Airport Detection on Optical Satellite Images Using Deep Convolutional Neural Networks

Peng Zhang, Xin Niu, Yong Dou, and Fei Xia

***Abstract*— This letter proposes a method using convolutional neural networks (CNNs) for airport detection on optical satellite images. To efficiently build a deep CNN with limited satellite image samples, a transfer learning approach had been employed by sharing the common image features of the natural images. To decrease the computing cost, an efficient region proposal method had been proposed based on the prior knowledge of the line segments distribution in an airport. The transfer learning ability on deep CNN for airport detection on satellite images had been first evaluated in this letter. The proposed method was tested on an image data set, including 170 different airports and 30 nonairports. The detection rate could reach 88.8% in experiments with seconds' computation time, which showed a great improvement over other the state-of-the-art methods.**

***Index Terms*— Airport detection, convolutional neural network (CNN), line segment detector (LSD), transfer learning.**

I. Introduction

AUTOMATIC airport detection in optical satellite images is a useful application in civil aviation planning [1]. However, airport detection remains a difficult work due to the complex target structures with highly changing surroundings.

As many other detection tasks, effective target features representation is also crucial in airport detection. According to the features extracted by the handcrafted or machine-learned approaches, previous studies of airport detection can be classified into two categories.

Handcrafted features, which attempt to represent the distinctive trait of the target by the expert knowledge, could present well proposals for further detections. For instance, the long straight runways were frequently explored as an efficient feature for airport detection [2], [3]. Nevertheless, such linear features still cause the confusions with rivers and roads. The weakness of the handcrafted features, including edge [4] and textures [5], lies in the difficulties to find proper representations through uncertain experience and limited expert knowledge. These features often represent targets in finite aspects.

Manuscript received July 24, 2016; revised December 6, 2016; accepted January 3, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61402507, Grant U1435219, and Grant 61572515, in part by the China Postdoctoral Science Foundation under Grant 2016M593023, and in part by the Natural Science Foundation of Hubei Province under Grant 2016CFB393.

P. Zhang, X. Niu, and Y. Dou are with the National Laboratory for Parallel and Distributed Processing, National University of Defense Technology, Changsha 410073, China (e-mail: niuxin@nudt.edu.cn).

F. Xia is with the Institute of Electronic Information Warfare, Electronic Engineering College, Naval University of Engineering, Wuhan 430033, China.

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.2017.2673118

On the other side, machine-learning approaches have shown great potentials to generate efficient image features. Especially, the success of deep convolutional neural network (CNN) in recent years has opened up prospects for superior image classification and detection [6]. In comparison with other conventional handcrafted or machine-learning approaches [7], [8], deep CNN could learn rich highly abstract image features to represent complex objects effectively. However, to obtain an effective deep CNN model, a large number of labeled data are usually required. For airport detection, however, acquisition of adequate labeled data from limited optical satellite images is often difficult. Nevertheless, optical satellite images still share common visual features in low and middle abstract level with natural images, which captured by general cameras. This phenomenon lays the foundation to explore the midlevel features of a CNN well trained from large natural image sets for airport detection using optical satellite data through proper transfer learning approach [9]. Besides the training difficulty, high computation expense makes another problem when applying a deep CNN model. Hence, efficient region proposal approaches are required for an efficient detection system. To this end, expert knowledge, such as the mentioned handcrafted runway indicators, could help.

Therefore, this letter introduces an airport detection approach from optical satellite images using deep CNN with an efficient proposal method by handcrafted features. The main contributions can be found in three aspects: 1) we proposed a robust method to detect airports based on the combination of hand-designed features and CNN features; 2) we proposed an efficient line-based region proposal approach; and 3) the transfer ability of CNN trained on nature images for optical satellite applications has been verified. To the best of our knowledge, this could be one of the first attempts to apply CNN in airport detection. The rest of this letter is organized as follows. In Section II, we give the methodology details. Experiments and analysis are shown in Section III. The conclusion is made in Section IV.

