Weakly Supervised Learning for Airplane Detection in Remote Sensing Images

Dingwen Zhang, Jianfeng Han, Dahai Yu, and Junwei Han

Abstract In contrast to the conventional approaches to learn geo-target classifier using fully supervised learning techniques which heavily rely on the artificial annotation in the training set of remote sensing images (RSIs), this paper attempts to develop a weakly supervised learning (WSL) approach for airplane detection in RSIs with cluttered background. The framework includes a novel WSL method to train airplane classifier using the training images with weak labels and an efficient detection scheme to localize the airplanes. The proposed WSL mainly consists of three components: the negative mining based training set initialization, the updating process for both the positive and negative training set, and the classifier evaluation mechanism that can efficiently terminate the updating process for the best performance. Comprehensive experiments on a large number of RSIs and comparisons with state-of-the-art fully supervised models demonstrate the effectiveness and efficiency of the proposed work.

Keywords Weakly supervised learning • Negative mining • Airplane detection

D. Zhang • J. Han
School of Automation, Northwestern Polytechnical University, Xi'an, China

J. Han (✉)
Computer Department, Information Engineering School, Tianjin University of Commerce, Tianjin, China
e-mail: hjf1208@sohu.com

D. Yu
School of Automation, Northwestern Polytechnical University, Xi'an, China

Tianjin Optical Electrical GaoSi Communication Engineering Technology Co., Ltd, Tianjin, China

B. Zhang et al. (eds.), *The Proceedings of the Second International Conference on Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 246, DOI 10.1007/978-3-319-00536-2_18,

1 Introduction

The advance of remote sensing technology leads to the dramatic growth of geospatial images in the amount and quantity. The high spatial resolution images can provide abundant spatial and contextual information for certain targets on the earth [1], which has necessitated the research into automatic analysis and understanding of RSIs. Nowadays, recent researches about target detection and recognition in high-resolution RSIs have become one of the most fundamental challenging tasks in this field. Typically, fully-supervised classifiers are used to fulfill the target detection task. However, a high-resolution RSI always contains complex textures that are hard for manually annotation. Since the target in RSIs only takes size between 1 and 5 % of the whole image area, people can hardly focus their attention on such small regions for detailed annotation. Moreover, since some special geo-targets are occluded or camouflaged, the artificial annotations to these targets become to be less precision and unreliable. Overall of above discussion, the manual annotation of geo-targets' locations is the most important part in fully-supervised detector training process which is tedious, time-consuming and inaccuracy. In order to solve this problem, it is essentially necessary to train target classifiers by weakly supervised methods [2–5].

In the process of weakly supervised learning (WSL) of target classifier, each image in the training set is annotated with a weak label which only indicates whether the image contains certain targets or not whereas locations and sizes are not necessarily provided. In this certain circumstance, a target model is training by using WSL that attempts to address two sub-problems simultaneously: localizing the targets in each positive training image (automated annotation) and training a model based on the automated annotation results (detector learning) [6]. To overcome these challenges and apply WSL method for airplane detection in RSIs, a novel framework is proposed as shown in Fig. 1. It includes a novel WSL method to train airplane classifier using the training images with weak label and an efficient detection scheme to localize the airplanes in RSIs. The WSL training process that is the most challenging task in this work consists of three major components. The first component is a negative mining based training set initialization which we can get from an initial training set. The second component is an updating process for both the positive and negative training set. In this process, our target classifier can improve its precision and accuracy gradually. The third component is a classifier evaluation mechanism that can efficiently terminate the updating process for the best performance.

