2015 IEEE International Conference on Multimedia Big Data

Negative Bootstrapping for Weakly Supervised Target Detection in Remote Sensing Images

Peicheng Zhou, Dingwen Zhang, Gong Cheng, Junwei Han
School of Automation
Northwestern Polytechnical University
Xi'an, China
{zpc19881119, zhangdingwen2006yyy, chenggong1119, junweihan2010}@gmail.com

***Abstract*—When training a classifier in a traditional weakly supervised learning scheme, negative samples are obtained by randomly sampling. However, it may bring deterioration or fluctuation for the performance of the classifier during the iterative training process. Considering a classifier is inclined to misclassify negative examples which resemble positive ones, comprising these misclassified and informative negatives should be important for enhancing the effectiveness and robustness of the classifier. In this paper, we propose to integrate Negative Bootstrapping scheme into weakly supervised learning framework to achieve effective target detection in remote sensing images. Compared with traditional weakly supervised target detection schemes, this method mainly has three advantages. Firstly, our model training framework converges more stable and faster by selecting the most discriminative training samples. Secondly, on each iteration, we utilize the negative samples which are most easily misclassified to refine target detector, obtaining better performance. Thirdly, we employ a pre-trained convolutional neural network (CNN) model named Caffe to extract high-level features from RSIs, which carry more semantic meanings and hence yield effective image representation. Comprehensive evaluations on a high resolution airplane dataset and comparisons with state-of-the-art weakly supervised target detection approaches demonstrate the effectiveness and robustness of the proposed method.**

Keywords-Remote sensing image (RSI); Target detection; Weakly supervised learning (WSL); Negative Bootstrapping; High-level feature

I. INTRODUCTION

Target detection in remote sensing images (RSIs) has become a fundamental problem faced for remote sensing images analysis. Especially, with the rapid development of remote sensing technology, the high-resolution RSIs contain richer visual information, which makes it possible to describe more surface appearance of the earth. However, how to reliably and effectively detect targets in cluttered scenes of RSIs is still a profound challenge in the field of remote sensing image analysis.

In the early studies, most methods employed unsupervised models to detect targets in RSIs [1-6], which mainly depended on the features used in their models and may be effective for detecting targets with simple appearance and small variations.

Recently, more approaches adopted supervised learning (SL) techniques to detect targets, which can take advantage of the prior knowledge achieved from training samples to train more robust object detector. For example, Cheng *et al.* [7] developed an object detection framework using a discriminatively trained mixture model. Han *et al.* [8] combined visual saliency and discriminative sparse coding for efficient and simultaneous multi-class targets detection from optical remote sensing images. Cheng *et al.* [9, 10] proposed a Collection of Part Detectors (COPD) method to detect multi-class geospatial objects on a public high-spatial-resolution RSIs data set containing 10-class objects[1]. In addition, other models such as Support Vector Machines (SVM) [4, 7, 11-14], *k*-nearest neighbour (*k*-NN) [15], Gaussian Mixture Models [16], Hough Forests [17], etc., have also been applied to target detection. However, good performance of the above approaches could be achieved only when the manually labeled samples are provided. To alleviate the tedious and unreliable manual annotation, some researchers adopted semi-supervised learning (SSL) methods to perform object detection [18, 19], in which only a few labeled training samples were used to train detectors and then new samples in training set were added from unlabeled data. However, these methods still require a comparative number of manual labeled positive examples.

To minimize the manual annotation while not deteriorate the performance of object detector significantly, many works employed weakly supervised learning (WSL) framework for target detection. Compared with SL and SSL, WSL only needs to indicate whether the images in the training set contain the to-be-detected targets or not, rather than annotates targets using accurate bounding boxes. For instance, Zhang et al. [20] developed an efficient target detection method by leveraging WSL in RSIs. Han *et al.* [21] proposed an improved WSL method which integrated saliency, intra-class compactness, and inter-class separability in a Bayesian framework.