II. Methodology

The proposed method is formed by mainly three steps, namely, region proposal, CNN identification, and localization optimization, as shown in Fig. 1. All the tested optical satellite images were collected from Google Earth with the resolution of 8 m $\times$ 8 m and the size of about 3000 $\times$ 3000.

To reduce unnecessary computation cost and potential misclassification with deep CNN, we first restricted the detection area to limited region proposals. The proposals were prepared by searching the airport runways using a

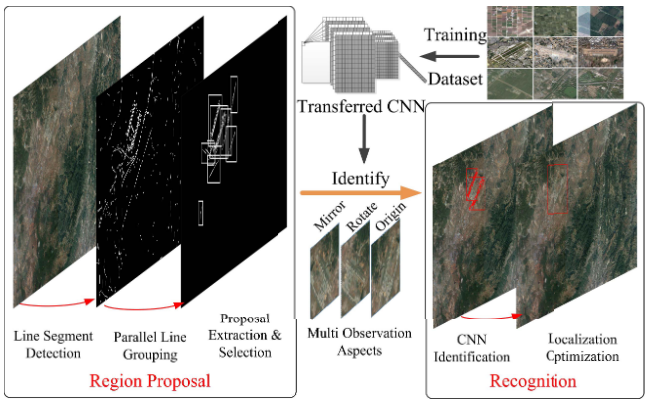

Fig. 1. Flowchart of the proposed method.

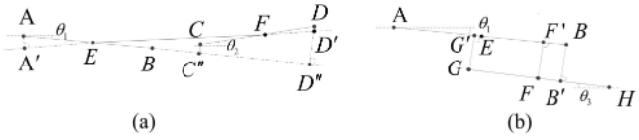

Fig. 2. Demonstration of the two to-be-linked line segments AB and CD in (a) and the two parallel adjacent line segments in (b).

hand-designed line detection approach. Second, a transfer learned CNN was employed to score the region proposals to identify the most possible airport patches. Finally, a location strategy was applied to indicate the optimal airport area based on the clusters of the CNN scored regions. Details of the above-mentioned steps are given in Sections II-A–II-C.

A. Region Proposal Through Line Segment Detection

The criterion of searching the region proposals is based on the prior knowledge that an airport commonly contains long straight parallel runways. Therefore, regions with long parallel lines were selected as the detection candidates. To this end, we first used an advanced line segment detector (LSD) [10] to obtain primitive line segments. Then, a linking process was applied to restore the truncated long straight line segments. Finally, we grouped neighboring parallel line segments and squared the spatial extent of each group as one region proposal. Existence of runways in a proposal was estimated according to the spatial relationship of the grouped lines.

1) Line Segment Detection and Linking: LSD has been widely used to detect line segment with low false alarm and high detection speed. However, as many other edge detection approaches based on image gradient, linear structures with intersections may be detected as a series of truncated line segments. To better indicate the airport runways, we proposed a linking approach to restore long straight line segments.

As shown in Fig. 2(a), given two line segments AB and CD and their middle points E and F, the restored line is marked as A′D′, where A' and D' are the projections of A and D on EF. Supposing AB is the initial line segment, and the tilt angles of AB and CD are θ_1 and θ_2, the linking rules are defined as follows:

$$\begin{cases} |\theta_1 - \theta_2| < T_{\text{angle}} \text{ or } |\theta_1 - \theta_2| < \pi - T_{\text{angle}} \\ 0 < |\text{BC}''| < T_1 \text{ and } 0 < |\text{CC}''| + |\text{DD}''| < T_2 \end{cases} \quad (1)$$

