Incorporating the Completeness and Difficulty of Proposals Into Weakly Supervised Object Detection in Remote Sensing Images

Xiaoliang Qian, *Member, IEEE*, Yu Huo, Gong Cheng, *Member, IEEE*, Xiwen Yao, *Member, IEEE*, Ke Li, Hangli Ren, and Wei Wang

***Abstract*—Weakly supervised object detection (WSOD) in remote sensing images (RSI) only require image-level labels to detect various objects. Most of the WSOD methods incline to capture the most discriminative parts of object rather than the entire object, and the number of easy and hard samples is imbalanced. To address the first problem, a novel metric named objectness score (OS) is proposed and incorporated into the training loss of our WSOD model. The OS is consisted of the traditional class confidence score (CCS) and the object completeness prior score (OCPS). The CCS can provide the probability that a proposal belongs to a certain class, and the OCPS can quantify the completeness that a proposal covers the entire object. Therefore, the samples which cover the entire object with high class confidences will be assigned large weight in the training loss through OS. To handle the second problem, a novel metric named difficulty evaluation score (DES) is proposed and also incorporated into the training loss. The DES is calculated by using the entropy of confidence score vector of each proposal and is used to quantify how difficult a proposal can be identified correctly, consequently, the hard samples will also be assigned large weight in the training loss through DES. The ablation experiments on two RSI datasets verify the effectiveness of the proposed OS and DES. The comprehensive quantitative and subjective evaluations demonstrate that our method inclines to detect the entire object accurately, and surpasses seven state-of-the-art WSOD methods.**

***Index Terms*—Difficulty evaluation score (DES), object completeness prior score, remote sensing image (RSI), weakly supervised object detection (WSOD).**

I. Introduction

OBJECT detection is a key task for the understanding of remote sensing images (RSI), and has been widely applied in various practical tasks [1]–[10]. Thanks to the development of deep learning [11]–[18], the performance of object detection in RSI has made a significant improvement [19], [20]; however, most of the existing object detection methods in RSI are fully supervised [21]–[23], which require the class and bounding box annotations of objects, apparently, acquiring these annotations is expensive and time-consuming. To reduce the labeling costs, some weakly supervised object detection (WSOD) methods are proposed, which only require image-level class labels during training. Comparing with the bounding box annotations, the cost of image-level labeling is greatly reduced, consequently, the WSOD in RSI has become an important research direction [24]–[26].

Most of the existing methods considered that each image was consisted of a bag of latent object instances, then the WSOD task was converted into a multiple instance learning problem [27]–[33]. Based on aforementioned works, two-stage WSOD methods [34]–[36] were proposed and achieve better performance. The online instance classifier refinement (OICR) [34] was a milestone work among them. Specifically, the deep feature vectors of proposals were fed into the basic multiple instance detector [31] to train an image-level classifier. Then, multiple instance classifier refinement (ICR) branches were simultaneously trained, and the pseudo labels of each ICR branch were produced by its upper ICR branch except the first ICR branch of which the pseudo labels were derived from the basic multiple instance detector.

Although existing WSOD methods have made substantial progress, two challenges still restrict the further improvement. First of all, most of the existing WSOD methods usually select the following proposals as positive samples: The proposals with the highest confidence score in each class and the proposals which have high overlap with the aforementioned proposals. So the first challenge is that the predicted results are inclined to capture the most discriminative parts of object rather than the entire object, especially for the RSI with complex background regions.

The second challenge is that the number of hard and easy samples is imbalanced. As a matter of fact, the number of easy samples is much more than that of hard samples, therefore, the accumulated loss of easy samples will dominate the total loss even if the loss of each easy sample is small. Consequently,

Manuscript received December 1, 2021; revised February 6, 2022; accepted February 8, 2022. Date of publication February 14, 2022; date of current version March 2, 2022. This work was supported in part by the National Science Foundation of China under Grant 62076223, Grant 41871322, and Grant 62003311, in part by the Key R&D Program of Shaanxi Province under Grant 2021ZDLGY01-08, and in part by the Natural Science Foundation of Henan Province under Grant 202300410503. *(Corresponding authors: Gong Cheng; Wei Wang.)*

Xiaoliang Qian, Yu Huo, Hangli Ren, and Wei Wang are with the College of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China (e-mail: qxl_sunshine@163.com; 1828598588@qq.com; 2019050@email.zzuli.edu.cn; wangwei-zzuli@zzuli.edu.cn).

Gong Cheng and Xiwen Yao are with the School of Automation, NorthWestern Polytechnical University, Xi'an 710072, China (e-mail: gcheng@nwpu.edu.cn; yaoxiwen@nwpu.edu.cn).

Ke Li is with the Zhengzhou Institute of Surveying and Mapping, Zhengzhou 450052, China (e-mail: like19771223@163.com).

