Ellipse Encoding for Arbitrary-Oriented SAR Ship Detection Based on Dynamic Key Points

Fei Gao, Yiyang Huo, Jinping Sun, *Member, IEEE*, Tao Yu, Amir Hussain, and Huiyu Zhou

***Abstract*— In recent years, there has been growing interest in developing oriented bounding box (OBB)-based deep learning approaches to detect arbitrary-oriented ship targets in synthetic aperture radar (SAR) images. However, most existing OBB-based detection methods suffer from boundary discontinuity problems for bounding box angle prediction and key point regression challenges. In this article, we present a novel OBB-based detection algorithm that utilizes ellipse encoding to effectively exploit the geometric and scattering properties of ship targets. Specifically, the ship contour is fit by an OBB inscribed ellipse that is encoded as a set of distances between dynamic key points on the bow and target center. By combining the bow angle interval and the decoding process, the negative impact of the boundary discontinuity problem is avoided. In addition, we propose an elliptical Gaussian distribution heatmap and a pooling strategy termed double-peak max-pooling (DPM) to deal with the challenge of separating densely distributed ships in inshore scenes. The former can enhance the heatmap's ship-side score gap between neighboring ship targets, while the latter can solve the problem of target center responses being suppressed after max-pooling. Simulation experiments conducted on the benchmark Rotating SAR Ship Detection Dataset (RSSDD) and Rotated Ship Detection Dataset in SAR Images (RSDD-SAR) demonstrate the superior performance of our method for ship target detection compared to several state-of-the-art OBB-based algorithms. Ablation experiments show that elliptical Gaussian distribution heatmap and DPM can further improve the inshore detection performance.**

***Index Terms*— Arbitrary-oriented, boundary discontinuity, ellipse encoding, ship detection, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) is one of the most effective space-based Earth observation means, which is widely used as an effective remote sensing data source for ground and sea surface targets. It is ideally suited for a wide range of applications on account of its all-day, all-weather imaging, large observation area, high imaging resolution, and other attractive characteristics. In terms of visual and statistical properties, high-resolution SAR images are more robust than natural optical images. Ship detection using SAR images is one of the important research areas in SAR image interpretation, which is widely employed in different fields, such as marine target monitoring, maritime rescue, port security, and monitoring [1], [2].

Traditional SAR ship target detection algorithms frequently rely on established statistical models of sea clutters that are constrained by threshold settings and human feature extraction, resulting in poor portability [3], [4]. On the other hand, these algorithms include complex stages, such as land–sea segmentation, image preprocessing, and target prescreening [5], making it challenging to match the detection accuracy and speed requirements of practical applications.

In recent years, target detection algorithms based on the deep convolutional neural network (DCNN) have achieved considerable success due to the development of deep learning theories and the improvement of processing capacity [6]. These algorithms are mainly based on the horizontal bounding box (HBB). For oriented ship targets with a large aspect ratio, which are varied in size and densely distributed in the inshore scenes, HBBs may contain a lot of background clutter and overlapping ships, making the subsequent interpretation process difficult. To improve the detection performance of direction-sensitive targets, such as ships, researchers have concentrated on more effective oriented bounding boxes (OBBs). In the case of ship targets, the OBB can not only represents the ship target's position in the complicated background more accurately but also collect its heading information. This makes sense for more advanced tasks, such as track prediction and situational estimation.

OBB-based detection algorithms have been thoroughly investigated in recent years; they can be divided into two types: anchor-based OBB detection algorithms and anchor-free OBB-based algorithms. Anchor-based algorithms typically have two stages. After first broadly dispersing the anchor boxes, the detectors regress the offsets between the target box and the anchor box parameters to provide region proposals. The second stage is combining the region of interest (ROI) data in order to enhance the box parameters and classify the different item categories. For instance, Jiang et al. [7] and Xia et al. [8] deliver ROI in the horizontal direction to predict the rotated bounding box, Ma et al. [9] and R(3)Det [10] generate rotated ROI (R-ROI) by presetting multiangle anchors, and Ding

Manuscript received 9 June 2022; revised 4 September 2022 and 26 November 2022; accepted 1 December 2022. Date of publication 7 December 2022; date of current version 20 December 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61771027 and Grant 61071139. The work of Amir Hussain was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant Ref. EP/M026981/1, Grant EP/T021063/1, and Grant EP/T024917/1. The work of Huiyu Zhou was supported in part by the Royal Society-Newton Advanced Fellowship under Grant NA160342 and in part by the European Union's Horizon 2020 Research and Innovation Program under Marie Sklodowska Curie Grant 720325. *(Corresponding author: Jinping Sun.)*

Fei Gao, Yiyang Huo, and Jinping Sun are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: feigao2000@163.com; hyy1604018154@163.com; sunjinping@buaa.edu.cn).

Tao Yu is with the Taiyuan Satellite Launch Center, Xinzhou 36300, China (e-mail: 597878178@qq.com).

Amir Hussain is with the Cyber and Big Data Research Laboratory, Edinburgh Napier University, EH11 4BN Edinburgh, U.K. (e-mail: A.Hussain@napier.ac.uk).

Huiyu Zhou is with the School of Computing and Mathematical Sciences, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: hz143@leicester.ac.uk).

Digital Object Identifier 10.1109/TGRS.2022.3227260

et al. [11] regress the horizontal ROI to the refined R-ROI. However, the settings for the anchor box must be manually specified for these algorithms, and they are challenging to calibrate. In addition, an imbalanced distribution of positive and negative samples might result in a long learning curve and suboptimal performance. To address the drawbacks, anchor-free OBB-based algorithms have been developed and can be further divided into two types: angle prediction- and key point regression-based approaches. State-of-the-art methods based on angle prediction achieve target OBB positioning by regressing the center point position, length, width, and rotation angle [12]. The so-called key point regression-based algorithms represent OBB as a set of key points [13], [14], [15] and localize the target by regressing the key point position.

Although the shortcomings related to anchor boxes are effectively solved, the above approaches frequently suffer from the problem of boundary discontinuity [16], [17], [18], [19], which primarily includes periodicity of angle (PoA) and exchangeability of edges (EoE). The former is owing to the bounded periodicity of the angle parameters, whereas the latter is due to the way the bounding box is defined. The primary cause is typically a high loss at the boundary when the predicted value exceeds the defined range, influencing the convergence consistency of the regression process. Researchers have also extensively studied the boundary discontinuity problem [17], [20], [21], [22]. For example, the angle regression problem has been transformed into an angle classification challenge by Yang and Yan [17] and Yang et al. [20]. The boundary discontinuity problem can be efficiently avoided by combining the long-edge definition approach and the circular smooth label (CSL). However, for angle classification, it is challenging to strike a compromise between branch efficiency and classification accuracy. He et al. [21] propose an OBB detection approach based on polar coordinate vector group encoding, while the key points' selection does not effectively combine the geometric and scattering characteristics of SAR ship targets.

