electronics

MDPI

Article

High-Quality Instance Mining and Weight Re-Assigning for Weakly Supervised Object Detection in Remote Sensing Images

Peixu Xing [1,2], Mengxing Huang [2], Chenhao Wang [3,*] and Yang Cao [1,*]

[1] Zhengzhou University of Light Industry, Zhengzhou 450002, China; 2003071@zzuli.edu.cn
[2] School of Information and Communication Engineering, Hainan University, Haikou 570228, China; huangmx09@hainanu.edu.cn
[3] School of Aerospace Science and Technology, Xidian University, Xi'an 710126, China
* Correspondence: wang_chenhao@stu.xidian.edu.cn (C.W.); 2015063@zzuli.edu.cn (Y.C.)

Abstract: Weakly supervised object detection (WSOD) in remote sensing images (RSIs) achieves high-value object classification and localization by only using image-level labels. However, two problems limit its performance. Firstly, adjacent instances are often misclassified because their pseudo-labels are determined solely based on the spatial distances between them and their corresponding seed instances. Secondly, most WSOD methods assign the highest weight to the instance that covers the discriminative part of an object, thereby urging WSOD models to focus on the discriminative part rather than the whole object. To handle the first problem, a high-quality instance mining (HQIM) module, which incorporates the feature similarities between instances into the label propagation process, enabling some misclassified adjacent instances to be removed, is proposed. To tackle the second problem, a weight re-assigning (WRA) strategy, which redistributes the loss weights of instances, is proposed. Specifically, the loss weights of instances focusing on the discriminative part are exchanged with those of instances that broadly cover the whole object. Ablation studies demonstrate the effectiveness of HQIM and WRA, while comparisons with popular models on two RSI benchmarks further verify the effectiveness of our model.

Keywords: remote sensing image; weak supervised object detection; high-quality instance mining; weight re-assigning

check for updates

Citation: Xing, P.; Huang, M.; Wang, C.; Cao, Y. High-Quality Instance Mining and Weight Re-Assigning for Weakly Supervised Object Detection in Remote Sensing Images. *Electronics* **2024**, *13*, 4753. https://doi.org/10.3390/electronics13234753

Academic Editor: Beiwen Li

Received: 25 October 2024
Revised: 27 November 2024
Accepted: 29 November 2024
Published: 1 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Remote sensing images exhibit a wide range of applications across diverse fields, such as object detection [1,2], classification [3,4], and anomaly detection [5,6]. Among these, object detection in remote sensing images (RSIs) stands out as particularly compelling. This technology boasts a diverse array of applications in both military and civilian domains, including but not limited to surveillance and monitoring, disaster response, urban planning, and agricultural management. With advances in remote sensing technology and deep learning [7,8], fully supervised object detection (FSOD) models [9,10] for RSIs have achieved remarkable outcomes by leveraging instance-level labels. However, in RSIs with sophisticated backgrounds, annotating all objects with instance-level labels is an exceptionally labor-intensive and time-consuming task, and some objects may even be impossible to annotate accurately. Conversely, annotating the image-level labels is straightforward. Consequently, weakly supervised object detection (WSOD) in RSIs holds significant practical value and has garnered considerable attention.

The weakly supervised deep detection networks (WSDDNs) [2], as a path-breaking work, initially integrated multiple-instance learning with deep learning to train an object detector that solely uses image-level labels. The trained detector predicts the class scores (CSs) of proposals to accomplish object localization and classification.

The Online Instance Classifier Refinement (OICR) [11] model, which builds upon the WSDDN framework, incorporates several instance classifier refinement (ICR) branches to

iteratively refine the CSs of proposals. The instance-level labels of the current ICR branch, which are generated by using the CSs of the upper ICR branch, are used to train the ICR branch. The pseudo-instance-level labels include positive instances (seed instances and neighboring instances) and negative instances (background instances). Specifically, the instance with the highest CS is regarded as the seed instance of each class. The remaining instances are subsequently divided into neighboring instances and background instances based on spatial distances between the seed instance and other instances.

A majority of WSOD [12–23] methods adopt OICR as the baseline to improve their performance. These methods can be broadly divided into two classes. The first class is designed to alleviate the problem of missing instances by mining more seed instances. For example, the proposal cluster learning (PCL) model [13] utilizes the *k*-means algorithm to divide the proposals into multiple clusters; then, the clusters centers are selected as seed instances. The neighboring instances are determined solely based on the spatial distances between the seed instances and other instances. Other similar models include the complete and invariant instance classifier refinement model [20], the multiple-instance self-training (MIST) model [24], and the weakly supervised contrastive loss (WSCL) [21]. The second class aims to guide the model in recognizing the whole object, rather than just its distinguishing parts. For example, the multi-scale image splitting-based feature enhancement model [17] enforces the spatial attention maps of positive instances to approximate the maximum spatial attention maps, thereby encouraging the WSOD model to pay more attention to the whole object. Other similar works include the semantic segmentation-guided pseudo-label mining [12] model, the instance re-detection comprehensive attention self-distillation model [14], and the pseudo-instance soft labels mining [25] model. Although the aforementioned methods have obtained notable performance in WSOD, there are still two problems that require attention and resolution.

For the first problem, during the label propagation process, the pseudo-labels of adjacent instances are determined solely based on the spatial distances between them and their corresponding seed instance. Consequently, some adjacent instances are often misclassified. As depicted in Figure 1a, the basketball court is incorrectly identified as the tennis court by the baseline model, as the basketball court is an adjacent instance of tennis court.

For the second problem, most WSOD methods tend to assign a higher loss weight to the instance that only covers the discriminative part of an object, whereas instances that broadly cover the whole object obtain a lower loss weight. Consequently, the inappropriate loss weights encourage the WSOD model to detect the discriminative part of an object rather than the whole object. As illustrated in Figure 1b, the baseline model only detects the most discriminative part of an airplane rather than the whole airplane.

To overcome the first problem, a high-quality instance mining (HQIM) strategy is proposed. Firstly, the seed instances are identified by selecting the top p percent proposals from the sorted proposals. Then, these seed instances are refined through the local Non-Maximum Suppression (NMS) operation. Secondly, the initial neighboring instances are determined based on the spatial distances between them and their corresponding seed instances; furthermore, these initial neighboring instances are refined by using the feature similarities between them and the seed instances to remove some misidentified adjacent instances. As illustrated in Figure 1a, the combination of the baseline and the HQIM module can avoid the misclassification of adjacent instances.

To overcome the second problem, we propose a weight re-assigning (WRA) strategy. First, the instance that fully covers the object is identified by using the spatial distances between positive instances. Then, we re-assign the loss weight of this comprehensive instance based on the CS of the instance covering the object's discriminative part. Conversely, the loss weight of the instance focusing on the discriminative part is adjusted by using the CS of the comprehensive instance. Consequently, as illustrated in Figure 1b, the baseline model with the proposed WRA strategy tends to capture the whole object.