Digital Object Identifier 10.1109/JSTARS.2022.3150843

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

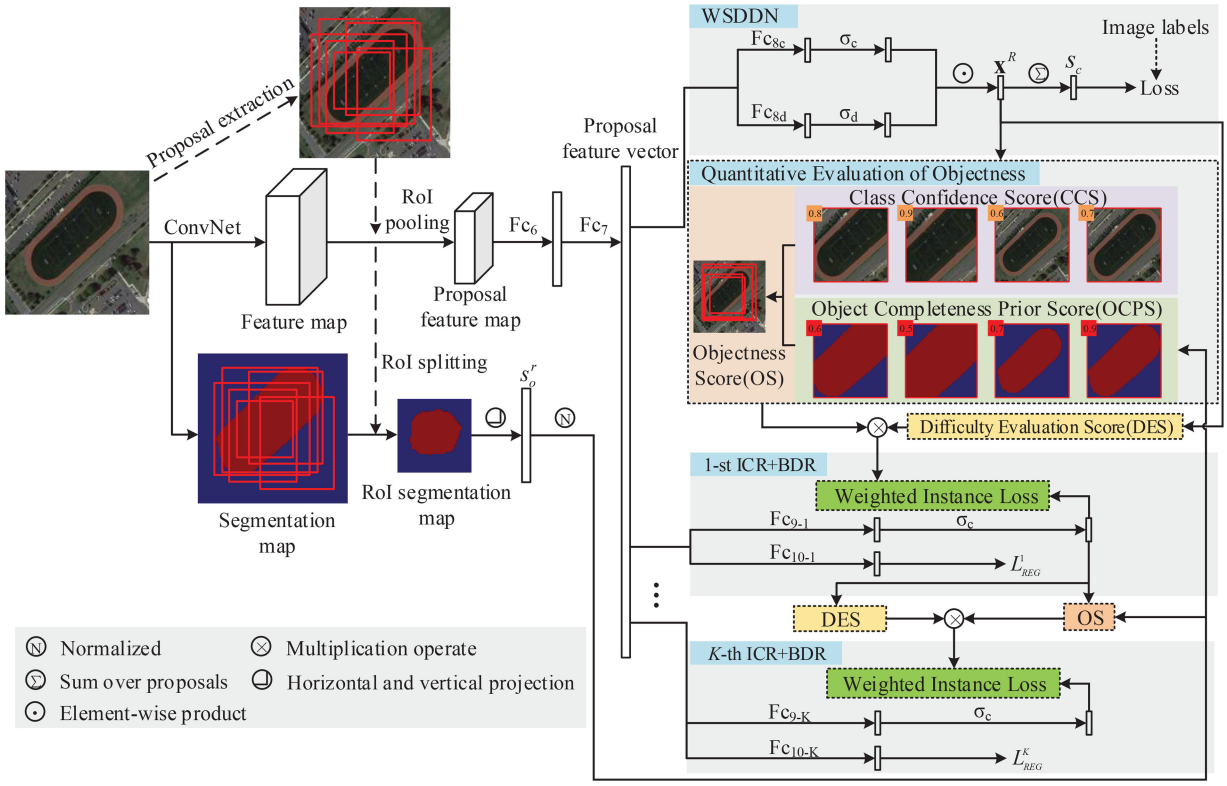


Fig. 1. Framework of our WSOD model. Two novel quantitative metrics: OS and DES are introduced and incorporated into the training loss of each ICR branch.

the training of WSOD model will guided by the easy samples, which will restrict the capability of detecting hard instances.

To address the first challenge, a novel quantitative metric named objectness score (OS) is proposed to comprehensively evaluate the objectness of proposals, which contains two parts: Traditional class confidence score (CCS) and the object completeness prior score (OCPS). The CCS can provide the probability that each proposal covers certain class of object, and the OCPS is used to quantify the completeness that each proposal cover the entire object. Our WSOD model in RSI is inclined to detect the entire object by incorporating the OS into the training loss.

To handle the second challenge, a novel entropy-based quantitative metric named difficulty evaluation score (DES) is proposed to enhance the capability of detecting the hard instances. As a matter of fact, the confidence score vector of each proposal can be regarded as a discrete probability density distribution (PDF) which represents the probabilities that a proposal belongs to each class, and the entropy of the PDF is employed to evaluate the difficulty that each proposal can be correctly identified (except for the hard negative instance, the details can be seen in Section III-B). Consequently, the DES will assign large weight to hard samples in the training loss in order to increase the proportion of accumulate loss of hard samples.

In summary, the contributions of this article can be summarized as follows.

1) A novel objectness score is proposed and incorporated into the training loss to address the problem that the existing WSOD models incline to capture the most discriminative parts of object rather than the entire object. The OS is consisted of traditional CCS and OCPS, where OCPS is used to quantify the completeness that a proposal covers the entire object, consequently, the OS can assign large weight to the samples which can cover entire object by introducing the OCPS.
2) A novel difficulty evaluation score is proposed and incorporated into the training loss to overcome the problem that the number of hard and easy samples is imbalanced. The DES is derived from the entropy of confidence score vector of a proposal (except for the hard negative instances), and can quantify the difficulty that a proposal can be identified correctly, consequently, the hard samples can be assigned large weight in training loss by using DES.

II. Related Work

A. Online Instance Classifier Refinement Network

The OICR [34] is a milestone work, which integrated multiple ICR branches into the weakly supervised deep detection networks (WSDDN) [31] to enhance the performance of WSOD. The WSDDN is used to provide initial supervised information for the subsequent ICR branches. The main contents of WSDDN and ICR branches are presented as follows.

As shown in Fig. 1, first of all, a set of proposals $R = \{rp_1, rp_2, \ldots, rp_{|R|}\}$ are generated from the input image through the selective search (SS) method [37], where $|R|$ denotes

the cardinality of *R*, i.e., the number of proposals. Second, the input image and proposals are fed into the backbone network to obtain the feature vector of each proposal which are output from the fully connected layer Fc7. Third, the feature vectors of proposals are fed into two branches: Classification stream and detection stream to acquire two matrices $\mathbf{x}^c, \mathbf{x}^d \in \mathbb{R}^{C\times|R|}$ outputted from the Fc8c and Fc8d, where *C* denotes the number of classes. The classification and detection score matrices $\sigma_{\rm c}(\mathbf{x}^c), \sigma_{\rm d}(\mathbf{x}^d) \in \mathbb{R}^{C\times|R|}$ are then obtained by using the softmax operator, which can be formulated as follows:

$$\sigma_{\rm c}(\mathbf{x}^c)_{ij} = \frac{e^{x^c_{ij}}}{\sum_{k=1}^{C} e^{x^c_{kj}}}, \sigma_{\rm d}(\mathbf{x}^d)_{ij} = \frac{e^{x^d_{ij}}}{\sum_{b=1}^{|R|} e^{x^d_{ib}}} \tag{1}$$

where $\sigma_{\rm c}(\mathbf{x}^c)_{ij}$ and $\sigma_{\rm d}(\mathbf{x}^d)_{ij}$ denote the classification and detection score of *j* proposal in class *i*, respectively. The final score matrix of total proposals $\mathbf{x} \in \mathbb{R}^{C\times|R|}$ are obtained through the following equations:

$$\mathbf{x} = \sigma_{\rm c}(\mathbf{x}^c) \odot \sigma_{\rm d}(\mathbf{x}^d) \tag{2}$$

where $\odot$ denotes the elementwise product. Then, the image-level score of class *c* (S_c) can be obtained as follows:

$$S_c = \sum_{r=1}^{|R|} x_{cr} \tag{3}$$

where $x_{cr} \in \mathbf{x}$ denotes the confidence score that proposal *r* belongs to class *c*. Finally, the WSDDN is optimized by the cross entropy loss

$$L_B = -\sum_{c=1}^{C} (y_c \log S_c + (1-y_c)\log(1-S_c)) \tag{4}$$

where L_B denotes training loss of WSDDN, $y_c \in \{0,1\}$ denotes the image-level label of class *c*. The $y_c = 1$ if the image contains the objects belonging to class *c*, otherwise, $y_c = 0$.

Afterward, the *K* ICR branches are combined with the WSDDN. Specifically, the feature vectors of proposals are first sent to *K* ICR branches to obtain *K* score matrices $\{\mathbf{x}^k \in \mathbb{R}^{(C+1)\times|R|}\}_{k=1}^{K}$, where the (*C*+1)th row of $\mathbf{x}^k$ denotes the score of background. The *k*th ICR branch is consisted of a fully connected layer and a softmax classifier, which is supervised by the instance-level pseudo labels derived from the $(k-1)$th ICR branch. The deriving of pseudo labels are as follows.