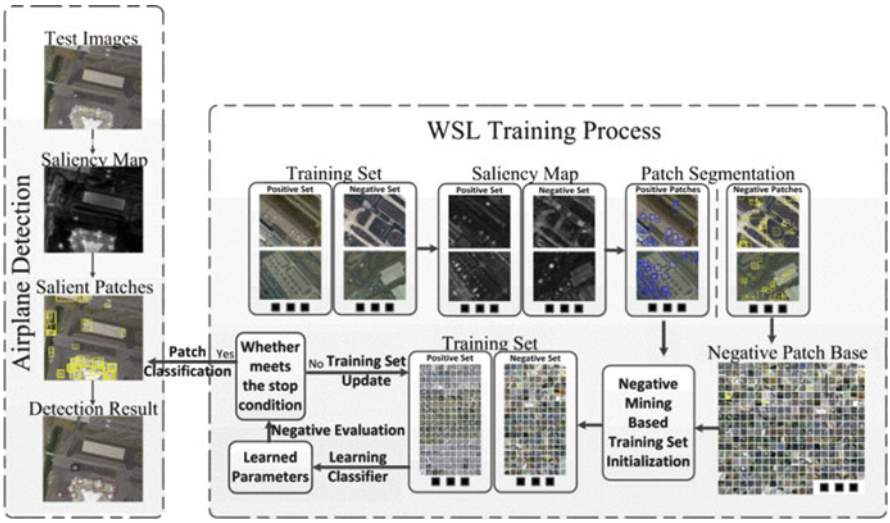

Fig. 1 Framework of the proposed algorithm

2 Training Set Initialization

In the training set initialization process, we use a novel saliency computation model to obtain the saliency maps of training images, and then apply a self-adaptive multi-threshold segmentation to obtain both positive and negative patch bases. Finally, we adopt negative mining to generate initial training set.

2.1 Saliency Model

Inspired by the characteristic of visual attention mechanism, different computational models [7, 8] have been proposed to detect salient object. In this paper, we propose a salient feature fusion process. For each RSI, the low- and mid-level features are extracted for every pixel of down-sampled image and they are used to generate the final saliency map like [9]. The low-level features are local contrast of intensity, orientation and color, color value in each channel, and global color contrast [9]. The mid-level features are SR [7], GBVS [8], FT [6], WSCR [10], and SDS [11]. All of these silent features have already been demonstrated by previous works to correlate with visual saliency or be biological plausible. For the weakly labeled training set, we do not know the locations of the targets. Duo to this reason, we cannot use artificial annotation to train a set of coefficients to fuse the salient features in our salient computation model. For this regard, we treat the weight of each salient feature equally which means that if a pixel is salient in most

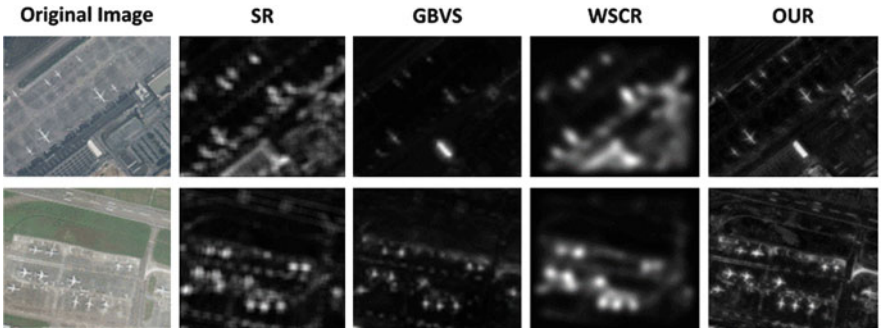

Fig. 2 Saliency map comparison

salient features, it is salient in the final saliency map. Figure 2 shows that our salient computation model can outperform most state-of-the-art methods.

2.2 Self-Adaptive Segmentation

Saliency model can calculate a salient value for each pixel in every RSI. Then, the next step is to obtain candidate patches which may contain certain targets based on the saliency map. An efficient way is to segment the saliency map by one or more adaptive threshold as (1) in [6].

$$thresh = \frac{k}{W \times H} \times \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} S(x, y) = k \times mean(S) \tag{1}$$

where W and H are the width and height of the saliency map in pixels respectively, and $S(x, y)$ is the saliency value of the pixel at position (x, y), k is a parameter set manually.

Afterwards, the candidate patches will be re-cropped to obtain the patch which should be the similar size as one target rather than some approximate region including large areas such as background. To achieve this goal, we prefer to use the technique in [4] by finding the area enclosing majority of its edge energy. Finally we obtain the external rectangle of certain area as one candidate or salient patch. In this paper we set $k = 1.5$ and 3 to acquire the refined patches which are stable for the next process.

2.3 Negative Training Set Initialization

Negative training set is the salient patches in negative RSIs. As we mentioned before, the negative RSIs are weakly labeled RSIs without any targets. In order to generate the negative patch base, we calculate their saliency map and use our self-adaptive segmentation method to obtain salient patches. Afterwards, we randomly select negative patches with the same number of patches in the positive training set to form the negative training set.