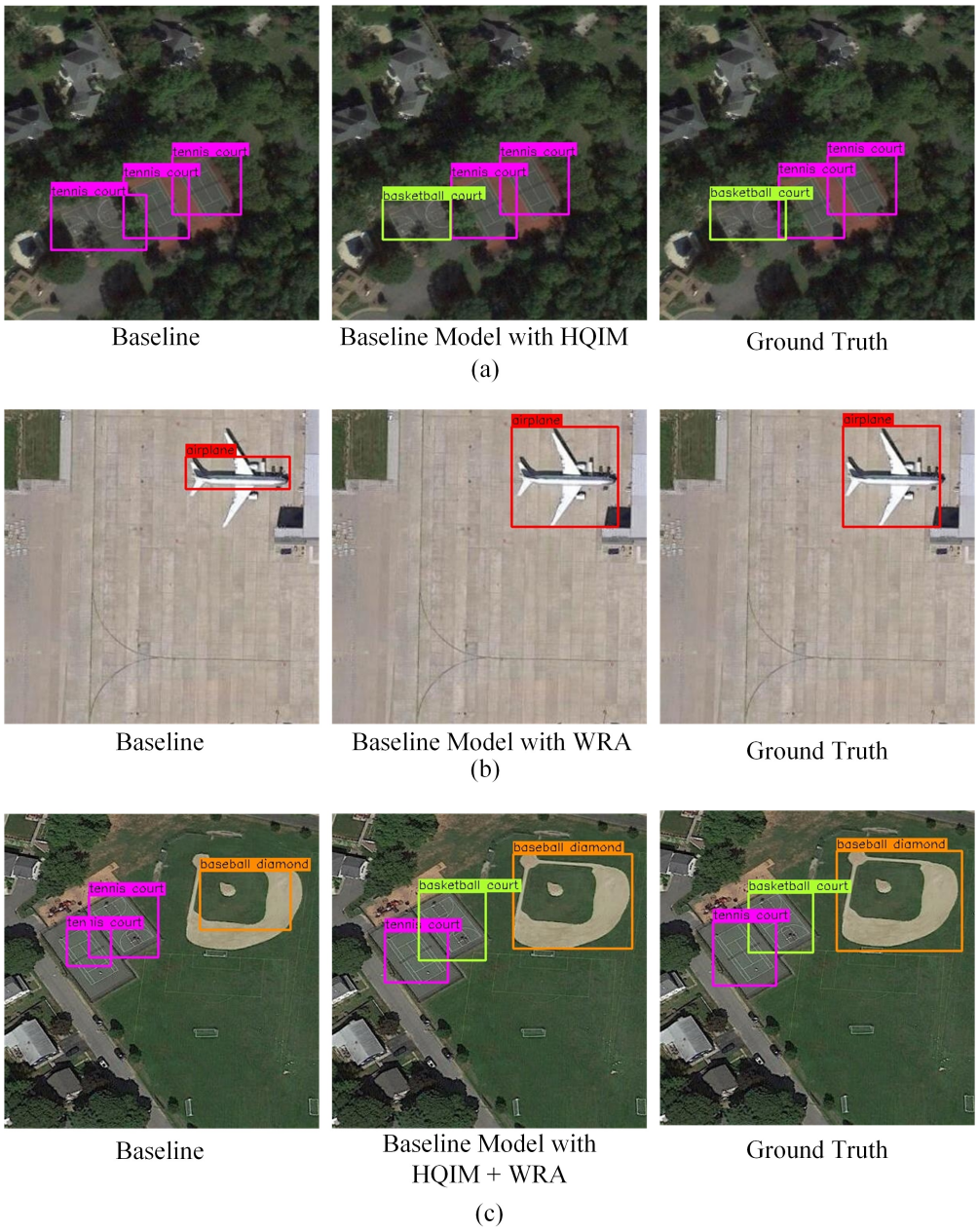


Figure 1. Illustration of problems and contributions of (**a**) HQIM, (**b**) WRA, and (**c**) HQIM + WRA.

Overall, as shown in Figure 1c, the baseline model with HQIM + WRA not only avoids the misclassification of adjacent instances but also encourages the model to focus on the whole object. The summarized contributions are outlined as follows:

1. During the label propagation process, the label of a neighboring instance is determined solely based on the spatial distance between it and its corresponding seed instance. Inevitably, this leads to some neighboring instances being misclassified. To address this issue, the HQIM module is proposed. This module utilizes feature similarity between seed instances and their neighboring instances to further refine the neighboring instances, thereby removing the misclassified neighboring instances.
2. Most WSOD models often assign higher loss weights to instances focusing on the discriminative part of an object, compared with those covering the entire object. Consequently, they tend to detect the discriminative part of an object. To address this issue, we propose the WRA strategy, which exchanges the loss weights between these two types of instances.

2. Related Works

OICR [11] is used as the baseline of our model. To enhance the localization capacity of the WSOD model, many researchers incorporate bounding box regression (BBR) branches into the OICR model. Therefore, the WSDDN, OICR, and BBR branches are introduced in turn.

2.1. Weakly Supervised Deep Detection Network

Initially, as illustrated in Figure 2, a variety of proposals $PR = \{pr_1, \dots pr_m, \dots pr_M\}$ are generated by using the selective search algorithm [26] for the input image I, where M denotes the quantity of proposals. Secondly, the feature vectors of all proposals, denoted by $\{f_m\}_{m=1}^{M}$, are obtained by importing I and its proposals into backbone networks, region of interest (RoI) pooling layer, and two fully connected (Fc) layers. Thirdly, the $\{f_m\}_{m=1}^{M}$ are imported into classification branch to obtain a classification score vector of class c, denoted by $\chi_c^{cls} \in \mathbb{R}^{1\times M}$. Similarity, the detection score vector of class c, denoted by $\chi_c^{dec} \in \mathbb{R}^{1\times M}$, is obtained by importing the $\{f_m\}_{m=1}^{M}$ into the detection branch. The classification and detection branches both include an FC layer and a softmax classifier. Then, the instance-level predicted CS vector of class c, denoted by $\chi_c \in \mathbb{R}^{1\times M}$, is acquired as follows:

$$\chi_c = \chi_c^{cls} \odot \chi_c^{dec}, \tag{1}$$

where $\odot$ represents the Hadamard product. The image-level predicted CS of class c, denoted by ϕ_c, is obtained as follows:

$$\phi_c = \sum_{m=1}^{M} x_c^m, \tag{2}$$

where $x_c^m \in \chi_c$ denotes the CS of the mth proposal in class c. Finally, the loss of the WSDDN, denoted by L_{WSDDN}, is obtained as follows:

$$L_{WSDDN} = -\sum_{c=1}^{C} (y_c \log \phi_c + (1 - y_c) \log(1 - \phi_c)), \tag{3}$$

where C denotes the quantity of classes and $y_c = 1$ or 0 represents that I contains or does not contain category c.

2.2. Online Instance Classifier Refinement

To improve the performance of the WSDDN, the K ICR branches are incorporated into the WSDDNs, where the ICR branch contains an FC layer and a softmax classifier. The feature vectors of the proposals are imported into the kth ICR branch to obtain the predicted CS vectors of all classes, denoted by $\{S_c^k \in \mathbb{R}^{1\times M}\}_{c=1}^{C+1}$, $k \in 1, 2, \cdots K$, where the (C+1)th dimension denotes the background class. The pseudo-instance-level labels of the kth ICR branch, denoted by $\mathbf{y}^k$, are generated by using $\{S_c^{k-1} \in \mathbb{R}^{1\times M}\}_{c=1}^{C+1}$, and the pseudo-labels of the first ICR branch are generated by using $\{\chi_c\}_{c=1}^{C}$. Specifically, the instance with the highest prediction CS is considered the seed instance, and the neighboring instances are determined by using the spatial distances between the seed instances and other instances. The loss of the kth ICR branch, denoted by L_{ICR}^k, is obtained by the following equation:

$$L_{ICR}^k = -\sum_{m=1}^{M}\sum_{c=1}^{C+1} w_m^k y_{c,m}^k s_{c,m}^k, \tag{4}$$

where $y_{c,m}^k \in \{0, 1\}$ represents the element in the mth column and cth row of $\mathbf{y}^k$, $s_{c,m}^k$ denotes the CS of the kth ICR branch of the mth proposal in class c. The $w_m^k = s_{c,m}^{k-1}$ denotes the loss weight of the mth proposal in the kth ICR branch, and $s_{c,m}^{k-1}$ denotes the CS of $(k-1)$th ICR branch of the mth proposal in class c.

