooling

For the cross shape of aircrafts and periodic intensities changing around aircrafts, we develop a CRM to score the candidate aircraft regions. As shown in Fig. 3, the scoring of the CRM consists of two parts: a convolutional layer and a circle pooling layer.

The convolutional layer has $r_n \times \alpha_n \times \mathrm{cls}_n$ convolutional kernels. The input feature map is fed into this layer to produce $(r_n \times \alpha_n \times \mathrm{cls}_n) \times w \times h$ circle response tensor. The $r_n \times \alpha_n \times \mathrm{cls}_n$ channels in the circle response tensor represent different class score maps in a specific radius and a central angle range, respectively. The circle pooling layer cumulates the scores from different channels and the sectors of circle response tensor. First, we transform each RoI rectangle to an RoI concentric circle with a radius $r_{\mathrm{roi}} = \max\{w_{\mathrm{roi}}, h_{\mathrm{roi}}\}/2$. The RoI circle is divided into $r_n \times \alpha_n$ sectors, as shown in Fig. 3. The range of radius and central angle for each sector limits in r_{roi}/r_n and $(2 \cdot \pi)/\alpha_n$, respectively. Then, we define a circle pooling operation that pools the (r_i, α_j) response tensor

$$s_c(r_i, \alpha_j \mid \Theta) = \frac{1}{n} \sum_{(r,\alpha) \in \mathrm{Sec}(i,j)} F_{i,j,c}(x_0 + r\cos\alpha, y_0 + r\sin\alpha \mid \Theta). \quad (3)$$

Here, $s_c(r_i, \alpha_j \mid \Theta)$ represents pooled response in the (i, j)th sector for the aircraft or background. Θ is the learnable parameters in convolutional layer. $F_{i,j,c}$ denotes a score map in $(r_n \times \alpha_n \times \mathrm{cls}_n) \times w \times h$ circle response tensor. (x_0, y_0) is the center of an RoI. n is the number of sampling pixels in circle response tensor. $\mathrm{Sec}(i, j)$ spans $\lfloor i \cdot (r_{\mathrm{roi}}/r_n) \rfloor \le r < \lceil (i+1) \cdot (r_{\mathrm{roi}}/r_n) \rceil$ and $\lfloor j \cdot (2 \cdot \pi/\alpha_n) \rfloor \le \alpha < \lceil (j+1) \cdot (2 \cdot \pi/\alpha_n) \rceil$. Based on $s_c(r_i, \alpha_j \mid \Theta)$, we will vote the final scores by averaging these pooled responses.

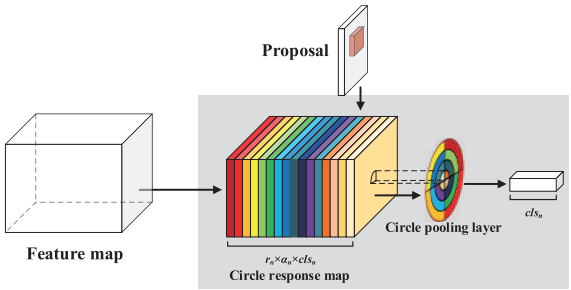

Fig. 3. CRM and circle pooling layer. The channels with different colors in CRM represent class score maps in a specific radius and central angle range. Circle pooling layer cumulate scores from different channels and sectors of CRM at proposal area.

III. Experiments

A. Experiment Setup

We validated our proposed method on two data sets, namely, RSOD data set [8] and a large-scale optical remote sensing (LSORS) data set. The LSORS data set is collected from Google Earth, including 97 different airports at a spatial resolution of 0.6 m/pixel. There are 72 airport area images using for training and validation, and the rest are used as the test data set. The 72 large-scale RSIs are cropped into 7348 patches with 512×512 due to the limitation of GPU memory, which contains 14577 aircrafts. These patches are randomly divided into 70% for training and 30% for validation. The training set is augmented by rotating the patches with different orientations. The size of image in test set is above 5500×7000, and the entire airport image covers an area of 13–87 km^2. The number of aircrafts in test set is 2718. For the RSOD data set, 70% images are used for training and validation, rest are used for test. We adopted the same training-validation split criterion with the LSORS data set.

For a fair comparison, we used three scales with box areas of 32^2, 64^2, and 128^2 pixels, and three aspect ratios of 0.5, 1, and 2 to construct nine handcraft anchors for R-FCN [11] and R-P-Faster R-CNN [9]. The number of clusters using in K-means and OEM is also set to nine. The IoU t using in metadata-based mining is set to 0.7. In addition, we set $r_n = 7$, $\alpha_n = 7$ which is equal to the number of spatial grids in the position-sensitive RoI pooling layer [11]. All networks are initialized by a pretraining model learned from natural images, and then fine-tuned for 40k iterations with a momentum of 0.9, a weight decay of 0.0005, and a learning rate of 0.001.