2.4 Positive Training Set Initialization

Positive training set is obtained from the positive RSIs. After getting positive patch base by using our saliency model and self-adaptive segmentation method, we need to adopt negative mining to select several positive patches with highest probability of containing the targets to form the positive training set.

As described in [12], negative mining is an approach of mining the nearest negative patches. Unlike [3, 5], it relies on the abundance of known negative patches and does not require optimization of intra-class cost function. The training image set consists of a set of positive images I_i^+ that contain the object of interest and a set of negative images I_i^- which do not. We consider a set of positive patches $x_{i,j=1..n}$ in each image i. Each patch $x_{i,j}$ is represented by a bag-of-words (BOW) histogram. The goal of this step is to score every patch in each positive image I^+ and select some highest scored patches as the initial positive training set. The negative mining algorithm accomplishes this by selecting the patches with large distance to the nearest neighbor in the negative patch base [see (2)] where $\|\cdot\|_1$ is the L_1 norm and $N(x_{i,j}^+)$ refers to the negative nearest neighbor of $x_{i,j}^+$.

$$Dist\left(x_{i,j}^+\right) = ||x_{i,j}^+ - N\left(x_{i,j}^+\right)||_1 \tag{2}$$

3 Training Set Updating

In order to implement the target detection task precisely, a training set updating process is applied to train the classifier iteratively. The motivation of this process is to achieve a strong classifier ultimately that can improve the performance of the classifier. It is important to notice that we use BOW features to train the target classifier because it is invariant to the target rotation, shape and size variation. After training the BOW classifier by using the initial training set, the obtained target classifier is then run on each positive image. Next, a set of patches with highest classification score are selected as the new positive training set. We then randomly select the same number of negative patches from the negative patch base to generate

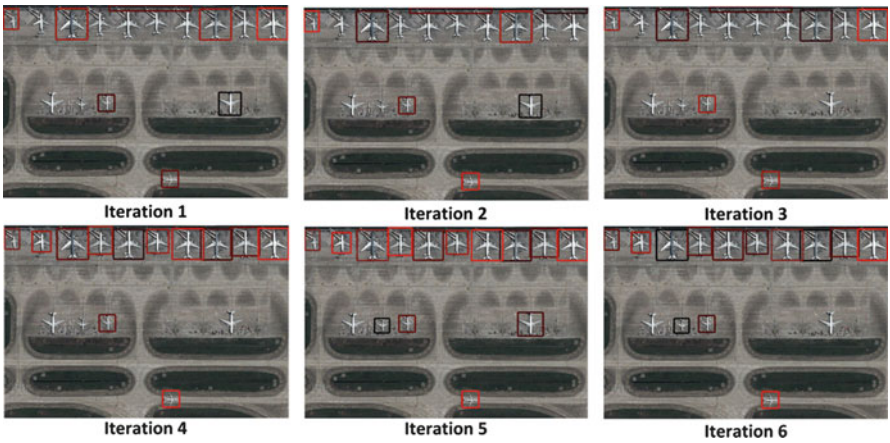

Fig. 3 Performance of training set updating

Table 1 Performance of negative evaluation

	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5	Iteration 6
F-measure	0.7361	0.7452	0.7441	0.7650	0.7741	0.7699
False rate	0.1412	0.0619	0.0387	0.0295	0.0290	0.0348

the new negative training set. Figure 3 shows the gradually improvement of the target classifier in each iteration and achieve the best performance in iteration 5.

