

PAPER • OPEN ACCESS

Fast Detection of Airports on Remote Sensing Images with Single Shot MultiBox Detector

To cite this article: Fei Xia and HuiZhou Li 2018 *J. Phys.: Conf. Ser.* **960** 012024

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

You may also like

- [Is It Small-scale, Weak Magnetic Activity That Effectively Heats the Upper Solar Atmosphere?](#)

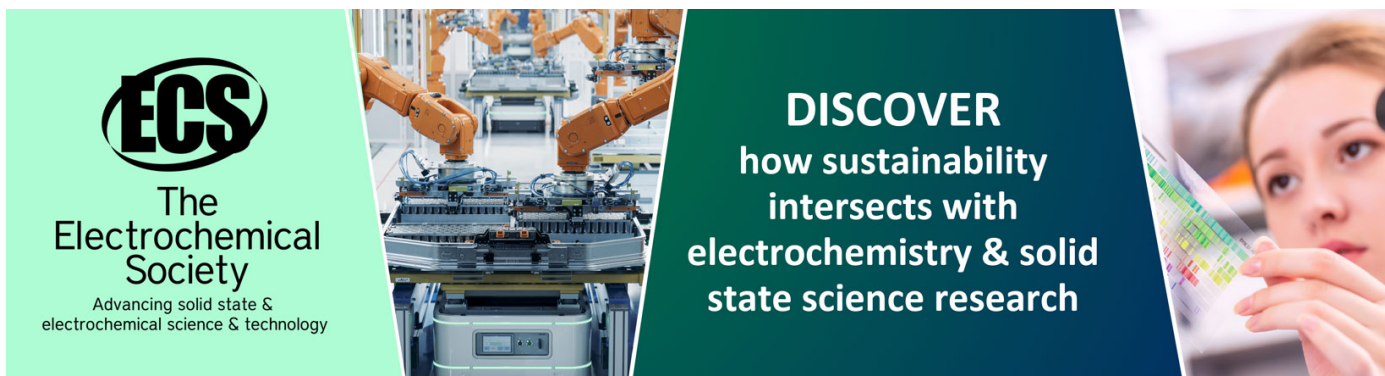
K. J. Li, J. C. Xu and W. Feng

- [Speeded-Up Robust Features-based image mosaic method for large-scale microscopic hyperspectral pathological imaging](#)

Qing Zhang, Li Sun, Jiangang Chen et al.

- [The interplay of policy and energy retrofit decision-making for real estate decarbonization](#)

Ivalin Petkov, Christof Knoeri and Volker H Hoffmann



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

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Fast Detection of Airports on Remote Sensing Images with Single Shot MultiBox Detector

Fei Xia¹ and HuiZhou Li^{1,*}

¹Institute of Electronic Information Warfare, Naval University of Engineering, Wuhan, China

*xycphoenix@nudt.edu.cn

Abstract. This paper introduces a method for fast airport detection on remote sensing images (RSIs) using Single Shot MultiBox Detector (SSD). To our knowledge, this could be the first study which introduces an end-to-end detection model into airport detection on RSIs. Based on the common low-level features between natural images and RSIs, a convolution neural network trained on large amounts of natural images was transferred to tackle the airport detection problem with limited annotated data. To deal with the specific characteristics of RSIs, some related parameters in the SSD, such as the scales and layers, were modified for more accurate and rapider detection. The experiments show that the proposed method could achieve 83.5% Average Recall at 8 FPS on RSIs with the size of 1024*1024. In contrast to Faster R-CNN, an improvement on AP and speed could be obtained.

