electronics

MDPI

Article

High-Quality Instance Mining and Dynamic Label Assignment for Weakly Supervised Object Detection in Remote Sensing Images

Li Zeng , Yu Huo *, Xiaoliang Qian and Zhiwu Chen *

College of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China

* Correspondence: yuhuo_henry2022@163.com (Y.H.); chenyyj@163.com (Z.C.)

Abstract: Weakly supervised object detection (WSOD) in remote sensing images (RSIs) has attracted more and more attention because its training merely relies on image-level category labels, which significantly reduces the cost of manual annotation. With the exploration of WSOD, it has obtained many promising results. However, most of the WSOD methods still have two challenges. The first challenge is that the detection results of WSOD tend to locate the significant regions of the object but not the overall object. The second challenge is that the traditional pseudo-instance label assignment strategy cannot adapt to the quality distribution change of proposals during training, which is not conducive to training a high-performance detector. To tackle the first challenge, a novel high-quality seed instance mining (HSIM) module is designed to mine high-quality seed instances. Specifically, the proposal comprehensive score (PCS) that consists of the traditional proposal score (PS) and the proposal space contribution score (PSCS) is designed as a novel metric to mine seed instances, where the PS indicates the probability that a proposal pertains to a certain category and the PSCS is calculated by the spatial correlation between top-scoring proposals, which is utilized to evaluate the wholeness with which a proposal locates an object. Consequently, the high PCS will encourage the WSOD model to mine the high-quality seed instances. To tackle the second challenge, a dynamic pseudo-instance label assignment (DPILA) strategy is developed by dynamically setting the label assignment threshold to train high-quality instances. Consequently, the DPILA can better adapt the distribution change of proposals according to the dynamic threshold during training and further promote model performance. The ablation studies verify the validity of the proposed PCS and DPILA. The comparison experiments verify that our method obtains better performance than other advanced WSOD methods on two popular RSIs datasets.

Keywords: weakly supervised object detection; remote sensing images; proposal comprehensive score; dynamic label assignment

Citation: Zeng, L. ; Huo, Y.; Qian, X.; Chen, Z. High-Quality Instance Mining and Dynamic Label Assignment for Weakly Supervised Object Detection in Remote Sensing Images. *Electronics* **2023**, *12*, 2758. https://doi.org/10.3390/electronics12132758

Academic Editor: George A. Tsihrintzis

Received: 23 May 2023
Revised: 17 June 2023
Accepted: 19 June 2023
Published: 21 June 2023

1. Introduction

Object detection in RSIs is a pivotal task of imagery interpretation, its purpose is to identify and locate high-value geographical objects in RSIs. Object detection in RSIs has wide applications in various fields, such as environmental monitoring [1,2], urban planning [3], agriculture [4,5], anomaly detection [6,7], and so on. With the progression of machine learning [8–14], object detection acquires satisfactory performance. The advanced performance is obtained by the fully supervised object detection (FSOD) [15–19] methods. However, the FSOD method needs category and location labels for instances to drive model training. Obviously, manually annotating the location labels for each instance of each RSI is laborious. In order to alleviate the burdensome annotated costs, weakly supervised object detection (WSOD) methods [20,21] have gradually entered the view of researchers because WSOD methods only require image-level category labels to drive model training.

At present, most of the WSOD models are trained based on the paradigm of multiple instance learning (MIL) [22–25]. Specifically, the training image is treated as a bag of latent instances, and then the latent instances are utilized to train the instance detector under the MIL constraints. Among these, a pioneering weakly supervised deep detection network (WSDDN) [26] has been developed, which first introduces MIL into the WSOD model. On the basis of WSDDN, an online instance classifier refinement (OICR) model [27] is developed by adding K instance classifier refinement (ICR) branches, which further improves the performance of the WSOD model. Subsequently, some works have been developed to further enhance the performance of WSOD through employing spatial correlation [28], initialization models [29], collaborative learning [30], etc.

Although the performance of classical WSOD has made significant progress, there are still two main challenges to be solved. The first challenge is that most of the WSOD methods [27,31] merely employ the proposal score (PS) to mine seed instances, however, high PS usually locates the remarkable region of an object but not the overall object. Unfortunately, these methods will obtain worse performance in RSIs with noisy background. The second challenge is that the traditional pseudo-instance labels assignment (PILA) strategy [27,31] cannot adapt to the quality distribution change of proposals during training. Specifically, the traditional PILA strategy sets a fixed label assignment threshold to determine the attribute (i.e., belonging to a positive or negative instance) of each instance. However, along with the training, the fixed threshold setting and dynamic model training are not matched, which is not conducive to training high-quality instances.

In order to tackle the first challenge, a novel high-quality seed instances mining (HSIM) module is designed to mine high-quality seed instances, as shown in Figure 1. Specifically, the proposal comprehensive score (PCS) is first designed and is composed of traditional proposal score (PS) and proposal space contribution score (PSCS). The PS indicates the probability that a proposal pertains to one category; the PSCS is calculated by considering the spatial relationships between top-scoring proposals and is utilized to measure the extent to which the proposal locates an object. Consequently, seed instances mined by PCS can better locate an object than traditional mined strategy, which merely utilize the PS.

In order to tackle the second challenge, an innovative dynamic pseudo instance label assignment (DPILA) strategy is developed to better adapt to the quality distribution change of proposals during training and, meanwhile, raise the number of positive instances in the initial training stage. Specifically, a label assignment threshold is dynamically calculated via elaborately designing a function that increases with the number of iterations. Consequently, the DPILA strategy can dynamically assign pseudo instance label for each instance, and further improves the performance of WSOD.

Our contributions can be summed up as follows. The first contribution is that a novel HSIM module is designed to mine high-quality seed instances. Specifically, a PCS is first designed, which is composed of traditional PS and proposed PSCS, where the PSCS is calculated by considering the spatial relationships between top-scoring proposals to estimate the wholeness with which the proposal locates an object. The seed instances mined by PCS can more completely locate an object than traditional mined strategies, which merely utilizes the PS; The second contribution is that a DPILA strategy is proposed to better adapt to the quality distribution change of proposals during training. Specifically, a dynamic label assignment threshold is defined by elaborately designing a function that increases with the number of iterations. The proposed DPILA strategy can dynamically assign a pseudo-instance label for each instance, which is conducive to model training; The third contribution is that the ablation studies verify the validity of PCS and DPILA. The comparison experiments display that our method obtains higher performance than other advanced WSOD methods on two popular RSIs datasets. Specifically, our method surpasses separately the state-of-the-art WSDDN, OICR, PCL, and MELM methods by 12.2% (8.3%), 12.8% (5.1%), 7.9% (3.4%) and 5.0% (2.9%) in terms of mAP on the NWPU VHR-10.v2 (DIOR) dataset, and surpasses them by 23.2% (11.9%), 18.4% (9.5%), 13.3% (2.8%) and 8.5% (1.0%) in terms of CorLoc on the NWPU VHR-10.v2 (DIOR) dataset.