In our experiments, the aircraft detection is considered to be correct if the IoU ratio between the detection box and the ground-truth box exceeds 0.5. We employ the precision, recall, and F1-measure to quantitatively evaluate the performance of our OE-FCN method. These metrics are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1} - \text{measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP, FP, and FN denote the number of true positives, false positives, and false negatives, respectively. The experiments were carried on the Ubuntu with an Intel(R) Core(TM) i7-6800 K CPU @ 3.40 GHz and a NVIDIA GeForce GTX 1080 Ti GPU.

B. Verification on Exemplar Mining and Circle Response Map

To evaluate the effectiveness of the OEM and the CRM, we make comparison on R-FCN [11], R-FCN using CRM, R-FCN with K-means anchors (KAs), R-FCN with OEM, and our OE-FCN. In R-FCN using CRM, CRM substitutes for position-sensitive score map. R-FCN with KA adopts K-means clustering on the training set to configure the scales and the aspect ratios of anchors. For R-FCN using OEM, it replaces the handcraft anchors with the OEM, which combines the metadata with high-level features to mine exemplars. All comparisons are validated on the LSORS data set.

TABLE I
Verification of Exemplar Mining and CRM

Method	R-FCN [11]	R-FCN +CRM	R-FCN +KA	R-FCN +OEM	Ours
Precision	71.5%	81.0%	84.4%	87.4%	92.0%
Recall	85.6%	90.7%	91.4%	92.4%	92.5%
F1-measure	77.9%	85.6%	87.7%	89.9%	92.2%

As shown in Table I, the R-FCN using handcraft anchors have a worse performance than that using KAs or OEM, since there is much more man-made interference caused by handcraft anchors setting. By introducing the high-level features into exemplar mining, both precision and recall are boosted by 3% and 1% in R-FCN using OEM, respectively. Actually, the anchor-based method can also be considered as a special case of exemplar mining. Anchors are generated to represent different exemplars only based on the statistics of the scales and the aspect ratios of the targets bounding boxes in the training set. However, the proposed method in this letter combines the labeled information with the high-level features of the targets to achieve more efficient exemplar mining and improve the performance of network training. Besides, comparing ours with the R-FCN using OEM, it proves that using the CRM to identify aircraft candidate regions is more accurate than the position-sensitive score map.

C. Comparison With the State-of-the-Arts

We compare the OE-FCN with two traditional state-of-the-art aircraft detection methods and two CNN-based methods: Adaboost using HoG [2], Adaboost using ACF [7], CNN with unsupervised score based bounding box regression [8], and R-P-Faster R-CNN with Visual Geometry Group-16 [9].

It can be shown from Table II and Fig. 4 that our method can achieve high accuracy and recall simultaneously, which is better than the other aircraft detection methods. Method [7] has the worst precision than other methods, since the sliding-window framework in [7] generates enormous spurious target regions and imposes a heavy burden on the classifier. Compared with the CNN-based methods, these two traditional methods perform poorly for the limited expressive capability of handcraft features. The reason why [8] has a lower recall is that the selective search method cannot extract plenty of aircraft regions, and it causes a great quantity of miss detection.

To avoid missing any aircraft regions, method [9] develops a region proposal network based on anchors principles. However, the anchor selection still relies on the prior information, such as scales and aspect ratios, which is unrobust to the intraclass variation of aircraft in RSI. Besides, the full-connected layer in [9] dramatically slows down the speed of aircraft detection. Instead of depending on the prior information of aircrafts, the proposed method in this letter combines the aircraft labeled information with the high-level features to automatically mine exemplars that represent different intra-class characteristics of aircrafts. In addition, it is also beneficial to enhance the adaptability of our network. Moreover, the full connection layer in aircraft identification is replaced by the CRM to speed up the aircraft detection and achieve higher accuracy.

TABLE II
COMPARISON OF THE DIFFERENT AIRCRAFT DETECTION METHODS

Dataset	LSORS dataset					RSOD dataset [8]				
Method	[2]	[7]	[8]	[9]	**Ours**	[2]	[7]	[8]	[9]	**Ours**
Precision	65.5%	30.8%	77.3%	76.8%	92.0%	85.1%	58.8%	85.4%	88.2%	95.6%
Recall	61.4%	75.1%	56.2%	91.0%	92.5%	59.6%	70.1%	65.6%	92.4%	93.0%
F1-measure	63.4%	43.7%	65.1%	83.3%	92.2%	70.1%	63.9%	74.2%	90.2%	94.3%
running time (s)	680.7	23.1	878.7	52.2	37.5	13.48	0.28	17.19	0.39	0.30