The OICR marked the proposal j_c which has highest score in class *c* as positive sample of class *c*. The proposals which are highly overlapped with proposal j_c are also marked as positive samples. If the intersection over union (IoU) between any proposal and proposal j_c lies in [0.1, 0.5), then the proposal is marked as negative sample. To discover more positive samples, the strategy of mining pseudo samples proposed by MIST [38] is adopted in this article, which selected the proposals of which the scores lies in top *p* percent and the IoU between each other was less than a defined threshold as the positive samples. Similarly, the proposals which were highly overlapped with above positive samples were also marked as positive samples.

Using above pseudo labels, the *k*th ICR branch can be optimized through the cross entropy loss

$$L^k_{\rm ICR} = -\frac{1}{|R|}\sum_{r=1}^{|R|}\sum_{c=1}^{C+1} w^k_r y^k_{cr} \log x^k_{cr} \tag{5}$$

where $L^k_{\rm ICR}$ denotes the loss function of *k*th ICR branch, $x^k_{cr} \in \mathbf{x}^k$ denotes the predicting score that proposal *r* belongs to class *c* for *k*th ICR branch, y^k_{cr} denotes the pseudo labels derived from the $(k-1)$th ICR branch, $w^k_r = x^{k-1}_{cj^{k-1}_c}$ denotes the weight of loss function, proposal j^{k-1}_c has highest score in class *c* for $(k-1)$th ICR branch.

In the inference stage, the mean output of *K* ICR branches is adopted as the predicting result.

B. Integrating the Bounding Box Regression Branches Into the OICR Network

Some existing works [38], [39] demonstrate that mining pseudo ground truth boxes and regressing them can significantly improve the performance of WSOD. Consequently, the *K* bounding box regression (BDR) and *K* ICR branches are used in pairs in this article. Following the fast region-based convolutional neural network (R-CNN) [40], the loss function of *k*th BDR branch $L^k_{\rm REG}$ is defined as follows:

$$L^k_{\rm REG} = \frac{1}{|R_{\rm pos}|}\sum_{r=1}^{|R_{\rm pos}|} {\rm smooth}_{L1}(t_r, \hat{t}_r) \tag{6}$$

where $R_{\rm pos}$ denotes the assembly of positive samples, $|R_{\rm pos}|$ denotes cardinality of $R_{\rm pos}$, i.e., the number of positive samples, ${\rm smooth}_{L1}(\cdot)$ denotes the smooth L1 loss function [40], t_r denotes the predicting offset of locations and sizes of proposal *r*, and $\hat{t}_r$ denotes the true offset of locations and sizes.

The OICR network with BDR branches are denoted as OICR+BDR, and is adopted as the baseline model of the proposed method.

III. Proposed Method

As shown in Fig. 1, the popular OICR architecture [34] is adopted as the basic architecture of our WSOD model. Two novel quantitative metrics: OS and DES are proposed and incorporated into the training loss of each ICR branch. The OS contains two parts, one is traditional CCS, and another one is OCPS which is used to quantify the completeness of each proposal. The samples which cover the entire object with high class confidences can be assigned large weight in the training loss by using the OS. The DES which is calculated through the entropy of distribution of confidence scores over classes, is used to evaluate the difficulty of each proposal. The hard samples can be assigned higher weight by incorporating the DES into the training loss, consequently, our WSOD model gives more attention on the hard samples. The details of OS, DES, and the training loss are presented as follows.

A. Objectness Score of Proposals

The formulation of OS is as follows:

$$OS_r^k = \beta \text{CCS}_r^k + (1-\beta)\text{OCP}S_r \tag{7}$$

where $OS_r^k, \text{CCS}_r^k \in [0,1]$ denote the OS and CCS of proposal *r* for *k*th ICR branch, respectively, CCS_r^k is equal to w_r^k with reference to (5), $\text{OCPS}_r \in [0,1]$ denotes the OCPS of proposal *r*, $\beta \in [0,1]$ is a weight parameter for controlling the ratio between CCS and OCPS.

1) *Object Completeness Prior Score (OCPS):* The OCPS can be calculated by using the contextual information of semantic segmentation [33], [41], [42]. The semantic segmentation algorithm [43] is first imposed on the input image to obtain the *C* segmentation maps corresponding to the *C* classes. Afterward, the contextual information of each proposal is extracted from the *C* segmentation maps [42], and is employed to calculate the OCPS. The details are as follows.
   First of all, the *C* segmentation maps of proposal *r* are split from the *C* segmentation maps of input image according to the spatial location of proposal *r*. The total projection score of proposal *r* along row and column direction S_p^r is calculated using following equation:

$$\begin{aligned} S_p^r &= \text{Avg}(S_x^r) + \text{Avg}(S_y^r) \\ S_x^r &= \text{Max}_x(S_r),\ S_y^r = \text{Max}_y(S_r) \end{aligned} \tag{8}$$

   where $S_r \in \mathbb{R}^{M\times N\times C}$ denotes the segmentation map of proposal *r*, and the *M* and *N* denote the height and width of proposal. $\text{Max}_x(\cdot)$ and $\text{Max}_y(\cdot)$ denote the maximizing operation along the row and column directions of S_r, respectively. $S_x^r \in \mathbb{R}^{1\times N\times C}$ and $S_y^r \in \mathbb{R}^{M\times 1\times C}$ denote the projection score of proposal *r* along row and column direction, respectively. $\text{Avg}(\cdot)$ denotes the averaging operation. Similarly, the projection score of the surrounding region of proposal *r*, denoted as SuS_p^r, can also be calculated following the routing of (8) to mine the contextual information, where the size of surrounding region is 120% of the size of proposal. The initial OCPS of proposal *r*, denoted as S_o^r, can be obtained through the following equation:

$$S_o^r = S_p^r - SuS_p^r. \tag{9}$$

   The final OCPS of proposal *r* is the normalized version of S_o^r

$$\text{OCP}S_r = \frac{S_o^r - \text{Min}\{S_o^r\}_{r=1}^{|R|}}{\text{Max}\{S_o^r\}_{r=1}^{|R|} - \text{Min}\{S_o^r\}_{r=1}^{|R|}} \tag{10}$$

   where $\text{Max}\{\cdot\}$ and $\text{Min}\{\cdot\}$ denote the maximizing and minimizing operation, respectively.
   The motivation behind the calculating scheme of OCPS is that a high-quality proposal should compactly cover an object. It can be seen from (9) and (10) that the OCPS_r is proportional to the S_p^r and inversely proportional to the SuS_p^r. If a proposal is oversize, the background pixels will reduce the value of S_p^r. On the contrary, if a proposal is small than target box, the pixels belonging to foreground objects will increase the value of SuS_p^r. Consequently, only the OCPS of high-quality proposals can obtain large value. Above calculating scheme of OCPS is not the only choice, the proposed schemes by Diba *et al.* [33] and Wei *et al.* [41], etc., can also be used for the same purpose.
2) *Weight Parameter* β : The β is used to control the ratio between CCS and OCPS, the value of β is defined as follows:

$$\beta = \begin{cases} \frac{n}{N}, & n < T \\ \frac{T}{N}, & n \geq T \end{cases} \tag{11}$$

   where n, N, and T denote the current number, total number, and threshold number of iterations. The motivation of (11) is as follows. In the preliminary stage of training, the CCS predicted by ICR branch is not reliable, therefore, the OCPS should take the dominant status in the OS, i.e., the β should be assigned a small value., The reliability of CCS increases gradually with the training going on, consequently, the value of β should be increased accordingly. The β is assigned a fixed value when the number of iterations exceed the threshold *T*.

