Adaptively Center-Shape Sensitive Sample Selection for Ship Detection in SAR Images

Yilong Lv, Min Li, Yujie He, Yu Song, and Weidong Du

Abstract—With the wide application of synthetic aperture radar in maritime surveillance, a ship detection method has been rapidly developed. However, there is still a key problem common in most methods, i.e., how to select positive and negative samples. The mainstream MaxIoUAssign has inherent problems, such as a fixed threshold and rough classification, resulting in the low quality of the positive samples. To solve these problems, we propose a new sample selection method called adaptively center-shape sensitive sample selection. The proposed method introduces shape similarity between proposal boxes and ground truth as one of the evaluation criteria and collaborates with intersection of union (IoU) to measure the quality of the proposal boxes. Meanwhile, the center distance between proposal boxes and ground truth is used to control the influence degree of IoU and shape similarity. In this way, the quality score of the proposal boxes can be determined through IoU, shape similarity, and center position, making sample selection more comprehensive. Additionally, to avoid a fixed threshold, the standard deviation of the quality score is used as a variable to form the adaptive threshold. Finally, we conducted extensive experiments on the benchmark SAR ship detection dataset (SSDD) and high-resolution SAR images datasets (HIRSID) datasets. The experimental results demonstrated the superiority of our method.

Index Terms—Adaptive threshold, center distance, samples selection, shape similarity, ship detection, synthetic aperture radar (SAR).

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) is a high-resolution image radar. As an active microwave imaging sensor, its microwave imaging process has a certain penetration effect on ground targets and is less affected by the environment. Thus, it can effectively detect various hidden targets. At the same time, its all-weather advantages enable it to complete exploration missions in all extreme conditions. Because of these characteristics, SAR has been widely used in ship detection [1], [2], [3], [4], [5], [6].

Traditional SAR image ship detection methods mainly infer the ship’s location and classification by observing the difference between the hull and background. There are three methods based on: 1) statistical features; 2) threshold; 3) transformation. For example, Iervolino and Guida [7] considered the marine clutter and signal backscattering in SAR images and proposed a generalized likelihood ratio test detector. Lang et al. [8] proposed a spatial enhanced pixel descriptor to realize the spatial structure information of the ship target and improve the separability between the ship target and ocean clutter. Leng et al. [9] defined the area ratio invariant feature group to modify the traditional detector. Among them, the constant false alarm rate [10], [11], [12] detection method and its improved version are the most widely studied. However, the traditional SAR ship detection method is not very reliable, and it is difficult to achieve accurate detection based on the difference between the hull and background.

Recently, convolutional neural networks (CNNs) have also been developed in object detection owing to the enhancement of deep learning and graphic processing unit (GPU) computing capability. Meanwhile, the detection performance of the SAR ship based on deep CNNs has been significantly improved. In particular, an accurate location is of great significance to SAR ship detection.

Currently, the precise location work mainly focuses on improving the network model, such as proposing a better network architecture or better strategy to extract reliable local features to obtain more accurate boundary regression. Specifically, these works are reflected in the category of the object detection algorithm. The first work divides the algorithm into anchor-based and anchor-free algorithms, which improve detection performance by constantly improving the design of the framework. The second work divides the algorithm into one-stage and two-stage by adjusting the training strategy.

The difference between the anchor-based and anchor-free algorithms lies in the generation method of the proposal boxes. The former generates some proposal boxes based on the anchor. The anchor needs to be manually designed according to the statistical characteristics of the datasets. Current mainstream anchor-based object detection algorithms include Faster R-CNN [13], RetinaNet [14], and you only look once (YOLO) [15], which search proposal boxes through the anchor and finally determine the target position. Then, the latter generates proposal boxes based on key or central points, which tries to eliminate an artificial anchor setting to reduce artificial interference. Current mainstream anchor-free algorithms include fully convolutional one-stage object detector (FCOS) [16], CornerNet [17], and CenterNet [18]. In addition to the above differences, their training strategies are pretty much the same, i.e., the proposal boxes will be divided into positive and negative samples using the sample selection method. Finally, positive and negative samples are used for the regression of ground truth.
The above process demonstrates that if the selected positive samples are very close to ground truth in center distance and shape, boundary regression will converge faster and the prediction accuracy will be higher. For example, the anchor-free FCOS algorithm distributes anchor points evenly in the image according to CNN’s downsample rate, and each anchor point predicts ground truth within a certain range. Once the target center point is close enough to an anchor point, the anchor point generates proposal boxes. The advantage of the FCOS algorithm is that the proposal box is closer to ground truth in center distance, but its shape is not accurate. Additionally, the anchor-based RetinaNet algorithm uses an artificial anchor as a proposal box to obtain positive samples with good position and shape. However, the anchor does not necessarily cover targets well, and many small targets correspond to fewer proposal boxes, which will be difficult to detect, as shown in Fig. 1.

The main difference between one-stage and two-stage algorithms is whether proposal boxes carry out subsequent second processing. In the one-stage algorithm, the proposal boxes do not perform preliminary screening, but they are directly used for selecting the sample, leading to the low quality of the positive samples in location and shape. Current mainstream one-stage object detection algorithms include single shot multi-box detector (SSD) [19], RetinaNet, and YOLO. By contrast, the two-stage algorithm first filters out some proposal boxes without appropriate positions and shapes and then uses the remaining proposal box for sample selection. Current mainstream two-stage object detection algorithms include Cascade Region-CNN (Cascade RCNN) [20], Libra RCNN [21], and region-based fully convolutional networks (R-FCN) [22]. The former training speed is relatively fast. Meanwhile, the latter has a slow speed but relatively high detection accuracy. This is because, after two-stage, the positive sample is closer to ground truth in position and shape, making up for the low quality of the positive sample. However, the influence of different sample selection methods on detection performance is almost not discussed in either single-stage or two-stage detection algorithms. Our experimental results showed that different sample selection methods influenced the model to select the best-quality positive samples.

In this article, we analyze the anchor-based and anchor-free algorithms, and one-stage and two-stage algorithms. Consequently, we conclude that each algorithm focuses on how to acquire high-quality proposal boxes. Thus, assuming that each algorithm can obtain high-quality proposal boxes, the performance gap between different algorithms will be reduced. However, although improving the network model can improve the quality of the proposal boxes, it will also bring problems, such as the network architecture being difficult to unify and a complex model. Additionally, more precise predictions generally require more model parameters and training time. Thus, it is not the most economical.