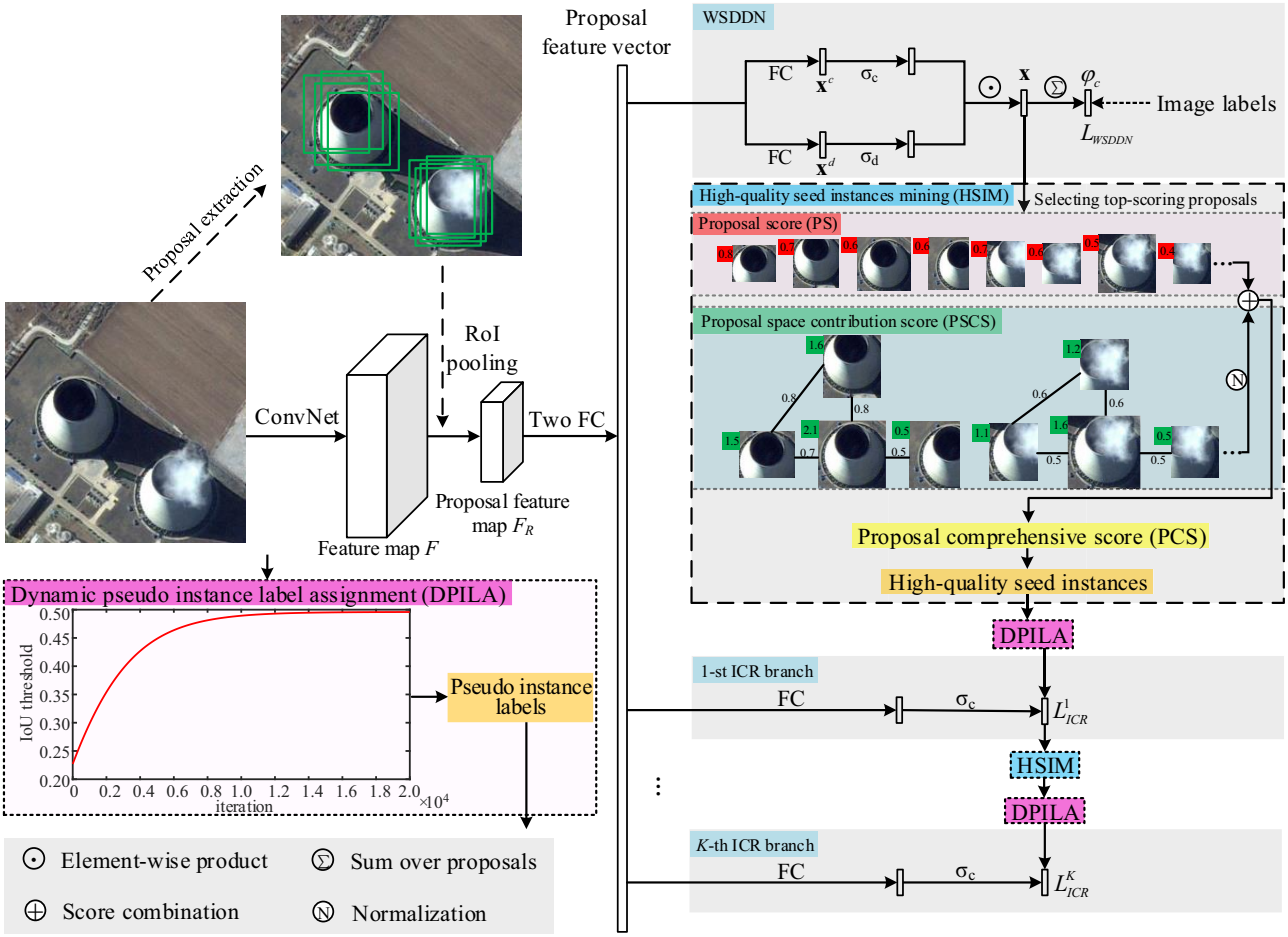

Figure 1. The overall framework of our method, which is established on the OICR network [27] by introducing two proposed modules including high-quality seed instance mining (HSIM) module and dynamic pseudo instance labels assignment (DPILA) strategy. Here, the HSIM is designed to mine high-quality seed instances. The DPILA strategy is proposed to better adapt to the quality distribution change of proposals during training.

2. Related Work

2.1. State-of-the-Art Weakly Supervised Object Detection Methods

Fully supervised object detection (FSOD) methods have achieved satisfactory performance. However, it needs category and location labels to drive model training, which is time-consuming to annotate with these precise labels. WSOD methods, which only require image-level labels to drive model training, have gradually entered the view of researchers. For example, Feng et al. [32] proposed a progressive contextual instance refinement strategy that can highlight more object parts and relieve the part domination problem. Yao et al. [33] proposed a dynamic curriculum learning strategy to robustly improve the performance. Feng et al. [34] proposed a triple context-aware network that can learn complementary and discriminative features and improve the performance of WSOD. Chen et al. [30] introduced the collaborative learning strategy into the WSOD model to improve its performance of WSOD. Feng et al. [35] proposed a self-supervised adversarial and equivariant network, that could learn complementary and consistent instance features, and promote the performance of WSOD. Chen et al. [36] proposed a full-coverage collaborative network, which could enhance the ability of multiscale feature extraction for WSOD detector.

2.2. Pseudo Instance Labels Mining

There are no instance-level labels to drive the model training in the WSOD. Therefore, it is a challenge to mine pseudo-instance labels for each instance. The current mainstream

pseudo-instance labels mining strategy can be divided into two steps, namely, seed instances mining and pseudo-instance label assignment. The details of the two steps are as follows.

2.2.1. Seed Instances Mining

Most of the seed instance mining strategies [27,37,38] select the proposal with the highest score in category c as seed instance. However, the strategy ignores the plain fact that RSIs usually contain multiple instances in the same category, and it is unreasonable to only select the proposal with the highest score as the seed instance in category c. Therefore, some improvements have been proposed. For instance, Tang et al. [39] use the k-means method to split the proposals into several clusters according to proposal score, select the proposal with the highest score in each cluster, and then utilize graph-based method to choose multiple seed instances with same category. Lin et al. [40] consider that the same category instance should have a similar appearance feature. Specifically, by selecting the highest-score proposal as a seed instance in category c, then calculating the similarity between the seed instance and other instances, if the similarity of a certain proposal is greater than the pre-set threshold, the proposal is selected as another seed instance. Cheng et al. [41] proposed a self-guided proposal generation strategy to generate directly high-quality seed instances. Qian et al. [42] proposed a novel seed instance mining strategy by employing the supplemental segmentation information. Ren et al. [31] sort all of the proposals from high to low according to the PS of existing categories in an image and then select proposals with the top p% score as the candidate seed instances. Finally, a similar non-maximum suppression (NMS) [43] operation is utilized to choose ultimate seed instances.

2.2.2. Pseudo-Instance Labels Assignment

Most of the WSOD methods [27,31,39,44] assign a pseudo-instance labels according to the fixed labels assignment threshold. Concretely, suppose an image contains category label c, the seed instance R_{si} belonging to category c can be mined according to the abovementioned methods. Furthermore, the R_{si} is labeled category c, i.e., $y^k_{cR_{si}} = 1$ and $y^k_{c'R_{si}} = 0, c \neq c'$, where k indicates the k-th ICR branch. Inspired by the reality that the proposals that have high spatial coverage with the seed instance should be assigned the same label. Specifically, if the maximum intersection over union (IoU) between a certain proposal and seed instances is greater than the fixed label assignment threshold of 0.5, then the proposals as neighbor positive instances are also labelled to category c and denote it to R_{npi}, namely, $y^k_{cR_{npi}} = 1$ and $y^k_{c'R_{npi}} = 0, c \neq c'$, otherwise the proposals are labelled to background instance and denote it to R_{bi}, namely, $y^k_{(C+1)R_{bi}} = 1$ and $y^k_{cR_{bi}} = 0, c \neq C+1$.

However, aforementioned methods merely employ the PS to mine seed instances, which leads to the mined instances inclining to locate discriminative regions of objects rather than overall objects. In addition, the fixed label assignment strategy cannot adapt to the quality distribution change of proposals, which is not conducive to training high-quality instances. These are also the problems to be solved in this paper.

3. Materials and Methods

As shown in Figure 1, the OICR framework [27] is employed as the baseline framework of the proposed method. On the basis of OICR, a novel high-quality seed instance mining (HSIM) module is designed to mine high-quality seed instances. Specifically, the PCS is first designed, which is composed of traditional PS and PSCS. The PS indicates the probability that a proposal pertains to a certain category; the PSCS is calculated by considering the spatial relationships between top-scoring proposals, which is utilized to measure the extent to which the proposal locates an object. In addition, a novel dynamic pseudo instance labels assignment (DPILA) strategy is proposed to better adapt to the quality distribution change of proposals during training and, meanwhile, raise the number of positive instances in the

initial training stage. Specifically, a label assignment threshold is dynamically calculated by elaborately designing a function that increases with the number of iterations.

3.1. Basic Weakly Supervised Object Detection Network

Bilen et al. [26] put forward a path-breaking weakly supervised deep detection network (WSDDN), which is the footstone of WSOD. The details of the WSDDN are as follows. Firstly, preset an image I and image-level category labels $Y = [y_1, \ldots y_c, \ldots, y_C]$, where $y_c \in \{1, 0\}$ denotes present or absent object category c in an image, and C expresses the quantity of object category. For each image, a range of proposals $R = \{r_1, r_2, \ldots, r_{|R|}\}$ are produced via employing edge boxes (EB) [45] or selective search (SS) [46] algorithms, where $|R|$ expresses the quantity of proposals. Secondly, the feature maps $F \in \mathbb{R}^{W \times H \times C}$ are obtained by sending the image I into the convolutional network (ConvNet), where C, H, and W indicate the channels, height, and width of the feature maps F. Thirdly, the feature maps F and the proposals R are sent into the region of interest (RoI) pooling layer to obtain the proposal feature maps F_R with a fixed size. Fourthly, the proposal feature vectors are obtained via two fully connected (FC) layers. These proposal feature vectors are then sent into two side-by-side branches, i.e., classification branch and detection branch, to produce two matrices $\mathbf{x}^c, \mathbf{x}^d \in \mathbb{R}^{C \times |R|}$ through respective FC layers. The classification score and detection score of each proposal are obtained by performing a softmax operation on the two matrices $\mathbf{x}^c$, $\mathbf{x}^d$ along different directions; the details are as follows:

$$[\sigma(\mathbf{x}^c)]_{cr} = \frac{e^{x^c_{cr}}}{\sum_{c'=1}^{C} e^{x^c_{c'r}}}, [\sigma(\mathbf{x}^d)]_{cr} = \frac{e^{x^d_{cr}}}{\sum_{r'=1}^{|R|} e^{x^d_{cr'}}} \tag{1}$$

where $[\sigma(\mathbf{x}^c)]_{cr}$ indicates the probability that the proposal r pertains to category c, $[\sigma(\mathbf{x}^d)]_{cr}$ represents the dedication of the proposal r to category c. The 'dedication' indicates the contribution of a proposal r to the image being classified in category c. Therefore, the $[\sigma(\mathbf{x}^d)]_{cr}$ also belongs to the probability to a certain extent; namely, the higher the $[\sigma(\mathbf{x}^d)]_{cr}$ value, the greater the probability of belonging to a positive instance. The proposal score is calculated via element-wise product between $\sigma(\mathbf{x}^c)$ and $\sigma(\mathbf{x}^d)$, which is denoted as follows:

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

where $\mathbf{x} \in \mathbb{R}^{C \times |R|}$ represents the proposal score. Furthermore, image-level prediction score φ_c of category c can be acquired by the sum of all proposals as follows:

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

Finally, the loss function L_{WSDDN} of WSDDN is defined as follows:

$$L_{WSDDN} = -\sum_{c=1}^{C} (y_c \log \varphi_c + (1 - y_c) \log(1 - \varphi_c)) \tag{4}$$

where $y_c \in \{1, 0\}$ expresses the image-level category label, which indicates present or absent object category c in an image.

To further promote the performance of the WSOD model, Tang et al. [27] introduced multi-stage instance classifier refinement (ICR) branches to improve the WSOD network. Specifically, we added K parallel ICR branches on the WSDDN, and each ICR branch consists of a FC layer and a softmax layer, and the output $(C + 1)$ dimension score matrix $\mathbf{x}^k \in \mathbb{R}^{(C+1) \times |R|}$, where $k \in 1, 2, \ldots, K$, and the $(C + 1)$-th dimension denotes background. The k-th ICR branch is supervised through the previous $(k - 1)$-th branch, excluding the

1-st ICR branch from WSDDN (i.e., $\mathbf{x}$). Finally, K ICR branches are trained by utilizing the cross-entropy loss, which is formulated as follows:

$$L_{ICR}^{k} = -\frac{1}{|R|}\sum_{r=1}^{|R|}\sum_{c=1}^{C+1} w_r^k y_{cr}^k \log x_{cr}^k \tag{5}$$

where the w_r^k denotes the loss weight, the $y_{cr}^k \in \{1, 0\}$ indicates the pseudo instance label. For more details, please refer to [27].

However, most of the existing methods [27,31,39] merely employ the proposal score (PS) of proposal to mine seed instances, where the PS indicates the probability that a proposal pertains to one category. Specifically, the proposal with the highest PS in a certain category is selected as the seed instance. However, the proposal (seed instance) with the highest PS usually locates the remarkable region of object but not the overall object. Therefore, existing methods are not able to mine high-quality seed instances.

3.2. High-Quality Seed Instance Mining Guided by Proposal Comprehensive Score

To overcome the above challenge, the proposal comprehensive score (PCS) is designed, which comprehensively considers the traditional proposal score (PS) and the proposed proposal space contribution score (PSCS). The PSCS is calculated by considering the spatial relationships between top-scoring proposals and is utilized to measure the extent to which the proposal locates an object. Consequently, seed instances mined by PCS can more completely locate an object than the traditional mined strategies, which merely utilize the PS. The details of PCS are as follows.

Firstly, the proposals are sorted from high to low based on their corresponding PS in the existing category. Secondly, the proposals with the top p% PS in category c are selected as top-scoring proposals and defined them as an assembly $R_c' = \{r_1', \ldots, r_n', \ldots, r_N'\}$, where the N expresses the quantity of top-scoring proposals in class c. Thirdly, the PSCS of each top-scoring proposal is calculated pursuant to the spatial relationship between the top-scoring proposals. Fourthly, the PCS is calculated by combining the PS and PSCS, which are defined as follows:

$$PCS_{cn} = \alpha PS_{cn} + (1-\alpha) PSCS_{cn} \tag{6}$$

where PS_{cn} indicates proposal score of the n-th proposal r_n' in category c, $PSCS_{cn}$ denotes the proposal space contribution score of r_n' in category c, α is the hyper-parameter to balance the contribution of PS and PSCS. The details of PSCS are as follows.

The undirected weighted graph $G_c^s = (V_c^s, E_c^s)$ is first constructed according to the spatial correlation of R_c', where the vertexes V_c^s denotes top-scoring proposals, each edge $E_c^s = \{\sigma_c^{nn'}\}$ denotes the spatial correlation between vertexes. As shown in Figure 2, the weight of each edge is obtained via calculating the IoU between vertexes, which is defined as follows:

$$\sigma_c^{r_n' r_{n'}'} = \begin{cases} \mathrm{IoU}(r_n', r_{n'}'), & \text{if } \mathrm{IoU}(r_n', r_{n'}') \geq T \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where the T indicates hyper-parameter, the $\mathrm{IoU}(r_n', r_{n'}')$ indicates the IoU value between r_n' and $r_{n'}'$, $n \neq n'$. Based on this, the $PSCS_{cn}$ can be calculated as follows:

$$PSCS_{cn} = N(\sum_{r_{n'}' \in R_c'} \sigma_c^{r_n' r_{n'}'}), n \neq n' \tag{8}$$

where $N(\cdot)$ indicates the normalization operator. Finally, following the mining strategy [31], the PCS is utilized to mine high-quality seed instances, and denotes them as a assemble $R_c^s = \{r_1^s, \ldots, r_m^s, \ldots, r_M^s\}$, where the M denotes the number of R_c^s in category c.

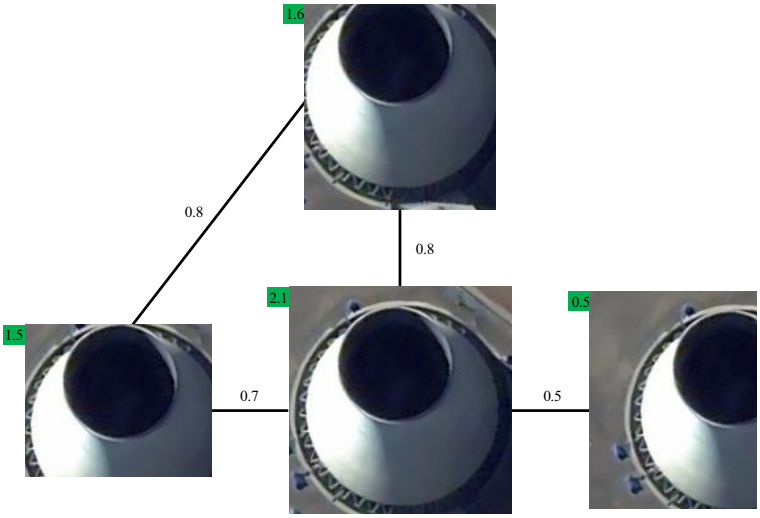

Figure 2. The details of weighted graph. Here, the graph is not undirected but has weighted. Specifically, the vertexes of graph denote top-scoring proposals, each edge denotes the spatial correlation (i.e., IoU) between vertexes.

3.3. Dynamic Pseudo Instance Label Assignment for Each Instance

Most of the WSOD methods usually set a fixed instance label assignment threshold (i.e., IoU value) to determine whether a certain proposal belongs to the positive or negative instance. If the IoU value between the proposal r and its nearest seed instance r_m^s greater than or equal to the default threshold T_{IoU}, the proposal is labeled as a positive instance; otherwise, the proposal is assigned a negative instance. Specifically, the label is defined as follows:

$$\text{label} = \begin{cases} 1, & \text{if IoU}(r, r_m^s) \geq T_{IoU} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $r \notin R_c^s$ indicates a certain proposal, T_{IoU} is a fixed value and usually set to 0.5, which cannot adapt to the quality distribution change of proposals. In addition, setting a high T_{IoU} may lead to the loss of some potential positive instances at the early stage of model training.