B. Difficulty Evaluation Score of Proposals

The motivation of designing the DES is to deal with the problem that the number of easy and hard samples is imbalanced. The accumulating loss of hard samples can be increased by incorporating the DES into the training loss. The definition of DES is as follows:

$$\begin{aligned} \text{DES}_r^k &= \begin{cases} e^{\text{entropy}(\mathbf{x}_r^k)}, & r \in R \,\&\, r \notin R_{\text{hns}} \\ e^{1-x_{(C+1)r}^k}, & r \in R_{\text{hns}} \end{cases} \\ \text{entropy}(\mathbf{x}_r^k) &= -\sum_{c=1}^{C+1} x_{cr}^k \log(x_{cr}^k) \end{aligned} \tag{12}$$

where DES_r^k denotes the DES of proposal *r* for *k*th ICR branch, $\mathbf{x}_r^k \in \mathbb{R}^{C+1}$ denotes the confidence score vector of proposal *r* for *k*th ICR branch, $x_{(C+1)r}^k$ denotes the confidence score of proposal *r* in background class for *k*th ICR branch, $\text{entropy}(\cdot)$ denotes the Shannon entropy, and R_{hns} denotes the assemble of hard negative samples. The details of (12) is given as follows.

1) *Proposal is Not a Hard Negative Sample:* The $\mathbf{x}_r^k$ can be considered as a discrete PDF of proposal *r* over each class. The PDF has unimodal characteristics if proposal *r* can be easily identified, i.e., the entropy of the PDF is small, on the contrary, the entropy of the PDF is large. Consequently, the entropy of $\mathbf{x}_r^k$ can be used to evaluate the difficulty of proposal *r*.
2) *Proposal is a Hard Negative Sample:* The hard negative sample is defined as follows: A proposal has the highest score in class *c*, however, the labels of input image does not contain the class *c* [24], [36]. Following the analysis of previous paragraph, the hard negative samples will be considered as the easy samples in the view of entropy, therefore, the entropy-based solution is not appropriate for the hard negative samples. As shown in (12), the confidence scores in background class are directly used to

Fig. 2. Subjective comparisons between baseline and baseline+OS.

calculate the DES of hard negative samples in this article. As a matter of fact, the confidence score of proposal r in background class is small if proposal r is a hard negative sample, i.e., the value $e^{1-x^k_{(C+1)r}}$ is large.

C. Training Loss

The proposed OS and DES defined in (7) and (12) are incorporated into the training loss of each ICR branch defined in (5) in the form of weight

$$L^k_{\text{ICR}-OS-\text{DES}} = -\frac{1}{|R|}\sum_{r=1}^{|R|}\sum_{c=1}^{C+1} OS^k_r \text{DES}^k_r y^k_{cr} \log x^k_{cr} \quad (13)$$

where $L^k_{\text{ICR}-OS-\text{DES}}$ denotes the final training loss of kth ICR branch.

The training loss of K ICR and BDR branches defined in (13) and (6) are combined with the basic training loss of WSDDN defined in (4) to form the final training loss of our WSOD model L

$$L = L_B + \sum_{k=1}^{K}(L^k_{\text{ICR}-OS-\text{DES}} + L^k_{\text{REG}}). \quad (14)$$

IV. Experiments

A. Experiment Setup

1) *Datasets:* The proposed method is evaluated on two RSI datasets: NWPU VHR-10.v2 [44], [45] and DIOR [21] datasets. The NWPU VHR-10.v2 dataset contains 10 object classes, 2775 object instances, and 1172 images with the size of 400 × 400, and is divided into three subsets: The training, validation, and testing sets, which include 679, 200, and 293 images, respectively. The DIOR dataset contains 20 object classes, 192472 object instances, and total 23463 images, and is also divided into three subsets: The training, validation, and testing sets, which include 5862, 5863, and 11738 images, respectively. Following the general setting of WSOD, only image-level labels of training and validation sets are utilized for training.
2) *Evaluation Metrics:* The mean average precision (mAP) and correct localization (CorLoc) [46] are utilized to evaluate the proposed method on testing and training sets, respectively. Following the setting of PASCAL VOC, the predicting bounding box is considered to be localized correctly if the IoU between ground truth box and it exceeds 0.5.
3) *Implementation Details:* The architecture of OICR [34] is adopted as the basic architecture of our WSOD model of which the backbone network is VGG-16 model [47] pretrained on ImageNet [48]. The number of ICR branch is 3, i.e., K=3, and each ICR branch is followed by a BDR branch. The strategy proposed by MIST [38] for mining the positive and negative pseudo samples is adopted by our method, and the top p percent and IoU threshold are set to 0.15 and 0.12, respectively. About 2000 proposals are obtained from each image by utilizing SS method [37]. The nonmaximum suppression (NMS) [49] strategy is employed to remove redundant bounding boxes, and the threshold of IoU is set to 0.3 [25], [26].

Our WSOD model is optimized by stochastic gradient descent (SGD) strategy with the momentum of 0.9 and the weight decay of 0.0001. The initial learning rate and batch size are set to 0.01 and 8, respectively. For NWPU VHR-10.v2 dataset, the number of iterations is 30 K, and decay the learning rate to one tenth of previous value at 20 and 26 K iterations. For DIOR dataset, the number of iterations is 60 K, and decay the learning rate to one tenth of previous value at 50 and 56 K iterations. The threshold T defined in (11) is set to 25 000 and 50 000 on NWPU VHR-10.v2 and DIOR dataset, respectively. All training samples are augmented by horizontally flip, rotating 90° and rotating 180° [24], [25]. The input image is resized to multiscale {480, 576, 688, 864, 1200} in both training and testing stage, and the longest side is restricted to 2000 [31], [34].

All the experiments are implemented on PyTorch framework, and running on the eight GPUs platform (NVIDIA TITAN RTX, i.e., a total of 8 × 24 GB memory).

TABLE I
Ablation Experiments on NWPU VHR-10.V2 and DIOR Test Sets in Terms of mAP

Method	NWPU VHR-10.v2	DIOR
Baseline (OICR+BDR)	0.5016	0.2013
Baseline+OS	0.5686	0.2584
Baseline+DES	0.5854	0.2617
Baseline+OS+DES(Ours)	**0.6149**	**0.2752**

Bold entities denote best results.