The typical WSL schemes focus on how to precisely select the initial positive training examples and annotate the new positives on each refining iteration [22-24]. There are few considerations have been made for the selection of negative examples, and random sampling is actually a widely adopted technique in the literatures. However, this may bring deterioration or fluctuation for the performance of the

[1] http://pan.baidu.com/s/1c0w8h3q

978-1-4799-8688-0/15 $31.00 © 2015 IEEE
DOI 10.1109/BigMM.2015.13

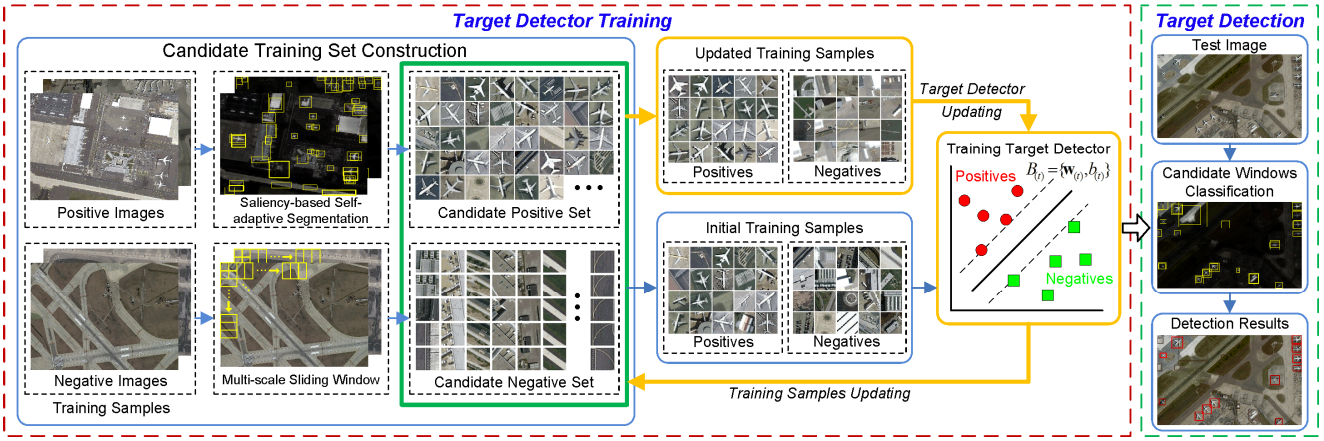

Figure 1. The framework of the proposed method.

classifier during the iterative training process and maybe drop into local optimization rather than global optimization. As can be seen from the works of [25, 26], a classifier is inclined to misclassify negative examples which are visually similar to positive ones, so including these partially overlapped and informative negatives should be important for enhancing the effectiveness and robustness of the classifier. In this paper, we propose to integrate Negative Bootstrapping scheme into weakly supervised learning to train object detector, which can obtain informative negatives substantially and automatically.

Furthermore, we utilize Caffe [27] toolbox to explore more semantic representation from RSIs, which is one of the prevalent convolutional neural network (CNN) frameworks. Compared with traditional hand-designed features, features extracted by CNN can carry more semantic meanings and have been demonstrated to be effective for solving challenging tasks in computer vision field.

The comprehensive experiments on an airplane RSI dataset and comparison with state-of-the-art target detection approaches demonstrate the effectiveness and robustness of the proposed method.

II. Proposed Framework

Fig. 1 gives the flowchart of our developed framework. It mainly consists of two stages: target detector training and target detection. In the model training stage, given the image-level labels indicating the existence of targets in the training images, we firstly initialize training samples by generating the most likely positive samples and the most relevant negative samples in proper sequence. Then we use these initialized training samples to train classifiers iteratively. On each iteration, we classify image patches from positive and negative RSIs by current classifier respectively. After several iterations we can obtain the optimal detector. In the target detection stage, given a testing RSI, we first employ saliency-based self-adaptive segmentation method [20] to predict a comparable small set of candidate windows. Then, using the target detector trained in the former stage to classify each window and obtain their corresponding responses. Finally, a post-processing is used to eliminate repeated detections via non-maximum suppression.

A. Candidate Training Set Construction

An important step of WSL is training samples initialization, so we need to construct a candidate training set for training samples initialization. The candidate training set should contain more positive samples with high probability and negative samples with diversity and non-redundancy.