To address the boundary discontinuity problem in OBB detection and the shortcomings of previous methods, in this article, we propose an SAR ship detection approach based on ellipse encoding, which utilizes the geometric features and scattering characteristics of SAR ship targets. Specifically, we initially employ a set of dynamic key points in the bow region to estimate the bounding box. The selection of the dynamic key points is based on the fact that ship targets in SAR images comprise strong scattering points, which aids in determining the ship's direction and position. For this purpose, we fit the ship contour with an OBB inscribed ellipse and encode the ellipse as a series of ordered key point distances from the center point. The location consistency of the key points and the strong scattering points is ensured in this manner. For the implementation of the decoding process, we introduce CSL to classify the bow angle interval. The position of the dynamic key points is determined by decoding the set of the ordered key point distances and the angle interval together, and then, the predicted ellipse and the target OBB are derived. Through the encoding and decoding processes of the OBB inscribed ellipse and the classification of bow angle interval by CSL, the negative effects of PoA and EoE on network convergence and detection performance can be effectively prevented. In addition, for the challenge of distinguishing densely distributed ships in the inshore scenes, we further present a heatmap based on elliptical 2-D Gaussian distributions and double-peak max-pooling (DPM) to predict the target center. By lowering the confidence score of the ship-side region in the heatmap, the former improves the score gap between nearby ships, while the latter can greatly enhance recall in the inshore scenes by selecting a score threshold with nearly no increase in false alarms. We evaluate the effectiveness of our dynamic key point-based detection algorithm, including the proposed elliptical heatmap and DPM, on the benchmark Rotating SAR Ship Detection Dataset (RSSDD) [23] and Rotated Ship Detection Dataset in SAR Images (RSDD-SAR) [24]. Comparative results of our method against the other existing state-of-the-art OBB-based detection methods reveal that our method outperforms the others in terms of detection performance.

The main contributions of this article are summarized as follows.

1) A novel detection approach is proposed for arbitrary-oriented SAR ships based on ellipse encoding and dynamic key points. The approach utilizes geometric features and scattering points, and effectively prevents the negative effects of PoA and EoE on network convergence and detection performance.
2) An elliptical Gaussian distribution heatmap and a pooling strategy termed DPM are designed to deal with the challenge of separating densely distributed ships, which can significantly improve the network responses to target centers and the overall detection performance in the inshore scenes.
3) Experiments are conducted on two benchmark RSSDD and RSDD-SAR datasets. Comparative results with state-of-the-art OBB-based detection methods show that our proposed method outperforms state-of-the-art OBB-based detection methods in terms of detection performance.

The rest of this article is divided into four parts. Section II details the detection method proposed in this article. Section III describes comparative experiments carried out on the RSSDD and RSDD-SAR datasets. Section IV analyses the influence of the CSL window radius and the superiority of sampling on the OBB inscribed ellipse. Finally, Section V presents some concluding remarks.

II. Related Work

Fig. 1 demonstrates a summary of the state-of-the-art detection algorithms based on HBB and OBB. We will go into more detail about these methods in this section.

A. Algorithms Based on HBB

HBB has a clear geometric meaning and is simple to describe parametrically, allowing accurate detection of targets

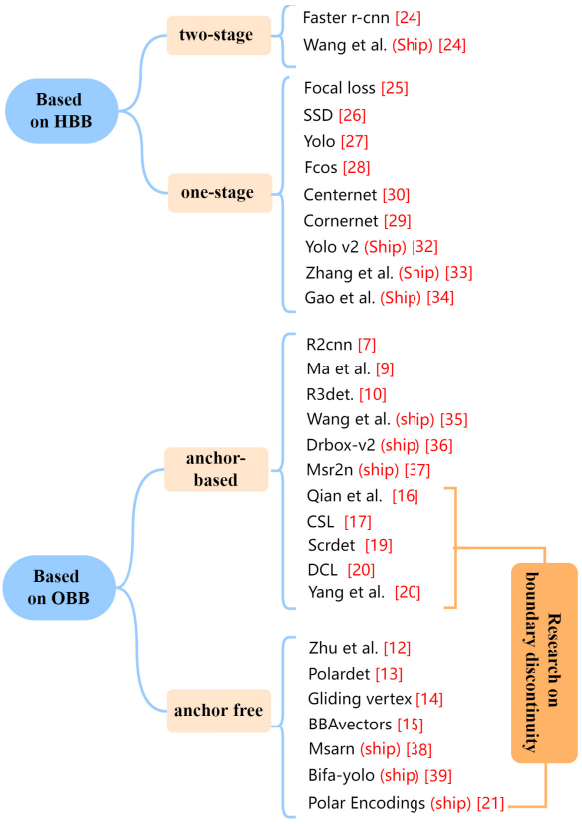


Fig. 1. Summary of detection algorithms based on HBB and OBB.

in natural scenes that are more uniformly oriented and square-like. Conventional DCNN-based horizontal detection methods are generally separated into one- and two-stage detection algorithms. They can automatically and efficiently extract layer-level characteristics, filter candidate regions, and be trained in an end-to-end manner. Typical two-stage algorithms adopt the region proposal network (RPN) to extract candidate regions [25], while one-stage algorithms discard the step of extracting candidate regions and are more concise [26], [27]. Anchor-free methods are the current research hotspot of one-stage detection algorithms that rely on deep features retrieved by the network to directly regress the target position, effectively lowering the model complexity. You Only Look Once (YOLO) version 1 is one of the first proposed anchor-free algorithms [28] with a significant performance gap in detection accuracy compared to anchor-based methods. More recently, a fully convolutional one-stage (FCOS) object detection scheme has been developed that introduces a multiscale matching strategy and a center-ness branch to effectively improve the detection accuracy [29]. Subsequent development of detection algorithms based on key points' detection, such as CenterNet and CornerNet, has provided new ideas for research in this area [30], [31]. HBB-based detection approaches have also been widely used in the field of SAR ship detection. Wang et al. [32] introduce focal loss to alleviate the category imbalance problem, achieving excellent detection accuracy on multiresolution and complicated background datasets. To overcome the problem of reduced detection performance caused by costly computing, Zhao et al. [13] use a simplified YOLO version 2 (YOLOv2) network for the detection of SAR ship targets. Chang et al. [33] design a high-resolution feature pyramid to make full use of both high-level features and low-level features, and then generate a high-resolution feature map that successfully lowers the network's false alarms in the inshore scenes. Zhang et al. [34] use a lightweight feature extraction network to improve YOLO version 3 (YOLOv3) for real-time applications, such as ocean monitoring and ship rescue, significantly increasing model identification speed with little loss of detection accuracy. Gao et al. [35] present a densely connected attention aggregation network that uses feature reuse to enhance the network's generalization performance, achieving satisfactory SAR ship detection performance, whereas the ship targets are frequently arranged randomly and packed closely together. Consequently, adopting the HBB to detect would cause the bounding boxes to be out of alignment with the objects.