B. Ablation Experiments

Ablation experiments are implemented to validate the effectiveness of the proposed OS and DES. As shown in Table I, the OICR+BDR model presented in Section II-B is adopted as the baseline model of proposed method. The baseline model with OS alone is denoted as Baseline+OS. The baseline model with DES alone is denoted as Baseline+DES. The Baseline+OS+EDS denotes the proposed method. Above four methods are quantitatively compared with each other on the NWPU VHR-10.v2 and DIOR datasets to demonstrate the effectiveness of proposed OS and DES.

As shown in the second and third rows of Table I, the baseline method can achieve an absolute increase of 6.7% and 8.38% on the NWPU VHR-10.v2 and DIOR datasets, respectively,

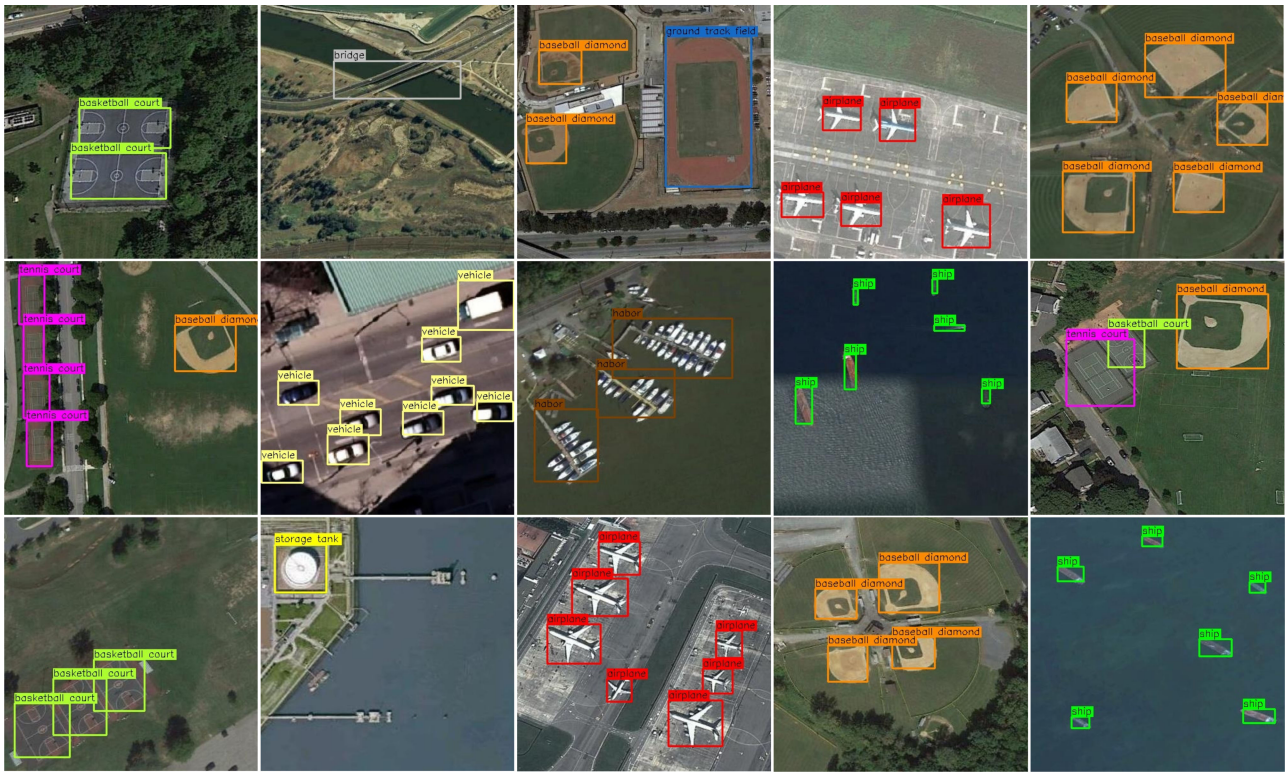


Fig. 3. Illustration of some detection results given by our method on NWPU-VHR-10.v2 testing set.

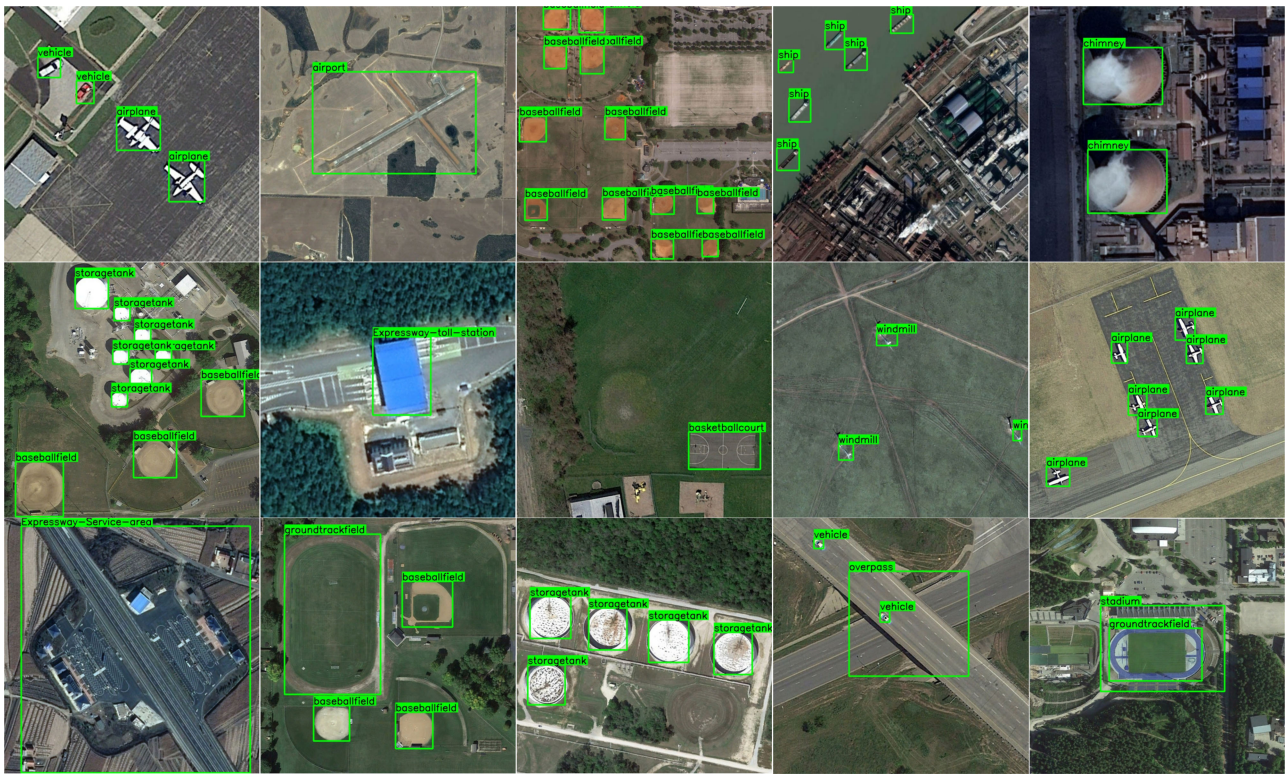


Fig. 4. Illustration of some detection results given by our method on DIOR testing set.