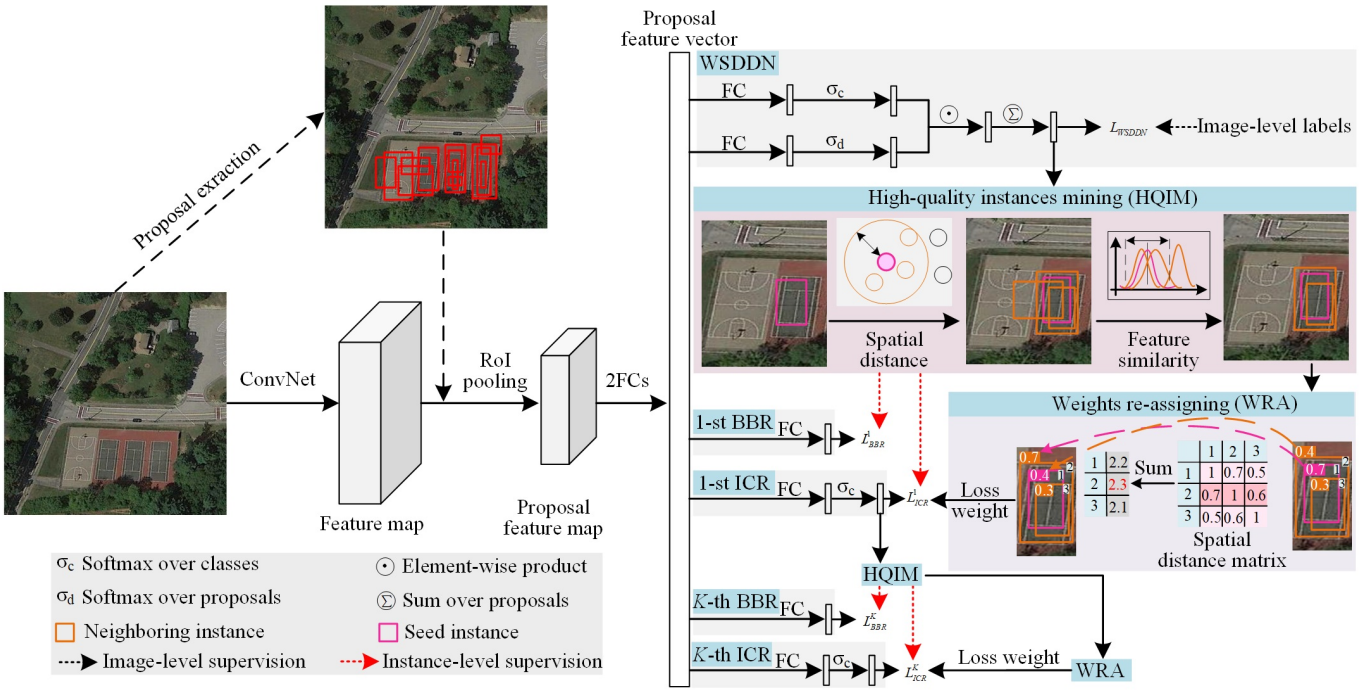


Figure 2. The overview of our model. The RSI and its proposals are imported into a ConvNet, followed by the RoI pooling layer and two FC layers, to extract feature vectors for all proposals. The CSs of these proposals, obtained from the WSDDN, are then imported into the HQIM module. This module effectively mines positive instances by leveraging spatial distances and feature similarities among instances. Furthermore, the WRA strategy redistributes the loss weights of positive instances to encourage the model to focus on the whole object.

2.3. Incorporating the Regression Branches into OICR

To improve the localization performance of OICR, K bounding box regression (BBR) branches are incorporated into OICR, where each BBR branch includes an FC layer. The loss of the kth BBR branch, denoted by L^k_{BBR}, is obtained as follows:

$$L^k_{BBR} = \frac{1}{|P^k|} \sum_{j=1}^{|P^k|} (smooth_{L1}(g_j\ ,\ \hat{g}_j)), \tag{5}$$

where P^k denotes the set of positive instances, $|P^k|$ denotes the number of positive instances, and $\hat{g}_r$ and g_r denote the target offset and prediction offset of the jth positive instance, respectively.

3. Proposed Method

3.1. Overview

As shown in Figure 2, our proposed method integrates the HQIM module and WRA strategy into the OICR model. Firstly, the HQIM module utilizes the sorted CSs of the proposals to mine seed instances. The initial neighboring instances are selected by using the spatial distances between seed instances and other instances, which are further refined by using the feature similarities between them and seed instances to remove some misclassified neighboring instances. Secondly, the WRA strategy employs spatial distances between positive instances to identify the one that broadly covers the whole object. The loss weight of this comprehensive instance is then re-assigned based on the CS of an instance that covers the object's discriminative part. Conversely, the loss weight of the instance focusing solely on the discriminative part is adjusted by using the CS of the comprehensive instance.

3.2. High-Quality Instance Mining Module

Popular WSOD models propagate the labels from the real image level to the pseudo-instance level for supervising the ICR branch. The performance of the WSOD model relies on the quality of pseudo-instance-level labels. The pseudo-instance-level labels include positive instances (seed instances and neighboring instances) and negative instances (background instances). Once the positive instances are identified, the rest of the instances are automatically classified as negative instances. Consequently, the quality of the positive instances ultimately determines the overall quality of the pseudo-instance-level labels.

In most WSOD models, the neighboring instances are determined by using the spatial distances between seed instances and other instances. This often results in the misclassification of neighboring instances, thereby limiting the quality of positive instances. To address this issue, the HQIM strategy is proposed to eliminate the misclassified neighboring instances. Firstly, the seed instances are mined from the sorted CSs of proposals. Then, to remove small and redundant seed instances, the NMS operation is applied to refine the seed instances. Secondly, the spatial distances between the seed instance and other instances are used to determine the initial neighboring instances. Subsequently, as shown in Figure 2, the feature distribution curves of the seed instance (represented by pink lines) and the neighboring instances (represented by orange lines) are illustrated. The feature distributions of misclassified neighboring instances are not similar to those of their corresponding seed instances. Therefore, the feature similarities between the seed instances and their neighboring instances are used to further refine the initial neighboring instances, removing the misclassified neighboring instances. The details of the HQIM module are as follows.

Considering the multiple objects of the same category in RSIs, firstly, we assume that the input image I contains category c; the initial set of seed instances of the kth ICR branch in class c, denoted by ISI_c^k, is determined by picking the top p percent proposals from the sorted S_c^{k-1}. The set of seed instances of class c in the kth ICR branch, denoted by SI_c^k, is obtained by implementing NMS among ISI_c^k.

Secondly, as shown in Figure 2, the initial set of neighboring instances of the ith seed instance $si_{c,i}^k \in SI_c^k$, denoted by $INI_{si_{c,i}^k}$, is obtained by selecting the instances that have high spatial distances (e.g., IoU > 0.5) between them and $si_{c,i}^k$. However, some neighboring instances barely cover the object. To remove these low-quality neighboring instances, feature similarity is proposed to evaluate the similarity between each seed instance and its neighboring instances, allowing us to eliminate those low-quality neighboring instances.

The definition of the feature similarity $fs_{si_{c,i}^k}^{ini_j}$ between seed instance $si_{c,i}^k$ and initial neighboring instance ini_j is as follows:

$$fs_{si_{c,i}^k}^{ini_j} = sim(f_{si_{c,i}^k}, f_{ini_j}), \tag{6}$$

where $sim(\cdot,\cdot)$ denotes the dot product between inputs, $f_{si_{c,i}^k}$ denotes the feature vector of seed instances $si_{c,i}^k$, $ini_j \in INI_{si_{c,i}^k}$ denotes the jth element of $INI_{si_{c,i}^k}$, and f_{ini_j} denotes the feature vector of ini_j. The initial neighboring instance ini_j is regarded as a neighboring instance if $fs_{si_{c,i}^k}^{ini_j} > \delta$, where δ denotes the threshold of feature similarity, which determines whether or not an initial neighboring instance belongs to the neighboring instances. Therefore, the set of neighboring instances of seed instance $si_i^{k,c}$, denoted by $NI_{si_{c,i}^k}$, is obtained. Finally, the neighboring instances and seed instances form the high-quality positive instances.

3.3. Weight Re-Assigning Strategy

The loss weights of instances not only ensure the stability of training but also encourage the model to focus on the instances with higher loss weights. Therefore, the reasonable assignment of instance loss weights greatly impacts the WSOD model's performance.

However, the instance that covers the discriminative part of object usually has a higher CS, and its loss weight is obtained by copying its CS. Conversely, the instance broadly covers the whole object, and both its CS and loss weight tend to be relatively small. Therefore, the trained object detector tends to detect the discriminative part of the object. To handle the problem, the WRA module is proposed. As shown in Figure 2, firstly, the spatial distance matrix between positive instances is calculated to select the instance that broadly covers the whole object. Secondly, the WRA strategy switches the loss weights of the instance that covers the distinguishing region of an object and the instance that broadly covers the whole object, to encourage the WSOD model to pay more attention to the whole object. The details of the WRA strategy are as follows.

As shown in Figure 2, firstly, the seed instance and its neighboring instance set are constructed as a positive instance cluster $SN = \{r_1, \ldots r_n, \ldots, r_N\}$, where r_n denotes the nth instance in the cluster, and N denotes the number of elements in the cluster. Secondly, the spatial distance matrix, denoted by $G \in \mathbb{R}^{N \times N}$, is constructed for the instances in the cluster according to their spatial distance. $g_{n,n'} \in G$ denotes the spatial distance between the r_nth and $r_{n'}$th instances in the cluster SN, which is obtained as follows:

$$g_{n,n'} = \frac{|r_n \cap r_{n'}|}{|r_n \cup r_{n'}|}. \tag{7}$$

Then, the spatial distance vector, denoted by $V \in \mathbb{R}^{N \times 1}$, is obtained as follows:

$$V = \sum_{n=1}^{N} G. \tag{8}$$

Finally, the index of the instance that broadly covers the whole object is selected as follows:

$$z = \arg\max V. \tag{9}$$

The index of the instance that has the highest CS in the cluster SN is obtained as follows:

$$p = \arg\max\{s_{c,r_n}^{k-1}\}_{n=1}^{N}, \tag{10}$$

where s_{c,r_n}^{k-1} denotes the CS of (k-1)th ICR branch of the r_nth positive instance in class c. The loss weights of instances r_z and r_p are assigned as follows:

$$w_{r_z} = s_{c,r_p}^{k-1}, \tag{11}$$

$$w_{r_p} = s_{c,r_z}^{k-1}, \tag{12}$$

where w_{r_z} (w_{r_p}) denotes the loss weight of the r_zth (r_pth) positive instance and s_{c,r_p}^{k-1} (s_{c,r_z}^{k-1}) denotes the CS of the (k-1)th ICR branch of the r_zth (r_pth) positive instance in class c. Therefore, the WRA strategy encourages the WSOD model to detect the whole object.