1) Candidate Positive Set: Considering there is no prior information about the position, shape and scale of targets in positive images I_{pos}, in this paper, we employed a saliency-based self-adaptive segmentation method [20] to construct candidate positive set S^+. To be specific, for a positive image, we first adopted the saliency model of [20] to yield an overall saliency map by linearly combining some normalized low-level and mid-level features for each pixel of the positive image. Then a self-adaptive segmentation was performed on the saliency map to obtain candidate positive regions under multiple thresholds. Finally, S^+ was formed by the patches, which were labeled by bounding boxes on the segmented regions. Other salienct object detection model [28] and segmentation method [29] can be also used for candidate positive set construction.

2) Candidate Negative Set: In WSL framework, negative images definitely do not contain any target. Therefore, we can easily collect negative training samples from negative images. To accommodate the purpose of mining informative negatives, we need a candidate negative set which contains diverse and non-redundant negative samples as much as possible. The construction of candidate negative set is implemented by the following steps.

- Adopt multi-scale window mechanism [8] to collect a large number of negative samples in negative RSIs I_{neg}. These negative samples form an unrefined negative set U^-.
- To exclude redundant negative samples and maintain the diversity of negative samples in U^-, we employ *k*-means clustering over these samples according to a

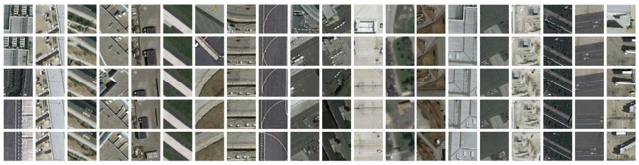

Figure 2. 20 randomly selected clusters from negative candidate set, where each column corresponds to one cluster's five top ranked negative samples.

predefined cluster number. We can also use other clustering methods such as adaptive clustering [30].

- Combine the top ranked n samples in each cluster to form the candidate negative set.

After obtaining the unrefined negative set U^- train cluster centers $\mathbf{D} = \{d_i, i = 1, 2, \cdots, K\}$ for a predefined cluster number K, where d_i is the i-th cluster of $\mathbf{D}$. Then we can obtain nK samples as the candidate negative set $S^- = \{s_{ij}^-, j = 1, 2, \cdots, n\} \subset U^-$ by selecting the top ranked n samples in each cluster, where s_{ij}^- is the feature representation of the j-th samples in the i-th cluster, which is extracted by a pre-trained CNN using Caffe toolbox [27]. Fig. 2 shows 20 randomly selected clusters, where each cluster has five top ranked negative samples. As can be seen from Fig. 2, (1) Almost all samples within each cluster are visual consistent, which can be regarded as redundant and should be excluded. (2) Samples between different clusters are visual different, so we should preserve at least one sample in each cluster. After these operations, we can obtain a set of diverse and non-redundant negatives.

B. Training Samples Initialization

1) Positive Training Samples: We adopted negative mining strategy to obtain initial positive training samples from the candidate positive set S^+, under the observation that targets are regularly different from negative samples in visual appearance. To be specific, let $S_{(1)}^+ = \{s_p^+, p = 1, 2, \cdots, n_p\} \subset S^+$ denote initial positive samples, where s_p^+ is the feature representation of p-th positive sample, n_p is the number of initial positive samples. The negative mining algorithm was implemented as follows.

$$S_{(1)}^+ = \left\{ s_p^+ \mid dist(s_p^+) > \tau, s_p^+ \in S^+ \right\} \tag{1}$$

$$dist(s_p^+) = \min_{i \in \{1,K\}, j \in \{1,n\}} \left\| s_p^+ - s_{ij}^- \right\|_1 \tag{2}$$

where $\|\bullet\|_1$ is the L1-norm and τ is a threshold used for excluding some noisy positive samples that is similar to negatives.