4 Classifier Evaluations

Evaluating trained target classifier in each iteration using weakly labeled data is difficult because we don't know what the target looks like and where the exact target locations are in positive images? To solve this problem, we propose to use a negative evaluation mechanism to measure the effectiveness of current classifier in iteration since negative instances do not contain any target. Specifically we use the classifier to classify each negative patch in negative patch base and calculate the false rate in iteration. Continually, the false rate decreases in the first several iterations and then begins to increase. Therefore, we setup the iteration with the local minimal false rate is the stop iteration, and the target classifier trained in this iteration will be selected as the final classifier which is preferred as the best performance in the updating process. This intuitive measure is proved base on the performance of the trained classifier in our experiments evaluation (see Table. 1). From Table. 1, we can observe that the local minimal false rate locates in iteration

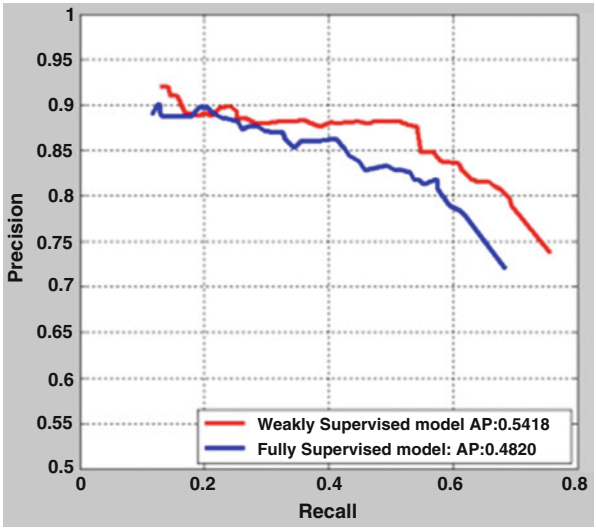

Fig. 4 Precision-recall curve

5 and the F-measure which is popularly used in object annotation evaluation increases to the maximum.

5 Airplane Detection

For the airplane detection, a salient patch based target detection scheme is adopted (see Fig. 1). We calculate the saliency maps of RSIs at first, and use our self-adaptive segmentation method to get their salient patches. Because our approach is robust to get salient patches which would contain the airplane targets in the entire test images, it appears to be more efficient to detect targets by classifying the salient patches in one RSI instead of the sliding windows. In this regard we use the classifier trained by our WSL method to accomplish the weakly supervised airplane detection.

6 Experimental Results

For this experiment, we collect 120 high resolution RSIs from ten different airports in different countries on Google Earth. The resolution of them is about 0.5 m per pixel, the scale of an image is about 1,000 by 800 pixels, and the size of an airplane target is from 700 pixels to 25,488 pixels. Hence, the result of our experiment can demonstrate that our algorithm is able to deal with multi-size target in large scale RSIs with cluttered background.