To overcome this issue, a dynamic pseudo instance label assignment (DPILA) strategy is proposed. The dynamic means that the label assignment threshold changes as the training progresses. Specifically, a growth function is designed to gradually adjust the IoU threshold as training goes on. The dynamic IoU threshold T_{IoU}^d is defined as follows, and its variation curve is also demonstrated in Figure 3.

$$T_{IoU}^d = \frac{1}{1 + e^{-l \times t - m}} - 0.5 \tag{10}$$

where l and m denote hyper-parameters, t indicates the number of current iterations. Therefore, the label is redefined as follows:

$$\text{label} = \begin{cases} 1, & \text{if IoU}(r, r_m^s) \geq T_{IoU}^d \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

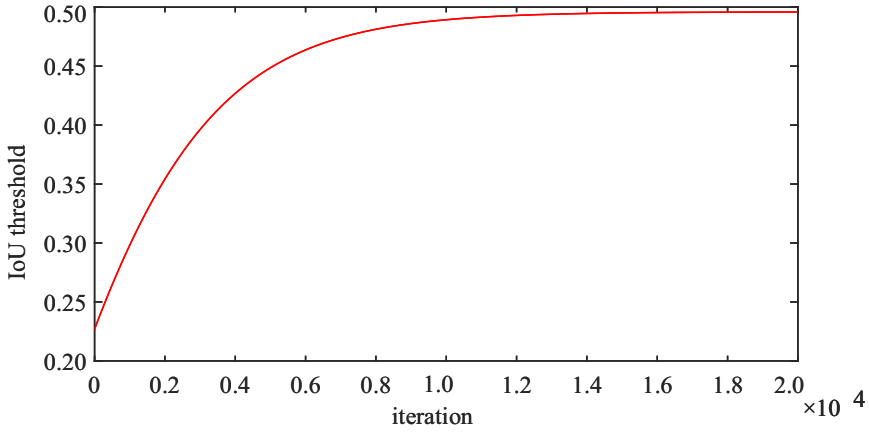

Figure 3. The variation curve of dynamic IoU threshold. The horizontal axis represents the number of iterations, the vertical axis represents the IoU threshold.

During testing, the DPILA strategy is discarded (i.e., all experiment results are from the mean output of 3 ICR branches), and the threshold is a fixed value (i.e., 0.5) following the WSOD criterion [27,31,39].

4. Experiment

4.1. Experiment Setup

4.1.1. Datasets

Extensive experiments are implemented to measure the validity of the proposed methods on the NWPU VHR-10.v2 dataset [47,48] and DIOR dataset [49]. The NWPU VHR-10.v2 dataset comprises 1172 images, each with dimensions of 400 × 400 pixels, which has 879 trainval images and 293 test images and includes 10 object categories and 2775 instances. The DIOR dataset has a greater level of difficulty and includes 23,463 images, each with dimensions of 800 × 800 pixels. The DIOR dataset is partitioned into a trainval set, consisting of 11,725 images, and a testing set, comprising 11,738 images, which includes 20 object categories and 192,472 instances.

4.1.2. Evaluation Metric

We employed two standard metrics to evaluate the performance of our method, which are widely used and accepted evaluation metrics in WSOD, namely, mean average precision (mAP) and correct localization (CorLoc) [50], where mAP evaluates the accuracy of detection on the testing set and CorLoc assesses the accuracy of localization on the trainval set. The two evaluation metrics comply with the PASCAL protocol.

4.1.3. Implementation Details

The OICR network serves as the baseline framework for the proposed method. Similar to refs. [27,39,51], the VGG-16 [52] is utilized as the backbone network, which has undergone pre-training on the large-scale ImageNet dataset [8], in accordance with standard practice. The quantity of ICR branches is configured as 3. Following the standard of WSOD, merely image-level category labels of the trainval set are employed to train our model. We utilized the stochastic gradient descent (SGD) strategy to optimize our WSOD model, configuring values of 0.9 and 0.0001 for the momentum and weight decay hyperparameters, respectively. The initial learning rate and batch size is separately set at 0.01 and 8. We conducted a total of 20K and 60K training iterations on the NWPU VHR-10.v2 and DIOR datasets, respectively. The decay weight of the learning rate is set to 0.1, and the step size are separately set at 18K and 50K iterations on the NWPU VHR-10.v2 and DIOR datasets. The hyper-parameters l, m and p are separately set to 0.0002, 1 and 15. For data augmentation, all training images are augmented via rotating 90°, 180° and horizontal flipping [32,33]. In addition, following the mainstream methods [27,39], the images are resized into five

distinct scales {480, 576, 688, 864, and 1200} for training and testing. Inferential results are post-processed via implementing NMS operation, whose threshold is set at 0.3 [32,39,53,54]. The training details can also be seen in Table 1. The region proposals are generated via using the image segmentation algorithm (i.e., the selective search algorithm [46]). Specifically, the algorithm consists of the following three steps: (1) Initial segmentation: the image is segmented into small regions based on pixel intensity and texture similarity. (2) Similarity measure: all adjacent region pairs are combined and assigned a similarity score based on color, texture, size, and shape differences. (3) Proposals generation: the most similar regions are merged repeatedly until the desired number of proposals is obtained. Following the paradigm of WSOD, about 2000 region proposals are generated via a selective search algorithm. The scale of image segmentation is not fixed, which is determined according to the merger of similar regions in step (3).

All experiments are implemented on 8 TITAN RTX GPUs with the PyTorch framework.

Table 1. The training details of our method, which includes training setting and parameter setting.

	Learning Rate	**Batch Size**	**Momentum**	**Weight Decay**	**Iteration Numbers**
Training Setting	0.01	8	0.9	0.0001	20 K/60 K
Parameter setting	K 3	l 0.0002	m 1	p (%) 15	NMS threshold 0.3

4.2. Parameter Analyses

4.2.1. Parameter Analysis of α

As previously discussed, the parameter α plays a critical role in determining the relative contributions of PS and PSCS. To objectively assess this relationship, we conducted a quantitative analysis of the DIOR dataset. As demonstrated in Figure 4, our approach achieved the highest mAP when α is 0.5. Based on these results, we adopted α = 0.5 as the optimal value for this paper.

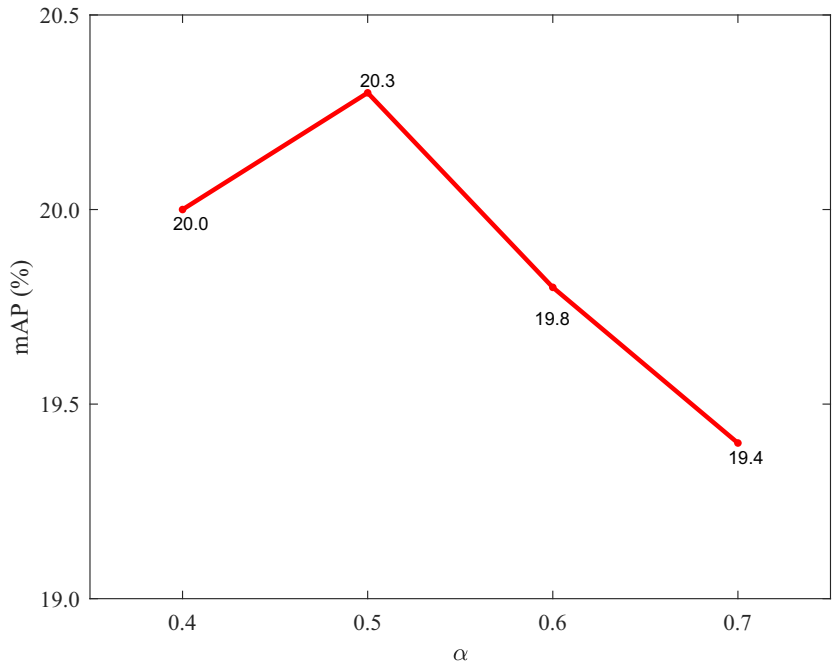

Figure 4. Parameter analysis of α on the DIOR dataset. The horizontal axis represents different α values, the vertical axis represents the mAP values.

4.2.2. Parameter Analysis of T

As mentioned before, T is the threshold to determine the value of $\sigma_c^{r'_n r'_{n'}}$, which is analyzed quantitatively on the DIOR dataset. As demonstrated in Figure 5, our approach achieved the highest mAP when T is 0.7. Based on these results, we adopted $T = 0.7$ as the optimal value for this paper.

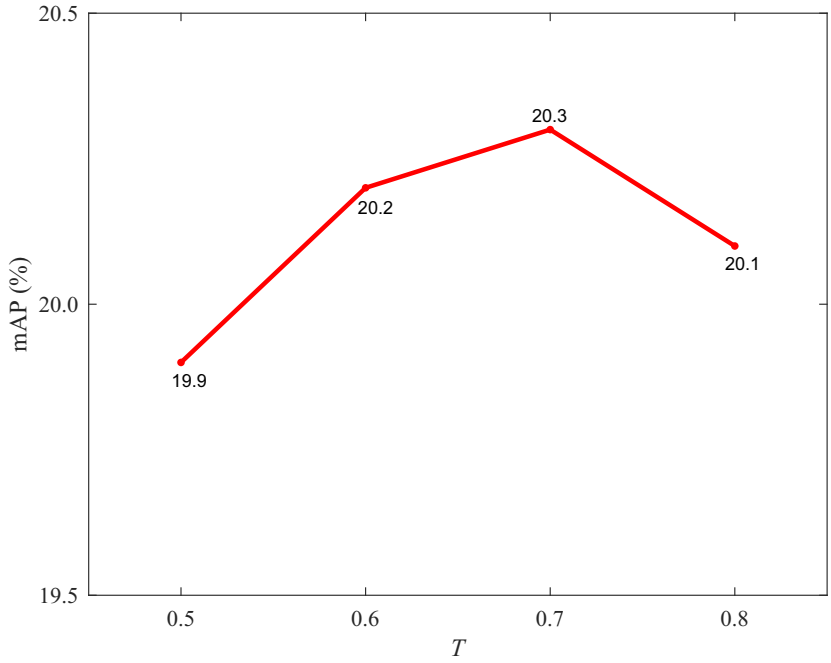

Figure 5. Parameter analysis of T on the DIOR dataset. The horizontal axis represents different T values, the vertical axis represents the mAP values.