3.4. Overall Training Loss

The overall loss of the proposed method, represented as L_{ALL}, is obtained as follows:

$$L_{ALL} = L_{WSDDN} + \sum_{k=1}^{K} (L_{ICR}^{k} + L_{BBR}^{k}). \tag{13}$$

During the inference stage, the feature vectors of all proposals are imported into the K trained ICR and BBR branches to obtain prediction scores and prediction offsets. Subsequently, the prediction scores and prediction offsets are averaged separately to jointly infer the initial detection results. Finally, the detection results are derived by applying NMS operation to the initial detection results.

4. Materials, Data, and Experiments

4.1. Materials and Data

The source codes of our model are designed to run on the Ubuntu 16.04 platform, leveraging eight GPUs (Titan RTX). These codes have been implemented within the PyTorch framework, ensuring compatibility and efficiency for deep learning tasks.

The NWPU VHR-10.v2 dataset [27,28] contains 1172 RGB images, each with a uniform size of 400 × 400 pixels. The spatial resolutions of these images vary from 0.08 m to 2 m. The images are sourced from Google Earth and the Vaihingen dataset [29]. Additionally, the dataset is divided into three subsets: training, validation, and test. Specifically, the training subset comprises 679 images, the validation subset includes 200 images, and these are used to train our model. Meanwhile, the test subset, which consists of 293 images, is utilized to validate the performance of our trained model. Furthermore, the NWPU VHR-10.v2 dataset encompasses 10 classes, totaling 2775 object instances.

The DIOR dataset [30] contains 23,463 RGB images, each with a uniform size of 800 × 800 pixels. The spatial resolutions of these images vary from 0.5 m to 30 m. The images are sourced from Google Earth. Additionally, the dataset encompasses 20 classes, totaling 192,472 object instances. Furthermore, the dataset is divided into three subsets: training, validation, and test. Specifically, the training subset comprises 5862 images, the validation subset includes 5863 images, and these are used to train our model. Meanwhile, the test subset, which consists of 11,738 images, is utilized to validate the performance of our trained model.

The PASCAL VOC 2007 dataset [31] contains 9963 RGB images. The dataset encompasses 20 classes, totaling 24,640 object instances. The PASCAL VOC 2007 dataset is also divided into a training subset (2501 images), a validation subset (2510 images), and a test subset (4952 images). The training and validation subsets are used to train our model, while the test subset is used to test the trained model.

4.2. Experiments

4.2.1. Implementation Details

Our method builds on the OICR [11] model and adapts the VGG16 [32] pre-trained on the ImageNet dataset [33] as the background. The count of ICR branches is configured to be 3, i.e., $K = 3$; the top p percent is set to 0.15, i.e., $p = 0.15$; the threshold of feature similarity is set to 0.3, i.e., $\delta = 0.3$; and the threshold of NMS during the refinement of seed instances is set to 0.1. In the inference stage, the NMS [34] threshold is established at 0.3.

The NWPU VHR-10.v2, DIOR, and PASCAL VOC 2007 datasets were augmented by applying horizontal flipping as well as rotations of 90° and 180°. Similar to almost all WSOD models [2,11,17,21,35], all images were resized to one of the following dimensions: {480, 576, 688, 864, 1200}. The model was optimized by using the stochastic gradient descent algorithm, and the momentum, weight decay, and batchsize were 0.9, 0.005, and 8, respectively. The number of iterations was 30 K for the NWPU VHR-10.v2 and PASCAL VOC 2007 datasets. The initial learning rate was 0.1, and it was reduced to 10% of the previous value at 20 K and 26 K for the NWPU VHR-10.v2 and PASCAL VOC 2007 dataset, respectively. Additionally, for the DIOR dataset, the number of iterations was 60 K. The initial learning rate was 0.1 and was decreased to 10% of its previous value at 50 K and 56 K iterations for the DIOR dataset.

The detection accuracy of the trained model in the test subset was assessed by the mean average precision (mAP), and the localization capability of our model was assessed by correct localization (CorLoc) in the training and validation subsets. CorLoc focuses more on the accuracy of object localization, while mAP considers a broader range of model performance indicators, including precision, recall, and confidence scores.

4.2.2. Parameter Analysis

The parameter p in the HQIM module determines the extent of seed mining, which was analyzed in terms of mAP on both the NWPU VHR-10.v2 and DIOR datasets. As

shown in Figure 3a,b, our model attained the maximum mAP at $p = 0.15$, while the mAP experiences a slight decline at $p = 0.16$. Consequently, this article adopted $p = 0.15$.

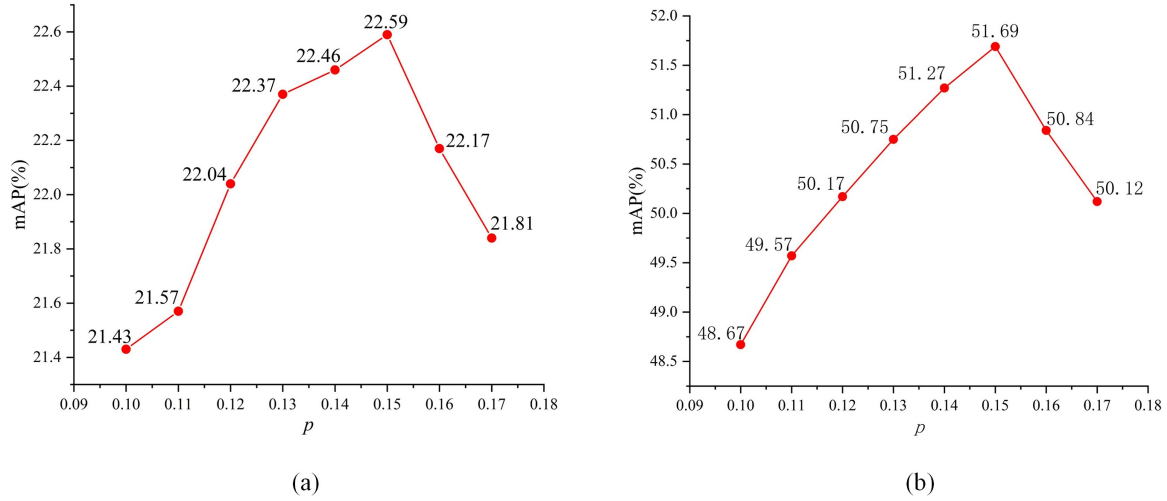


Figure 3. Parameter analysis of p in terms of mAP on the DIOR (**a**) and NWPU VHR-10.v2 (**b**) datasets.

The parameter δ in the HQIM module determines whether or not an initial neighboring instance belongs to a neighboring instance, which was analyzed in terms of mAP on both the NWPU VHR-10.v2 and DIOR datasets. As shown in Figure 4a,b, our model attained the maximum mAP at $\delta = 0.3$, while the mAP experiences a slight decline at $\delta = 0.4$. Consequently, this article adopted $\delta = 0.3$.