A widely overlooked improvement is how to effectively select the positive and negative samples from the mixed proposal box. As long as a remarkable selection strategy is used to select high-quality proposals, it is not necessary to make laborious changes to the network structure. Currently, the mainstream sample selection method is MaxIoUAssign, but this method can only roughly evaluate the quality of the proposal boxes. MaxIoUAssign is not fully competent because of the fixed threshold value and complex proposal box distribution. In view of these situations, Zhang et al. [23] proposed an adaptive training sample selection (ATSS) method to investigate the differences between anchor-based and anchor-free algorithms. It adaptively adjusts the threshold according to the statistical characteristics of the proposal box intersection of union (IoU). Additionally, Zhu et al. [24] proposed the auto-assign that adopts a confidence weighting module to modify the positive and negative confidences of the locations in the spatial and scale dimensions. Zhang et al. [25] proposed a free anchor that adopts a learning-to-match approach and selects positive and negative samples through network training, thus eliminating manual design. Kim and Lee [26] proposed probabilistic anchor assignment that fits a Gaussian mixed distribution according to the training state of the model and uses the distribution to adaptively separate proposal boxes into positive and negative samples. However, these methods do not consider the validity of the IoU-based evaluation criteria, which is the problem identified in this article.

We found that using IoU to evaluate proposal boxes is very rough. So, IoU does not describe the importance of a proposal box uniquely and does not effectively describe some of the situations that often occur in a sample selection, as shown in Fig. 2(a). Intuitively, because B is more like the ground truth, we should choose B over A. However, their IoU is near, which means that they are the same. Fig. 2(b) shows that A is the proposal box completely covered by ground truth. Because A only contains a part of the object, it is not easy to predict the entire target. However, the proposal box B consists of part of ground truth and background. It is difficult to predict the whole object based on part of ground truth, but background information can assist the network model to accurately predict the coordinates. Therefore, A should be abandoned, and B should be selected as a positive sample. However, A will have a larger IoU than B. These situations result in the suboptimal performance of the model.
To select high-quality positive samples from the proposal box, we propose a novel sample selection strategy called adaptively center-shape sensitive sample selection (AC4S). Compared with ATSS, autoassign, and other methods, our method not only relieves the disadvantages of the conventional MaxIoUAssign method but also does not add new hyperparameters and does not need to modify the network structure. First, it uses the shape similarity and IoU between the proposal boxes and the ground truth as the evaluation criteria of sample quality. Compared with the MaxIoUAssign method, our method can refine the evaluation of sample quality, thus improving the quality of positive samples. Second, in order to balance the influence of shape similarity and IoU, we introduce the center distance between the proposal box and the ground truth as the weight factor. Additionally, owing to the few positive samples of small targets, we adopt adaptive thresholds to increase the number of positive samples of small targets and reduced the number of positive samples of large targets. Furthermore, we conducted extensive experiments on the benchmark SAR ship detection dataset (SSDD) and high-resolution SAR images datasets (HRSID) datasets. The experimental results verified the effectiveness of the proposed method.

The main contributions of our work can be summarized as follows.

1) By observing the experimental phenomena of the current mainstream sample selection methods, we conducted a detailed analysis and found that the IoU-based evaluation criteria in sample selection are rough and the samples corresponding to different sizes of targets are unbalanced.

2) To solve the common problems in the current mainstream positive and negative sample selections, we propose the AC4S method. By using basic data from datasets, such as center location and shape similarity, the proposed method can select high-quality positive samples from a large number of proposal boxes without increasing model parameters. At the same time, fixed thresholds were replaced with adaptive ones to balance the samples of different targets.

3) We conducted extensive experiments on the benchmark SSDD and HRSID datasets to prove the effectiveness of the proposed method. The experimental results confirmed that the proposed method is effective.

The rest of this article is organized as follows. Section II illustrates the proposed method in detail. Next, the experimental results on several dataset and the corresponding analysis are provided in Section III. Finally, Section IV concludes this article.

II. METHODOLOGY

This section introduces the proposed AC4S method, which is divided into three components: 1) center-distance evaluation criteria; 2) shape-similarity evaluation criteria; 3) adaptive threshold. First, we introduce the current mainstream MaxIoUAssign method and the method proposed. Second, we analyze IoU. Next, we introduce the construction of the center-distance evaluation criteria. Then, we present the shape-similarity evaluation criteria. Finally, we introduce the structure of the adaptive threshold.

A. MaxIoUAssign

MaxIoUAssign method is one of the most widely used positive and negative sample selection methods. It is based on a fixed threshold, i.e., the IoU threshold between proposal boxes and ground truth. First, IoU between proposal boxes and ground truth is calculated one by one, and ground truth corresponding to the maximum IoU is taken as the corresponding target of the proposal box. Once the maximum IoU is greater than the fixed IoU threshold, it is regarded as a positive sample of the target; otherwise, it is a negative sample.

This method is generally suitable for most methods, including Faster-RCNN, YOLO and RetinaNet. However, this approach also has some inherent shortcomings. First, the quality of proposal boxes is not solely determined by IoU, which leads to the fact that even if some proposal boxes have the same IoU, it does not mean that they all have the same quality. Therefore, we should consider the quality of proposal box from many aspects.

Additionally, the proposal box corresponding to a small target is less than that of a large target. Thus, the IoU of the corresponding proposal box is inevitably low, so the fixed IoU threshold is not very friendly to a small target, and it may even have no positive samples. In this case, some small targets will not be detected because they cannot participate in the training, directly leading to the algorithm being not sensitive to small targets. Therefore, an appropriate sampling method should be adopted to compensate for the imbalance between large and small targets.

In this article, we used a Faster RCNN as the baseline method to verify the effectiveness of the proposed method, and its structure is shown in Fig. 3. As a two-stage target detection method, we first use the MaxIoUAssign method through RPN to extract RoI and obtain more accurate proposal boxes in the first stage. Next, RoI uses MaxIoUAssign again to extract positive and negative samples to calculate the loss function in the second stage. It is worth noting that the target of MaxIoUAssign in
the first stage is an anchor, which is manually set. Because the anchor in different positions has the same aspect ratio, it leads to failure to reflect the role of shape similarity. In view of this phenomenon, we do not improve MaxIoUAssign in the first stage, but we focus on the second stage. In the second stage, RoI will have an irregular position and shape similarity after the adjustment in the first stage. Therefore, the proposed method can be used to replace MaxIoUAssign in the second stage.