2) Negative Training Samples: The informative negative samples are considered to be samples which are most likely misclassified. However, on the first iteration, we do not have pre-trained target detector used for predicting which samples are most likely misclassified. To improve the performance of target detector which is traditionally trained by randomly sampled negative samples in the first round, we initialize the negative samples to be those that are most similar to the initial positive samples by measuring their distances in CNN feature space. The distance between negative samples in S^- and the initial positive samples $S_{(1)}^+$ can be calculated by

$$Dist(S^-, S_{(1)}^+) = \left\{ dist(s_q^-, s_p^+), s_q^- \in S^-, s_p^+ \in S_{(1)}^+ \right\} \tag{3}$$

$$dist(s_q^-, s_p^+) = \min_{p \in \{1, n_p\}} \left\| s_q^- - s_p^+ \right\|_1 \tag{4}$$

We then rank all negative samples in S^- by $Dist(S^-, S_{(1)}^+)$ in ascending order and select top ranked samples as our initial negative samples. To balance the number of training samples, the negative samples should be selected with the same number of $S_{(1)}^+$. Let $S_{(1)}^-$ be the initial negative samples, we can generate it by

$$S_{(1)}^- \leftarrow select\ top\ samples(S^-, Dist(S^-, S_{(1)}^+), |S_{(1)}^+|) \tag{5}$$

where $|\bullet|$ denotes the cardinality of a given data set.

Algorithm 1 Training Procedure of Target Detector

Input: Initial training samples $S_{(1)}^+$ and $S_{(1)}^-$, candidate training set S^+ and S^-, and the number of learning iteration T.

Output: Target detector $B_{(T)} = \{\mathbf{w}_{(T)}, b_{(T)}\}$

1. Train an initial target detector $B_{(1)} = \{\mathbf{w}_{(1)}, b_{(1)}\}$
2. **For** $t = 2$ to T **do**
   (1) **Update training samples**
   (a) Calculate scores of candidate samples by $B_{(t-1)} = \{\mathbf{w}_{(t-1)}, b_{(t-1)}\}$
   $Score_{(t)}(s_p^+) = \mathbf{w}_{(t-1)}^T s_p^+ + b_{(t-1)}, s_p^+ \in S^+$
   $Score_{(t)}(s_q^-) = \mathbf{w}_{(t-1)}^T s_q^- + b_{(t-1)}, s_q^- \in S^-$
   (b) Select new training samples:
   $S_{(t)}^+ = \{s_p^+ \mid Score_{(t)}(s_p^+) > \sigma, s_p^+ \in S^+\}$
   $S_{(t)}^- \leftarrow select\ top\ samples(S^-, Score_{(t)}(S^-), |S_{(t)}^+|)$
   (2) **Update target detector**
   Use $S_{(t)}^+$ and $S_{(t)}^-$ to train a new target detector $B_{(t)} = \{\mathbf{w}_{(t)}, b_{(t)}\}$
end

C. Iterative Target Detector Training

Algorithm 1 gives the procedure of target detector training.

1) Target detector training: After obtaining the initial training samples $S^+_{(1)}$ and $S^-_{(1)}$, we can train an initial target detector $B_{(1)}$, and then update training samples and optimize target detector iteratively until the convergence is reached. Let T denote the total number of iterations, $t = 2, \cdots, T$ be the iteration index, $S^+_{(t)}$ and $S^-_{(t)}$ are the updated training samples on the t-th iteration, $B_{(t)}$ be the target detector trained on them. In this paper, we adopt the linear SVM to train target detector, which is formulated as

$$\min_{\mathbf{w}_{(t)}, b_{(t)}} \frac{1}{2}\mathbf{w}^T_{(t)}\mathbf{w}_{(t)} \quad s.t. \;\; y_m(\mathbf{w}^T_{(t)}s_m + b_{(t)}) - 1 \geq 0 \tag{6}$$

where $s_m \in S^+_{(t)} \cup S^-_{(t)}$ is the m-th training samples, $y_m \in \{1, -1\}$ is the label of s_m. The target detector can be represented as $B_{(t)} = \{\mathbf{w}_{(t)}, b_{(t)}\}$. For predicting the score of one sample, we regard it as a classification problem, which is formulated as

$$Score_{(t+1)}(s_m) = \mathbf{w}^T_{(t)}s_m + b_{(t)} \tag{7}$$

2) Training Samples Updating: On the t-th iteration, the informative positive training samples $S^+_{(t)}$ and negative training samples $S^-_{(t)}$ can be updated by $B_{(t-1)}$, which is trained on the (t-1)-th iteration. Specifically, we use the target detector $B_{(t-1)}$ to score positive and negative candidate samples respectively. Their scores can be calculate by (7) and represented by