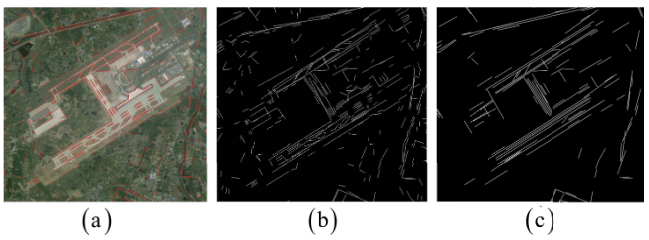

Fig. 3. Results of line segment detection and linking. (a) LSD. (b) Linking method in [8]. (c) Proposed linking method.

where T_{angle}, T_1, and T_2 are thresholds. Thresholds in the linking rules (1) and the following grouping rules (2) were estimated by the statistics of the airport features from the experimental images with specific resolution. Specifically, the angle difference of the linked lines should be less than $T_{\text{angle}}(5°)$. The smaller T_{angle} is, the fewer truncated lines will be involved in restoration and the smaller line groups will be formed. Besides, adjacent linked segments are supposed to be closed in a range [$< T_1(45)$]. $T_2(4)$ guaranteed the restored lines not to deflect too much from the linked lines. Overall, the linking rules ensure that the linked lines are close and collinear.

The linking process greedily merges line segment pairs in iteration. In each iteration, an initial line AB is selected from the current line list. AB grows by continually merging a line segment until there is no segment meeting the linking rules with current AB. Each time, one merging candidate CD is selected to merge with AB. CD should meet the linking rules and lead to the smallest crossing angle between AB and the restored A′D′(or the growed AB). By this linking process, the angle difference between a restored line and all its linked line segments can be controlled within a given limit. Therefore, the linked long straight line could better represent the real edge of runways (Fig. 3).

2) Grouping of Local Parallel Line Segments: According to the prior knowledge that airports runways are usually parallel and close to each other, we attempt to represent a potential airport region by the spatial extent of a local parallel line group. As shown in Fig. 2(b), the grouping rules can be summarized as follows:

$$\begin{cases} |\theta_1 - \theta_3| < T_{\text{angle}} \text{ or } |\theta_1 - \theta_3| > \pi - T_{\text{angle}} \\ V_{\min} < |\text{BB}'| < V_{\max} \text{ and } V_{\min} < |\text{GG}'| < V_{\max} \\ 0 < |\text{GB}'| \le |\text{GH}| \text{ and } 0 < |G'B| \le |\text{AB}| \end{cases} \quad (2)$$

where AB is a line segment in a group and GH is a candidate line to be added in. |BB′| and |GG′| are the vertical distances from AB to GH. The $V_{\min}(2)$ is set to exclude lines detected from narrow roads but not runways and $V_{\max}(60)$ gives the empirical upper limit of the width of the runway array in a airport. Three rules in (2) guarantee the parallelism and vicinity of AB and GH, and the last rule restricts them only to be abreast. If GH meets the grouping rules with AB, the grouping algorithm adds GH into AB's group.

The groups are formed one by one. Each group starts from an initial line randomly selected from the ungrouped lines. A group grows with adjacent lines and ends to a certain spatial extent. To avoid confusing proposals of urban and agriculture

areas which consist of short parallel lines, segments with the length under 60 will not to be selected as the initial lines.

3) Selection of Region Proposals: Specifically, an airport region proposal was given by the minimum bounding rectangle of a line group. To measure the possibility of a region proposal from the perspective of the grouped line distribution, we have defined a proposal score as follows:

$$\text{Score}^{k}_{\text{proposal}} = \sum_{\substack{i,j\in R_k \\ i\neq j}} (W(|\text{CL}_{i,j}|)\times\exp(-\text{Dis}(i,j)^2)/(2\sigma^2)) \tag{3}$$

$$W(|\text{CL}_{i,j}|) = |\text{CL}_{i,j}| \times (f(|\text{CL}_{i,j}|) + 2\times|\text{CL}_{i,j}|/(|i|+|j|) \tag{4}$$