B. Analysis of IoU

To illustrate the problems with MaxIoUAssign, we explain how it works and why it is important. IoU is calculated according to the location information of the proposal box \((x_1, y_1, x_2, y_2)\) and ground-truth\((x_1', y_1', x_2', y_2')\). Its calculation formula is as follows:

\[
\text{IoU} = \frac{\min(x_2, x_2') - \max(x_1, x_1')\min(y_2, y_2') - \max(y_1, y_1')}{(x_2 - x_1)(y_2 - y_1) + (x_2' - x_1')(y_2' - y_1') - \uparrow}
\]

where \(\min(\cdot)\) and \(\max(\cdot)\) represent the maximum and minimum values, respectively. \(\uparrow\) indicates copying the numerator above. To explore how the IoU function is affected by center distance and shape, \((x_1, y_1, x_2, y_2)\) to \((x, y, 2w, 2h)\), \((x_1', y_1', x_2', y_2')\) to \((x', y', 2w', 2h')\) is performed, where the results of the transformation represent the central x, y coordinates, width, and height of a box. We continue to transform the components in formula (1) as shown

\[
\min(x_2, x_2') - \max(x_1, x_1')
\]

\[
\min(x', w, x + w) - \max(x', w, x + w).
\]

There are two possible values of \(\min(x' + w, x + w)\) and \(\max(x' - w, x - w)\). We present all possible situations to analyze how the center distance and shape similarity affect this formula. All possible situations are as follows:

\[
(x' + w') - (x' - w') = 2w' \tag{3}
\]

\[
(x + w) - (x' - w') = (x - x') + (w + w') \tag{4}
\]

\[
(x' + w') - (x - w) = (x' - x) + (w' + w) \tag{5}
\]

\[
(x + w) - (x - w) = 2w \tag{6}
\]

In the case of formulas (3) and (6), the central x, y coordinates of the proposal box and ground truth will not participate in calculating IoU. This detailed analysis shows that in the case of formulas (3) and (6), the center points of the proposal box and ground truth have no obvious positional relationship. At this time, IoU can no longer meet the needs of positive and negative sample selections. Therefore, the proposed method adds an evaluation criterion to compensate for the above vacancy. It is obvious that the smaller the center distance \(L_{cen}\) is, the faster the boundary regression will converge.

We continue to study the influence of shape on IoU. For the convenience of our calculation, by assuming that \(L_{cen}\) has reached the optimal: \((x - x') = (y - y') = 0\), then formula (1) will be converted to formula (7). The results in the following:

\[
\text{IoU} = \frac{C_1}{4w} \frac{1}{2w' + 2h'} \left(2w, 2w' + w, 2w \right) C_3 \left(2h', h + h', 2h \right)
\]

\[
= \frac{C_1}{4w} \frac{1}{w + h} \left(2w, 2w + h, 2w \right) C_3 \left(2h, h + 1, 2h \right).
\]

Formula (7) shows that IoU is determined by \(\frac{w}{w'}\) and \(\frac{h}{h'}\), namely, shape similarity \(L_{shape}\), indicating that \(L_{shape}\) is an important criterion for evaluating the quality of proposal boxes. Therefore, to obtain higher quality positive samples, \(L_{shape}\) is taken as an important evaluation criterion.

We comprehensively evaluates the quality of proposal box from three aspects of \(L_{cen}\), \(L_{shape}\), and IoU.

For large targets, there are many proposal boxes that meet the requirements of the IoU threshold, and their center point and shapes are rich. In general, the larger the \(L_{cen}\) is, the more attention is paid to \(L_{shape}\). On the contrary, when \(L_{cen}\) is small, comprehensive consideration should be given to the IoU of the proposal boxes because shape similarity will be less important. For small targets, owing to the small number of its proposal boxes, the method of using a fixed threshold will lead to the imbalance of samples corresponding to small targets, affecting the training of small targets. Therefore, a method to balance the size of the target sample should be considered.

To solve the above problem, we propose the AC4S method, and its process is shown in Algorithm 1. We inherited and transcended MaxIoUAssign method. We also investigated the influence of center distance and shape similarity on experimental results.
Algorithm 1: Adaptively Center-Shape Sensitive Sample Selection.

**Input:**
1. \( G \in R^{n \times 4} \) is a set of ground-truth on the images
2. \( A \in R^{k \times 4} \) is a set of all anchor
3. \( \text{thre} \) is a mutual threshold of sample selection

**Output:**
4. \( S_{pos} \) is a set of proposal positive samples on the image
5. \( S_{neg} \) is a set of proposal negative samples on the image
6. \( \text{IoU}() \) is a function of calculating IoU
7. \( \text{Cen}() \) is a function of calculating Center-Distance
8. \( \text{Shape}() \) is a function of calculating Shape-Similarity
9. \( T_{score} \) is a Tensor to store Quality-Score for each proposal box
10. \( \text{neg} \) is a set of proposal negative samples on the image
11. \( \text{pos} \) is a set of all anchor
12. \( \text{var} \) is a variance
13. \( \text{thre} \) is a threshold
14. \( \text{dim} \) is a dimension
15. \( \text{std} \) is a standard deviation

**Variables:**
- \( k \): number of anchor
- \( n \): number of ground-truth
- \( \text{score} \): quality score
- \( \text{index} \): index of \( \text{score} \)
- \( \text{variance} \): variance
- \( \text{threshold} \): threshold

**Algorithm:**

```python
Input: G \in R^{n \times 4} is a set of ground-truth on the images
A \in R^{k \times 4} is a set of all anchor
thre is a mutual threshold of sample selection
Output: S_{pos} is a set of proposal positive samples on the image
S_{neg} is a set of proposal negative samples on the image
IoU() is a function of calculating IoU
Cen() is a function of calculating Center-Distance
Shape() is a function of calculating Shape-Similarity
T_{score} is a Tensor to store Quality-Score for each proposal box

for i \leftarrow 1 \text{ to } k do
    for j \leftarrow 1 \text{ to } n do
        iou \leftarrow \text{IoU}(A[i], G[j])
        center \leftarrow \text{Cen}(A[i], G[j])
        shape \leftarrow \text{Shape}(A[i], G[j])
        T_{score}[i, j] \leftarrow (1 - center) \times iou + center \times shape
    end for
end for

score = \text{Max}(T_{score}, \text{dim} = 1)
index = \text{Argmax}(T_{score}, \text{dim} = 1)
compute each ground-truth variance: var = \text{std(score[index])}
compute each ground-truth threshold: thre = 0.5 + var

build an empty set for proposal positive and negative samples of the ground-truth
S_{pos} \leftarrow \emptyset, S_{pos} \leftarrow \emptyset

for l \leftarrow 1 \text{ to } n do
    for each proposal box \( s \in \text{score[index] == l} \) do
        if \( s > \text{thre} \) then
            S_{pos} \leftarrow S_{pos} \cup s
        else
            S_{neg} \leftarrow S_{neg} \cup s
        end if
    end for
end for
return \( S_{pos}, S_{neg} \)
```

C. Center Distance

Center distance is a measure of the difference between the positions of two boxes. Considering the boundary regression task of target detection, the closer the center point of the proposal boxes is to the center point of ground truth, the closer the predicted value is to 0, making it easier for the boundary regression to converge to the label. Therefore, when selecting positive and negative samples, center distance is an important criterion for evaluating positive and negative samples. In particular, the proposal boxes around ground truth must be considered.