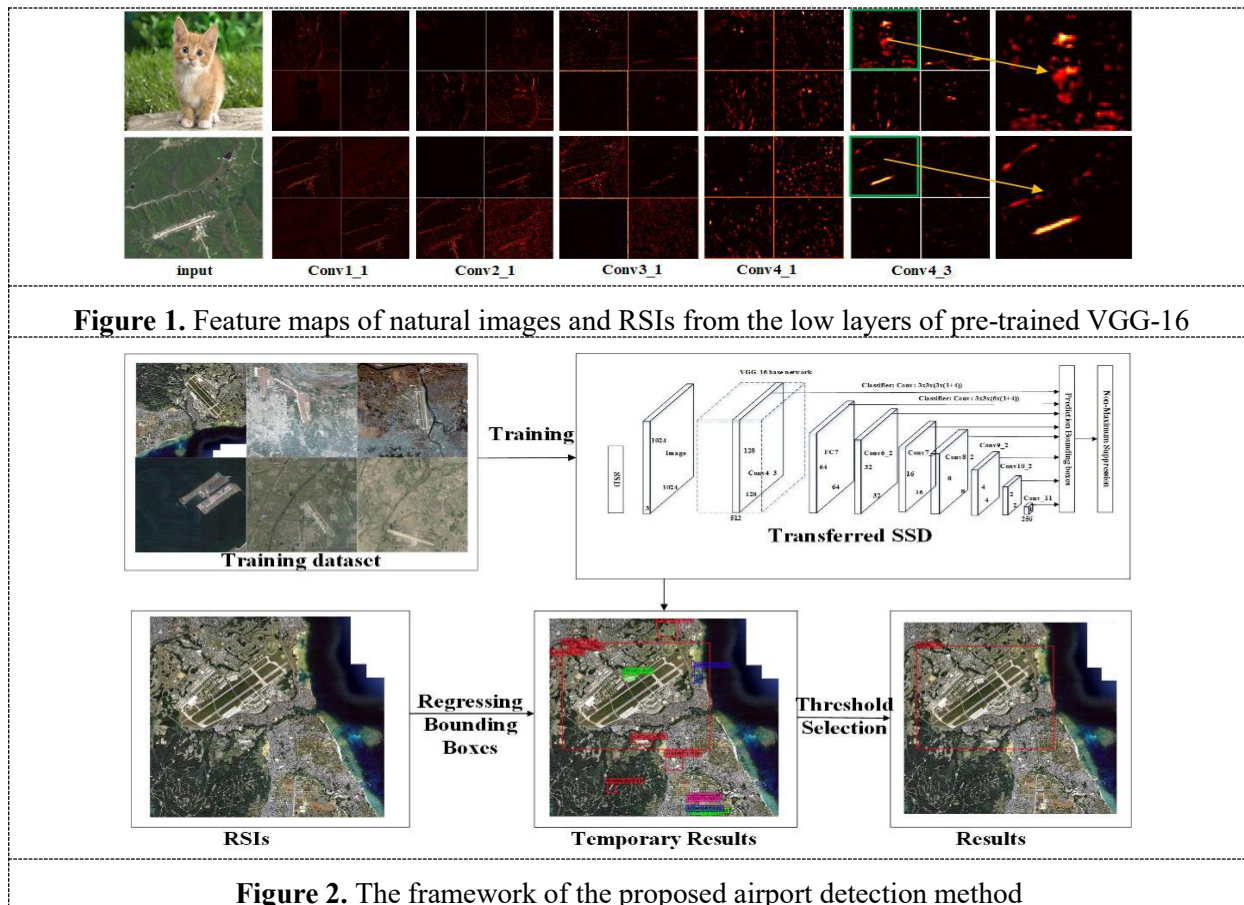
1. Introduction

Object detection on remote sensing images plays a more and more important role in our lives, e.g. urban planning, disaster prevention, etc. Bridges, ships and aircrafts are becoming significant objects in these applications. With great developments of remote sensing technologies in recent years, airport detection has caused widespread concern due to its civil and military value. Simultaneously, the various structures of airports and complex backgrounds make airport detection challenging [1].

The usual airport detection methods are composed of two steps: region proposal generation and feature recognition. In both steps, airport features influence the quality of region proposals and the performance of airport recognition directly. Besides, these two steps are processed independently. As a result of different features used in two steps, the airport could not be recognized exactly without accurate region proposals and computing different features used in these two steps cost too much. Only using one kind of robust feature and making two steps cooperate are helpful to improve the airport detection performance.

As for the features used in previous works on airport detection, they can be divided into two classes: hand-designed features and machine-learning features. hand-designed features are some designed feature descriptors based on the nature of images and the expert knowledge, such as lines, textures, scale-invariant feature transform (SIFT [2]) and so on. For example, many methods extracted the long straight lines as the runways to generate the region proposals for further detection [3, 4]. However, rivers and roads produced wrong proposals due to the same linear features. It is the same with other hand-designed features because the target object is hard to exactly describe from limited knowledge and finite angles. As a result of many differences between the human perception and the potential feature, the hand-designed features may decrease the detection performance when the application scene changed.