4.3. Ablation Studies

Ablation studies are constructed to verify the validity of the proposed PCS and DPILA. Specifically, as shown in Table 2, the baseline, baseline+PCS, baseline+DPILA, and baseline+PCS+DPILA experiments are implemented on the DIOR dataset.

Table 2. Ablation studies of our method on the DIOR dataset.

Baseline (OICR)	PCS	DPILA	DIOR	
			mAP	**CorLoc**
✓			16.5	34.8
	✓		20.3	42.2
		✓	18.9	41.0
	✓	✓	**21.6**	**44.3**

Bold entities denote best results.

4.3.1. Influence of PCS

The baseline+PCS experiment is constructed to validate the influence of the proposed PCS. As shown in Table 2, the baseline+PCS method obtains 20.3% mAP and 42.2% CorLoc on the DIOR dataset, which surpasses the baseline method 3.8% mAP and 7.4% CorLoc. Therefore, the validity of PCS is verified obviously. The major reason for performance enhancement is that the proposed PCS can effectively guide the WSOD model to mine high-quality seed instances, which further encourage model to locate more complete object.

4.3.2. Influence of DPILA

The baseline+DPILA experiment is constructed to validate the influence of the proposed DPILA. As shown in Table 2, the baseline+DPILA method obtains 18.9% mAP and

41.0% CorLoc, which outperforms the baseline method 2.4% mAP and 6.2% CorLoc on the DIOR dataset. Therefore, the validity of DPILA is verified obviously. The major reason for performance enhancement is that the proposed DPILA strategy can adapt to the quality distribution change of proposals during training and mine some potential positive instances at the early stage of model training. Consequently, the DPILA strategy can dynamically assign a pseudo-instance label for each instance, which further improves the performance of WSOD.

The baseline+PCS+DPILA experiment is constructed to verify the influence of the combination of PCS and DPILA. As shown in Table 2, the baseline+PCS+DPILA method obtains 21.6% mAP and 44.3% CorLoc on the DIOR dataset, which outperforms the other three methods. Therefore, the validity of the combination of PCS and DPILA is verified effectively.

4.4. Comparison with Other Advanced WSOD Methods

To further validate the integrated performance of our method, we reported the comprehensive results and provided comparisons with seven WSOD methods and four fully supervised object detection (FSOD) methods on two popular RSIs datasets. Specifically, the 4 WSOD methods, including WSDDN [26], OICR [27], min-entropy latent model (MELM) [53], and proposal cluster learning (PCL) [39], were compared with our method on two RSIs datasets. The other 3 WSOD methods, including dynamic curriculum learning (DCL) [33], full-coverage collaborative Network (FCC-Net) [36], and collaborative learning-based network (CLN) [30], were compared with our method on the DIOR dataset. The 4 FSOD methods include region-based convolutional neural networks (R-CNN) [55], Fast R-CNN [56], Faster R-CNN [57], and rotation-invariant convolutional neural networks (RICNN) [47].

4.4.1. Comparison in Terms of mAP

Tables 3 and 4 demonstrate the comparison in terms of mAP between our approach and other advanced WSOD methods. Specifically, as shown in Table 3, our approach obtains 47.3% mAP on the NWPU VHR-10.v2 dataset. Compared with other advanced WSOD methods, our method significantly exceeds the WSDDN, OICR, PCL, and MELM by 12.2%, 12.8%, 7.9%, and 5.0% in terms of mAP, respectively, on the NWPU VHR-10.v2 dataset. As shown in Table 4, our method obtains 21.6% mAP on the DIOR dataset. Compared with the other advanced WSOD methods, our method significantly exceeds the WSDDN, OICR, PCL, MELM, DCL, FCC-Net and CLN-RSOD methods on the DIOR dataset, with an increase in mAP of 8.3%, 5.1%, 3.4%, 2.9%, 1.4%, 3.3% and 3.3%, respectively. Compared with the FSOD methods, our approach further decreases the performance gap between FSOD method and WSOD method.

Table 3. Comparisons with other advanced methods in terms of AP (%) and mAP (%) on the NWPU VHR-10.v2 dataset.

Method	Airplane	Ship	Storage Tank	Baseball Diamond	Tennis Court	Basketball Court	Ground Track Field	Harbor	Bridge	Vehicle	mAP
R-CNN [55]	85.4	88.9	62.8	19.7	90.7	58.2	68.0	79.9	54.2	49.9	65.8
RICNN [47]	88.7	78.3	86.3	89.1	42.3	56.9	87.7	67.5	62.3	72.0	73.1
Fast R-CNN [56]	90.9	90.6	89.3	47.3	100.0	85.9	84.9	88.2	80.3	69.8	82.7
Faster R-CNN [57]	90.9	86.3	90.5	98.2	89.7	69.6	100.0	80.1	61.5	78.1	84.5
WSDDN [26]	30.1	41.7	35.0	88.9	12.9	23.9	99.4	13.9	1.9	3.6	35.1
OICR [27]	13.7	67.4	57.2	55.2	13.6	39.7	92.8	0.2	1.8	3.7	34.5
PCL [39]	26.0	63.8	2.5	89.8	64.5	76.1	77.9	0.0	1.3	15.7	39.4
MELM [53]	80.9	69.3	10.5	90.2	12.8	20.1	99.2	17.1	14.2	8.7	42.3
Ours	77.9	32.0	48.1	90.9	28.5	62.4	88.6	40.2	1.2	3.6	**47.3**

Bold entities denote best results.

Table 4. Comparisons with other advanced methods in terms of AP (%) and mAP (%) on the DIOR dataset.

Method	Airplane	Airport	Baseball Field	Basketball Court	Bridge	Chimney	Dam	Expressway Service Area	Expressway Toll Station	Golf Field	
R-CNN [55]	35.6	43.0	53.8	62.3	15.6	53.7	33.7	50.2	33.5	50.1	
RICNN [47]	39.1	61.0	60.1	66.3	25.3	63.3	41.1	51.7	36.6	55.9	
Fast R-CNN [56]	44.2	66.8	67.0	60.5	15.6	72.3	52.0	65.9	44.8	72.1	
Faster R-CNN [57]	50.3	62.6	66.0	80.9	28.8	68.2	47.3	58.5	48.1	60.4	
WSDDN [26]	9.1	39.7	37.8	20.2	0.3	12.2	0.6	0.7	11.9	4.9	
OICR [27]	8.7	28.3	44.1	18.2	1.3	20.2	0.1	0.7	29.9	13.8	
PCL [39]	21.5	35.2	59.8	23.5	3.0	43.7	0.1	0.9	1.5	2.9	
MELM [53]	28.1	3.2	62.5	28.7	0.1	62.5	0.2	28.4	13.1	15.2	
DCL [33]	20.9	22.7	54.2	11.5	6.0	61.0	0.1	1.1	31.0	30.9	
FCC-Net [36]	20.1	38.8	52.0	23.4	1.8	22.3	0.2	0.6	28.7	14.1	
CLN [30]	10.1	33.2	43.9	23.4	0.8	38.8	0.7	1.1	19.3	11.6	
Ours	10.5	32.4	64.2	28.0	1.1	13.3	0.3	0.3	29.9	50.9	
Method	**Ground Track Field**	**Harbor**	**Overpass**	**Ship**	**Stadium**	**Storage Tank**	**Tennis Court**	**Train Station**	**Vehicle**	**Windmill**	**mAP**
R-CNN [55]	49.3	39.5	30.9	9.1	60.8	18.0	54.0	36.1	9.1	16.4	37.7
RICNN [47]	58.9	43.5	39.0	9.1	61.1	19.1	63.5	46.1	11.4	31.5	44.2
Fast R-CNN [56]	62.9	46.2	38.0	32.1	71.0	35.0	58.3	37.9	19.2	38.1	50.0
Faster R-CNN [57]	67.0	43.9	46.9	58.5	52.4	42.4	79.5	48.0	34.8	65.4	55.5
WSDDN [26]	42.4	4.7	1.1	0.7	63.0	4.0	6.1	0.5	4.6	1.1	13.3
OICR [27]	57.4	10.7	11.1	9.1	59.3	7.1	0.7	0.1	9.1	0.4	16.5
PCL [39]	56.4	16.8	11.1	9.1	57.6	9.1	2.5	0.1	4.6	4.6	18.2
MELM [53]	41.1	26.1	0.4	9.1	8.6	15.0	20.6	9.8	0.0	0.5	18.7
DCL [33]	56.5	5.1	2.7	9.1	63.7	9.1	10.4	0.0	7.3	0.8	20.2
FCC-Net [36]	56.0	11.1	10.9	10.0	57.5	9.1	3.6	0.1	5.9	0.7	18.3
CLN [30]	48.9	19.6	9.5	13.0	54.5	10.8	10.3	0.5	9.2	6.7	18.3
Ours	55.4	12.4	15.0	34.0	33.9	30.0	1.3	4.1	14.8	0.8	**21.6**

Bold entities denote best results.