B. Anchor-Based Algorithms Based on OBB

Compared with HBB, OBB indicates the target's position with the smallest circumscribed rectangle, thereby reducing redundant information in the detection box and improving the signal-to-noise ratio (SNR). OBB-based detection algorithms have been thoroughly investigated in recent years. There are three types of anchor-based OBB detection algorithms. The first paradigm delivers ROI in the horizontal direction. For example, Jiang et al. [7] predict the minimum target rotated bounding box by extracting multiscale pooling features of the horizontal ROI. Xia et al. [8] achieve OBB-based target detection by anticipating the error between the horizontal ROI and the true bounding box vertices. The second paradigm of algorithms generates R-ROI by presetting multiangle anchors; for example, Ma et al. [9] propose rotated anchors comprising six types of angles, three aspect ratios, and three scales. R(3)Det proposed by Ma et al. [9] combines the advantages of the high recall of horizontal anchors and the adaptability of rotated anchors to dense scenes. R(3)Det introduces rotated anchors to optimize the horizontal ROI, achieving superior detection results on the benchmark DOTA [10] dataset across numerous categories. The third paradigm of anchor-based OBB algorithms generates R-ROI by horizontal anchors. For example, Ding et al. [11] design the R-ROI learner, which regresses the horizontal ROI to the refined R-ROI using ROI pooling and ROI align. This is shown to deliver accurate R-ROIs without increasing the number of anchors. In the case of SAR ship detection, Wang et al. [36] introduce improved SSD to simultaneously predict the position and ship angle. An et al. [37] devise a multilayer anchor preconditioning mechanism ideal for small target detection, as well as a combination of hard negative mining (HNM) and focal loss to address the problem of positive and negative sample imbalance. Pan et al. [38] use a multilevel network to optimize the localization of the bounding box and improve the detection accuracy. The imbalance between positive and negative anchor boxes is a problem for such anchor-based solutions. Slow training and poor detection performance would result from the problem.

C. Anchor-Free Algorithms Based on OBB

Anchor-free OBB-based algorithms can be further divided into two paradigms: angle prediction- and key point regression-based approaches. State-of-the-art methods based on angle prediction achieve target OBB positioning by regressing the center point position, length, width, and rotation angle. For example, R-CenterNet augments the anchor-free detection network CenterNet with an angle branch to achieve OBB detection. To address the angle periodicity problem, Zhu et al. [12] propose encoding the OBB angle with a 2-D periodic vector. The period of the different dimensions of the vector differs by $\pi/2$, and the angle prediction values at the boundary can be distinguished by the coded difference of each dimension [12]. The so-called key point regression-based algorithms represent OBB as a set of key points. For example, PolarDet regresses the target center, center offset, angle between the four polar diameters and the reference y-axis, smaller angle between the width of the minimum bounding rectangle and its height, and aspect ratio between the shorter sides and polar diameters to generate the target OBB [13]. Xu et al. [14] select the four vertices of the bounding box as the key points and localize the target by regressing the distances of the four OBB vertices from the corresponding HBB vertices. Yi et al. [15] use the junction of the two perpendicular hull centerlines and the hull as the key points. In the case of the Cartesian coordinate system, the four key points fall in four different quadrants of the Cartesian coordinates [15]. The OBB is divided into two types: the HBB and the strict OBB, and their detection is performed by length–width regression and key point regression, respectively. Conventional methods based on angle prediction and key point regression have also been extensively investigated for SAR ship detection. Chen et al. [39] propose a multiscale adaptive recalibration network to detect multiscale and oriented ships in complicated scenes. The sensitivity of the network to the target angle is improved by recalibration of the retrieved multiscale features with global information, which increases the accuracy of localization. To efficiently collect multiscale features, Sun et al. [40] offer a bidirectional information interaction module. A random rotation stitching data augmentation method is also being developed to address the issue of imbalanced angular distribution in SAR datasets [40]. However, the above approaches frequently suffer from the problem of boundary discontinuity.

D. Problem of Boundary Discontinuity

PoA and EoE are the two main components of boundary discontinuity and have received substantial research [16], [17], [18], [19]. The angle regression problem has been transformed into an angle classification challenge by Yang and Yan [17] and Yang et al. [20]. They offer a CSL and a densely coded label (DCL) to boost the error tolerance of neighboring angles. The boundary discontinuity problem can be efficiently avoided by combining the long-edge definition approach and CSL. However, CSL directly uses angles as its classification labels, and the number of labels is proportional to the angle range and angle interval. Although a small angle interval helps lower the inaccuracy between the angle prediction and the ground truth, it also increases the thickness of the angle classification layer and the network parameters. DCL uses gray code to densely encode the angle values, yet the encoded labels are uncorrelated to the angle distance, reinstating the PoA problem. He et al. [21] propose an OBB detection approach based on polar coordinate vector group encoding, which uses unique encoding and decoding processes to avoid the harmful impacts of PoA and EoE on network convergence. However, the key points' selection does not effectively combine the geometric and scattering characteristics of SAR ship targets. Yang et al. turn the OBB into a 2-D Gaussian distribution and devise the Gaussian Wasserstein distance loss function, which is rotational intersect over union (IOU) loss suited for gradient backpropagation [22]. The periodicity of the Gaussian distribution, on the other hand, avoids the border discontinuity problem created by the OpenCV representation and the long-edge representation. However, the 2-D Gaussian distribution of the square-like target is a Gaussian normal distribution with a covariance of zero. The loss generated by the Gaussian Wasserstein distance loss function is always zero; hence, the network regression cannot be properly guided.