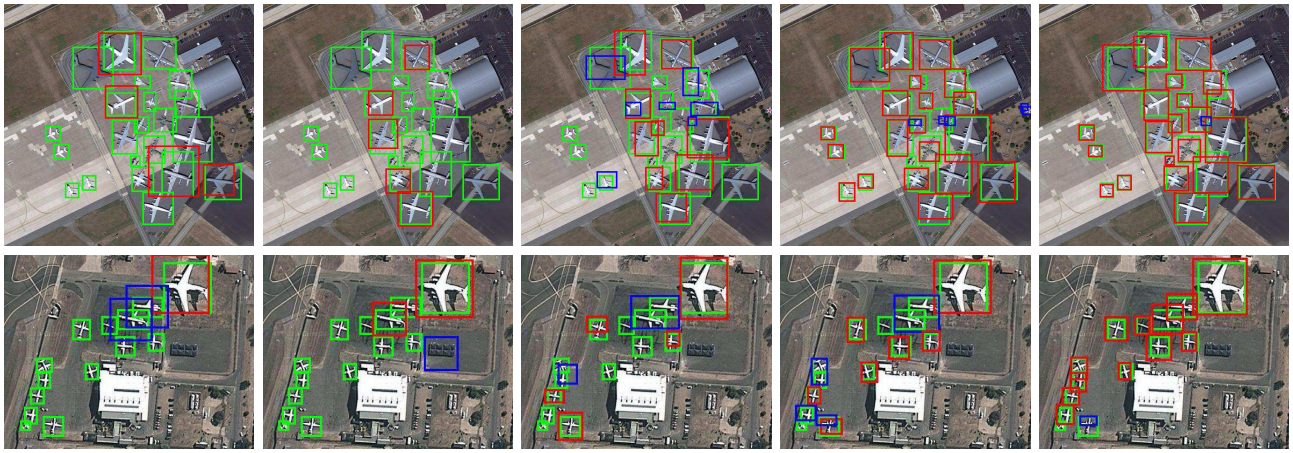

Fig. 4. Detection results of different methods. The true positives, false positives, and ground-truths are indicated by red, blue, and green rectangles, respectively. (a) [2]. (b) [7]. (c) [8]. (d) [9]. (e) Ours.

IV. CONCLUSION

In this letter, we have presented an end-to-end OE-FCN framework to tackle the intraclass variation of aircraft and achieve a rapid detection in RSIs. The online exemplar-mining mechanism is introduced into the CNN, which adopts the exemplars to represent different intraclass characteristics of aircrafts. In addition, according to the cross shape of aircrafts, the CRM is developed to construct our FCN. It accomplishes an accurate and speed-up aircraft detection. The experiments demonstrated that the OE-FCN framework could obtain an outstanding performance on aircraft detection in large-scale RSIs. In the future, we will further study an OEM based on weakly supervised learning, since the manual annotation in large-scale RSIs is generally expensive and unreliable for small targets.

REFERENCES

[1] G. Cheng *et al.*, "Object detection in remote sensing imagery using a discriminatively trained mixture model," *ISPRS J. Photogramm. Remote Sens.*, vol. 85, no. 9, pp. 32–43, 2013.

[2] Z. An, Z. Shi, X. Teng, X. Yu, and W. Tang, "An automated airplane detection system for large panchromatic image with high spatial resolution," *Optik-Int. J. Light Electron Opt.*, vol. 125, no. 12, pp. 2768–2775, 2014.

[3] W. Zhang, X. Sun, H. Wang, and K. Fu, "A generic discriminative part-based model for geospatial object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 99, pp. 30–44, Jan. 2015.

[4] F. Zhang, B. Du, L. Zhang, and M. Xu, "Weakly supervised learning based on coupled convolutional neural networks for aircraft detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5553–5563, Sep. 2016.

[5] 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.

[6] W. Diao, X. Sun, X. Zheng, F. Dou, H. Wang, and K. Fu, "Efficient saliency-based object detection in remote sensing images using deep belief networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 137–141, Feb. 2016.

[7] A. Zhao *et al.*, "An effective method based on ACF for aircraft detection in remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 744–748, May 2017.

[8] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.

[9] X. Han, Y. Zhong, and L. Zhang, "An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery," *Remote Sens.*, vol. 9, no. 7, p. 666, 2017.

[10] Y. Zhong, A. Ma, Y. S. Ong, Z. Zhu, and L. Zhang, "Computational intelligence in optical remote sensing image processing," *Appl. Soft Comput.*, vol. 64, pp. 75–93, Mar. 2018.

[11] 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.

[12] Y. Zhong, F. Fei, Y. Liu, B. Zhao, H. Jiao, and L. Zhang, "SatCNN: Satellite image dataset classification using agile convolutional neural networks," *Remote Sens. Lett.*, vol. 8, no. 2, pp. 136–145, 2017.

[13] X. Han, Y. Zhong, L. Cao, and L. Zhang, "Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification," *Remote Sens.*, vol. 9, no. 8, p. 848, 2017.

[14] B. Zhao, B. Huang, and Y. Zhong, "Transfer learning with fully pretrained deep convolution networks for land-use classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1436–1440, Sep. 2017.

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

[16] D. Sculley, "Web-scale k-means clustering," in *Proc. ACM 19th Int. Conf. World Wide Web*, 2010, pp. 1177–1178.