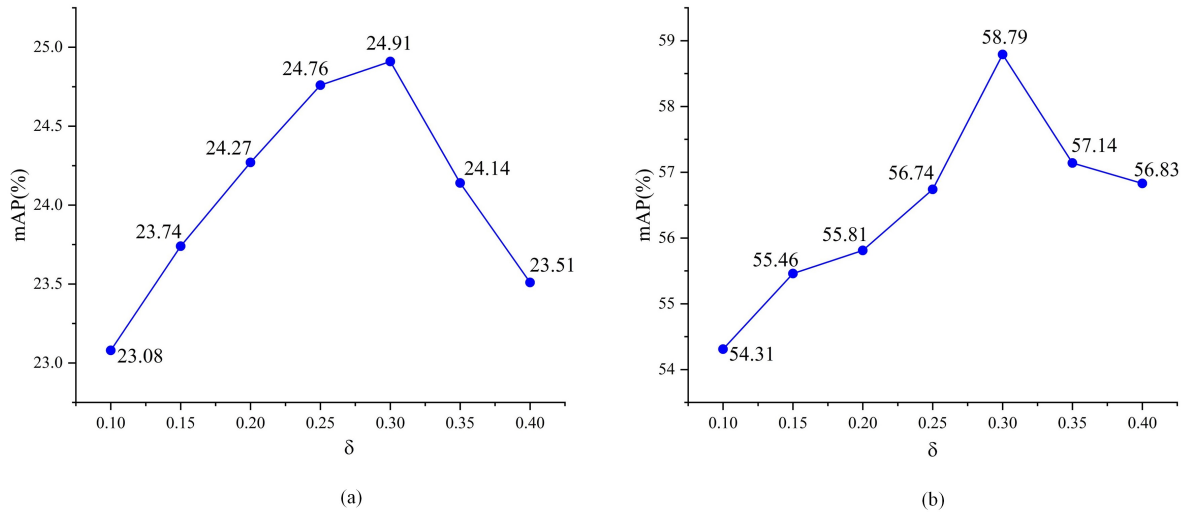


Figure 4. Parameter analysis of δ in terms of mAP on the DIOR (**a**) and NWPU VHR-10.v2 (**b**) datasets.

The parameter K in our model determines the number of ICR branches, which was analyzed in terms of mAP on both the NWPU VHR-10.v2 and DIOR datasets. As illustrated in Figure 5a,b, our model attained the maximum mAP at $K = 3$, while the mAP experiences a slight decline at $K = 4$. Consequently, this article adopted $K = 3$.

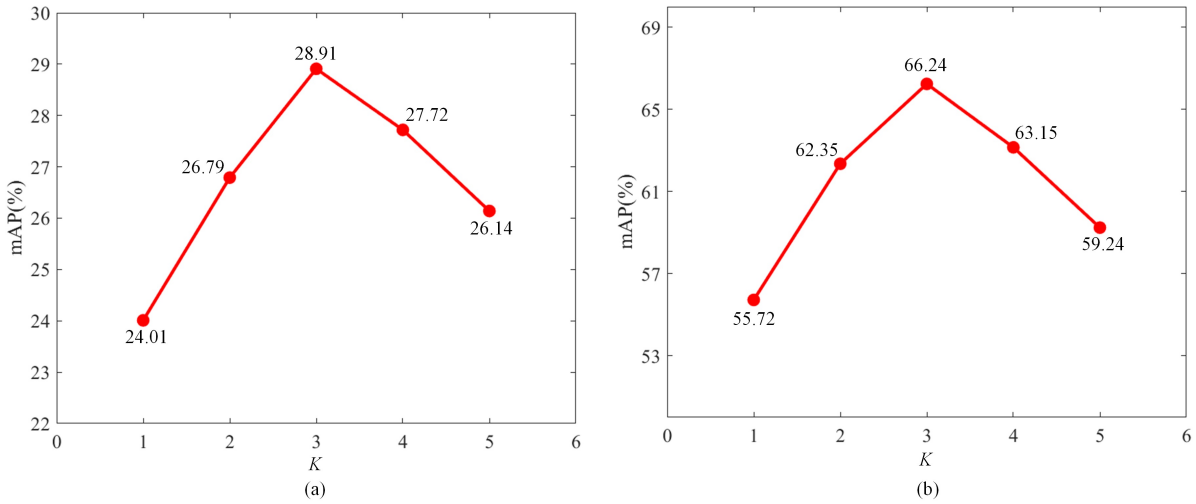


Figure 5. Parameter analysis of K in terms of mAP on the DIOR (**a**) and NWPU VHR-10.v2 (**b**) datasets.

4.2.3. Ablation Study

The ablation studies, conducted on the NWPU VHR-10.v2 dataset and more challenging DIOR dataset, were used to validate the efficacy of HQIM, WRA, and their combination. Particularly, to enhance the localization ability of the WSOD model, like numerous WSOD models [12,16,20,21,24], the combination of the OICR and BBR branches was employed as the baseline model.

Quantitative ablation study: As shown in Table 1, compared with the baseline, the mAP (CorLoc) of baseline + HQIM increased by 9.46% (7.97%) and 4.81% (5.75%) on the NWPU VHR-10.v2 and DIOR datasets, which validates the effectiveness of the HQIM module. Furthermore, compared with the baseline, the mAP(CorLoc) of baseline + WRA increased by 12.55% (11.40%) and 7.13% (8.10%) on the NWPU VHR-10.v2 and DIOR datasets, which validates the effectiveness of the WRA strategy. In addition, compared with the baseline, the mAP (CorLoc) of baseline + HQIM + WRA increased by 17.28% (15.35%) and 8.81% (11.13%) on the NWPU VHR-10.v2 and DIOR datasets, which validates the effectiveness of our method.

Table 1. Ablation study of HQIM module and WRA strategy on NWPU VHR-10.v2 and DIOR datasets.

Baseline	HQIM	WRA	NWPU VHR-10.v2		DIOR	
			mAP	CorLoc	mAP	CorLoc
✓			48.96	61.54	20.10	42.79
✓	✓		58.79	69.51	24.91	48.54
✓		✓	61.51	72.94	27.23	50.89
✓	✓	✓	66.24	76.89	28.91	53.92

Subjective ablation study: To further validate the effectiveness of the HQIM module, the WRA strategy, and their combination, subjective ablation studies were conducted on the NWPU VHR-10.v2 and DIOR datasets. Firstly, as illustrated in Figure 6a, subjective comparisons were conducted between the detection results of the baseline and those of baseline + HQIM. Obviously, baseline + HQIM demonstrated the ability to reduce the misclassification of neighboring instances, effectively addressing the first problem outlined in the Introduction. Secondly, as shown in Figure 6b, subjective comparisons were conducted between the detection results of the baseline and those of baseline + WRA. Apparently, compared with the baseline, baseline + WRA captured the whole object, rather than just its discriminative parts, effectively tackling the second problem introduced in the Introduction. Thirdly, as shown in Figure 6c, subjective comparisons were conducted between the detection results of the baseline and those of baseline + HQIM

+ WRA. Obviously, compared with the baseline, baseline + HQIM + WRA model not only avoided the misclassification of neighboring instances but also encouraged the model to focus on the whole object.