On the other hand, machine learning features make good use of some prior knowledge, automatically learning low-level and high-level image features that humans do not know how to represent. Since 2012, due to the success of deep CNN Alexnet [5], deep learning has shown up the excellent potential for superior image classification and object detection. However, large amounts of labeled data are required to train an ideal CNN model while the annotated RSIs are in scarce. Based on the common features between RSIs and natural images, transferring the feature representation of a CNN model trained on natural images to airport detection in RSIs is feasible. The method [6] used the Alexnet to extract and recognize the airports achieved a good performance. However, the existing airport detection methods using CNN still separate the region proposal generation and feature recognition.

As for object detection algorithms using CNN on natural images are consisted of two aforementioned steps. The methods of generating region proposal are BING [7], EdgeBox [8] or Selective Search [9], that all of them spent high computation to process large-sized RSIs without the guarantee of achieving high-quality proposals. Appeared with SPP-NET [10] and the subsequent Faster R-CNN [11], region proposal generation with hierarchical convolutional layers and ROI pooling for decreasing the computation of each region proposal become the mainstream. However, the two steps in these methods are still independent. SSD [12], the state-of-the-art detection algorithm, used a single deep neural network and generated object positions and categories from the network directly. SSD combined two separate steps into one and was faster than YOLO [13] with high performance. In the paper, SSD was employed in airport detection on RSIs.

SSD used a group of default boxes, which are in different scales and aspect ratios, to discretize the output space of bounding boxes for per feature map location by a fully convolutional network. During the airport detection task, the structures of airports are flexible and the sizes are various. By combining the feature maps attained from different layers at different scales, we can achieve the fixed-length airport features with different scales without considering the airport sizes. SSD was firstly designed to

detect objects in natural images and some modifications were necessary to deal with the differences between RSIs and natural images. At first, the size of remote sensing images is much larger than the natural images, while RSIs usually covers a wide area and the natural images only covers a small place. Moreover, the proportion the target took up in natural images is much bigger than RSIs. Secondly, the remote sensing images, which were taken from a very high altitude point along with various uncertainties, contain complex backgrounds with the illumination intensity changing. Nevertheless, the low-level visual features in RSIs were in common with natural images and the pre-trained VGG-16 was used in transferred SSD.

As far as we know, it may be the first study which introduced SSD into airport detection on RSIs. The paper proceeds as follows. We detailed the airport detection method in Section II. Section III shows some experiments and analyses. In section IV, we make the conclusion.

2. Methodology

The proposed airport detection method used the convolutional network to produce the bounding boxes of airports and scores directly. Based on SSD, some auxiliary layers were added to process RSIs with large sizes, without cutting slices. Benefiting from the pre-trained neural model VGG-16 [14] and the end-to-end way of detection, the airport features extracted were robust enough and the detection is fast.

2.1. Framework

Based on the pre-trained model VGG-16, the model could extract stronger features from airports than training a new one. Training a deep CNN model effectively is to compute the suitable values for millions of parameters, which depended on a big amount of annotated data for training. Though the data of RSIs is abundant, the annotated images are in scarce. Because of that RSIs are usually very large and the targets (airports) are much smaller than the image itself, manual labeling costs very much work. It is a fact that the rich low and middle level feature representation learned by a trained convolutional neural networks could be transferred to other visual recognition tasks [15]. Figure 1 contained several visual feature maps generated by the convolution layers from *conv1_1* to *conv4_3* from a RSIs and a natural image with the pre-trained VGG-16 network. Based on the hypothesis that RSIs and natural images share plenty of common low and middle level visual features, we transferred the object detection framework SSD and VGG-16 from natural images to RSIs domain. Figure 2 shows the transferred airport detection framework. The transferred SSD architecture built on a normal feed-forward CNN. And the feature representations of low and middle level are generated by the pre-trained VGG-16 network, which was called as *Base network*. Then a small amount of annotated RSIs was used to finetune the initial CNN and estimate the extra added parameters, that would guide the model learn better features of airports.

2.2. Modification for large input

In the structure of VGG-16, the convolution layers were kept and the full-connected layers were re-implemented with convolution. Therefore, both *fc6* and *fc7* would produce feature maps with specific scales. In the original SSD, extra convolutional layers were added to generate the coordinates and confidences of bounding boxes. The sizes of each extra feature maps, which decreased with the increasing number of convolution layers, corresponded to different scales of the original RSIs. The fixed structure of network decided the sizes of input images. As for SSD, there are two different network structures processing the input size of 300 and 512 respectively. SSD-512 contained an extra *Conv_10* branch, which guaranteed that the last layer of feature maps is in the size of 1x1.