TABLE I
DEFINITION OF THE MATHEMATICS SYMBOLS

Mathematics			
$\in$	belong to	$\lfloor\rfloor$	round down
$\cap$	intersection	$\sum()$	accumulate
$\cup$	union	$\prod()$	multiply
max()	maximum	min()	minimum
sin()	sine function	cos()	cosine function
arctan()	arctangent function	sqrt()	square root
exp()	exponential function	abs()	absolute value
$\cdot$	multiply	$*$	convolution
Indices			
P	heatmap	O	offset map
D	disance map	A	angle map
M	angle intervals number	N	key points number
α	azimuth	θ	sampling range
Θ	angle set	β	sampling step

III. METHODOLOGY

A summary of the definition of the mathematics symbols in this article is given here in Table I.

Fig. 2 illustrates the overall architecture of the proposed method in this article, which can be divided into three parts: the feature extractor, the feature refinement network, and the center-point-based OBB detector. The feature extractor uses the SAR ship image as input to extract multilevel features, which are subsequently fused in the feature refinement network via long skip connections (LSCs) [41]. An attention module is embedded into the process of upsampling shallow features to generate the final high-resolution feature map with strong semantic information. Next, the high-resolution feature map is processed by four branches in the OBB detectors, from which we can obtain the center heatmap P, the offset map O, the distance map D, and the angle map A. In the training stage, the losses are calculated according to multibranch outputs and the ground truth to jointly guide the network regression. In the inference stage, we generate the fit ellipse

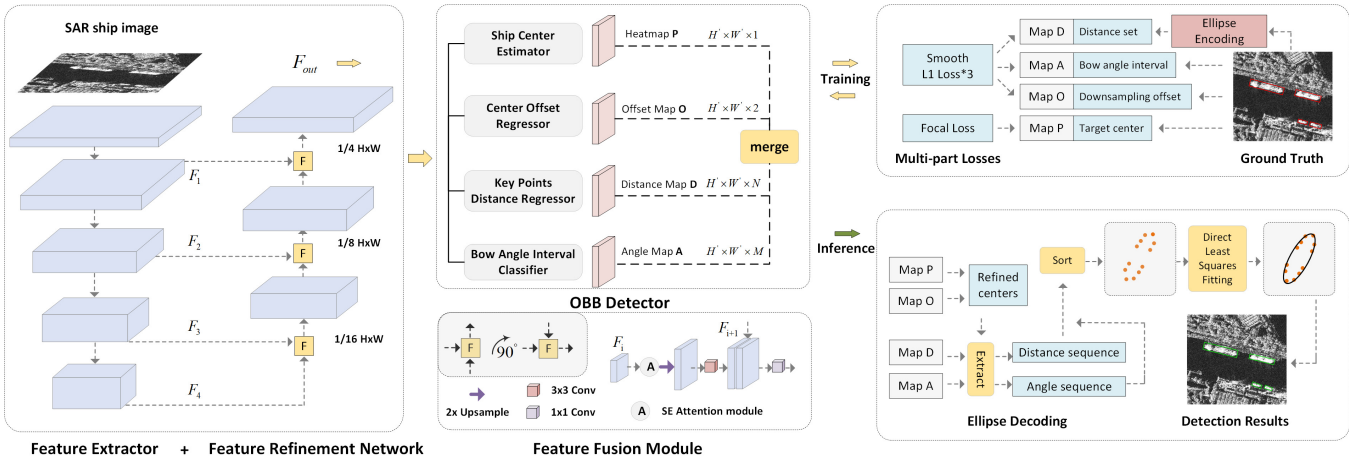


Fig. 2. Overall architecture of our method. The network structure can be divided into three parts: the feature extractor, the feature refinement network, and the OBB detector. The four branches of the OBB detector output the center heatmap P, the offset map O, the distance map D, and the angle map A.

and circumscribed bounding box through the dynamic key point positions, which are calculated by the combination of the multibranch predictions. Finally, the target detection OBBs are obtained using confidence thresholding. In the ellipse encoding process, we divide M copies uniformly in the range of $[0, \pi)$ to obtain M angle intervals and M sampling angles. The CSL then smooths the bow angle interval α to form the ground truth in the bow angle interval prediction, and the sampling interval $[\alpha - \theta, \alpha + \theta)$ is obtained. Next, for the OBB inscribed ellipse, the key points are sequentially sampled at N sampled angles in the interval $[\alpha - \theta, \alpha + \theta)$. The distances of the N key points from the center point are used as the distance encoding values. Due to the central symmetry of the ellipse, these N key points can roughly fit the shape of the entire ellipse. In the ellipse decoding process, the compensated center points are first obtained from the center heatmap and the offset map. Following this, the predictions of key point distance and bow angle interval are retrieved based on the location of center points in the distance and angle maps. The predicted ellipse is, finally, restored in two steps: 1) sorting the key points using the bow angle interval and finding the entire $2N$ boundary point sets and 2) using a direct least-squares fitting approach to obtain the fit ellipse of the point sets. The circumscribed bounding box of the ellipse is the predicted ship OBB.

Next, we present the network structure and loss function in further depth.

A. Network Structure and Loss Function

The overall network structure of our method can be divided into three parts: the feature extractor, the feature refinement network, and the OBB detector. The detailed structure of each part is described in the following.

1) *Feature Extractor:* The network depth is crucial to the model performance. To effectively extract the multiscale features of ship targets in SAR images from shallow to deep, we adopt ResNet-101 [42] as the feature extraction backbone. ResNet-101 superimposes identity mapping on the shallow network. In this way, without introducing additional parameters and computational complexity, it can effectively avoid model performance degradation when the network depth increases. The extractor is composed of multiple convolution stages. With the stage going deeper, the contained semantic information is gradually enhanced, and the receptive field increases. For input SAR image $I \in \mathbb{R}^{H \times W \times 3}$, the extractor generates multiscale features $\{F_1, F_2, F_3, F_4\}$.
2) *Feature Refinement Network:* Deep features with low resolution typically have strong semantic information, which helps with target classification but limits detection. By contrast, shallow features in high resolution contain more low-level features, which damages the representation ability of target recognition. To generate high-resolution feature maps with strong semantic information, we introduce LSC to fuse multilevel features in the feature refinement network. Specifically, we first upsample the features in lower resolution and perform a 3×3 convolution, after which the high-resolution features are channelwise concatenated. In the upsampling process of LSC, the squeeze-and-excitation (SE) attention module is adopted to further strengthen the saliency feature representation ability. Then, a 1×1 convolution is employed to reduce the output channel dimension and generate the output features. The high-resolution fused feature size finally output by the feature refinement network is one-fourth of the input image.
3) *OBB Detector and Loss Function:* The OBB detector is composed of four branches: the center prediction branch, the center offset regression branch, the key point distance regression branch, and the bow angle prediction branch. The center prediction and the center offset regression branches both consist of a 3×3 convolution concatenated with a 1×1 convolution, which separately outputs the center heatmap $\mathrm{P} \in \mathbb{R}^{\tilde{H} \times \tilde{W} \times 1}$ and the offset map $\mathrm{O} \in \mathbb{R}^{\tilde{H} \times \tilde{W} \times 2}$. The key point distance regression and the bow angle prediction branches both contain two cascaded 3×3 convolutions and output the distance map $\mathrm{D} \in \mathbb{R}^{\tilde{H} \times \tilde{W} \times N}$ and the angle map $\mathrm{A} \in \mathbb{R}^{\tilde{H} \times \tilde{W} \times M}$, respectively, where N represents the key points' number and M denotes the number of angle