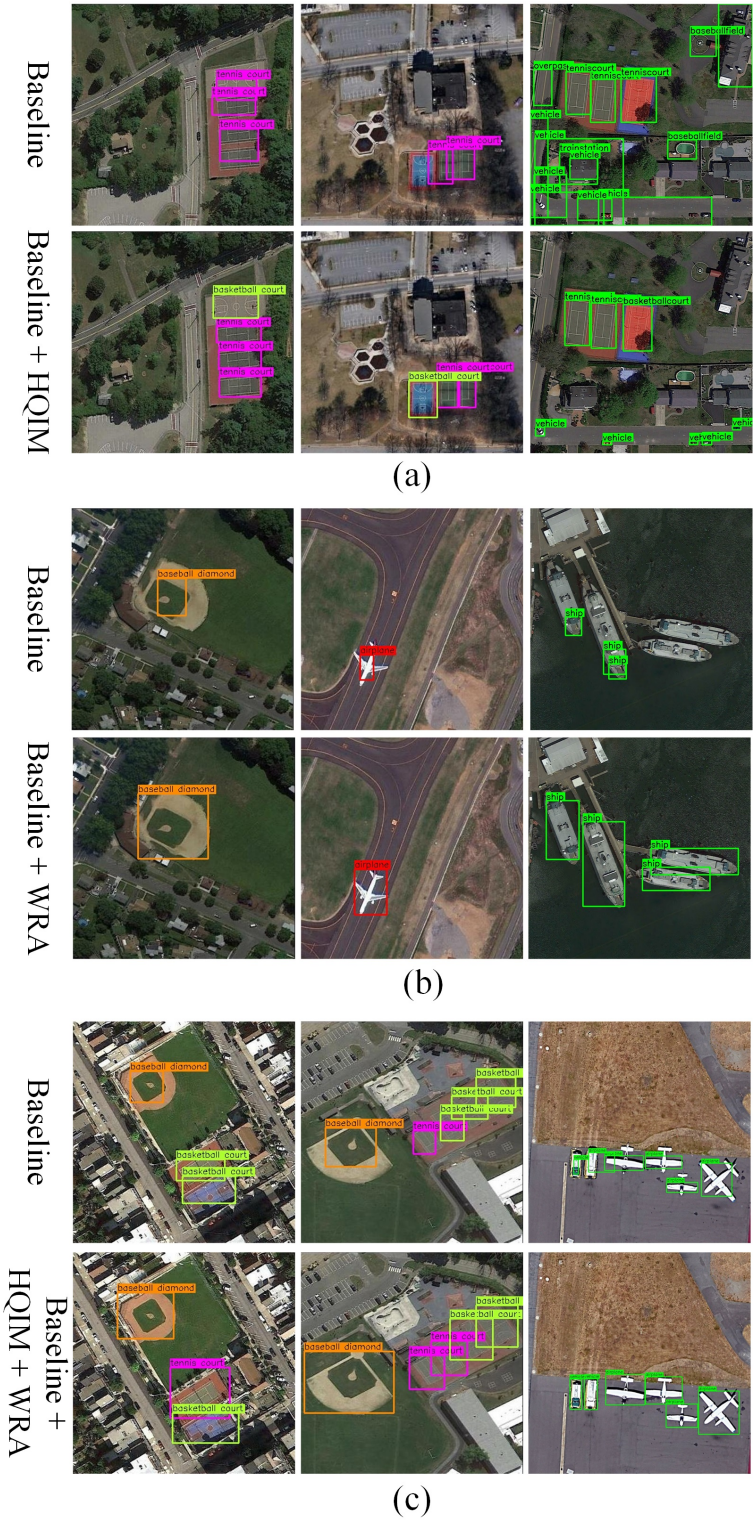


Figure 6. A subjective validation of the HQIM module (**a**), the WRA strategy (**b**), and their combination (**c**) was conducted on the NWPU VHR-10.V2 and DIOR datasets. Specifically, the RSIs enclosed within green rectangles were selected from the DIOR dataset, and the remaining RSIs were selected from the NWPU VHR-10.V2 dataset.

4.2.4. Quantitative Comparison with Popular Models

To verify the performance of our model, two classical FSOD models (Fast R-CNN [1] and Faster R-CNN [9]) and eleven popular WSOD models, which included the WSDDN [2], OICR [11], MIST [24], the dynamic curriculum learning (DCL) model [16], the progressive contextual instance refinement (PCIR) model [19], the multiple-instance graph (MIG) model [36], the triple context-aware (TCA) model [35], the self-supervised adversarial and equivariant (SAE) model [37], the self-guided proposal generation (SPG) model [22], the PISLM model [25], and the SGPLM model [12], were quantitatively compared with our model on the NWPU VHR-10.v2 and more challenging DIOR datasets. Particularly, the results of the compared methods are cited from previous publications, and the DCL model without available source codes lacks class-by-class CorLoc metrics in the NWPU VHR-10.v2 and DIOR datasets.

As shown in Table 2, the table lists different methods (columns) and the corresponding average precision (AP) for various object categories (rows) in the NWPU VHR-10.v2 dataset. The mAP of each model is provided in the last column of Table 2, which represents the mean AP across all categories for each model. Our method achieves the mAP of 66.24% on the NWPU VHR-10.v2 dataset, outperforming the benchmark model OICR by 31.72%, thereby demonstrating its overall effectiveness. Furthermore, our model's mAP surpasses the following models by the respective percentages: WSDDN by 31.12%, MIST by 14.72%, DCL by 14.13%, PCIR by 11.27%, MIG by 10.29%, TCA by 7.42%, SAE by 5.52%, SPG by 3.44%, PISLM by 2.44%, and SGPLM by 1.04%.

Table 2. Comparisons with popular models on the 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 [1]	90.91	90.60	89.29	47.32	100.00	85.85	84.86	88.22	80.29	69.84	82.72
Faster R-CNN [9]	90.90	86.30	90.53	98.24	89.72	69.64	100.00	80.11	61.49	78.14	84.51
WSDDN [2]	30.08	41.72	35.98	88.90	12.86	23.85	99.43	13.94	1.92	3.60	35.12
OICR [11]	13.66	67.35	57.16	55.16	13.64	39.66	92.80	0.23	1.84	3.73	34.52
MIST [24]	69.69	49.16	48.55	80.91	27.08	79.85	91.34	46.99	8.29	13.36	51.52
DCL [16]	72.70	74.25	37.05	82.64	36.88	42.27	83.95	39.57	16.82	35.00	52.11
PCIR [19]	90.78	78.81	36.40	90.80	22.64	52.16	88.51	42.36	11.74	35.49	54.97
MIG [36]	88.69	71.61	75.17	94.19	37.45	47.68	100.00	27.27	8.33	9.06	55.95
TCA [35]	89.43	78.18	78.42	90.80	35.27	50.36	90.91	42.44	4.11	28.30	58.82
SAE [37]	82.91	74.47	50.20	96.74	55.66	72.94	100.00	36.46	6.33	31.89	60.76
SPG [22]	90.42	81.00	59.53	92.31	35.64	51.44	99.92	58.71	16.99	42.99	62.89
PISLM [25]	87.60	81.00	57.30	94.00	36.40	80.40	100.00	56.90	9.80	35.60	63.80
SGPLM [12]	90.70	79.90	69.30	97.50	41.60	77.50	100.00	44.40	17.20	33.50	65.20
Ours	90.81	80.52	73.42	96.59	47.35	78.94	100.00	43.89	18.35	33.51	66.24

As shown in Table 3, the table lists different methods (columns) and the corresponding CorLoc for various object categories (rows) in the NWPU VHR-10.v2 dataset. The CorLoc of each model is provided in the last column of Table 3, which represents the mean CorLoc across all categories for each model. The CorLoc of our model achieves 76.89% on the NWPU VHR-10.v2 dataset. Furthermore, our model's CorLoc surpasses the following models by the respective percentages: WSDDN by 41.65%, OICR by 36.88%, MIST by 6.59%, DCL by 7.19%, PCIR by 5.02%, MIG by 6.73%, TCA by 4.13%, SAE by 3.43%, SPG by 3.48%, PISLM by 2.59%, and SGPLM by 1.49%.

As shown in Table 4, the table lists different methods (columns) and the corresponding AP for various object categories (rows) in the DIOR dataset. The mAP of each model is provided in the last column of Table 4, which represents the mean AP across all categories for each model. Our methodology achieves the mAP of 28.91% on the more challenging DIOR dataset, surpassing the benchmark model OICR by 12.41%. Moreover, our model's mAP outperforms the following models by the respective percentages: WSDDN by 15.65%,

MIST by 6.73%, DCL by 8.72%, PCIR by 3.99%, MIG by 3.80%, TCA by 3.09%, SAE by 1.81%, SPG by 3.14%, PISLM by 0.31%, and SGPLM by 0.41%.

Table 3. Comparisons with popular models on the NWPU VHR-10.v2 dataset in terms of CorLoc (%).

Method	Airplane	Ship	Storage Tank	Baseball Diamond	Tennis Court	Basketball Court	Ground Track Field	Harbor	Bridge	Vehicle	mAP
WSDDN [2]	22.32	36.81	39.95	92.48	17.96	24.24	99.26	14.83	1.69	2.89	35.24
OICR [11]	29.41	83.33	20.51	81.76	40.85	32.08	86.60	7.41	3.70	14.44	40.01
MIST [24]	90.20	82.50	80.30	98.60	48.50	87.40	98.30	66.50	14.60	35.80	70.30
DCL [16]	-	-	-	-	-	-	-	-	-	-	69.70
PCIR [19]	100.00	93.06	64.10	99.32	64.79	79.25	89.69	62.96	13.26	52.22	71.87
MIG [36]	97.79	90.26	87.18	98.65	54.93	64.15	100.00	74.07	12.96	21.57	70.16
TCA [35]	96.91	91.78	95.13	88.65	66.90	62.83	95.98	54.18	19.63	55.50	72.76
SAE [37]	97.06	91.67	87.81	98.65	40.86	81.13	100.00	70.37	14.81	52.22	73.46
SPG [22]	98.06	92.67	70.08	99.65	51.86	80.12	96.20	72.44	12.99	60.02	73.41
PISLM [25]	94.40	86.60	68.50	97.80	69.80	87.50	100.00	68.60	16.00	56.60	74.60
SGPLM [12]	98.20	93.80	89.30	99.10	50.20	88.90	100.00	71.00	12.30	51.20	75.40
Ours	99.29	94.79	91.68	97.95	58.43	91.67	100.00	68.78	13.64	56.68	76.89