$$Score_{(t)}(S^+) \leftarrow \left\{ Score_{(t)}(s^+_p) = \mathbf{w}^T_{(t-1)}s^+_p + b_{(t-1)}, s^+_p \in S^+ \right\} \tag{8}$$

$$Score_{(t)}(S^-) \leftarrow \left\{ Score_{(t)}(s^-_q) = \mathbf{w}^T_{(t-1)}s^-_q + b_{(t-1)}, s^-_q \in S^- \right\} \tag{9}$$

Then, the updated positive samples can be obtained by selecting the positive samples with their scores above a given threshold σ in S^+.

$$S^+_{(t)} = \left\{ s^+_p \mid Score_{(t)}(s^+_p) > \sigma, s^+_p \in S^+ \right\} \tag{10}$$

The updated negative samples $S^-_{(t)}$ can be obtained by selecting the top ranked negative samples with the same number of $S^+_{(t)}$.

$$S^-_{(t)} \leftarrow select\ top\ sample(S^-, Score_{(t)}(S^-), |S^+_{(t)}|) \tag{11}$$

D. Target Detection

To detect targets in RSIs efficiently and accurately, we employ the candidate-patch-based target detection scheme [20] rather than the conventional sliding window methods. For a given RSI, the candidate windows can be obtained by saliency-based self-adaptive segmentation method following previous works [8, 20]. Then, we use the target detector $B_{(T)} = \{\mathbf{w}_{(T)}, b_{(T)}\}$ to obtain their responses to determine whether these candidate windows contain targets or not. Finally, a non-maximum suppression scheme is adopted to eliminate repeated detections.

III. EXPERIMENTS

A. Experimental Setup

1) Data set description: We quantitatively evaluate the proposed method using an airplane RSI benchmark from Zhang et al. [20], which includes 170 high-spatial-resolution images from Google Earth, where 120 RSIs contain airplanes serving as positive images and 50 RSIs do not contain any airplanes serving as negative images. To objectively and fairly evaluate this method, we use the same data selection strategy as in [20, 21]. In brief, the positive RSIs are separated into two non-overlap sets: one set contains 70 images used for training and the other set contains 50 images used for testing. The resolution of these images is 1000 by 800 pixels, the spatial resolution of these images ranges from 0.5 m to 2 m, and the area of targets in these images varies from 700 pixels to 25488 pixels.

2) Feature extraction: We employ the Caffe toolbox [22] to extract features with a pre-trained model. Specifically, we firstly resize each image patch from their original pixel size to a uniform 227×227 pixel size as the inputs of Caffe. Then, we feed the raw data into a forward propagating neural network with five convolutional layers and two fully connected layers to extract a 4096-dimensional feature vector for each image patch. Finally, we convert each 4096-dimensional feature to a lower 1024-dimensional feature using principal component analysis (PCA) to avoid the curse of dimensionality in the detector training process. It should be noted that the contribution of the principal component is over 98% when we reduce the dimensionality of orginal feature from 4096 to 1024 and other efficient dimensionality reduction methods can also be used here (e.g., Folded-PCA [31]).

3) Implementation: To construct the candidate negative set, we collected negative samples with three window scales (i.e., 60×60, 80×80, 100×100), and refined them using k-means clustering with cluster number $K = 5000$. Then we selected the top-1 sample in each cluster to form our candidate negative set. In experiments, we empirically set the threshold $\tau = 0.85$ in training samples initializing and $\sigma = 0.95$ in training samples updating. The binary classifier was trained by LibSVM toolbox [32] with a linear kernel. The number of iterations used in all the experiments was set to $T = 100$.