We designed an evaluation function as a criterion to calculate the distance between two center points. Its form is shown in formula (8). Intuitive understanding is shown in Fig. 4.

\[
L_{cen} = \sqrt{\min(x_1 - x'_1, x_2 - x_1) \times \min(y_1 - y'_1, y_2 - y_1)}
\]

\[
(I \text{f } x_1 - x'_1 > 0, x'_2 - x_1 > 0, y_1 - y'_1 > 0, y'_2 - y_1 > 0. \text{else.})
\]

(8)

Here, \( L_{cen} \) represents center distance. We use the \( L_{cen} \) evaluation criteria to select proposal boxes. Fig. 5 shows that the selected proposal boxes are concentrated near the label.

The value of the evaluation function is always greater than or equal to 0 and less than 1, which meets our basic properties for an evaluation function. Fig. 5 shows that the \( L_{cen} \) is closer to 1 when the center point is closer to the label. It is worth noting that when the center point of the proposal box is not in ground truth, \( L_{cen} \) is set to 0. We make up for the situation in formulas (3) and (6) by setting the center-distance evaluation criteria, which play an important role in selecting high-quality positive and negative samples.
Fig. 6. Statistical characteristics about \( L_{\text{shape}} \). Figures A, B, and C represent the situation of small, large, and dense targets, respectively. The vertical axis represents the amount of RoI. The horizontal coordinate represents the value of \( L_{\text{shape}} \).

D. Shape Similarity

To determine the shape distribution of the proposal boxes, we selected three images with different target characteristics from the SSDD dataset to observe their distribution, including small, large, and dense targets. Their statistical characteristics regarding \( L_{\text{shape}} \) are shown in Fig. 6. Faster RCNN collected a total of 600 RoIs, and it can be seen from Fig. 6 that different targets vary greatly. In A, RoI’s \( L_{\text{shape}} \) is mainly concentrated in 0.2–0.5. In B, RoI’s \( L_{\text{shape}} \) is mainly concentrated in 0–0.3, and in C, RoI’s \( L_{\text{shape}} \) is mainly concentrated in 0.6–1. These results showed that the shape of different targets has different influences on sample selection. Therefore, it is necessary to take \( L_{\text{shape}} \) as a separate evaluation criterion for the sample selection strategy.

An anchor can roughly cover all targets in the image by setting different positions and aspect ratios, and each target can usually find the anchor with a close distance. Therefore, even if the center point is very close to the center of ground truth, it cannot be directly regarded as a positive sample. We also need to pay attention to another important factor, shape similarity, which refers to height and width ratios between the proposal box and ground truth.

Shape similarity is also important for selecting positive and negative samples. From boundary regression loss function (9), it can be found that the model tries to predict the ratio of height \( \Delta h \) and width \( \Delta w \) between the proposal boxes and ground truth. As \( \Delta h \) and \( \Delta w \) usually carry out zero initialization, it can be determined from formula (9) that if \( \log \frac{w}{w'} \) and \( \log \frac{h}{h'} \) are small, they converge very quickly and are more stable after convergence.

\[
\text{Loss}_{\text{reg}} = \text{smooth}_l \left( \left( \Delta w, \log \frac{w}{w'} \right) + \left( \Delta h, \log \frac{h}{h'} \right) \right).
\]

It should be noted that in order to make the loss function converge faster, we must make \( L_{\text{shape}} \) consistent with \( \text{Loss}_{\text{reg}} \). Therefore, referring to the structure of the boundary regression loss function, the evaluation function we designed is shown

\[
L_{\text{shape}} = \sqrt{\frac{\min(w, w') \cdot \min(h, h')}{\max(w, w') \cdot \max(h, h')}}.
\]

Here, we use sqrt to slow down the drastic changes caused by the product. This evaluation method can limit \( L_{\text{shape}} \) within the range of 0 to 1. As shown in formula (10), when the shape between the proposal boxes and ground truth is similar, the \( L_{\text{shape}} \) value will approach 1; conversely, it will be near 0. Therefore, the performance of the proposal boxes in shape similarity can be evaluated using this method.

E. Quality Score

We study the performance of \( L_{\text{shape}} \) and \( L_{\text{cen}} \) on different targets. We conducted experiments on small, large, and dense targets on the SSDD dataset, and the experimental results are shown in Fig. 7. A total of 600 RoIs were collected using Faster RCNN. \( L_{\text{shape}} \) and \( L_{\text{cen}} \) of each target were used as coordinate axis labels. Mask area is used as the proposal region for positive samples. However, the RoI of different targets in this region varies greatly, which is not conducive to the balance of training samples. Therefore, it is not enough to select samples only from \( L_{\text{shape}} \) and \( L_{\text{cen}} \). Thus, we evaluate the factors influencing the IoU, \( L_{\text{shape}} \) and \( L_{\text{cen}} \) functions.
Specifically, when the center point of the proposal box is close to ground truth, a large weight is added to $L_{\text{shape}}$. When the center distance between the two boxes is far, we assign a large weight to IoU to comprehensively consider the position and shape of the proposal box. Therefore, we directly take $L_{\text{cen}}$ as the weight. Then, our quality score (QS) evaluation function is shown as follows:

$$ QS = (1 - L_{\text{cen}}) \times \text{IoU} + L_{\text{cen}} \times L_{\text{shape}}. $$ (11)

To further explore the difference between the proposed method and MaxIoUAssign, we conducted an experiment on the SSDD dataset, and the results are shown in Fig. 8. This figure shows the results obtained using QS and IoU for the same $L_{\text{shape}}$ and $L_{\text{cen}}$, respectively. When $L_{\text{shape}}$ and $L_{\text{cen}}$ are small, the difference between IoU and QS is not large because when $L_{\text{cen}}$ is 0, QS will degenerate into IoU. With the increase in $L_{\text{shape}}$ and $L_{\text{cen}}$, the difference between IoU and QS gradually becomes larger. The gap boundary between high-quality and low-quality RoIs becomes clearer by replacing IoU with QS, facilitating the separation of positive and negative samples.