where R_k is the kth region proposal and L_i and L_j are the ith and jth line segments in R_k. $\text{CL}_{i,j}$, which refers to G′B in Fig. 2(b), is the abreast part of two lines. The function f in Fig. 4 weights the length of $\text{CL}_{i,j}$. The higher $W(|\text{CL}_{i,j}|)$ is, the more likely two line segments are from the same linear structure such as a runway. $\text{Dis}(i, j)$ is the distance from L_i to L_j. If the midpoint of line L_j can be projected on L_i, $\text{Dis}(i, j)$ is the vertical distance of two parallel lines [FF′ in Fig. 2(b)]. Otherwise, $\text{Dis}(i, j)$ is the distance between the two midpoints of L_i and L_j (EF). Therefore, regions with similar long abreast lines would get higher score. Based on the experimental results, it was effective to eliminate most of the wrong proposals when setting the threshold of the proposal scores as 280. The proposal score will be used with the predication of CNN together to determinate the final airport locations.

B. Identification by Transferred CNN

To explore the natural image trained deep CNN for the optical satellite image detection, the widely used successful model AlexNet [6] has been used as the transfer source. To fine-tune this pretrained model, we have selected thousands of samples from 11 classes, including airport, runways, farm, brown land, green land, ridge, town, sea, land-and-water, roads, and rivers. By the argumentation of eight directions rotation and spatial translation, the total sample number had been expanded to 19 000. Samples of airports were cut from 40 scenes out of the totally 170 airport images with their real squared size. All the other samples were prepared in square patches with the size of 227 × 227 in pixels in the same scale. With the lower layers of the pretrained CNN fixed, we fine-tuned the top fully connected layers (FLs) by the training samples. In this way, we attempt to rebuild the high level features for airport detection by exploring the common low and middle level features. Specifically, to fine-tune the pretrained AlexNet from an FL, all the layers above FL were removed. A softmax classifier [11] with the output classes defined in our task was connected to FL. Then, the AlexNet was retrained with our labeled samples by the backpropagation approach [12].

Although with limited satellite samples, the superior classification performance of the transferred CNN in comparison with other conventional methods was observed in the experiments. Since the region proposal was the minimum bounding rectangle of the potential runways, the proposal usually only occupies the main part of the airport excluding lots of surrounding info. Thus, in the CNN detection, we have expanded the proposals by empirically 100 pixels (with the image resolution of 8 m × 8 m in the experiments) in four directions first. To avoid the distortion of data stretching, we kept the test samples as image patches collected from the minimum bounding squares of the proposals to fit the square shape of the training data.

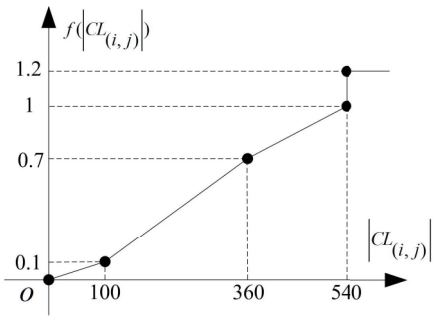

Fig. 4. Function f in (4).

To achieve a robust detection, we also collected transformations on the original test patch include rotation and mirroring to evaluate the possibility of the location from multiviews. The average scores of the original patch and its transformations were taken as the detection score of the region proposal by CNN.

C. Localization Optimization Based on CNN Results

Although the transferred CNN could bring well-trained low and middle level image features as the foundation for the further fine-tuning, due to the limited satellite image samples (especially the airport samples), particular airports, such as the smaller one with grass or within the city area, still have the risk to be recognized as farmland or urban by CNN. Moreover, some large airports could be identified as many separate smaller ones due to the imperfect proposals by the unrestored long straight lines. Therefore, we proposed a localization optimization approach combining the CNN predication and the prior possibility based on the grouped line distribution.