TABLE II
COMPARISON WITH STATE-OF-THE-ART METHODS ON NWPU VHR-10.V2 DATASET IN TERMS OF MAP

Method	Airplane	Ship	Storage tank	Baseball Diamond	Tennis Court	Basketball Court	Ground Track field	Harbor	Bridge	Vehicle	mAP
Fast R-CNN [40]	0.9091	0.9060	0.8929	0.4732	1.0000	0.8585	0.8486	0.8822	0.8029	0.6984	0.8272
Faster R-CNN [50]	0.9090	0.8630	0.9053	0.9824	0.8972	0.6964	1.0000	0.8011	0.6149	0.7814	0.8451
WSDDN [31]	0.3008	0.4172	0.3498	0.8890	0.1286	0.2385	0.9943	0.1394	0.0192	0.0360	0.3512
OICR [34]	0.1366	0.6735	0.5716	0.5516	0.1364	0.3966	0.9280	0.0023	0.0184	0.0373	0.3452
PCL [35]	0.2600	0.6376	0.0250	0.8980	0.6445	0.7607	0.7794	0.0000	0.0130	0.1567	0.3941
MIST [38]	0.6968	0.4916	0.4855	0.8091	0.2708	0.7985	0.9134	0.4699	0.0829	0.1336	0.5152
DCL [24]	0.7270	0.7425	0.3705	0.8264	0.3688	0.4227	0.8395	0.3957	0.1682	0.3500	0.5211
PCIR [25]	0.9078	0.7881	0.3640	0.9080	0.2264	0.5216	0.8851	0.4236	0.1174	0.3549	0.5497
TCA [26]	0.8943	0.7818	0.7842	0.9080	0.3527	0.5036	0.9091	0.4244	0.0411	0.2830	0.5882
Ours	0.8164	0.6833	0.6531	0.9344	0.3643	0.8154	0.9867	0.5377	0.0986	0.2587	**0.6149**

Bold entities denote best results.

TABLE III
COMPARISON WITH STATE-OF-THE-ART METHODS ON DIOR DATASET IN TERMS OF MAP

Method	Airplane	Airport	Baseball field	Basketball court	Bridge	Chimney	Dam	Expressway service area	Expressway toll station	Golf field
Fast R-CNN [40]	0.4417	0.6679	0.6696	0.6049	0.1556	0.7228	0.5195	0.6587	0.4476	0.7211
Faster R-CNN [50]	0.5028	0.6260	0.6604	0.8088	0.2880	0.6817	0.4726	0.5851	0.4806	0.6044
WSDDN [31]	0.0906	0.3968	0.3781	0.2016	0.0025	0.1218	0.0057	0.0065	0.1188	0.0490
OICR [34]	0.0870	0.2826	0.4405	0.1822	0.0130	0.2015	0.0009	0.0065	0.2989	0.1380
PCL [35]	0.2152	0.3519	0.5980	0.2349	0.0295	0.4371	0.0012	0.0090	0.0149	0.0288
DCL [24]	0.2089	0.2270	0.5421	0.1150	0.0603	0.6101	0.0009	0.0107	0.3101	0.3087
MIST [38]	0.3201	0.3987	0.6271	0.2897	0.0746	0.1287	0.0031	0.0514	0.1738	0.5102
PCIR [25]	0.3037	0.3606	0.5422	0.2660	0.0909	0.5859	0.0022	0.0965	0.3618	0.3259
TCA [26]	0.2513	0.3084	0.6292	0.4000	0.0413	0.6778	0.0807	0.2380	0.2989	0.2234
Ours	0.4110	0.4862	0.6748	0.3392	0.0432	0.3471	0.0074	0.1229	0.2433	0.5674

Method	Ground track field	Harbor	Overpass	Ship	Stadium	Storage tank	Tennis court	Train station	Vehicle	Windmill	mAP
Fast R-CNN [40]	0.6293	0.4618	0.3803	0.3213	0.7098	0.3504	0.5827	0.3791	0.1920	0.3810	0.4998
Faster R-CNN [50]	0.6700	0.4386	0.4687	0.5848	0.5237	0.4235	0.7952	0.4802	0.3477	0.6544	0.5548
WSDDN [31]	0.4253	0.0466	0.0106	0.0070	0.6303	0.0395	0.0606	0.0051	0.0455	0.0114	0.1326
OICR [34]	0.5739	0.1066	0.1106	0.0909	0.5929	0.0710	0.0068	0.0014	0.0909	0.0041	0.1650
PCL [35]	0.5636	0.1676	0.1105	0.0909	0.5762	0.0909	0.0247	0.0012	0.0455	0.0455	0.1819
DCL [24]	0.5645	0.0505	0.0265	0.0909	0.6365	0.0909	0.1036	0.0002	0.0727	0.0079	0.2019
MIST [38]	0.4948	0.0536	0.1224	0.2943	0.3553	0.2536	0.0081	0.0459	0.2222	0.0080	0.2218
PCIR [25]	0.5851	0.0860	0.2163	0.1209	0.6428	0.0909	0.1362	0.0030	0.0909	0.0752	0.2492
TCA [26]	0.5385	0.2484	0.1106	0.0909	0.4640	0.1374	0.3098	0.0147	0.0909	0.0100	0.2582
Ours	0.6355	0.0536	0.2311	0.2134	0.5744	0.2466	0.0085	0.0997	0.1834	0.0154	**0.2752**

Bold entities denote best results.

by incorporating the OS into the training loss of ICR branches, which can validate the effectiveness of OS. As a matter of fact, the OS is consisted of CCS and OCPS, and the CCS has been introduced in OICR, consequently, the main reason of improvement lies in the proposed OCPS. As shown in the second and fourth row, the baseline method can achieve an absolute increase of 5.71% and 6.04% on the NWPU VHR-10.v2 and DIOR datasets, respectively, which demonstrates that incorporating the DES into training loss is effective.

The combination scheme of OS and DES is also proved to be effective by comparing the last row with previous three rows in Table I. The subjective comparisons between Baseline and Baseline+OS on NWPU VHR-10.v2 dataset can intuitively demonstrate the effectiveness of OS. As shown in Fig. 2, the Baseline+OS method inclined to capture the entire object guided by OS.

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

To evaluate the overall performance of the proposed method, two state-of-the-art fully supervised object detection (FSOD) methods and seven WSOD methods are compared with the proposed method on NWPU VHR-10.v2 and DIOR datasets. The two classic FSOD methods include fast R-CNN [40] and faster R-CNN [50], and the seven WSOD methods include WSDDN [31], OICR [34], PCL [35], DCL [24], MIST [38], PCIR [25], and TCA [26].