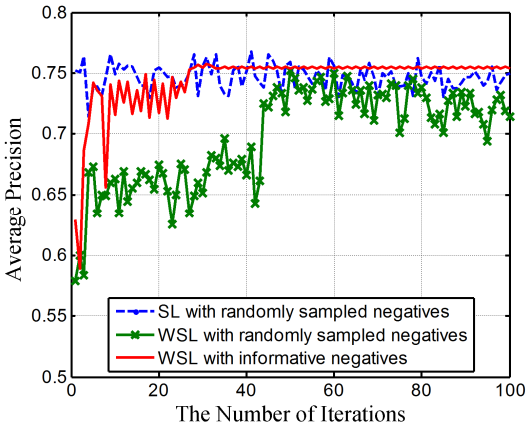

Figure 3. The performance comparison of two WSL methods and one SL method.

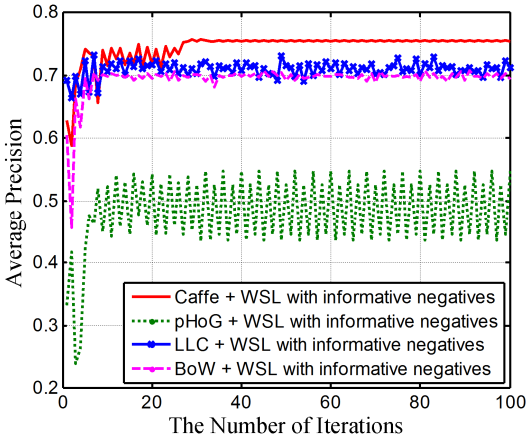

Figure 4. Performance comparison of target detector trained by four types of features.

4) Evaluation criterion: We employ Average Precision (AP) to quantitatively evaluate the performance of the proposed method, which is a standard criterion for evaluating target detection [7, 8, 20] and is measured by the area under Precision-Recall curve (PRC). A detection result is considered as a true positive if the overlap area between detection windows and ground truth is more than 50%.

B. Evaluations

1) The influence of informative negatives: The goal of this paper is enhancing the robustness and effectiveness of classifier training with informative negatives in WSL. Naturally, we compared this method with conventional WSL which obtained negative training samples by randomly sampling on each iteration. For fair comparison, the negative samples used in these methods were all selected from the same candidate negative set constructed in Section II. As shown in Fig. 3, after several iterations, the performance of WSL with informative negatives is stable and the average precision is fixed on 0.7539 after 100 iterations, while the performance of WSL with randomly sampled negatives is fluctuated through the whole iterative process and the average precision is 0.7147 finally. We also compared the proposed method with supervised learning strategy which trained the target detector using manually labeled positive samples and randomly sampled negatives from candidate negative set. As can be seen from Fig. 3, the proposed method is even more robust and effective than supervised learning method. The mean average precision of supervised learning method is 0.7470, while our proposed method is 0.7539 after 100 iterations. The results show that using informative negatives to train target detector is very important for the robustness and effectiveness in WSL.

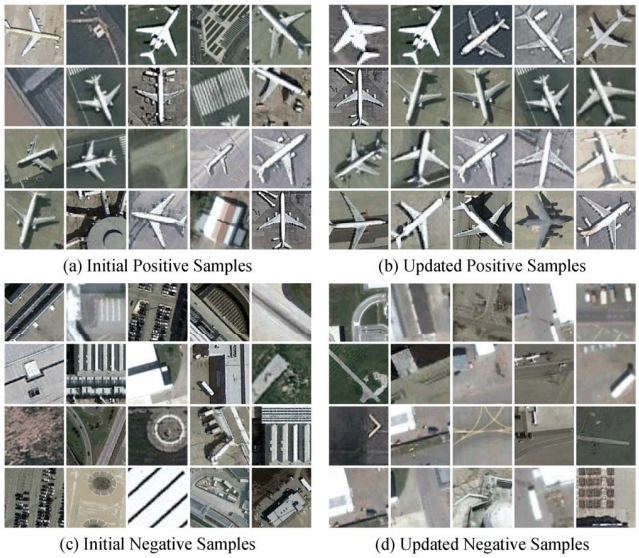

Figure 5. The initial positive samples and negative samples and their corresponding updated samples. (a) Initial positive samples obtained by saliency-based self-adaptive segmentation [20]. (b) Updated positives obtained automatically by the proposed method after 100 iterations. (c) Negative samples collected by random sampling. (d) Updated negatives obtained automatically by the proposed method after 100 iterations.