For localization, region proposals, which have been identified by CNN as airport or runways with CNN score higher than 0.7, have been selected as the detection candidates. To restore the complete airport region, overlapped candidate regions with colinear lines or intersecting lines were merged with the minimum bounding rectangle and the maximum CNN score of these merged candidates was assigned to the expanded region. With the proposal score and the CNN score, the final predication of the airport location, marked as R in (5), can be decided with the following rules:

$$R = \begin{cases} \{k|\text{Score}^{k}_{\text{CNN}} > 0.98\}, & \\ & \text{if } \text{Score}^{\max}_{\text{CNN}} > 0.98 \\ \{k|\text{Score}^{k}_{\text{proposal}} = \max\limits_{k\in K}\text{Score}^{k}_{\text{proposal}}\} & \\ & \text{where} \\ K = \{k|\text{Score}^{k}_{\text{CNN}} - \text{Score}^{\max}_{\text{CNN}} < 0.05\} & \\ & \text{else} \end{cases} \tag{5}$$

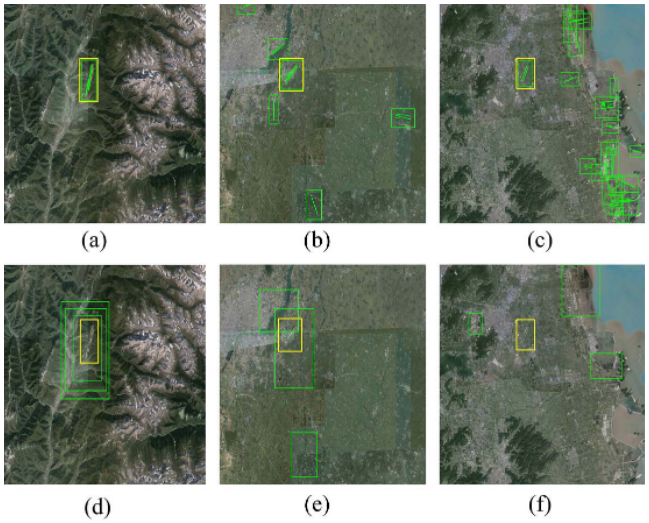

Fig. 5. Comparison of region proposals. (a)–(c) Results of our method. (d)–(f) Region proposals obtained by [8]. (Yellow boxes) True airport locations. (Green boxes) Region proposals.

where $\text{Score}_{\text{CNN}}^{k}$ is the CNN airport classification score of the kth candidate region and $\text{Score}_{\text{CNN}}^{\max}$ is the maximum CNN airport classification score over all the merged candidates. If there are no regions meeting the above-mentioned conditions, the detection algorithm reports no airport on this image.

III. Experiments

The proposed method was tested on the GoogleEarth images. The data set contains most of the civil airports in China, including several famous international airports. There were totally 170 images containing airports and 30 images with no airports. The computing platform consisted of a CPU of Intel i7-4790K and a GPU of NVIDIA GTX780 with 3-GB video memory. The algorithm was compiled in VS2012.

In our experiments, an airport is considered as being covered by a proposal if the INT-UION [13] score is no less than 0.5. The detection rate (DR) and the false alarm rate (FAR) are calculated by the method in [8].

A. Effect of Region Proposal

To demonstrate the effect of the proposed region proposal approach, a comparison has been made with a saliency-based approach, which was also used in the optical image airport detection.

As shown in Fig. 5, proposals generated by [8] are relatively fewer than the proposed method. However, the proposed method could generate more accurate proposal in both location and the extension. Furthermore, the nested proposals could be avoided by our method.

B. Capability of Transferred CNN

Our CNN detection treated all the training samples and the test regions in regular squared shape. The superiority of this transfer learning strategy will be demonstrated in Section III-B1 in comparison with that using irregular sample shape but stretching the data to square to fit the CNN input. Detection strategy of fusion of multiobservation aspects was also compared with that by the single observation aspect. Since different layer levels of the pretrained CNN have different transfer ability, we also conducted a comparison to check the classification performances using different fully connected top layers.