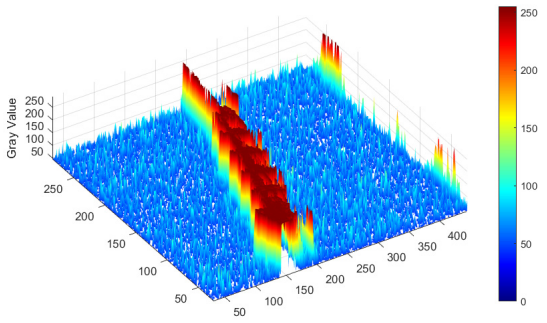


Fig. 3. Typical gray-scale value distribution map of the SAR image.

intervals. For the center prediction branch, we adopt focal loss for supervised training; the losses of the other three branches are calculated by smooth L1 loss. These four parts of losses are combined to form the overall loss function and guide the network regression. The four output prediction maps can be used to achieve the final detection results by an ellipse decoding method, which will be introduced in detail in the following.

B. Ellipse Encoding

1) Geometric Features and Scattering Characteristics of SAR Ship Targets: Different types of ships exhibit distinct geometric features and physical traits in SAR images due to regularities in their hull size and structure, superstructure, and functional equipment distribution. The followings are some examples of typical ship geometrical features.

1) *Contour:* It is the basis for extracting azimuth, dimensions, and other geometric features. The ship's contour can be approximately fit by an ellipse.
2) *Azimuth:* The angle between the direction of the longitudinal axis of the ship and the reference coordinate system.
3) *Aspect Ratio:* Ratio of hull length to width.
4) *Area:* Number of target pixels in the detection area.

The area and aspect ratio are the characteristics that are easiest to understand and compute for SAR ship targets. They are also the characteristics that are most frequently employed. Anchor-based algorithms typically necessitate predetermining the size and aspect ratio of the anchors based on these two geometric features. This leads to issues such as difficult parameter tuning, difficult detection of targets with large-scale variations, high computational consumption associated with anchor, and poor generalization ability. The ship azimuth is one of the important characteristics to describe the target attitude. The estimation process has many similarities with size estimation, and there is a correlation between the estimation accuracy of both. For ship targets with a large aspect ratio, the OBB IOU is sensitive to angle variation. Directly regressing the angles produces angular periodicity problems in either the 180° or 90° representation.

Scattering characteristics are an important physical characteristic of ship targets in SAR images. Fig. 3 shows a 3-D visualization of the ship's target gray-scale values in an SAR image. From Fig. 3, we can see the different scattering intensities of each part of the ship and the location distribution of the strong scattering points. It illustrates that, compared with the weak scattering region, the strong scattering points usually have higher gray-scale values.

The scattering intensity in each component of the ship, the distribution of the strong scattering centers, and the low scattering area all reflect the ship's structural characteristics to a considerable extent. For example, structures such as ship decks, which are nearly flat, are usually low-scattering areas. Structures such as the bow that contain right-angle reflective surfaces often contain strong scattering points, which appear in SAR images as a sequence of brilliant spots or bright lines of a specific shape, forming a unique "trailing cross." These strong scattering points facilitate the determination of the ship's target position and directional angle. Most existing OBB-based algorithms, on the other hand, directly regress the OBB length and width, or key locations on the OBB, without accounting for the position of the strong scattering points. At the same time, the key points' selection did not take use of the SAR ship targets' scattering characteristics.

To overcome these issues, the OBB detection network is designed by combining ship geometric features, such as ship contour, azimuth, and aspect ratio, and the scattering characteristics of the ship. First, we determine the position and orientation of the ship targets by sampling dynamic key point groups in the bow region based on the scattering characteristics of this region, which usually contains strong scattering points in SAR images. Second, to ensure the consistency between the strong scattering points and the key point positions, as well as the realizability of the OBB generation based on the key points' information, we use an ellipse to fit the hull contour of the ship and encode the ellipse as a set of distances from the dynamic key points' group to the ship center.

The specific design of our method in this article is described in the following.

2) Sampling Interval and Sampling Step: To ensure the coverage effectiveness of the dynamic key points on the bow area and the fitting accuracy of the OBB inscribed ellipse, we set the sampling step and sampling interval in terms of the ship azimuth and aspect ratio range. The visualization of sampling intervals under different azimuths is shown in Fig. 4. α represents the ship azimuth in the range of $[0, \pi)$, which is the included angle between the long side of OBB and the X-axis of the reference coordinate system. The sampling interval is defined as the 2θ angle interval centered on the ship azimuth. When sampling in $[0, \pi)$, we need to divide the sampling interval according to the ship azimuth. In particular, when the ship azimuth $\alpha \in [\theta - 2\theta/N, \pi - \theta)$, continuous sampling interval can be obtained in $[\alpha - \theta, \alpha + \theta)$, as shown in Fig. 4(a), where N denotes the number of the key points. In the other cases, the sampling boundary exceeds the range of $[0, \pi)$; the sampling interval will be divided into two parts, at the bow and stern, respectively. As shown in Fig. 4(b), when $\alpha \in [0, \theta - 2\theta/N)$, the lower bound of the sampling interval is less than zero, and discontinuous sampling interval can be expressed as $[0, \alpha + \theta) \bigcup [\alpha - \theta + \pi, \pi)$. Similarly, when the upper bound of the sampling interval equals to or exceeds π, the sampling interval will be divided as