Table 4. Comparisons with popular models on the 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 [1]	44.17	66.79	66.96	60.49	15.56	72.28	51.95	65.87	44.76	72.11
Faster R-CNN [9]	50.28	62.60	66.04	80.88	28.80	68.17	47.26	58.51	48.06	60.44
WSDDN [2]	9.06	39.68	37.81	20.16	0.25	12.28	0.57	0.65	11.88	4.90
OICR [11]	8.70	28.26	44.05	18.22	1.30	20.15	0.09	0.65	29.89	13.80
MIST [24]	32.01	39.87	62.71	28.97	7.46	12.87	0.31	5.14	17.38	51.02
DCL [16]	20.89	22.70	54.21	11.50	6.03	61.01	0.09	1.07	31.01	30.87
PCIR [19]	30.37	36.06	54.22	26.60	9.09	58.59	0.22	9.65	36.18	32.59
MIG [36]	22.20	52.57	62.76	25.78	8.47	67.42	0.66	8.85	28.71	57.28
TCA [35]	25.13	30.84	62.92	40.00	4.13	67.78	8.07	23.80	29.89	22.34
SAE [37]	20.57	62.41	62.65	23.54	7.59	64.62	0.22	34.52	30.62	55.38
SPG [22]	31.32	36.66	62.79	29.10	6.08	62.66	0.31	15.00	30.10	35.00
PISLM [25]	29.10	49.80	70.90	41.40	7.20	45.50	0.20	35.40	36.80	60.80
SGPLM [12]	39.10	64.60	64.40	26.90	6.30	62.30	0.90	12.20	26.30	55.30
Ours	42.19	65.01	66.15	25.74	6.70	60.15	1.29	13.48	25.31	57.81

Method	Ground Track Field	Harbor	Overpass	Ship	Stadium	Storage Tank	Tennis Court	Train Station	Vehicle	Windmill	mAP
Fast R-CNN [1]	62.93	46.18	38.03	32.13	70.98	35.04	58.27	37.91	19.20	38.10	49.98
Faster R-CNN [9]	67.00	43.86	46.87	58.48	52.37	42.35	79.52	48.02	34.77	65.44	55.49
WSDDN [2]	42.53	4.66	1.06	0.70	63.03	3.95	6.06	0.51	4.55	1.14	13.27
OICR [11]	57.39	10.66	11.06	9.09	59.29	7.10	0.68	0.14	9.09	0.41	16.50
PCL [13]	56.36	16.76	11.05	9.09	57.62	9.09	2.47	0.12	4.55	4.5	18.19
MELM [15]	41.05	26.12	0.43	9.09	8.28	15.02	20.57	9.81	0.04	0.53	18.65
MIST [24]	49.48	5.36	12.24	29.43	35.53	25.36	0.81	4.59	22.22	0.80	22.18
DCL [16]	56.45	5.05	2.65	9.09	63.65	9.09	10.36	0.02	7.27	0.79	20.19
PCIR [19]	58.51	8.60	21.63	12.09	64.28	9.09	13.62	0.30	9.09	7.52	24.92
MIG [36]	47.73	23.77	0.77	6.42	54.13	13.15	4.12	14.76	0.23	2.43	25.11
TCA [35]	53.85	24.84	11.06	9.09	46.40	13.74	30.98	1.47	9.09	1.00	25.82
SAE [37]	52.70	17.57	6.85	9.09	51.59	15.43	1.69	14.44	1.41	9.16	27.10
SPG [22]	48.02	27.11	12.00	10.02	60.04	15.10	21.00	9.92	3.15	0.06	25.77
PISLM [25]	48.50	14.00	25.10	18.50	48.90	11.70	11.90	3.50	11.30	1.70	28.60
SGPLM [12]	60.60	9.40	23.10	13.40	57.40	17.70	1.50	14.00	11.50	3.50	28.50
Ours	61.42	10.41	20.40	14.14	58.62	18.91	2.16	13.61	10.02	4.69	28.91

As shown in Table 5, the table lists different methods (columns) and the corresponding CorLoc for various object categories (rows) in the DIOR dataset. The CorLoc of each model is provided in the last column of Table 5, which represents the mean CorLoc across all categories for each model. Notably, our model achieves a CorLoc of 53.92% on the challenging DIOR dataset. Furthermore, our model's CorLoc surpasses the following models by significant margins: WSDDN by 21.48%, OICR by 19.15%, MIST by 10.32%, DCL by 11.72%, PCIR by 7.80%, MIG by 7.12%, TCA by 5.51%, SAE by 4.50%, SPG by 5.62%, PISLM by a narrow margin of 0.72%, and SGPLM by 0.72%.

Table 5. Comparisons with popular models on the DIOR dataset in terms of CorLoc (%)

Method	Airplane	Airport	Baseball Field	Basketball Court	Bridge	Chimney	Dam	Expressway Service Area	Expressway Toll Station	Golf Field
WSDDN [2]	5.72	59.88	94.24	55.94	4.92	23.40	1.03	6.79	44.52	12.75
OICR [11]	15.98	51.45	94.77	55.79	2.63	23.89	0.00	4.82	56.68	22.42
MIST [24]	91.60	53.20	93.50	66.30	10.80	30.70	1.50	14.03	35.20	47.50
DCL [16]	-	-	-	-	-	-	-	-	-	-
PCIR [19]	93.10	45.60	95.50	68.30	3.60	92.10	0.20	5.40	58.40	47.50
MIG [36]	76.98	46.86	95.39	63.61	23.00	95.07	0.21	16.96	57.88	50.77
TCA [35]	81.58	51.33	96.17	73.45	5.03	94.69	15.89	32.79	45.95	48.56
SAE [37]	91.20	69.37	95.48	67.52	18.88	97.78	0.21	70.54	54.32	51.43
SPG [22]	80.48	32.04	98.68	65.00	15.20	96.08	22.52	16.99	46.08	50.96
PISLM [25]	85.50	68.90	96.80	75.80	11.60	94.70	0.80	67.50	60.50	46.50
SGPLM [12]	92.20	58.30	97.80	74.20	16.20	95.20	0.30	51.30	56.20	52.30
Ours	994.14	59.34	98.12	70.46	17.59	94.61	4.80	52.19	54.27	53.47

Method	Ground Track Field	Harbor	Overpass	Ship	Stadium	Storage Tank	Tennis Court	Train Station	Vehicle	Windmill	CorLoc
WSDDN [2]	89.90	5.45	10.00	22.96	98.54	79.61	15.06	3.45	11.56	3.22	32.44
OICR [11]	91.41	18.18	18.70	31.80	98.28	81.29	7.45	1.22	15.83	1.98	34.77
MIST [24]	87.10	38.60	23.40	50.70	80.50	89.20	22.40	11.50	22.20	2.40	43.60
DCL [16]	-	-	-	-	-	-	-	-	-	-	42.20
PCIR [19]	88.60	15.80	5.20	39.50	98.10	85.60	13.40	56.50	9.70	0.60	46.10
MIG [36]	89.39	42.12	19.78	37.94	97.93	80.65	13.77	10.34	10.50	6.94	46.80
TCA [35]	85.26	38.91	20.17	30.63	84.59	91.46	56.28	3.79	10.45	1.25	48.41
SAE [37]	88.28	48.03	2.28	33.56	14.11	83.35	65.59	19.88	16.41	2.85	49.42
SPG [22]	89.18	49.45	22.00	35.16	98.61	90.04	32.56	12.73	9.98	2.34	48.30
PISLM [25]	75.20	50.50	28.30	39.70	92.60	77.00	55.10	10.10	20.90	5.60	53.20
SGPLM [12]	91.70	48.60	23.00	32.70	98.80	89.30	43.50	19.50	18.30	4.00	53.20
Ours	93.04	49.71	22.04	34.81	99.01	90.42	44.13	18.31	17.52	10.51	53.92

In addition, our model significantly reduces the disparity with respect to the FSOD models on the NWPU VHR-10.v2 and DIOR datasets. Notably, our model's performance in specific classes, e.g., airplane, ground track field, and baseball field, approaches that of the two classical FSOD models.

Furthermore, to validate the generalization capability of our method, we also conducted experiments on the PASCAL VOC 2007 dataset. Specifically, our model achieved the mAP of 57.25% on the PASCAL VOC 2007 dataset, surpassing the benchmark model OICR by 16.05%. Additionally, our model's CorLoc reached 71.56%, outperforming OICR by 10.96%.