### F. Adaptive Threshold

To eliminate artificial interference and balance the positive and negative samples of different sizes, we adopt an adaptive threshold to replace a fixed threshold.

Fig. 9 shows that the median of QS is generally within the range of 0 to 0.4. The score distribution of large targets is scattered, and the standard deviation is large. Thus, there are more proposal boxes with a high score. Therefore, the threshold value should be appropriately increased to obtain fewer positive samples. Small targets have small QS, concentrated distribution, and small standard deviation, so proposal boxes with a high score will become fewer, and the threshold value should be appropriately reduced to obtain more positive samples. Therefore, we consider using the standard deviation of QS as the adaptive factor of the threshold, and the details are shown in formula (12). The threshold will be adjusted adaptively according to the differences of the target proposal boxes so that small targets can select high-quality positive samples

$$ \text{thre} = \alpha + \text{std}(\text{score}). $$ (12)

Here, $\alpha$ represents a hyperparameter, and $\text{std}(\cdot)$ represents the standard deviation.

### III. Experiment

In this section, to verify the validity of the proposed method, we conducted extensive experiments on SSDD and HRSID datasets. First, we introduced the dataset, evaluation criteria,
and experimental environment. Then, to compare the differences between MaxIoUAssign and our method, we conducted experiments on the SSDD dataset and analyzed their differences. Next, we performed ablation experiments to explore the setting of the hyperparameters in the evaluation criteria. Finally, the proposed method was compared with several state-of-the-art methods on the SSDD and HRSID datasets.

A. Dataset

To prove the superiority of this method, we conducted extensive experiments on the SSDD and HRSID datasets.

SSDD is the first SAR ship dataset established in 2017. It has been widely used by many researchers since its publication and has become the baseline dataset for SAR ship detection. The SSDD dataset contains many scenarios and ships and involves various sensors, resolutions, polarization modes, and working modes. Additionally, the label file settings of this dataset are the same as those of the mainstream PASCAL visual object classes (VOC) dataset, so training of the algorithms is convenient.

In using the SSDD dataset, researchers used to randomly divide training, validation, and test datasets. These inconsistent divisions often result in the absence of common evaluation criteria. As researchers gradually discovered this problem, they began to establish uniform training and test datasets. Currently, 80% of the total dataset are training datasets, and the remaining 20% are test datasets. There are 1160 images in the SSDD dataset. Therefore, the number of images in the training dataset is 921, and the number of images in the test dataset is 239. For further refinement, images whose names end with digits one and nine are set as test datasets. In this way, the performance of various detection algorithms can be evaluated in a targeted way.

The HRSID dataset is a dataset released by University of Electronic Science and Technology of China UESTC in January 2020. HRSID is used for ship detection, semantic segmentation, and instance segmentation tasks in high-resolution SAR images. The dataset contains 5604 high-resolution SAR images and 16,951 ship instances. Its label file settings are the same as those of the mainstream of the Microsoft common objects in context (MS COCO) dataset.

B. Evaluation Criteria

To evaluate the detection performance of the algorithm model, we adopted the evaluation criteria $AP$, $AP_{50}$, $AP_{75}$, $AP_s$, $AP_m$, and $AP_l$ in the MS COCO dataset. Average Precision ($AP$) is the area under the accuracy-recall curve. $AP$ is calculated by precision and recall, where precision and recall are shown in formula (13). It is important to note that $AP$ is the mean value with $IoU = 0.50 : 0.05 : 0.95$ (primary challenge measure), $AP_{50}$ is the $AP$ with $IoU = 0.5$ (PASCAL VOC measure), and $AP_{75}$ is the $AP$ with $IoU = 0.75$ (Strict measure). $AP_s$, $AP_m$, and $AP_l$ represent $AP$ of small target, medium target and large target respectively, where small target with an area less than $32^2$ pixels, medium target with an area between $32^2$ pixels and $96^2$ pixels, and large target with an area greater than $96^2$ pixels

$$P = \frac{TP}{TP + FP} \times 100\%$$  \hfill (13)

$$R = \frac{TP}{TP + FN} \times 100\%.$$  \hfill (14)

Table II: Effect of different weights of $IoU$ and $L_{shape}$

| $IoU$ | $L_{shape}$ | $three$ | $recall$ | $AP_{50}$ | $AP_{75}$ | $AP_s$ | $AP_m$ | $AP_l$ |
|-------|-------------|---------|----------|-----------|-----------|--------|--------|--------|
| 0.5   | 0.5         | 0.5     | 63       | 93.5      | 65.3      | 54.2   | 62.9   | 55.6   |
| 0.6   | 0.4         | 0.5     | 64       | 94.1      | 65.7      | 54.9   | 65.4   | 57.5   |
| 0.7   | 0.3         | 0.5     | 63.5     | 93.3      | 66.7      | 55.9   | 62.6   | 55.7   |
| 0.8   | 0.2         | 0.5     | 64.6     | 93.3      | 69.2      | 56.4   | 64.1   | 59.6   |
| 1     | 0           | 0.5     | 63.7     | 94.6      | 64.9      | 55.2   | 63.7   | 44.4   |
| $1 - L_{cen}$ | $L_{cen}$ | 0.5     | 65.2     | 95.5      | 67.3      | 56.8   | 65.3   | 60     |

Fig. 10. PR curves of different weights on SSDD.