1) *Quantitative Evaluation:* The quantitative comparisons on NWPU VHR-10.v2 and DIOR datasets in terms of mAP is shown in Tables II and III, respectively. The overall performance of our method is obviously superior to seven existing WSOD methods on two datasets. Specifically, our method exceeds the WSDDN, OICR, PCL, MIST, DCL, PCIR, and TCA by 26.37%, 26.97%, 22.08%, 9.97%, 9.38%, 6.52%, and 2.67%, respectively, on NWPU VHR-10.v2 dataset, and exceeds them by 14.26%, 11.02%, 9.33%, 7.33%, 5.34%, 2.6%, and 1.7%, respectively, on DIOR dataset. In addition, the proposed method has narrowed the gap between WSOD and FSOD methods, and the performance of our method is comparable with FSOD methods in certain class, such as Baseball Diamond and

TABLE IV
COMPARISON WITH STATE-OF-THE-ART METHODS ON NWPU VHR-10.V2 AND DIOR TRAINING SETS IN TERMS OF CORLOC

Method	NWPU VHR-10.v2	DIOR
WSDDN [31]	0.3524	0.3244
OICR [34]	0.4001	0.3477
PCL [35]	0.4506	0.4152
DCL [24]	0.6965	0.4223
MIST [38]	0.7034	0.4357
PCIR [25]	0.7187	0.4612
TCA [26]	0.7276	0.4841
Ours	**0.7368**	**0.4992**

Bold entities denote best results.

Basketball Court in Table II, and Airplane and Ground track field in Table III.

In addition, it can be found that the mAP scores on DIOR dataset are obviously lower than that on NWPU VHR-10.v2 dataset since the DIOR dataset is more challenging than NWPU VHR-10.v2 dataset.

The quantitative comparisons on two datasets in terms of CorLoc is also implemented to evaluate the localization accuracy of our method. As shown in Table IV, the overall performance of our method is better than seven WSOD methods on both datasets. Specifically, our method exceeds the WSDDN, OICR, PCL, MIST, DCL, PCIR, and TCA by 38.44%, 33.67%, 28.62%, 4.03%, 3.34%, 1.81%, and 0.92%, respectively on NWPU VHR-10.v2 dataset, and exceeds them by 17.48%, 15.15%, 8.4%, 7.69%, 6.35%, 3.8%, and 1.51%, respectively on DIOR dataset.

On the whole, the proposed method gets the best performance on two RSI datasets comparing with seven state-of-the-art WSOD methods.

2) *Subjective Evaluations:* To intuitively demonstrate the effectiveness of our method, some detection results given by our method on NWPU VHR-10.v2 and DIOR testing sets are illustrated in Figs. 3 and 4, respectively.

As shown in Fig. 3, the detected objects of different class on NWPU VHR-10.v2 dataset are enclosed by bounding boxes with different color. As shown in Fig. 4, the detected objects on DIOR dataset are enclosed by bounding boxes with green color. The class of each detected object is shown in the top left corner of bounding box. Obviously, our method inclines to detect the entire object accurately under the condition of complex background.

V. CONCLUSION

A novel metric OS is proposed to overcome the problem that the existing WSOD methods in RSI inclined to discover the most discriminative parts of object rather than the entire object. The OS is consisted of traditional CCS and OCPS, and is incorporated into the training loss of our WSOD model, where OCPS is used to quantify the completeness that each proposal covers the entire object. Therefore, the samples which can cover the entire object with high class confidence will be assigned large weight in the training loss through OS. Another novel metric DES is proposed to handle the problem that the number of easy and hard samples is imbalanced. The DES can quantitatively evaluate how difficult a proposal can be identified correctly, and is also incorporated into the training loss of our WSOD model. Consequently, the proportion of accumulating loss of hard samples in total training loss will be increased. The ablation experiments on NWPU VHR-10.v2 and DIOR datasets validate that incorporating the OS and DES can apparently improve the performance of baseline method. The comprehensive quantitative comparisons with state-of-the-art methods on two RSI datasets show that the overall performance of the proposed method surpasses the seven state-of-the-art WSOD methods. The subjective evaluations intuitively demonstrate that our method inclines to detect the entire objects accurately.

As mentioned before, the OCPS plays an important role in proposed method, and the accuracy of OCPS relies on the quality of segmentation maps. To increase the accuracy of OCPS, exploring a more robust weakly supervised segmentation algorithm is our future work.

REFERENCES

[1] Q. Chen, M. Huang, and H. Wang, "A feature discretization method for classification of high-resolution remote sensing images in coastal areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8584–8598, May 2021.

[2] X. Yao, Q. Cao, X. Feng, G. Cheng, and J. Han, "Scale-aware detailed matching for few-shot aerial image semantic segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, 2022, Art. no. 5611711.

[3] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens*, vol. 59, no. 5, pp. 4340–4354, Aug. 2020.

[4] B. Tu, W. Kuang, W. He, G. Zhang, and Y. Peng, "Robust learning of mislabeled training samples for remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5623–5639, Sep. 2020.

[5] X. Qian *et al.*, "Generating and sifting pseudolabeled samples for improving the performance of remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4925–4933, Aug. 2020.

[6] Y. Chen, D. Ming, X. Ling, X. Lv, and C. Zhou, "Landslide susceptibility mapping using feature fusion-based CPCNN-ML in Lantau Island, Hong Kong," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3625–3639, Mar. 2021.

[7] B. Tu, X. Yang, X. Ou, G. Zhang, J. Li, and A. Plaza, "Ensemble entropy metric for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5513617.

[8] L. Du *et al.*, "Multi-modal deep learning for landform recognition," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 63–75, Dec. 2019.

[9] X. Qian, J. Li, J. Cao, Y. Wu, and W. Wang, "Micro-cracks detection of solar cells surface via combining short-term and long-term deep features," *Neural Netw*, vol. 127, pp. 132–140, Jul. 2020.

[10] B. Tu, X. Yang, C. Zhou, D. He, and A. Plaza, "Hyperspectral anomaly detection using dual window density," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8503–8517, May 2020.

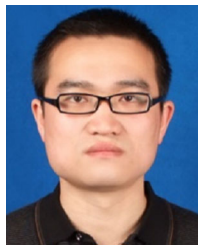
[11] D. Hong *et al.*, "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Nov. 2022, Art. no. 5518615..

[12] W. Kuo, B. Hariharan, and J. Malik, "Deepbox: Learning objectness with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2479–2487.

[13] 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.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2014.

[14] D. Ming, J. Li, J. Wang, and M. Zhang, "Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example," *ISPRS J. Photogramm. Remote Sens.*, vol. 106, pp. 28–41, Aug. 2015.