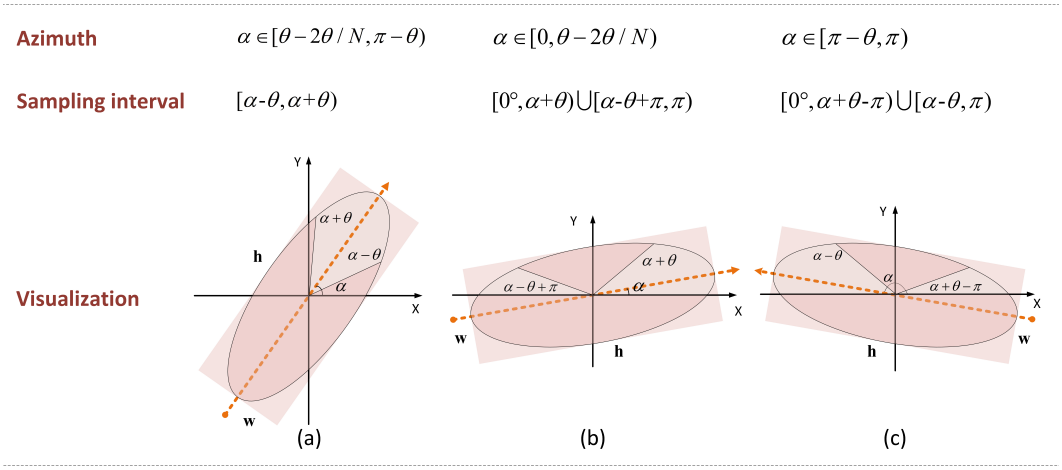


Fig. 4. Sampling interval under different azimuths. (a) Continuous sampling interval. (b) Discontinuous sampling interval. (c) Discontinuous sampling interval.

$[0, \alpha + \theta - \pi) \bigcup [\alpha - \theta, \pi)$. The number of key points is an important hyperparameter in the sampling process and is determined by the sampling interval and sampling step simultaneously. We set these parameters based on the aspect ratio information of the ship targets given in [23] and [24]. The aspect ratio distribution of ship targets in SAR images of the RSSDD and RSDD-SAR datasets is shown in Fig. 5.

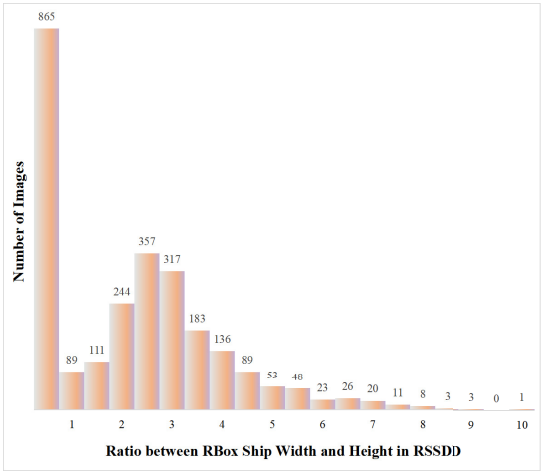
To guarantee that the sampling interval effectively covers strong scattering points and the validity of ellipse encoding, the ship azimuth α is taken as the central axis for sampling in the range of $\pi/6$, that is, $\theta = \pi/6$. Furthermore, at least one key point should be located in the spire area to ensure the realizability of ellipse fitting through the key points' position in the decoding process.

As shown in Fig. 6, in this article, the bow spire area is defined as the area surrounded by the connection between the OBB center point and the vertex of the short side. This area contains strong scattering points with rich feature information. Through the distance between the center point and the key points in this area, the long-axis information of the inscribed ellipse can be efficiently described. When N key points are sampled in the range of 2θ, the sampling step is $2\theta/N$. Given the OBB with height h and width w, (1) should be satisfied to ensure at least one key point falls in the bow spire area

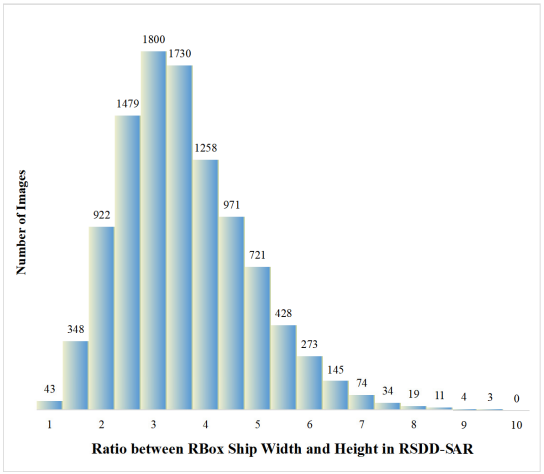
$$\arctan(w/h) = \frac{1}{2} \cdot \frac{2\theta}{N} = \frac{\theta}{N}. \tag{1}$$

For the ship OBBs with the largest aspect ratio in the SAR datasets with $w/h = 10$, the number of sampling points should meet $N \geq 6$. That is, when the sampling step equals $2\theta/N = \pi/18$, at least one key point can be sampled precisely in the spire area, ensuring the consistency of the positions of key points and strong scattering points. Thus, we set the number of the dynamic key points $N = 6$ to encode the inscribed ellipse of OBB for the model's effectiveness and efficiency.

3) Ellipse Encoding Process: To overcome the boundary discontinuity problem and effectively exploit the geometric and scattering characteristics of ship targets, we suggest fitting the ship contour through an OBB inscribed ellipse. This ellipse is encoded as a distance set of the ordered dynamic key points,



(a)



(b)

Fig. 5. Distribution of aspect ratio of ship targets in SAR images. (a) RSSDD. (b) RSDD-SAR.

which are located on the bow. Fig. 7 shows the process of encoding OBB inscribed ellipse in the case of continuous sampling interval, as shown in Fig. 4(a), when the number of sampling points $N = 6$ and the sampling step $\beta = 2\theta/N = \pi/18$. The specific process can be described as follows.

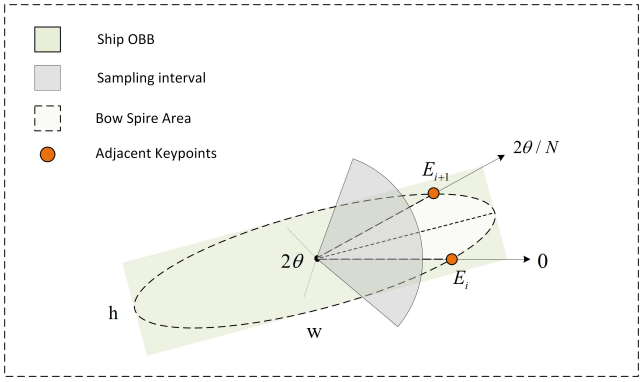


Fig. 6. Sampling interval setting principle.