4.4.2. Comparison in Terms of CorLoc

Tables 5 and 6 demonstrate the comparison in terms of CorLoc between our approach and other advanced WSOD methods. Specifically, as shown in Table 5, our approach acquires 58.4% CorLoc on the NWPU VHR-10.v2 dataset. Compared with the other advanced WSOD methods, our method significantly exceeds the WSDDN, OICR, PCL, and MELM methods on the NWPU VHR-10.v2 dataset, with an increase in CorLoc of 23.2%, 18.4%, 13.3%, and 8.5%, respectively. As shown in Table 6, our method obtains 44.3% CorLoc on the DIOR dataset. In comparison to other advanced WSOD methods, our approach significantly exceeds the WSDDN, OICR, PCL, MELM, DCL and FCC-Net methods by 11.9%, 9.5%, 2.8%, 1.0%, 2.1%, and 2.6% CorLoc, respectively, on the DIOR dataset.

Table 5. Comparisons with other advanced methods in terms of CorLoc (%) on the NWPU VHR-10.v2 dataset.

Method	WSDDN [26]	OICR [27]	PCL [39]	MELM [53]	Ours
NWPU VHR-10.v2	35.2	40.0	45.1	49.9	**58.4**

Bold entities denote best results.

Table 6. Comparisons with other advanced methods in terms of CorLoc (%) on the DIOR dataset. '-' denotes the CorLoc value has not been reported in their study.

Method	WSDDN [26]	OICR [27]	PCL [39]	MELM [53]	DCL [33]	FCC-Net [36]	CLN [30]	Ours
DIOR	32.4	34.8	41.5	43.3	42.2	41.7	-	**44.3**

Bold entities denote best results.

4.4.3. Subjective Comparison

In addition, to further evaluate our method, Four advanced WSOD methods that provide source codes are subjectively compared with our method on two RSI datasets in

Figures 6 and 7, respectively. Figure 6 shows the visual comparison results on the NWPU VHR-10.v2 dataset, and the objects with different categories are enclosed by utilizing the bounding boxes with different colors. Figure 7 displays the visual comparison results on the DIOR dataset, and the objects are enclosed by utilizing green bounding boxes. What is more, the category of object is attached to the bounding box. As shown in Figures 6 and 7, the detection results of our approach can completely locate and correctly identify objects.

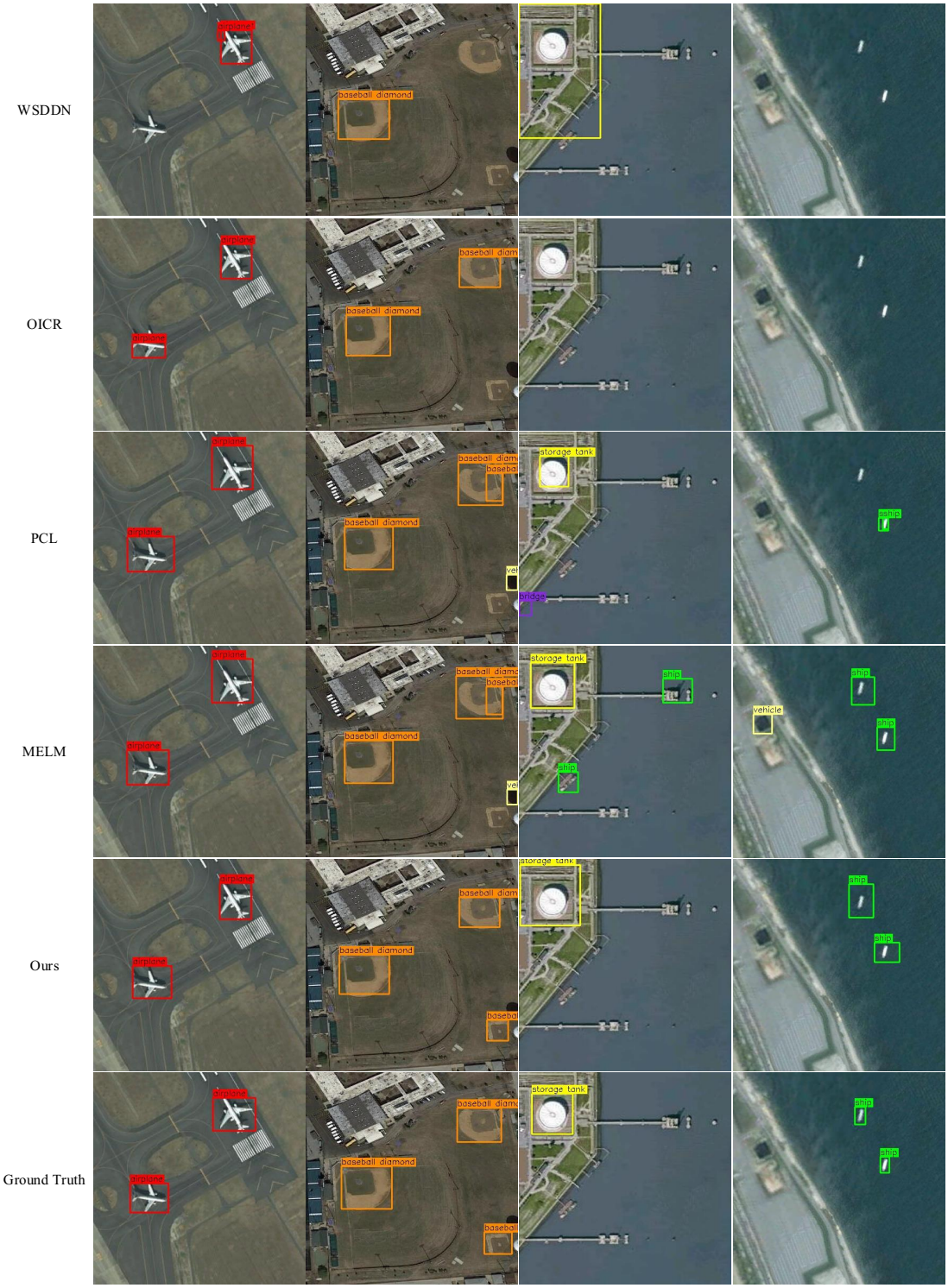

Figure 6. Four advanced WSOD methods that provide source codes are subjectively compared with our method on the NWPU VHR-10.v2 dataset.