In our method, the *Conv_11* branch, consisting of convolution layers and 'PriorBox' layer, was added to process larger size of input images (1024). Without changing the other layers, the *Conv_11* branch took the output of *Conv_10* as input and the results were combined with others for regression. The parameters of *Conv_11*, such as *kernel*, *pad*, was depended upon the size of the input image. The *min_size*, *max_size*, *step* were decided by the target size and we needed to add the extra *aspect ratio* in the network manually. It is different for the reception filed of feature maps in each layer. Combining these feature maps from different layers can help detect the multi-scale targets. As for each layer after

Conv4_3 (included), SSD used several convolutional filters with the size of 3x3 to generate a group of detection predictions with fixed number, including the offsets of the bounding boxes and the scores for each class. Airports in RSIs usually varied in size and structure. Figure 4 shows the statistic information of the airport lengths in RSIs, which indicates the distribution feature of the airport scales. The default aspect ratio need not be changed for airport detection. As seen from Figure 2, there are seven convolution layers for predicting the bounding boxes, called *Conv4_3*, *Conv6 (FC6)*, *Conv7 (FC7)*, *Conv8_2*, *Conv9_2*, *Conv10_2* and *Conv11_2*, which represent seven features in different scales. In the end, the step called non-maximum suppression (NMS) removed the redundant boxes and achieved the final detections.



Figure 3. Four detection samples by transferred SSD where airports were in mountains and cities.

Table.1 Evaluation results of transferred SSD and Faster R-CNN on training dataset.

Method	AP	Time (s)
Transferred SSD (8m GSD, Input 1024)	0.835	0.11
Transferred SSD (16m GSD, Input 512)	0.796	0.06
Transferred SSD (24m GSD, Input 512)	0.771	0.07
Faster R-CNN (8m GSD, Input 1024)	0.806	0.52s
Faster R-CNN (16m GSD, Input 512)	0.780	0.24s

3. Experiment

The proposed method was tested on RSIs downloaded from GoogleEarth. The dataset contained 217 RSIs, covering most Chinese airports and several international airports. All the RSIs were with 8m ground sample distance (GSD) and the sizes are about 3000 x 3000. The training set was made up of 800 images cropping from 100 RSIs with some augmentation. 17 RSIs were used as the validation set and the other 100 RSIs were the test set. We added the training data into VOC2007 for fine-tuning the SSD. The proposed method was conducted on a computer containing a CPU of Intel E5-2640 and NVIDIA TITAN X Pascal.

3.1. Airport Detection Using Transferred SSD

In the proposed transferred SSD, the sizes of input images were 1024x1024. All the original RSIs would be resized to the fixed size. The large input size could keep more airport features. Or the regions of airports in images were too small to extract features when resizing the RSIs to 512x512 or 300x300. To estimate the proposed method, the public criterion like Precision Rate, Recall Rate are defined as follows.

$$\begin{cases} PR = \frac{\text{number of detected airports}}{\text{number of detected objects}} \\ RR = \frac{\text{number of detected airports}}{\text{number of airports}} \end{cases} \quad (1)$$

The intersection-over-union (IoU) is calculated by the detected bounding box and the ground-truth of objects. If IOU is bigger than 0.5, we considered that the airport is detected. Figure 3 shows some airport detection results on the test dataset. The IoU scores of prediction results are high, which is above 0.6. In Figure 4, the airports are beside the mountains and rivers or alongside the beach. In Table 1, the evaluation results of airport detection using transferred SSD and Faster-RCNN are shown with the given recall rate. Figure 4 depicts the Recall-Precision-Curve of two methods on test dataset. AP and Time indicate the detection performance and speed respectively. The VGG-16 network was used as Base network for both methods. The experimental results indicate that the transferred SSD has an improvement than Faster R-CNN in terms of accuracy and time-consumption. At the speed of 8 FPS, the transferred SSD could get an average performance (AP) of 83.5%. Contrast to Faster R-CNN, the speed is much faster with 3% improvement achieved on AP.

For the airport detection task on RSIs, the sizes of airports are varied and some airports are so small only to take up a little area. And Faster R-CNN only employed the last layer's feature maps, which is different from SSD. The features of small airports may disappear and weaken in the process of information transmitting between so many layers, result in some misrecognitions. However, the transferred SSD utilized the feature maps from different layers and handle the multi-scale target properly. Different from Faster R-CNN which is consisted of region proposal extraction and feature recognition, SSD uses the single end-to-end CNN added with convolutional filters with small kernels to predict offsets of the default bounding boxes and classes directly, which reduced the computational cost at large.