4.2.5. Subjective Evaluation

Figures 7 and 8 give the visualization of some detection results of our model on the NWPU VHR-10.v2 and DIOR datasets, respectively. Obviously, our method intuitively demonstrates excellent performance on the 24 test images.



Figure 7. Visualization of some detection results of our model on the NWPU VHR-10.v2 dataset.



Figure 8. Visualization of some detection results of our model on the DIOR dataset.

In summary, the better performance of our method mainly contributes to two aspects. Firstly, the HQIM module removes the misclassified neighboring instances; thus, the quality of positive instances is improved. Secondly, the WRA strategy re-assigns a high loss weight to the instances that broadly cover the whole object, thereby urging our model to detect the whole object.

4.2.6. Evaluation of Computational Cost

As shown in Table 6, in order to evaluate the computational cost, our method was compared with two classical WSOD models (WSDDN and OICR) and five advanced WSOD models with available source codes (PCL, MELM, MIST, PISLM, and SPGLM) on the NWPU VHR-10.v2 dataset in terms of training time, frames per second (FPS), giga floating point operations (GFLOPs), and mAP.

Table 6. Comparisons with popular models on the NWPU VHR-10.V2 dataset in terms of mAP and computation cost.

Method	Training Time (H)	Inference Speed (fps)	GFLOPS (G)	mAP (%)
WSDDN [2]	2.89	4.99	287.61	35.12
OICR [11]	3.25	4.15	286.23	34.52
PCL [13]	4.85	4.71	287.14	39.41
MELM [15]	5.12	4.79	290.34	42.29
MIST [24]	6.50	4.01	290.43	51.52
PISLM [25]	9.83	3.58	296.43	63.80
SGPLM [12]	10.50	4.21	334.26	65.20
Ours	4.12	4.93	289.93	66.24

The proposed method exhibits a marginally longer training time and slightly higher GFLOPs compared with OICR. However, it significantly outperforms OICR in terms of mAP. Particularly, in practical application, the trained model is deployed within a workstation to infer the test image. Obviously, during the inference stage, our method attains an impressive 4.93 fps while maintaining a high mAP of 66.24%. In summary, these results underscore better overall efficiency of our method among the four evaluated approaches.

5. Conclusions

The primary contributions of this paper include two significant parts. On the one hand, the HQIM module is proposed to handle the problem of misclassification of neighboring instances. During the label propagating process, the labels of neighboring instances are assigned solely based on spatial distances. The HQIM module further refines the selected neighboring instances by utilizing feature similarities between them and their corresponding seed instances to remove some misclassified neighboring instances. On the other hand, the WRA strategy is proposed to handle the problem that the instance covering the discriminative part of object is assigned the highest loss weight. The loss weights of instances typically guide the model's attention; thus, WSOD models pay more attention to the instances that cover the discriminative part of an object rather than those that cover the whole object. The WRA strategy switches the loss weights between instances focused on the discriminative part of the object and instances that cover the whole object, encouraging the WSOD model to pay more attention to the whole object. The ablation studies demonstrate the efficacy of the HQIM module, the WRA strategy, and their combination. The quantitative comparisons with popular WSOD models show that our method achieves the mAP (CorLoc) of 66.24% (76.89%) on the NWPU VHR-10.v2 dataset, outperforming the benchmark model OICR by 31.72% (36.88%), further illustrating the superiority of our model.

The localization accuracy of our method currently depends on the proposals produced by the conventional SS algorithm, which has inherent limitations. Hence, the next step in refining our method involves improving the localization precision by adopting advanced segmentation techniques.

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

Funding: This work was supported by the National Natural Science Foundation of China (grant No. 62076223).

Data Availability Statement: The DIOR and NWPU VHR-10.v2 datasets are available at the following URLs: https://drive.google.com/drive/folders/1UdlgHk49iu6WpcJ5467iT-UqNPpx__CC (accessed on 27 March 2023) and https://drive.google.com/file/d/15xd4TASVAC2irRf02GA4LqYFbH7QITR-/view (accessed on 27 March 2023), respectively.

Acknowledgments: We are deeply grateful to the authors of the WSDDN and OICR for generously sharing their source codes, which have significantly facilitated our experiment. Additionally, the Vaihingen dataset was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [Cramer, 2010]: https://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html (accessed on 28 March 2023).

Conflicts of Interest: The authors explicitly state that they have no conflicts of interest to declare.

Abbreviations

The following abbreviations are used in this manuscript:

WSOD	weak supervised object detection
RSI	remote sensing image
CS	class score
HQIM	high-quality instance mining
WRA	weight re-assigning

References

1. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (ICCV 2015), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
2. Bilen, H.; Vedaldi, A. Weakly supervised deep detection networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2846–2854.
3. Cheng, G.; Han, J.; Lu, X. Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proc. IEEE* **2017**, *105*, 1865–1883. [CrossRef]
4. Li, M.; Li, W.; Liu, Y.; Huang, Y.; Yang, G. Adaptive Mask Sampling and Manifold to Euclidean Subspace Learning with Distance Covariance Representation for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5508518. [CrossRef]
5. Huo, Y.; Cheng, X.; Lin, S.; Zhang, M.; Wang, H. Memory-Augmented Autoencoder with Adaptive Reconstruction and Sample Attribution Mining for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5518118. [CrossRef]
6. 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*, 5504205. [CrossRef]
7. 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]
8. Cheng, X.; Huo, Y.; Lin, S.; Dong, Y.; Zhao, S.; Zhang, M.; Wang, H. Deep Feature Aggregation Network for Hyperspectral Anomaly Detection. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5033016. [CrossRef]
9. 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]
10. 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*, 5605209. [CrossRef]
11. 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 (CVPR), Honolulu, HI, USA, 21 July–26 July 2017; pp. 2843–2851.

12. 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]
13. 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.* **2020**, *42*, 176–191. [CrossRef] [PubMed]
14. Huang, Z.; Zou, Y.; Kumar, B.V.K.V.; Huang, D. Comprehensive Attention Self-Distillation for Weakly-Supervised Object Detection. In *Proceedings of the Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Sydney, Australia, 2020; Volume 33, pp. 16797–16807.
15. 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 (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 1297–1306.
16. 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]
17. Qian, X.; Wang, C.; Li, C.; Li, Z.; Zeng, L.; Wang, W.; Wu, Q. Multiscale Image Splitting Based Feature Enhancement and Instance Difficulty Aware Training for Weakly Supervised Object Detection in Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 7497–7506. [CrossRef]
18. Huo, Y.; Qian, X.; Li, C.; Wang, W. Multiple Instance Complementary Detection and Difficulty Evaluation for Weakly Supervised Object Detection in Remote Sensing Image. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6006505. [CrossRef]
19. 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]
20. Qian, X.; Wang, C.; Wang, W.; Yao, X.; Cheng, G. Complete and Invariant Instance Classifier Refinement for Weakly Supervised Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5627713. [CrossRef]
21. Seo, J.; Bae, W.; Sutherland, D.J.; Noh, J.; Kim, D. Object Discovery via Contrastive Learning for Weakly Supervised Object Detection. In *Proceedings of the Computer Vision—ECCV 2022*; Springer Nature: Cham, Switzerland, 2022; pp. 312–329.
22. 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*, 5625311. [CrossRef]
23. Xie, X.; Cheng, G.; Feng, X.; Yao, X.; Qian, X.; Han, J. Attention Erasing and Instance Sampling for Weakly Supervised Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5600910. [CrossRef]
24. 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.
25. Qian, X.; Huo, Y.; Cheng, G.; Gao, C.; Yao, X.; Wang, W. Mining High-Quality Pseudoinstance Soft Labels for Weakly Supervised Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5607615. [CrossRef]
26. 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]
27. 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]
28. Li, K.; Cheng, G.; Bu, S.; You, X. Rotation-Insensitive and Context-Augmented Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2337–2348. [CrossRef]
29. Cramer, M. The DGPF-Test on Digital Airborne Camera Evaluation Overview and Test Design. *Photogramm.-Fernerkund.-Geoinf.* **2010**, *2010*, 73–82. [CrossRef]
30. 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]
31. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
32. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
33. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
34. Hosang, J.; Benenson, R.; Schiele, B. Learning non-maximum suppression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21 July–26 July 2017; pp. 4507–4515.
35. 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]
36. Wang, B.; Zhao, Y.; Li, X. Multiple Instance Graph Learning for Weakly Supervised Remote Sensing Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5613112. [CrossRef]
37. 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]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.