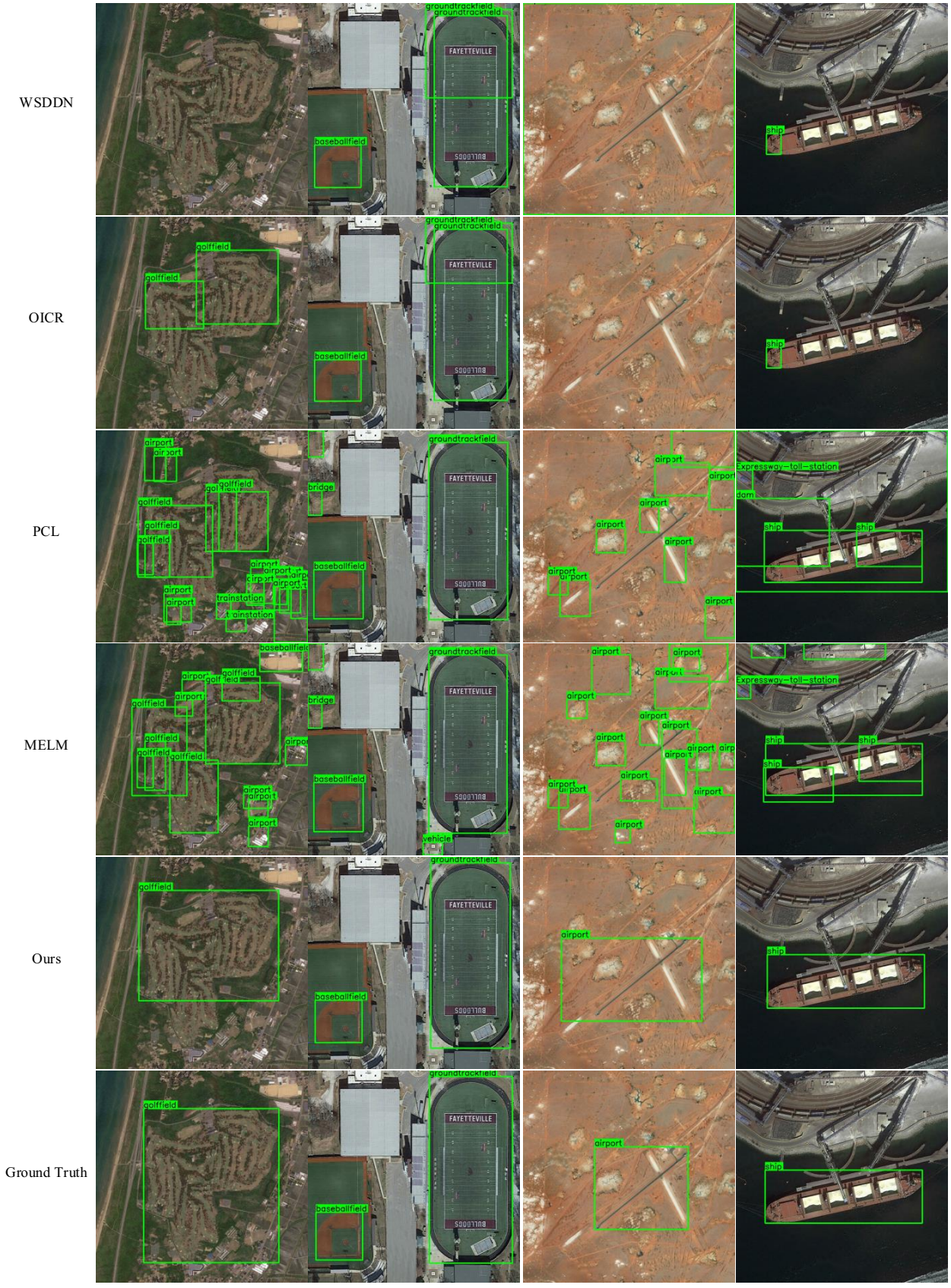

Figure 7. Four advanced WSOD methods that provide source codes are subjectively compared with our method on the DIOR dataset.

4.5. Runtime Analysis

In order to assess the practicality of the proposed approach in real-world scenarios, we further reported the runtime of the proposed method in terms of training and inference. As shown in Table 7, during training, compared with the baseline method, the computational time increases from 24.8 to 30.4 h by incorporating the HSIM into the baseline method. The additional complexity is mainly introduced because HSIM is added. Furthermore, when we incorporate the DPILA into the baseline method, the computational time increased from 24.8 to 25.0 h, which is caused by the calculation of DPILA. During inference, the HSIM module and calculation of DPILA are discarded; namely, all experiment results are from the mean output of 3 ICR branches (as shown in the lower right of Figure 1). Therefore, all methods have the same complexity, which costs the same inference time (i.e., 2.2 h) during inference. Although the training time of the baseline method is less than ours (24.8 versus 30.7 h), its performance is reduced by 5.1% compared with ours.

Table 7. The Complexity analysis of our method on the DIOR Dataset. All experiments are implemented on ubuntu16.04 and NVIDIA TITAN RTX GPU.

Method	Training Time (Hours)	Inference Time (Hours)	mAP (%)
Baseline (OICR)	24.8	2.2	16.5
+HSIM (PCS)	30.4	2.2	20.3
+DPILA	25.0	2.2	18.9
+HSIM+DPILA	30.7	2.2	21.6

5. Discussion

To tackle the first challenge, the detection results of WSOD tend to locate the significant regions of the object but not the overall object. The PCS, which consists of traditional PS and PSCS, is designed as a novel metric to mine high-quality seed instances. To tackle the second challenge, traditional pseudo-instance label assignment strategies cannot adapt to the quality distribution changes of proposals during training, which is not conducive to training a high-performance detector. A DPILA strategy is developed via dynamically setting the label assignment threshold to train high-quality instances. Consequently, collaborating on the proposed PCS with DPILA achieves better performance than other advanced WSOD methods on two popular RSIs datasets. Specifically, our method surpasses separately WSDDN, OICR, PCL, and MELM methods by 12.2% (8.3%), 12.8% (5.1%), 7.9% (3.4%), and 5.0% (2.9%) in terms of mAP on the NWPU VHR-10.v2 (DIOR) dataset, and surpasses separately WSDDN, OICR, PCL, and MELM methods by 23.2% (11.9%), 18.4% (9.5%), 13.3% (2.8%) , and 8.5% (1.0%) in terms of CorLoc on the NWPU VHR-10.v2 (DIOR) dataset.

6. Conclusions

In this paper, a novel HSIM module is designed to tackle the challenge that the detection results of WSOD detector tend to locate the significant regions of an object but not the overall object. Specifically, the PCS is first designed and is composed of traditional PS and proposed PSCS. The PSCS is utilized to evaluate the wholeness with which a proposal locates an object. Consequently, high PCS will encourage the WSOD model to mine high-quality seed instances. A DPILA strategy is developed to tackle the challenge that traditional pseudo-instance label assignment strategies cannot adapt to the quality distribution change of proposals during training. Specifically, a dynamic label assignment threshold is defined by elaborately designing a function that increases with the number of iterations. Consequently, the DPILA strategy can dynamically assign a pseudo instance label for each instance, which further improves the performance of WSOD. The ablation studies verify the validity of the proposed PCS and DPILA. The comparison experiments verify that our approach obtains better performance than other advanced WSOD detectors

on two popular RSIs datasets. The subjective comparison straightforwardly demonstrates that our method can completely locate and correctly identify objects.

The shortcomings of the proposed model are that it achieves poor performance in individual classes such as Dam, Windmill, etc. The possible reason is that our model is susceptible to interference from complex backgrounds. For instance, the Dam is disturbed by the large reservoir, so the reservoir is often mistakenly identified as Dam. The Windmill is disturbed by the shadow of Windmill, so the shadow of Windmill is often mistakenly identified as Windmill. To improve the anti-interference ability of our model, we plan to design a novel feature enhancement module to enhance the feature extraction ability of WSOD. The high-quality feature is conducive to correctly identifying the object and enhances the robustness of the WSOD model.

Author Contributions: Conceptualization, L.Z., Y.H. and X.Q.; methodology, L.Z. and Y.H.; software, Y.H.; validation, X.Q. and Z.C.; formal analysis, L.Z., X.Q. and Z.C.; resources, Z.C.; writing—original draft, Y.H.; writing—review and editing, L.Z.; supervision, Z.C.; project administration, Z.C.; funding acquisition, X.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 62076223, in part by the Key Science and Technology Program of Henan Province under Grant 232102211018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The NWPU VHR-10.v2 and DIOR datasets are available at following URLs: https://drive.google.com/file/d/15xd4TASVAC2irRf02GA4LqYFbH7QITR-/view (accessed on 15 October 2022) and https://drive.google.com/drive/folders/1UdlgHk49iu6WpcJ5467iT-UqNPpx__CC (accessed on 15 October 2022), respectively.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Z.; Ma, Z.; van der Kuijp, T.J.; Yuan, Z.; Huang, L. A review of soil heavy metal pollution from mines in China: Pollution and health risk assessment. *Sci. Total Environ.* **2014**, *468*, 843–853. [CrossRef]
2. Sanaei, F.; Amin, M.M.; Alavijeh, Z.P.; Esfahani, R.A.; Sadeghi, M.; Bandarrig, N.S.; Fatehizadeh, A.; Taheri, E.; Rezakazemi, M. Health risk assessment of potentially toxic elements intake via food crops consumption: Monte Carlo simulation-based probabilistic and heavy metal pollution index. *Environ. Sci. Pollut. Res.* **2021**, *28*, 1479–1490. [CrossRef]
3. Oliveira, V.; Pinho, P. Evaluation in urban planning: Advances and prospects. *J. Plan. Lit.* **2010**, *24*, 343–361. [CrossRef]
4. Wosner, O.; Farjon, G.; Bar-Hillel, A. Object detection in agricultural contexts: A multiple resolution benchmark and comparison to human. *Comput. Electron. Agric.* **2021**, *189*, 106404. [CrossRef]
5. Zhao, W.; Yamada, W.; Li, T.; Digman, M.; Runge, T. Augmenting crop detection for precision agriculture with deep visual transfer learning—A case study of bale detection. *Remote Sens.* **2020**, *13*, 23. [CrossRef]
6. Lin, S.; Zhang, M.; Cheng, X.; Wang, L.; Xu, M.; Wang, H. Hyperspectral anomaly detection via dual dictionaries construction guided by two-stage complementary decision. *Remote Sens.* **2022**, *14*, 1784. [CrossRef]
7. Cheng, X.; Zhang, M.; Lin, S.; Zhou, K.; Wang, L.; Wang, H. Multiscale superpixel guided discriminative forest for hyperspectral anomaly detection. *Remote Sens.* **2022**, *14*, 4828. [CrossRef]
8. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Conference and Workshop on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
9. Qian, X.; Zeng, Y.; Wang, W.; Zhang, Q. Co-Saliency Detection Guided by Group Weakly Supervised Learning. *IEEE Trans. Multimed.* **2023**, *25*, 1810–1818. [CrossRef]
10. Lin, S.; Zhang, M.; Cheng, X.; Zhou, K.; Zhao, S.; Wang, H. Dual Collaborative Constraints Regularized Low-Rank and Sparse Representation via Robust Dictionaries Construction for Hyperspectral Anomaly Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 2009–2024. [CrossRef]
11. Cheng, X.; Zhang, M.; Lin, S.; Zhou, K.; Zhao, S.; Wang, H. Two-Stream Isolation Forest Based on Deep Features for Hyperspectral Anomaly Detection. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
12. Kuo, W.; Hariharan, B.; Malik, J. DeepBox: Learning Objectness with Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015; pp. 2479–2487.