First, the long and short axis length E_a and E_b of the OBB inscribed ellipse can be calculated by (2) and (3), respectively, where h and w denote the height and width of the target OBB

$$\mathrm{E}_a = \max(h, w)/2 \tag{2}$$

$$\mathrm{E}_b = \min(h, w)/2. \tag{3}$$

Second, in the VOC annotation method, the ship target OBB is represented as the vertex position $P_0 = \{\mathbf{p}_i | \mathbf{p}_i = (x_i, y_i), i = 1, 2, 3, 4\}$, and the position of the target center can be calculated by $\mathbf{c} = (x_c, y_c) = \sum_{i=1}^{4} \mathbf{p}_i/4$. Then, in the range of $[0, \pi)$, the angle set to be sampled $\Theta = \{k \cdot \beta | k = 0, 1, \ldots, \pi \cdot N/2\theta\}$ and the sequence of actual sampling angle $\Theta = \{\Theta_i | \alpha - \theta \leq \Theta_i < \alpha + \theta\}$ can be obtained based on the sampling step and interval. For every sampling angle in sequence Θ, a key point is sampled from the OBB inscribed ellipse. The points' set containing N dynamic key points can be expressed as $P = \{\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_N\}$. For every key point E_i in P, the distance d_i from the center point to E_i can be calculated by

$$\mathrm{d}_i = \sqrt{\frac{(\mathrm{E}_a \cdot \mathrm{E}_b)^2}{(\mathrm{E}_a \cdot \sin(\Theta_i - \alpha))^2 + (\mathrm{E}_b \cdot \cos(\Theta_i - \alpha))^2}} \quad i \in \{1, 2, \ldots, N\} \tag{4}$$

where α denotes the ship azimuth same as above. The obtained distance set is served as the ground truth of the distance regression branch. Because of the central symmetry of the ellipse, the distance set $D = \{\mathrm{d}_1, \mathrm{d}_2, \ldots, \mathrm{d}_N\}$ can actually represent $2N$ key points $P = \{\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_N\} \cup \{\mathrm{E}_{1'}, \mathrm{E}_{2'}, \ldots, \mathrm{E}_{N'}\}$. Therefore, the key points can validly cover the bow and stern parts, so as to completely represent the shape of the ellipse, realizing the effective encoding of the OBB inscribed ellipse.

4) Continuity in the Boundary Cases: The process of ellipse encoding is equivalent to sampling from a periodic function with period π in the interval $[0, \pi)$. The discontinuity issue in the key point regression algorithms can be avoided by this periodicity. An analysis of the specific ideas that the ellipse encoding uses to solve the boundary discontinuity problem is given in the following.

The analysis of boundary continuity of ellipse encoding is illustrated in Fig. 8. The green ellipse denotes the inscribed ellipse of ground-truth OBB with the azimuth of zero. The ordered sequence of dynamic key points obtained by sampling in the range of $[0, \pi)$ is $P_{\mathrm{gt}} = \{\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_6\}$, and $D_{\mathrm{gt}} = \{\mathrm{d}_1, \mathrm{d}_2, \ldots, \mathrm{d}_6\}$ is the corresponding ground truth in the distance branch. The yellow points' set $P_c = \{\mathrm{E}_{1c}, \mathrm{E}_{2c}, \ldots, \mathrm{E}_{6c}\}$ shows the predicted key points' position on the basis of the predicted value of the distance branch and the predicted bow angle interval. The yellow ellipse represents the predicted ellipse fit by P_c with the azimuth of $\pi - \delta\theta \approx \pi$; in the boundary case, the prediction angle occurs a sudden change and generates distance errors $\Delta D_i = \{\mathrm{d}_i - \mathrm{d}_{ic} | i = 1, 2, \ldots, N\}$ between the predicted key points and the ground truth. As shown in Fig. 8, the distance losses of $\{\mathrm{E}_1, \mathrm{E}_2, \mathrm{E}_3, \mathrm{E}_5, \mathrm{E}_6\}$ maintain fine continuity. For the key point E_4, due to the symmetry of the inscribed ellipse, the distance between E_4 and the center point is equivalent to the distance between E_4'' and the center point. It can be seen that the distance between them also remains continuous. Therefore, in the boundary case when the ship azimuth changes from 0 to π, no sharp increase rises in the losses of the distance branch. On the whole, our method effectively avoids the aforementioned boundary discontinuity problems in key point regression detection methods.

The process of predicting the position of key points needs to combine the distance from the key point to the center point and the angle interval of the bow. Nowadays, the angle prediction methods based on regression encounter PoA to a certain extent. One of the main reasons is that the ideal prediction result exceeds the definition range, resulting in a large loss function value. In order to solve the boundary problem in angle prediction, some researchers treat the angle regression task as a classification problem, avoiding the occurrence of the predicted value beyond the defined range through limited classification results. Nevertheless, these methods bring obvious shortcomings. On the one hand, converting the angle regression into a classification task is a process from continuous to discrete, leading to an accuracy reduction in the conversion process. For example, in the case of 1°, it is impossible to predict the result in the step of 0.5, and there will be a large loss for ship targets IOU with a large aspect ratio. On the other hand, these methods are incapable of measuring the angular distance between the prediction and the ground-truth label.

In our methods, the ship angle prediction is transformed into the classification of the angle interval where the bow is located. When the sampling step is β, $[0, \pi)$ is divided into π/β angle intervals. Once the angle interval of the bow is confirmed, the sampling interval and the actual sampling angle sequence can be obtained sequentially. The classification of M categories for the angle interval of the bow can markedly avoid the accuracy reduction caused by the direct classification of the ship angle. In the meanwhile, we adopt CSL to measure the angular distance between the prediction and the ground-truth label in the classification process in order to estimate the angle label distance, and we address the PoA issue by including periodicity.