Table III: Effect of different threshold

| $\alpha$ | $std$ | $recall$ | $AP_{50}$ | $AP_{75}$ | $AP_s$ | $AP_m$ | $AP_l$ |
|----------|-------|----------|-----------|-----------|--------|--------|--------|
| 0.5      | ✓     | 65.2     | 95.5      | 67.3      | 56.8   | 65.3   | 60     |
| 0.55     | ✓     | 65.2     | 96.3      | 64.9      | 56.3   | 65.4   | 61.7   |
| 0.45     | ✓     | 64.1     | 95.4      | 64       | 55     | 63.6   | 57.7   |
| 0.6      | ✗     | 64.5     | 95.1      | 67.8      | 55.9   | 64.6   | 57.2   |
| 0.7      | ✗     | 64.1     | 94.6      | 66       | 55.5   | 64.7   | 53.6   |
| 0.8      | ✗     | 63.5     | 94.8      | 64.3      | 55.1   | 63.3   | 60.4   |
Fig. 11. Intuitive difference between IoU and QS. The (1), (2), and (3) rows represent small, large, and dense targets, respectively. The (a) column represents the original image. The (b) column represents the QS and IoU of RoI at different positions in Faster RCNN. The (c) column shows the smoothed distributions of different RoIs in the image. (d) column represents the density of positive samples in the image corresponding to (a) column when $QS > 0.65$. (e) column represents the density of positive samples when $IoU > 0.5$. The color at ship’s center point is red, and the position away from the center point is light. The contrast of color shows that our method will make the sample quality appear more outstanding.

ships correctly classified as negative. $AP$ is defined as

$$R = \int_0^1 P(R)dR$$

where $P$ represents precision and $R$ represents recall. $AP$ is equal to the area under the curve.

In addition, floating point operations (FLOPs) and Params are adopted in this article to evaluate the computational performance and the training parameters. FLOPs can be used to measure the complexity of the model. At the same time, Frames Per Second (FPS) is adopted in this paper to evaluate the running speed. FPS is used to evaluate the number of images processed per second or the time required to process an image to assess the detection speed. The shorter the time, the faster the speed.

C. Experimental Settings

All experiments were implemented in PyTorch 1.6.0, CUDA 11.2, and cuDNN 7.4.2 with an Intel intel(R) xeon(R) silver 4110 CPU and an NVIDIA Geforce TITAN RTX GPU. The PC operating system is Ubuntu 18.04. Table I presents the computer and deep learning environment configuration for our experiments.

The algorithm model in this article is based on the MMDe-
tection framework. We trained the proposed method based on Faster RCNN using the stochastic gradient descent algorithm for 12 epochs, with a total of two images per small batch.

The initial learning rate was set to 0.01, the weight decay was 0.0001, and the momentum was 0.9. Our code is available at https://github.com/LITTERWWE/AC4S.

D. Ablation

After the analysis in Section II, we determined three influencing factors that distinguish positive and negative samples: 1) $IoU$; 2) $L_{cen}$; 3) $L_{shape}$. In this section, we will study the influence of different influencing factors on the experimental results.

1) Selection of Weight: To verify the influence of different parameters on function construction, we set different weights for formula (16) on the SSDD dataset for the ablation experiment. Additionally, instead of using the adaptive threshold, we fixed the threshold at 0.5. The detection performance of the algorithm is presented in Table II

$$score = (1 - weight) \times IoU + weight \times L_{shape}. \quad (16)$$

First, the fifth row of Table II shows the experimental results of the original algorithm Faster RCNN. It can be clearly seen from Table II that after adding $L_{shape}$, a part of the hyperparameter setting can achieve better results than the original Faster RCNN algorithm in $AP_{75}$, $AP_s$, $AP_{m}$, $AP_{l}$, and Recall. We can also see that the detection performance of $L_{cen}$ is better than manual settings and the original Faster RCNN algorithm. Finally, to further illustrate the superiority of the $L_{cen}$ method to manual settings, we draw a $PR$ curve, as shown in the Fig. 10.
(black line) is superior to other manual setting methods and the original Faster RCNN algorithm at different recall rates. Meanwhile, when recall is greater than 0.8, our curve will decline more smoothly. These show that $L_{cen}$ can replace the manual setting and original Faster RCNN algorithm.

2) Effect of Adaptive Threshold: Because the fixed threshold is not friendly to the small target, adaptive positive and negative sample selections are performed in this article. Our adaptive threshold is similar to the method in ATSS. However, the difference is that ATSS takes the sum of the mean and standard deviation of IoU as the adaptive threshold. Meanwhile, the proposed method only uses standard deviation because the mean score is very small, leading to a small adaptive threshold and an excessive number of positive samples, thus affecting the training results.

We conducted extensive experiments on the SSDD dataset. The experiment was divided into two parts: 1) The first part did not use standard deviation; 2) the second part used standard deviation. The experimental results are shown in Table III. In the first part, the threshold was set from 0.6 to 0.8 due to the large mean value of $L_{shape}$. In the second part, we lowered $\alpha$ after using the variance because the variance was in the range of 0~0.4.

By comparing the effects of the two parts, we found that $AP_{50}$ was higher when variance was used as the adaptive threshold. In particular, when $\alpha$ was equal to 0.5, our method had improved
on all metrics compared with the original Faster RCNN algorithm.

We performed a detailed analysis of the experimental results. Because a small target has a small standard deviation, the proposed adaptive threshold can relatively reduce the threshold of the small target and can increase the number of positive samples with the small target. Additionally, with a large standard deviation for a large target, increasing the threshold of the large target can help the large target select positive samples of higher quality. Therefore, the detection performance can be improved through an adaptive threshold. According to the above analysis, our adaptive threshold method can play an effective role in sample selection. In the following work, $\theta$ is set to 0.55 and variance is used. Using this configuration, we achieved 96.3% AP on the SSDD dataset.

### E. MaxIoUAssign versus Our Method

We selected images from three scenarios, as shown in Fig. 11, where Figs. 11(1), (2), and (3) represent small, large, and dense targets, respectively. We observe the differences between the proposed method and MaxIoUAssign of RoIs on these images in three forms, as follows.

![Fig. 11(b) represents the scatter diagram of QS (red sphere) and IoU (blue sphere) at different positions of RoI in Faster RCNN. X and Y represent the coordinates of RoIs on the image, and Z represents the QS or IoU value. As the figure shows, the red sphere’s maximum value is larger than that of the blue sphere, while the minimum value is almost the same. This phenomenon shows that, on the one hand, high-quality RoIs show higher QS scores than IoU. On the other hand, low-quality RoIs performed almost identical on QS and IoU. It will lead to some high-quality RoIs standing out when QS is used because the gap between them and low-quality RoIs becomes larger, thus facilitating the screening of high-quality RoIs. Then, our method creates a clear dividing line between RoI QSs, which makes it easier for the model to select high-quality RoIs.

The phenomenon mentioned above may be difficult to see in scatter plots, so we smoothed the scatter diagram, and the result is shown in Fig. 11(c). In the figure, the colored surfaces are the IoU distribution surfaces of all RoIs. In contrast, the gray surfaces are the QS distribution surfaces of RoI. It is evident from the figure that the QS surface is significantly higher than the IoU surface at the center point. In addition, the closer to the central point, the more significant the gap between QS and IoU. However, the farther away from the center, the smaller the gap between QS and IoU, and finally almost overlapped.