13. Qian, X.; Cheng, X.; Cheng, G.; Yao, X.; Jiang, L. Two-stream encoder GAN with progressive training for co-saliency detection. *IEEE Signal Process. Lett.* **2021**, *28*, 180–184. [CrossRef]
14. Lin, S.; Zhang, M.; Cheng, X.; Zhou, K.; Zhao, S.; Wang, H. Hyperspectral Anomaly Detection via Sparse Representation and Collaborative Representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *16*, 946–961. [CrossRef]
15. Han, X.; Zhong, Y.; Zhang, L. An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery. *Remote Sens.* **2017**, *9*, 666. [CrossRef]
16. Qian, X.; Wu, B.; Cheng, G.; Yao, X.; Wang, W.; Han, J. Building a bridge of bounding box regression between oriented and horizontal object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–9. [CrossRef]
17. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J.; Lei, L.; Zou, H. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 3–22. [CrossRef]
18. Zhang, Y.; Ma, C.; Zhuo, L.; Li, J. Arbitrary-Oriented Object Detection in Aerial Images with Dynamic Deformable Convolution and Self-Normalizing Channel Attention. *Electronics* **2023**, *12*, 2132. [CrossRef]
19. Qian, X.; Lin, S.; Cheng, G.; Yao, X.; Ren, H.; Wang, W. Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion. *Remote Sens.* **2020**, *12*, 143. [CrossRef]
20. Fasana, C.; Pasini, S.; Milani, F.; Fraternali, P. Weakly Supervised Object Detection for Remote Sensing Images: A Survey. *Remote Sens.* **2022**, *14* 5362. [CrossRef]
21. Zhang, X.; Yu, W.; Ma, X.; Kang, X. Weakly Supervised Local-Global Anchor Guidance Network for Landslide Extraction With Image-Level Annotations. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6005505. [CrossRef]
22. Ren, W.; Huang, K.; Tao, D.; Tan, T. Weakly supervised large scale object localization with multiple instance learning and bag splitting. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 405–416. [CrossRef]
23. Wang, X.; Zhu, Z.; Yao, C.; Bai, X. Relaxed multiple-instance SVM with application to object discovery. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1224–1232.
24. Cinbis, R.G.; Verbeek, J.; Schmid, C. Weakly supervised object localization with multi-fold multiple instance learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 189–203. [CrossRef] [PubMed]
25. Hong, D.; Yokoya, N.; Chanussot, J.; Zhu, X.X. An augmented linear mixing model to address spectral variability for hyperspectral unmixing. *IEEE Trans. Image Process.* **2018**, *28*, 1923–1938. [CrossRef]
26. Bilen, H.; Vedaldi, A. Weakly supervised deep detection networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2846–2854.
27. Tang, P.; Wang, X.; Bai, X.; Liu, W. Multiple instance detection network with online instance classifier refinement. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2843–2851.
28. Kantorov, V.; Oquab, M.; Cho, M.; Laptev, I. Contextlocnet: Context-aware deep network models for weakly supervised localization. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 350–365.
29. Li, D.; Huang, J.B.; Li, Y.; Wang, S.; Yang, M.H. Weakly supervised object localization with progressive domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3512–3520.
30. Chen, S.; Wang, H.; Mukherjee, M.; Xu, X. Collaborative Learning-based Network for Weakly Supervised Remote Sensing Object Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2022, *Early access*. [CrossRef]
31. Ren, Z.; Yu, Z.; Yang, X.; Liu, M.Y.; Lee, Y.J.; Schwing, A.G.; Kautz, J. Instance-aware, context-focused, and memory-efficient weakly supervised object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10598–10607.
32. Feng, X.; Han, J.; Yao, X.; Cheng, G. Progressive contextual instance refinement for weakly supervised object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8002–8012. [CrossRef]
33. Yao, X.; Feng, X.; Han, J.; Cheng, G.; Guo, L. Automatic weakly supervised object detection from high spatial resolution remote sensing images via dynamic curriculum learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 675–685. [CrossRef]
34. Feng, X.; Han, J.; Yao, X.; Cheng, G. TCANet: Triple Context-Aware Network for Weakly Supervised Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 6946–6955. [CrossRef]
35. Feng, X.; Yao, X.; Cheng, G.; Han, J.; Han, J. SAENet: Self-Supervised Adversarial and Equivariant Network for Weakly Supervised Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5610411. [CrossRef]
36. Chen, S.; Shao, D.; Shu, X.; Zhang, C.; Wang, J. FCC-Net: A Full-Coverage Collaborative Network for Weakly Supervised Remote Sensing Object Detection. *Electronics* **2020**, *9*, 1356. [CrossRef]
37. Kosugi, S.; Yamasaki, T.; Aizawa, K. Object-aware instance labeling for weakly supervised object detection. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6064–6072.
38. Zeng, Z.; Liu, B.; Fu, J.; Chao, H.; Zhang, L. Wsod2: Learning bottom-up and top-down objectness distillation for weakly-supervised object detection. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8292–8300.
39. Tang, P.; Wang, X.; Bai, S.; Shen, W.; Bai, X.; Liu, W.; Yuille, A. Pcl: Proposal cluster learning for weakly supervised object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *42*, 176–191. [CrossRef]

40. Lin, C.; Wang, S.; Xu, D.; Lu, Y.; Zhang, W. Object instance mining for weakly supervised object detection. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 11482–11489. [CrossRef]
41. Cheng, G.; Xie, X.; Chen, W.; Feng, X.; Yao, X.; Han, J. Self-Guided Proposal Generation for Weakly Supervised Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]
42. Qian, X.; Huo, Y.; Cheng, G.; Yao, X.; Li, K.; Ren, H.; Wang, W. Incorporating the Completeness and Difficulty of Proposals Into Weakly Supervised Object Detection in Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 1902–1911. [CrossRef]
43. Hosang, J.; Benenson, R.; Schiele, B. Learning non-maximum suppression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4507–4515.
44. Huo, Y.; Qian, X.; Li, C.; Wang, W. Multiple Instances Complementary Detection and Difficulty Evaluation for Weakly Supervised Object Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* 2023, *Early access*. [CrossRef]
45. Zitnick, C.L.; Dollár, P. Edge boxes: Locating object proposals from edges. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 Sepetmber 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 391–405.
46. Uijlings, J.R.; Van De Sande, K.E.; Gevers, T.; Smeulders, A.W. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [CrossRef]
47. Cheng, G.; Zhou, P.; Han, J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [CrossRef]
48. Li, K.; Cheng, G.; Bu, S.; You, X. Rotation-insensitive and context-augmented object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 2337–2348. [CrossRef]
49. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]
50. Deselaers, T.; Alexe, B.; Ferrari, V. Weakly supervised localization and learning with generic knowledge. *Int. J. Comput. Vis.* **2012**, *100*, 275–293. [CrossRef]
51. Qian, X.; Li, C.; Wang, W.; Yao, X.; Cheng, G. Semantic segmentation guided pseudo label mining and instance re-detection for weakly supervised object detection in remote sensing images. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *119*, 103301. [CrossRef]
52. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–13.
53. Wan, F.; Wei, P.; Jiao, J.; Han, Z.; Ye, Q. Min-entropy latent model for weakly supervised object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1297–1306.
54. Wang, B.; Zhao, Y.; Li, X. Multiple instance graph learning for weakly supervised remote sensing object detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5613112. [CrossRef]
55. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
56. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
57. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef] [PubMed]