TABLE I

Comparisons of Results Based on Different Training Data and Different Observation Aspects

Methods	Detection Rate (DR)	False-Alarm Rate (FAR)
Single/Non-square	0.86	0.7
Single/Square	0.86	0.7
Multi/Square	0.89	0.53

TABLE II

Comparisons on Performance of AlexNet With Different Full-Connected Layers Learned

Full-connected Layers Learned	Convergence Iterations	Error Rate
FL1	66	0.139
FL2	27	0.051
FL3	16	0.039

1) Influence of Stretching Data: The input of AlexNet is a fixed image square with the size of 227 × 227. However, the shapes of real airports are usually irregular rectangles. In the training or recognizing steps, the image samples need to be stretched to such fixed sizes and the shapes of airports may be largely distorted. In this way, the pretrained AlexNet may report wrong prediction. To test the influence of the data stretching using irregular shaped samples, two training sets were prepared. The patches of airports in one of training sets are with the same sizes to the real airports. The patches in another are collected from the minimum bounding squares of airports. As shown in Table I, improvement by the fixed square sample strategy was not significant. But better airport location could be achieved by this strategy based on the detection results. In the following experiments, we still used such fixed sample shape strategy.

2) Fusion of Multiobservation Aspects: To demonstrate the robustness of detection using multiobservation aspects, we compared the detection on the original region proposal and that by the average of the detections on the original region proposal, the 90° rotation and the mirror of the same proposal. As shown in Table I, the multiobservation aspects detection could slightly increase DR and decrease the FAR. Therefore, in the following experiments, such multiobservation aspects strategy was applied.

3) Transferable Ability on Different CNN Layers: As mentioned earlier, reconstruction of the high level features which mostly represented by the top FLs were the focus in our CNN transfer learning. In AlexNet, there are three full-connected layers, named as FL1, FL2, and FL3 from low to high level toward the top of the network. To test the transferable ability of these layers, we compared the classification results on the target and background classes by fine-tuning the AlexNet from FL1 to FL3. As shown in Table II, with the higher level layer, more accurate classification performance could be achieved. In the following experiments, we used the FL3 fine-tuned AlexNet for comparisons.

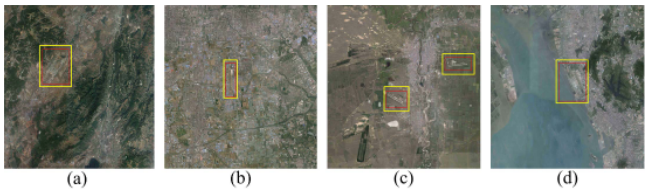

Fig. 6. Detection results by our method in (a) mountains, (b) and (c) cities, and (d) by the sea. In (c), two airports were detected correctly. Yellow boxes: true airport locations. Red boxes: detection results.

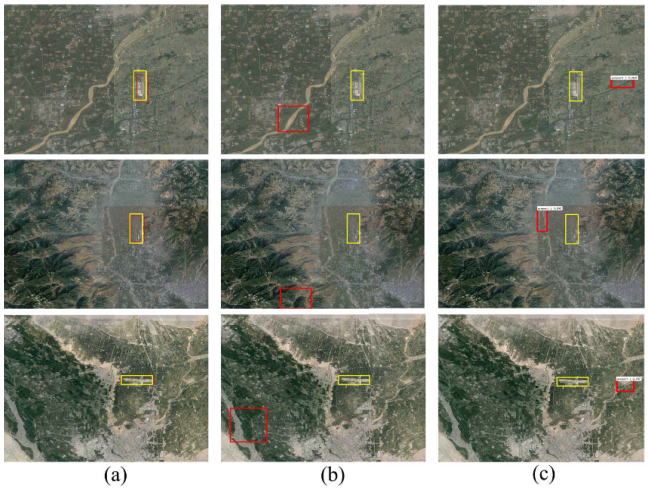

Fig. 7. Comparison with different methods in three scenes (a column as a group). (a) Our method. (b) Method in [8]. (c) Method in [14]. Yellow boxes: true airport locations. Red boxes: detection results.