It should be noted that the value range of QS and IoU is $[0–1]$. Although the difference between QS and IoU seems not evident in the Fig. 11(c), taking Fig. 11(1)(c) as an example, the difference between QS and IoU at the central point is between 0.15–0.2. This gap may seem small visually, but it is enough for the neural network to perfectly select high-quality RoI, thus making the model training more effective.

Finally, look at the heat map. The darker the color, the larger the QS or IoU. Take Fig. 11(1)(d) and (1)(e) as an example, Fig. 11(1)(d) is the proposed method. The color at the center point is red, while the position away from the center point is white. Fig. 11(1)(e) is the MaxIoUAssign method. The color at the center point is red, and the position away from the center point is light red. The contrast of color shows that our method will make the sample quality appear more transparent.

### F. Training and Inference Times

In order to prove that the proposed method hardly reduces the training speed and inference time while improving the detection performance, two recognized indicators, FLOP and Params, are adopted to evaluate the computational performance and complexity of the model. As for the training time, we counted the training time of our method and the baseline method Faster RCNN on SSDD and HRSID. For the inference time, FPS was
used as the evaluation criterion. The results are shown in the Table IV.

As seen from the table, Flops and Params corresponding to the two methods are the same. Meanwhile, on the SSDD and HRSID datasets, the training time of our method is approximately 30 s longer than that of the baseline method, which is almost negligible. Experimental results show that the proposed method has little effect on the training time and does not increase the model’s computational load and training parameters.

In addition, there is almost no difference in FPS between the two methods, which indicates that the inference time of the two methods is almost the same, proving that the proposed method hardly affects the inference time.

G. Experiment on SSDD

To prove the advancement of our method, we conducted extensive experiments on SSDD dataset. In this section, we still
take anchor-based target detection algorithm Faster RCNN as the baseline, and compare it with other two methods based on CNN, including Dynamic RCNN and neural architecture search (NAS) FCOS. Table V shows the test results on SSD. As can be seen from the table, compared with the original Faster RCNN algorithm, our method improves recall by 1.5%, $AP_{50}$ by 1.7%, $AP_s$, $AP_m$, and $AP_l$ by 1.1%, 1.7%, and 17.3%, respectively. In addition, by comparing our method with the current mainstream algorithms in target detection, we find that our method is almost superior to other algorithms. Specifically, compared with dynamic RCNN, our method improves recall by 2.1% for $AP_{50}$, 0.8% for $AP_s$, and 0.4% for $AP_l$. Compared with Cascade RCNN, our method improved recall by 2.9%, $AP_{50}$ by 5.2%, $AP_s$, $AP_m$ and $AP_l$ by 1.7%, 3.6% and 8.2%, respectively. Compared with NAS FCOS, our method improved recall by 9.0%, $AP_{50}$ and $AP_{75}$ by 11.5% and 1.7%, $AP_s$, $AP_m$ and $AP_l$ by 3.0%, 3.6% and 8.2%, respectively.

Fig. 12 also shows that the proposed method is superior to the other algorithms under different recalls. When a recall is greater than 0.8, the decline in the proposed method is more stable than that of dynamic RCNN and NAS FCOS. The first row in Fig. 13 shows that the proposed method can detect the target accurately and does not have the problem of repeated detection compared with Faster RCNN and NAS FCOS. Meanwhile, the second row in the figure shows that the proposed method avoids repeated detection and error detection compared with Faster RCNN and Dynamic RCNN. The third row shows that although none of the three methods can completely detect the ship, the error rate of the proposed method is lower than that of other algorithms. The above analysis demonstrates that all algorithms have problems of repeated, error, and missed detections, but the detection accuracy of the proposed method is relatively high. Therefore, the proposed method can achieve a better detection effect than the dynamic RCNN and NAS FCOS algorithms.

H. Contrast Experiment With Other Sample Selection Methods

To comprehensively evaluate the performance of the method in sample selections, we compared it with other sample selection methods on the HRSID dataset, including ATSS and AutoAssign algorithms. The results are shown in Table VI. Because AutoAssign is based on FCOS, we did not migrate it to Faster RCNN. ATSS and the proposed method were applied to the Faster RCNN algorithm. These methods focus on sample selection, but their operation and ideas are different, so it has the significance of comparison.

As presented in Table VI, compared with the Faster RCNN algorithm, our $AP_{50}$ and $AP_{75}$ improved by 0.9 and 0.4, respectively. Although the effect on HRSID was not as obvious as that on SSD, it still improved the original algorithm. Additionally, compared with the mainstream ATSS method, the proposed method improved $AP_{50}$, $AP_{75}$, $AP_s$, and $AP_m$ by 8.1%, 0.9%, 6.9%, and 3.4%, respectively, in the HRSID dataset. However, compared with the current mainstream AutoAssign, although ours method was 1.3% lower in $AP_{50}$, $AP_{75}$ was 5.8% higher, $AP_s$ was 5.2% higher, and $AP_m$ was 0.3% higher. Moreover, the algorithm complexity of AutoAssign is much higher than that of the proposed method. To sum up, our proposed method is advanced in sample selection.

I. Experiment on HRSID

To verify the robustness of the proposed algorithm, we conducted extensive experiments on the HRSID dataset. In this section, we still take anchor-based Faster RCNN as the baseline and compare it with other two methods based on CNN: 1) Dynamic RCNN; 2) NAS FCOS. Table VII presents the test results on the HRSID dataset. As presented in Table VII, compared with the original Faster RCNN algorithm, our method improved recall, $AP_{50}$, $AP_s$, $AP_m$, and $AP_l$ by 1.5%, 1.7%, 1.1%, 1.7%, and 17.3%, respectively. Additionally, our method is superior to current mainstream algorithms in target detections. Specifically, compared with Dynamic RCNN, our method improved recall by 2.1% for $AP_{50}$, 0.8% for $AP_s$, and 0.4% for $AP_l$. Compared with NAS FCOS, our method improved recall, $AP_{50}$, $AP_s$, $AP_m$, and $AP_l$ by 9.0%, 11.5%, 1.7%, 3.0%, 3.6%, and 8.2%, respectively.