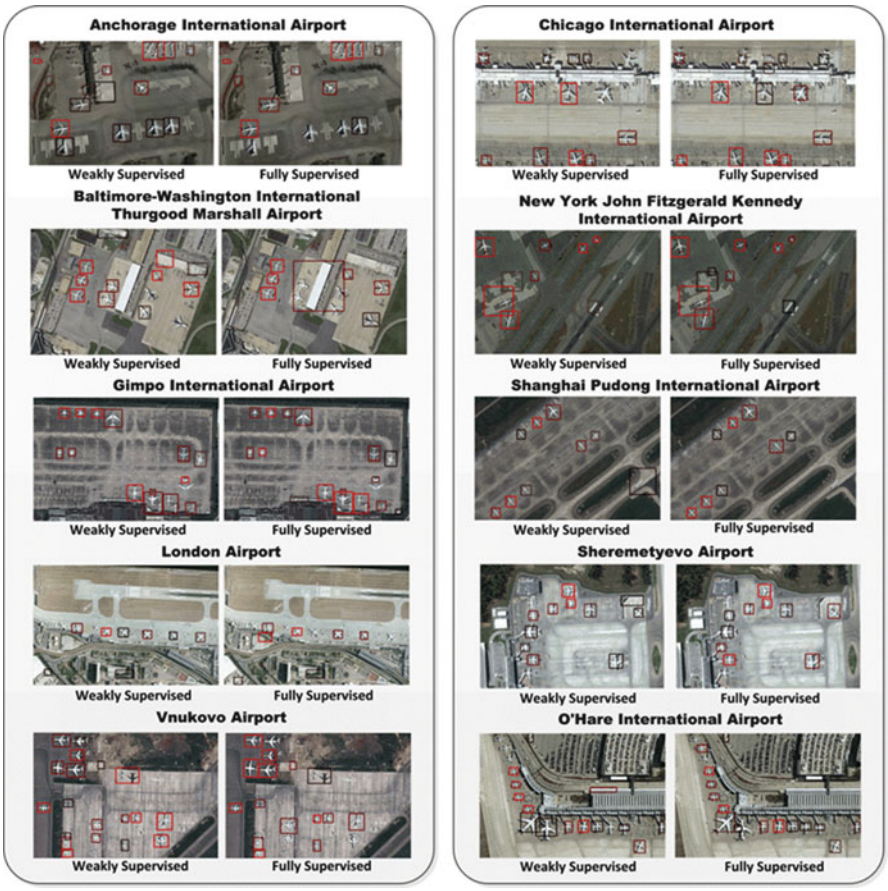

Fig. 5 Examples of detection results

In this evaluation we separate all RSIs into three sets: 50 of them as test set, 50 of them as positive training set, and 20 of them as negative training set. The training set is used to train airplane classifier by our WSL method, and the test set is used to compare the performance of the trained classifier by using proposed method and the classifier trained by fully supervised method.

In additional to demonstrate the airplane detection accuracy based on proposed WSL trained airplane classifier among the test data, we present the Precision-Recall curve (see Fig. 4) and several detection result examples (see Fig. 5). Both figures demonstrate that the proposed WSL method can achieve higher detection precision in more test RSIs and a mean AP of 0.5418 superior to a mean AP of 0.4820 of the FS approach. The mean AP has been proved to be the most accurate approximation to average precision and the most robust measure in the presence of incomplete detection. Note that the weakly supervised method is the proposed approach in this

paper and the fully supervised method is the classifier trained by artificial annotation with BOW feature to detect target in the same test image set.

7 Conclusions

In this paper, we present a framework for the challenging task of detecting airplane with classifier trained by WSL method in large scale RSIs with clustered background. The framework includes a novel WSL method to train airplane classifier using the training images with weak label and an efficient detection scheme to localize the airplanes in RSIs. The experimental results demonstrate that WSL method in remote sensing target detection is an efficient approach since weakly supervised learning target classifier can achieve better detection accuracy compared with a fully supervised classifier.

Acknowledgments This work is supported by graduate starting seed fund of Northwestern Polytechnical University under grant Z2013105.

References

1. Tuia D, Pacifici F, Kanevski M, Emery WJ (2009) Classification of very high spatial resolution imagery using mathematical morphology and support vector machines. IEEE Trans Geosci Rem Sens 47(11):3866–3879
2. Deselaers T, Alexe B, Ferrari V (2012) Weakly supervised localization and learning with generic knowledge. Int J Comput Vis 100(3):275–293
3. Deselaers T, Alexe B, Ferrari V (2010) Localizing objects while learning their appearance. ECCV Part IV. LNCS 6314:452–466
4. Pandey M, Lazebnik S (2011) Scene recognition and weakly supervised object localization with deformable part-based model. ICCV, Barcelona, 6–13 November, 2011, pp. 1307–1314
5. Siva P, Xiang T (2011) Weakly supervised object detector learning with model drift detection. ICCV, Barcelona, 6–13 November, 2011, pp. 343–350
6. Achanta R, Hemami S et al (2009) Frequency-tuned salient region detection. IEEE conference on computer vision and pattern recognition, Miami, FL, 20–25 June, 2009, pp. 1597–1604
7. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. IEEE conference on computer vision and pattern recognition, Minneapolis, MN, 17–22 June, 2007, pp. 1–8
8. Harel J, Koch C, Perona P (2007) Graph-based visual saliency. Advances in neural information processing systems. MIT Press, Cambridge, MA, pp 545–552
9. Judd T, Ehinger K, Durand F, Torralba A (2009) Learning to predict where humans look. Proc. IEEE 12th international conference on computer vision, Kyoto, 29 September–2 October, 2009, pp. 2106–2133.
10. Han B, Zhu H, Ding Y. (2011) Bottom-up saliency based on weighted sparse coding residual. ACM International Conference on Multimedia, Scottsdale, Arizona, pp. 1117–1120
11. Achanta R, Estrada F, Wils P, Süsstrunk S (2008) Salient region detection and segmentation. Int Conf Comput Vis Syst 5008:66–75
12. Siva P, Chris R, Tao X (2012) In defence of negative mining for annotating weakly labelled data. In ECCV. Springer, Berlin, pp 594–608