C. Comparisons With State-of-the-Art Methods

To demonstrate the efficiency of our proposed method, the state-of-the-art target detection approaches [8], [14] for optical data have also been compared.

The method in [8] explored the top–down and bottom–up saliency to extract proposal regions and used the handcraft scale-invariant feature transform (SIFT) features for support vector machine classification. Due to the widely dispersed saliency maps, the detected areas were frequently larger than real airports in complex scenes. Moreover, the effects of the saliency varied according to the image quality. Besides, SIFT and texture features were often observed not robust enough with changing conditions. In comparison, faster regions with CNN (RCNN) [14] extracted and recognized proposals by the shared CNN features. It could be noted that by only using CNN, the faster RCNN proposals were often not as great as expected. However, due to the features extraction strategy with region-of-interest pooling [15], the speed of faster RCNN was remarkable.

As shown in Table III, by combining region proposal based on prior knowledge and transfer learned CNN, the proposed method could significantly improve the DR with limited time consuming (about 20 s). Selected results of our proposed method are shown in Fig. 6.

Comparison examples with other methods are shown in Fig. 7. Generally, linear structures formed the major confusion source for the method [8], [14]. It could be noted that regions with river or road were often misclassified by these methods.

TABLE III
COMPARISONS OF DIFFERENT METHODS

Methods	DR	FAR	Time(s)
Proposals+Texture [4]	63.1%	50%	68.64
Proposals+SIFT	64.1%	50%	64.09
TD+BD+SIFT [8]	53.5%	56.7%	8.57
TD+BD+CNN	69.4%	36.7%	3.85
Faster RCNN [14]	67.6%	70%	0.076
Proposals+CNN(Ours)	**88.8%**	**53.3%**	**21.3**

Mountain ridges and farmlands caused many confusions for them as well.

IV. CONCLUSION

This letter proposed a method of airport detection combining transfer learned CNN and region proposal with prior knowledge. Due to the line-based region proposal, the computation of the CNN detection could be greatly reduced. By transfer learning on limited satellite samples, deep CNN could recognize the airports in high accuracy. Experimental results showed that our proposed method could detect airports with high accuracy on complex area in seconds.

REFERENCES

[1] C. Tao, Y. Tan, H. Cai, and J. Tian, "Airport detection from large IKONOS images using clustered SIFT keypoints and region information," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 128–132, Jan. 2011.

[2] Z. Li, Z. Liu, and W. Shi, "Semiautomatic airport runway extraction using a line-finder-aided level set evolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4738–4749, Dec. 2014.

[3] W. Wu *et al.*, "Recognition of airport runways in FLIR images based on knowledge," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 9, pp. 1534–1538, Sep. 2014.

[4] Y. Qu, C. Li, and N. Zheng, "Airport detection base on support vector machine from a single image," in *Proc. IEEE 5th Int. Conf. Inf. Commun. Signal Process.*, Dec. 2005, pp. 546–549.

[5] O. Aytekin, U. Zongur, and U. Halici, "Texture-based airport runway detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 471–475, May 2013.

[6] 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.

[7] X. Wang, Q. Lv, B. Wang, and L. Zhang, "Airport detection in remote sensing images: A method based on saliency map," *Cognit. Neurodyn.*, vol. 7, no. 2, pp. 143–154, 2013.

[8] D. Zhu, B. Wang, and L. Zhang, "Airport target detection in remote sensing images: A new method based on two-way saliency," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1096–1100, May 2015.

[9] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1717–1724.

[10] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "LSD: A fast line segment detector with a false detection control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 722–732, Apr. 2010.

[11] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[12] D. E. Rumelhart, "Leaning internal representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 318–362, 1986.

[13] M. Everingham *et al.* (2008). *The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results*. [Online]. Available: http://pascal-network.org/

[14] 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.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 346–361.