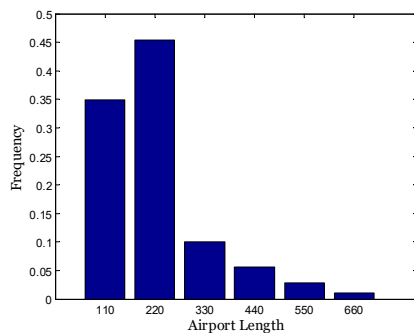


Figure 4. Statistic of the airport lengths in RSIs of 1024x1024.

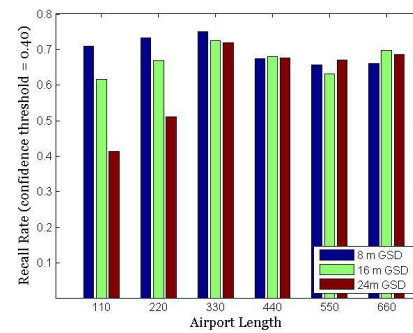


Figure 5. The Recall Rate of each set with different resolution.

3.2. Effect of Different Resolution on Airport Detection

To demonstrate the impact of different resolution on airport detection, we built other datasets by cropping the original RSIs into the size of 1024x1024 and 1500x1500 individually. And resizing the cropped images to the sizes of 512x512 and 1024x1024 respectively as the input of SSD. The original collected RSIs are 8 m GSD. After these resize operations, we gained three databases. In *Database1*, the images with the size of 1500 are resized to 512. In *Database2*, the images with size of 1024 are resized to 512. In *Database3*, the images with size of 1024 do not change. The resolution of the images in different databases is 24 m, 16m and 8m GSD individually. Table 1 shows that, with the spatial resolution increasing, the AP of airport detection declined at the same time. To make a further study on this phenomenon, the airports were grouped into 6 sets based on the length of airports and the recall rate of each set was calculated, which is shown in Figure 5. According to the chart, the recall rate of the groups in which the length is less than 330, decreased obviously. The recall rate declined gradually when the airport length is over 330. The regions with short airports were not large enough to

generate abundant features for transmitting in the deep network. It is further proved that the visual feature of small objects may lose when it passed through hierarchical layers.

4. Conclusion

This paper proposes a transferred SSD which transferred visual feature representations from natural images to RSIs. The method locates and recognizes the targets in one step. The transferred SSD predicts the offset of default boxes with feature maps from different layers, which help detect the multi-scale airports more accurately. In Comparison with other methods, this method utilizes a fully convolutional network and integrates the region proposal generation and feature extraction into one stage, which makes the computational cost decrease a lot. The experiments indicated that our method achieves better airport detection performance and satisfies time requirements of most applications.

Acknowledgement

We acknowledge support by the National Natural Science Foundation of China under Grant 61572515 and the China Postdoctoral Science Foundation of Grant 2016M593023.

References

- [1] 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
- [2] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2):91-110.
- [3] 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. 375
- [4] 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.
- [5] Krizhevsky, Alex, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks," *International Conference on Neural Information Processing Systems* Curran Associates Inc. 2012:1097-1105.
- [6] Zhang P, Niu X, Dou Y, et al. Airport Detection on Optical Satellite Images Using Deep Convolutional Neural Networks[J]. *IEEE Geoscience & Remote Sensing Letters*, 2017, PP(99):1-5.
- [7] M. M. Cheng, Z. Zhang, W. Y. Lin, et al. Cheng, Ming Ming, et al. "BING: Binarized Normed Gradients for Objectness Estimation at 300fps," (2014):3286-3293.
- [8] C. L. Zitnick and P. Dollár. "Edge Boxes: Locating Object Proposals from Edges," 8693(2014):391-405.
- [9] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders. "Selective Search for Object Recognition[J]," *International Journal of Computer Vision*, 2013, 104(2):154-171.
- [10] K. He, X. Zhang, S. Ren, et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," He, Kaiming, et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 37.9(2015):1904-1916.
- [11] S. Ren, K. He, R. Girshick, et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]," *Computer Science*, 2015:1-1.
- [12] W. Liu, D. Anguelov, D. Erhan, et al. "SSD: Single Shot MultiBox Detector[J]. 2016.
- [13] J. Redmon, S. Divvala, R. Girshick, et al. "You Only Look Once: Unified, Real-Time Object Detection[J]," *Computer Science*, 2015.
- [14] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. *Computer Science*, 2014.
- [15] 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 Recog.*, 2014, pp. 1717–1724.