[15] J. Han *et al.*, "Representing and retrieving video shots in human-centric brain imaging space," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2723–2736, Apr. 2013.
[16] X. Qian, X. Cheng, G. Cheng, X. Yao, and L. Jiang, "Two-stream encoder GAN with progressive training for co-saliency detection," *IEEE Signal Process. Lett.*, vol. 28, pp. 180–184, Jan. 2021.
[17] B. Tu, C. Zhou, D. He, S. Huang, and A. Plaza, "Hyperspectral classification with noisy label detection via superpixel-to-pixel weighting distance," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4116–4131, Jan. 2020.
[18] M. Huang *et al.*, "Remote sensing image fusion algorithm based on two-stream fusion network and residual channel attention mechanism," *Wireless Commun. Mobile Comput.*, vol. 2022, Jan. 2022, Art. no. 8476000.
[19] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, Jul. 2016.
[20] G.-S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.
[21] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogramm. Remote Sens.*, vol. 159, pp. 296–307, Jan. 2020.
[22] X. Qian, S. Lin, G. Cheng, X. Yao, H. Ren, and W. Wang, "Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion," *Remote Sens.*, vol. 12, no. 1, pp. 143–163, Jan. 2020.
[23] X. Han, Y. Zhong, and L. Zhang, "An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery," *Remote Sens.*, vol. 9, no. 7, pp. 666–687, Jun. 2017.
[24] X. Yao, X. Feng, J. Han, G. Cheng, and L. Guo, "Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 675–685, May 2020.
[25] X. Feng, J. Han, X. Yao, and G. Cheng, "Progressive contextual instance refinement for weakly supervised object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens*, vol. 58, no. 11, pp. 8002–8012, Apr. 2020.
[26] X. Feng, J. Han, X. Yao, and G. Cheng, " TCANET: Triple context-aware network for weakly supervised object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens*, vol. 59, no. 8, pp. 6946–6955, Oct. 2021.
[27] W. Ren, K. Huang, D. Tao, and T. Tan, "Weakly supervised large scale object localization with multiple instance learning and bag splitting," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 38, no. 2, pp. 405–416, Jul. 2015.
[28] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic annotation of high-resolution satellite images via weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3660–3671, Feb. 2016.
[29] X. Wang, Z. Zhu, C. Yao, and X. Bai, "Relaxed multiple-instance SVM with application to object discovery," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1224–1232.
[30] R. G. Cinbis, J. Verbeek, and C. Schmid, "Weakly supervised object localization with multi-fold multiple instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 189–203, Feb. 2016.
[31] H. Bilen and A. Vedaldi, "Weakly supervised deep detection networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2846–2854.
[32] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Nov. 2018.
[33] A. Diba, V. Sharma, A. Pazandeh, H. Pirsiavash, and L. Van Gool, "Weakly supervised cascaded convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 914–922.
[34] P. Tang, X. Wang, X. Bai, and W. Liu, "Multiple instance detection network with online instance classifier refinement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2843–2851.
[35] P. Tang *et al.*, "PCL: Proposal cluster learning for weakly supervised object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 176–191, Oct. 2018.
[36] G. Cheng, J. Yang, D. Gao, L. Guo, and J. Han, "High-quality proposals for weakly supervised object detection," *IEEE Trans. Image Process.*, vol. 29, pp. 5794–5804, Apr. 2020.
[37] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Apr. 2013.
[38] Z. Ren *et al.*, "Instance-aware, context-focused, and memory-efficient weakly supervised object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10598–10607.
[39] Z. Huang, Y. Zou, V. Bhagavatula, and D. Huang, "Comprehensive attention self-distillation for weakly-supervised object detection," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 16797–16807.
[40] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.
[41] Y. Wei *et al.*, "TS2C: Tight box mining with surrounding segmentation context for weakly supervised object detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 434–450.
[42] K. Yang *et al.*, "Rethinking segmentation guidance for weakly supervised object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 946–947.
[43] J. Ahn, S. Cho, and S. Kwak, "Weakly supervised learning of instance segmentation with inter-pixel relations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2209–2218.
[44] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens*, vol. 54, no. 12, pp. 7405–7415, Sep. 2016.
[45] K. Li, G. Cheng, S. Bu, and X. You, "Rotation-insensitive and context-augmented object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens*, vol. 56, no. 4, pp. 2337–2348, Dec. 2017.
[46] T. Deselaers, B. Alexe, and V. Ferrari, "Weakly supervised localization and learning with generic knowledge," *Int. J. Comput. Vis.*, vol. 100, no. 3, pp. 275–293, May 2012.
[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2012, pp. 1–13.
[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25rd Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[49] J. Hosang, R. Benenson, and B. Schiele, "Learning non-maximum suppression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4507–4515.
[50] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2016.



Xiaoliang Qian (Member, IEEE) received the Ph.D. degree in control science and engineering from the School of Automation, Northwestern Polytechnical University, Xi'an, China, in 2013.

He is currently an Associate Professor with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include remote sensing image understanding, deep learning, and computer vision.

Yu Huo received the B.S. degree in smart grid information engineering from the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China, in 2019, where he is currently working toward the M.S. degree.

His research interests include object detection in remote sensing images and deep learning.

Gong Cheng (Member, IEEE) received the B.S. degree in biomedical engineering from Xidian University, Xi'an, China, in 2007, and the M.S. and Ph.D. degrees in pattern recognition and intelligent systems from Northwestern Polytechnical University, Xi'an, China, in 2010 and 2013, respectively.

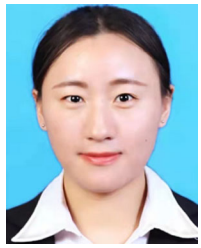
He is currently a Professor with Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.

Xiwen Yao (Member, IEEE) received the B.S. in information security and Ph.D. degrees in pattern recognition and intelligent systems from the Northwestern Polytechnical University, Xi'an, China, in 2010 and 2016, respectively.

He is currently an Associate Professor with Northwestern Polytechnical University. His research interests include computer vision and remote sensing image processing, especially on fine-grained image classification and object detection.

Ke Li received the Ph.D. degree in cartography and geographical information engineering from the Zhengzhou Institute of Surveying and Mapping, Zhengzhou, China, in 2008.

He is currently an Associate Professor with Zhengzhou Institute of Surveying and Mapping. He has authored or coauthored over 30 journal and conference papers in the related areas. His research interests include GIS, deep learning, and computer vision.



Hangli Ren received the Ph.D. degree in mathematics from Qufu Normal University, Qufu, China, in 2019.

She is currently a Lecturer with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. Her research interests include computer vision, switched systems, finite-time control, filtering and event-triggered control.

Wei Wang received the Ph.D. degree in microelectronics from Concordia University, Montréal, QC, Canada, in 2002.

He is currently a Professor with the School of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include artificial intelligence, remote sensing image understanding, computer vision, and tactile sensing.