2) High-level feature vs traditional features: To validate the effectiveness of high-level semantic feature extracted by Caffe toolbox [27], we applied the proposed framework to detect airplanes with four types of features. They are Caffe [27], pyramid histogram of oriented gradient (pHoG) [33], bag-of-words (BoW) [34], and locality-constrained linear coding (LLC) [35]. The parameters setting for each type of feature is the same as the works of [27, 34], [35], [33]. As can be seen from Fig. 4, the feature extracted by Caffe toolbox is much stronger than all other features extracted by pHOG [33], BoW [34], and LLC [35].

3) Qualitative analysis of the proposed method: To qualitatively evaluate the influence of the training samples on target detector training, we visualize some training samples used for target detector training in Fig. 5. As can be seen, after 100 times sample updating, the noisy samples in initial positive samples set are removed. The updated negatives in Fig. 5(d) obtained by the proposed method have similar visual representation to positives. They are deemed to be more informative and can be used to train more refined detector. On the contrary, the negatives in Fig. 5 (c) collected by random sampling do not have this characteristic.

IV. CONCLUSION

In this paper, we developed a framework for target detection in RSIs by integrating Negative Bootstrapping into weakly supervised learning. By utilizing easily misclassified negatives to train target detector, the effectiveness and robustness of target detector was improved significantly. In addition, we have proved that the high-level feature extracted by a pre-trained CNN model was effective for target representation in RSIs. In the future work, we will (1) extend the proposed target detector training framework to multi-class object detection on different RSI datasets; (2) integrate some discriminative information between positives and negatives to train more effective target detector; (3) fine-tune the pre-trained CNN model using RSIs to enhance domain adaptation.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation of China under Grants 61473231, 61401357, and China Postdoctoral Science Foundation under Grant 2014M552491.

REFERENCES

[1] X. Li, S. Zhang, X. Pan, P. Dale, and R. Cropp, "Straight road edge detection from high-resolution remote sensing images based on the ridgelet transform with the revised parallel-beam Radon transform," Int. J. Remote Sens., vol. 31, pp. 5041-5059, 2010.

[2] M. Tello, C. López-Martínez, and J. J. Mallorqui, "A novel algorithm for ship detection in SAR imagery based on the wavelet transform," IEEE Geosci. Remote Sens. Lett., vol. 2, pp. 201-205, 2005.

[3] B. Sirmacek and C. Unsalan, "Urban area detection using local feature points and spatial voting," IEEE Geosci. Remote Sens. Lett., vol. 7, pp. 146-150, 2010.

[4] P. Li, H. Xu, and J. Guo, "Urban building damage detection from very high resolution imagery using OCSVM and spatial features," Int. J. Remote Sens., vol. 31, pp. 3393-3409, 2010.

[5] B. Sirmacek and C. Unsalan, "Urban-area and building detection using SIFT keypoints and graph theory," IEEE Trans. Geosci. Remote Sens., vol. 47, pp. 1156-1167, 2009.

[6] Y. Li, X. Sun, H. Wang, H. Sun, and X. Li, "Automatic target detection in high-resolution remote sensing images using a contour-based spatial model," IEEE Geosci. Remote Sens. Lett., vol. 9, pp. 886-890, 2012.

[7] G. Cheng, J. Han, L. Guo, X. Qian, P. Zhou, X. Yao, *et al.*, "Object detection in remote sensing imagery using a discriminatively trained mixture model," ISPRS J. Photogramm. Remote Sens., vol. 85, pp. 32-43, 2013.

[8] J. Han, P. Zhou, D. Zhang, G. Cheng, L. Guo, Z. Liu, *et al.*, "Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding," ISPRS J. Photogramm. Remote Sens., vol. 89, pp. 37-48, 2014.

[9] G. Cheng, J. Han, P. Zhou, and L. Guo, "Scalable multi-class geospatial object detection in high-spatial-resolution remote sensing images," in IGARSS, 2014.

[10] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," ISPRS J. Photogramm. Remote Sens., vol. 98, pp. 119-132, 2014.