To intuitively observe the effect of the proposed method, we marked the detection results in the image, as shown in Fig. 14. HRSID is larger than the SSD dataset, and there are more dense small targets, so the detection is more difficult, and the detection effect is worse. However, the proposed method is still superior to the other three detection algorithms. Fig. 14 shows that other algorithms often have an error and repeated detections for dense small targets, but the proposed method performs better than them. Fig. 15 shows that the results of the proposed method are better than those of other algorithms at different recall rates, and our curves can almost cover other curves. Therefore, the above analysis demonstrates that the proposed method can also show relatively good effects on the HRSID dataset.

IV. CONCLUSION

In this article, we proposed a new sample selection algorithm for SAR ship detection. To select high-quality proposal boxes whose shape is similar to ground truth, we retained IoU and introduced shape similarity as the evaluation criterion of sample quality. Center distance was used as a weight to balance IoU and shape similarity, which was conducive to obtaining proposal...
boxes of higher quality. Furthermore, to avoid the fixed threshold, the standard deviation of QS was taken as the variable to regulate the threshold, which promoted the balance of samples. The experimental results showed that the proposed AC4S can effectively improve the performance of target detection and is better than other algorithms.

REFERENCES

[1] L. Du, H. Dai, Y. Wang, W. Xie, and Z. Wang, “Target discrimination based on weakly supervised learning for high-resolution SAR images in complex scenes,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 1, pp. 461–472, Jan. 2020.

[2] M. Shahzad, M. Maurer, F. Fraundorfer, Y. Wang, and X. Z. Zhu, “Buildings detection in VHR SAR images using fully convolution neural networks,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 2, pp. 1100–1116, Feb. 2019.

[3] L. Huang et al., “OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 1, pp. 195–208, Jan. 2018.

[4] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, “Complex-valued convolutional neural network and its application in polarimetric SAR image classification,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 12, pp. 7177–7188, Dec. 2017.

[5] G. Yang, H.-C. Li, W. Yang, K. Fu, Y.-J. Sun, and W. J. Emery, “Unsupervised change detection of SAR images based on variational multivariate gaussian mixture model and shannon entropy,” IEEE Geosci. Remote Sens. Lett., vol. 16, no. 5, pp. 826–830, May 2019.

[6] C. H. Gierull, “Demystifying the capability of sublook correlation techniques for vessel detection in SAR imagery,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 4, pp. 2031–2042, Apr. 2019.

[7] P. Iervolino and R. Guida, “A novel ship detector based on the generalized likelihood ratio test for SAR imagery,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 8, pp. 3616–3630, Aug. 2017.

[8] H. Lang, Y. Xi, and X. Zhang, “Ship detection in high-resolution SAR images by clustering spatially enhanced pixel descriptor,” IEEE Trans. Geosci. Remote Sens. vol. 57, no. 8, pp. 5407–5423, Aug. 2019.

[9] X. Leng, K. Li, X. Xing, S. Zhou, and H. Zou, “Area ratio invariant feature group for ship detection in SAR imagery,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 7, pp. 2376–2388, Jul. 2018.

[10] S. Wang, Wang, S. Yang, and L. Jiao, “New hierarchical saliency filtering for fast ship detection in high-resolution SAR images,” IEEE Trans. Geosci. Remote Sens., vol. 55, no. 1, pp. 351–362, Jan. 2017.

[11] T. Li, Z. Liu, R. Xie, and L. Ran, “An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 1, pp. 184–194, Jan. 2018.

[12] W. Ao, F. Xu, Y. Li, and H. Wang, “Detection and discrimination of ship targets in complex background from spaceborne ALOS-2 SAR images,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 2, pp. 536–550, Feb. 2018.

[13] S. RenK. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time target detection with region proposal networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, 1137–1149, Jun. 2017.

[14] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 318–327, Feb. 2020.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 779–788.

[16] Z. Tian, C. Shen, H. Chen, and T. He, “FCOS: Fully convolutional one-stage object detection,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 9626–9635.

[17] H. Law and J. Deng, “CornerNet: Detecting objects as paired key-points,” Int. J. Comput. Vis., vol. 128, no. 3, pp. 642–656, 2020.

[18] X. Zhou, D. Wang, and P. Krahenbuhl, “Objects as points,” Apr. 2019, arXiv:1904.07850v2. [Online]. Available: https://arxiv.org/abs/1904.07850.

[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed, “SSD: Single shot multibox detector,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 21–37.

[20] Z. Cai and N. Vasconcelos, “Cascade R-CNN: Delving into high quality object detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 6154–6162.

[21] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, “Libra R-CNN: Towards balanced learning for object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 821–830.

[22] J. Dai, Y. Li, K. He, and J. Sun, “R-FCN: Object detection via region-based fully convolutional networks,” in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 379–387.

[23] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, “Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020, pp. 9756–9765.

[24] B. Zhu et al., “AutoAssign: Differentiable label assignment for dense object detection,” 2020, arXiv:2007.03496. [Online]. Available: https://arxiv.org/abs/2007.03496.

[25] X. Zhang, F. Wan, C. Liu, R. Ji, and Q. Ye, “FreeAnchor: Learning to match anchors for visual object detection,” in Proc. Adv. Neural Inf. Process. Syst., 2019, pp. 147–155.

[26] K. Kim and H. S. Lee, “Probabilistic anchor assignment with iou prediction for object detection,” 2020, arXiv:2007.08103. [Online]. Available: https://arxiv.org/abs/2007.08103.

[27] H. Zhang et al., “Dynamic R-CNN: Towards high quality object detection via dynamic training,” in Proc. Eur. Conf. Comput. Vis., Cham, 2020, pp. 260–275.

[28] N. Wang et al., “NAS-FCOS: Efficient search for object detection architectures,” Int. J. Comput. Vis., vol. 129, no. 12, pp. 3299–3312, 2021.

Yilong Lv received the B.S. degree in remote sensing science and technology from the Xi’an Institute of High Technology, Xi’an, China, in 2016. He is currently working toward the M.S. degree in computer science and technology with Xi’an Hi-Tech Research Institute, Xi’an, China.

His main research interests include remote sensing, synthetic aperture radar (SAR) image processing, target detection and recognition, and feature extraction.

Yu Song received the bachelor’s and master’s degrees in electronics and communication engineering from the College of Communications and Command, Urbana, IL, USA, in 2004 and 2009, respectively. His research interests include object detection, sparse signal representation, infrared target tracking, and cryptography.

Wei Li received the B.S. and M.S. degrees in information security from the Engineering University of PAP, Xi’an, China, in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree in computer science and technology with Xi’an Hi-Tech Research Institute, Xi’an, China.

His research interests include information security and cryptography.