[11] H. Sun, X. Sun, H. Wang, Y. Li, and X. Li, "Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model," IEEE Geosci. Remote Sens. Lett., vol. 9, pp. 109-113, 2012.

[12] F. Bi, B. Zhu, L. Gao, and M. Bian, "A visual search inspired computational model for ship detection in optical satellite images," IEEE Geosci. Remote Sens. Lett., vol. 9, pp. 749-753, 2012.

[13] Z. Li and L. Itti, "Saliency and gist features for target detection in satellite images," IEEE Trans. Image Process., vol. 20, pp. 2017-2029, 2011.

[14] 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, pp. 128-132, 2011.

[15] G. Cheng, L. Guo, T. Zhao, J. Han, H. Li, and J. Fang, "Automatic landslide detection from remote-sensing imagery using a scene classification method based on boVW and pLSA," Int. J. Remote Sens., vol. 34, pp. 45-59, 2013.

[16] S. Bhagavathy and B. S. Manjunath, "Modeling and detection of geospatial objects using texture motifs," IEEE Trans. Geosci. Remote Sens., vol. 44, pp. 3706-3715, 2006.

[17] Z. Lei, T. Fang, H. Huo, and D. Li, "Rotation-invariant object detection of remotely sensed images based on texton forest and hough voting," IEEE Trans. Geosci. Remote Sens., vol. 50, pp. 1206-1217, 2012.

[18] Q. Liu, X. Liao, and L. Carin, "Detection of unexploded ordnance via efficient semisupervised and active learning," IEEE Trans. Geosci. Remote Sens., vol. 46, pp. 2558-2567, 2008.

[19] L. Capobianco, A. Garzelli, and G. Camps-Valls, "Target detection with semisupervised kernel orthogonal subspace projection," IEEE Trans. Geosci. Remote Sens., vol. 47, pp. 3822-3833, 2009.

[20] D. Zhang, J. Han, G. Cheng, Z. Liu, S. Bu, and L. Guo, "Weakly Supervised Learning for Target Detection in Remote Sensing Images," IEEE Geosci. Remote Sens. Lett., vol. 12, pp. 701-705, April 2015.

[21] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning," IEEE Trans. Geosci. Remote Sens., in press.

[22] P. Siva and T. Xiang, "Weakly supervised object detector learning with model drift detection," in ICCV, 2011, pp. 343-350.

[23] Z. Shi, T. M. Hospedales, and T. Xiang, "Bayesian joint topic modelling for weakly supervised object localisation," in ICCV, 2013.

[24] P. Siva, C. Russell, and T. Xiang, "In defence of negative mining for annotating weakly labelled data," in ECCV 2012.

[25] X. Li, C. G. Snoek, M. Worring, and A. W. Smeulders, "Social negative bootstrapping for visual categorization," in ACM ICMR, 2011.

[26] X. Li, C. G. Snoek, M. Worring, D. Koelma, and A. W. Smeulders, "Bootstrapping visual categorization with relevant negatives," IEEE Trans. Multimedia, vol. 15, pp. 933-945, 2013.

[27] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in ACM MM, 2014, pp. 675-678.

[28] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background Prior Based Salient Object Detection via Deep Reconstruction Residual," IEEE Trans. Circuits Syst. Video Technol, in Press.

[29] Y. Feng, J. Ren, and J. Jiang, "Object-based 2D-to-3D video conversion for effective stereoscopic content generation in 3D-TV applications," IEEE Trans. Broadcasting, vol. 57, pp. 500-509, 2011.

[30] J. Ren and J. Jiang, "Hierarchical modeling and adaptive clustering for real-time summarization of rush videos," IEEE Trans. Multimedia, vol. 11, pp. 906-917, 2009.

[31] J. Zabalza, J. Ren, M. Yang, Y. Zhang, J. Wang, S. Marshall, *et al.*, "Novel Folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," ISPRS J. Photogramm. Remote Sens., vol. 93, pp. 112-122, 2014.

[32] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, p. 27, 2011.

[33] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in ACM ICIVR, 2007, pp. 401-408.

[34] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in ECCV, 2004.

[35] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in CVPR, 2010.