When the ship azimuth is in the range of $[0, N/2\theta)$, the azimuth label corresponding to the bow is zero. Fig. 9 illustrates the loss results of the angle branch when one-hot label and CSL are adopted, respectively. For the one-hot label, the angle prediction task is regarded as a classification task, including $M = \pi/\beta$ categories. At this time, the cross-entropy

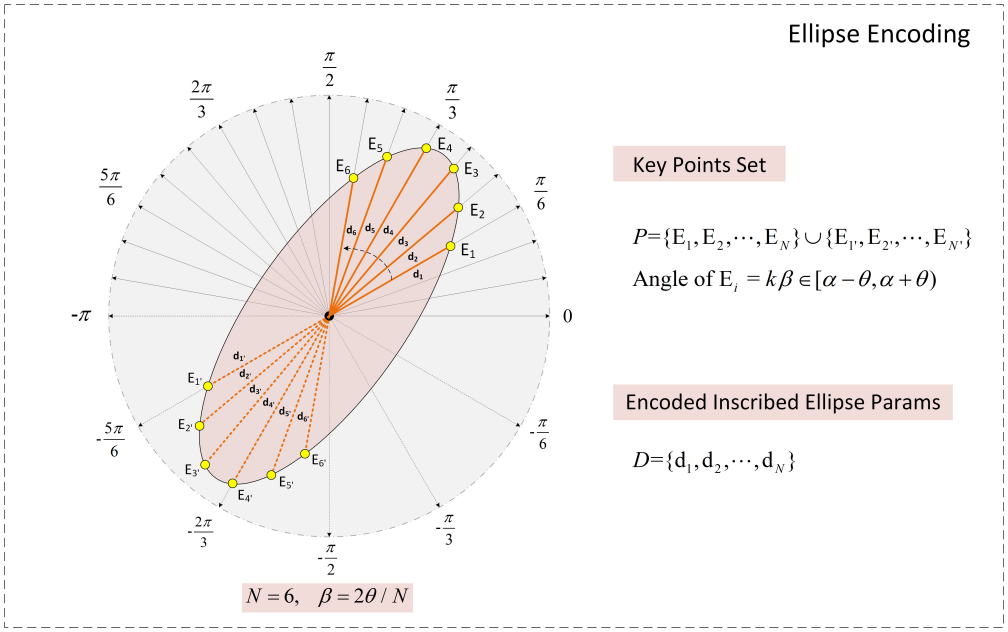


Fig. 7. OBB inscribed ellipse encoding process.

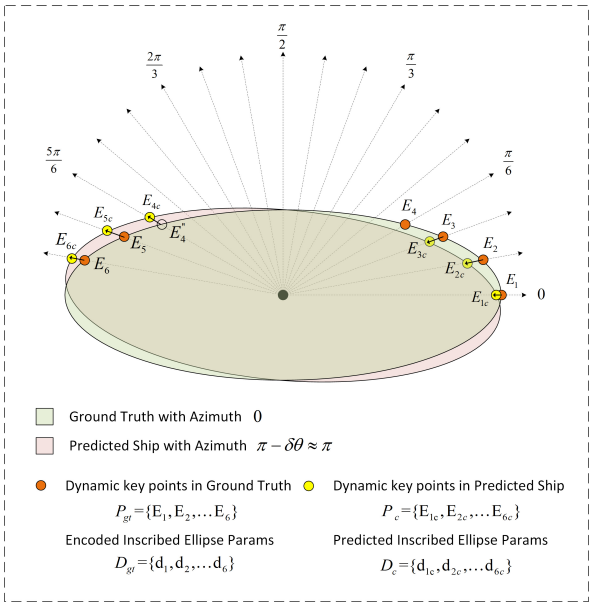


Fig. 8. Continuity in the boundary cases of ellipse encoding.

loss is used to calculate the loss between the prediction and ground truth. As shown in Fig. 9, when the bow angle interval is predicted to be one and two, respectively, the loss calculated by the cross-entropy loss function is approximately equal. In this case, it is impossible to estimate the angle label distance and effectively guide the regression of the angle prediction branch. Therefore, we introduce CSL to assist in judging the angle distance without introducing the problem of PoA. First, we use a Gaussian window function to smooth the one-hot label of the azimuth. The specific calculation process can be described as follows:

$$\begin{aligned} \mathrm{d}_i &= \{\min(|i-\mathrm{L}_A|, M-1-|i-\mathrm{L}_A|)|i\in\{1,2,\ldots,M\}\} \\ \mathrm{L}_{\text{smooth}} &= \left\{e^{-\left(\frac{\mathrm{d}_i}{\sqrt{2}\sigma}\right)^2}|i\in\{1,2,\ldots,M\}\right\} \end{aligned} \tag{5}$$

where L_A is the azimuth label, d_i indicates the distance between the ith position and L_A in one-hot label, σ denotes the Gaussian window radius, and $\mathrm{L}_{\text{smooth}}$ represents the smoothed label. For different prediction values, the loss calculated by smooth L1 loss raises with the increase of angle label distance and remains continuous at the boundary. CSL can effectively judge the label distance and significantly improve the network performance in predicting the angle interval of the bow.

C. Ellipse Decoding

The overall diagram of the ellipse decoding process is shown in Fig. 10. The decoding process extracts the information from each branch of the OBB detector and combines them to obtain the final OBB detection results. Specifically, K target center points' set $C=\{(x_j,y_j)|j=1,2,\ldots,K\}$ with the highest confidence are first extracted from the center heatmap $P\in\mathbb{R}^{\tilde{H}\times\tilde{W}\times 1}$, where K denotes the preset maximum number of targets in a single image. For each center point in C, its position is compensated by the predicted downsampling quantization errors $(\Delta x_i,\Delta y_i)$, which can be obtained from the corresponding location of the offset map $O\in\mathbb{R}^{\tilde{H}\times\tilde{W}\times 2}$. The refined center points' set can be represented as $C_o=\{(x_j+\Delta x_i,y_j+\Delta y_i)|j=1,2,\ldots,K\}$. Similarly, for each refined center point in C_o, the values of N channels corresponding to the position of center points in the distance regression map $D\in\mathbb{R}^{\tilde{H}\times\tilde{W}\times N}$ are extracted as the key point distance prediction vector $D_c=\{\mathrm{d}_{1c},\mathrm{d}_{2c},\ldots,\mathrm{d}_{Nc}\}$. At the same time, the values of M channels in the angle map $A\in\mathbb{R}^{\tilde{H}\times\tilde{W}\times M}$ are obtained as the angle prediction denoted as $A_c=\{a_{1c},a_{2c},\ldots,a_{Mc}\}$.

Then, the maximum value among the M channels in A_c is selected, and its subscript i indicates that the ship bow is in the ith angle interval, that is, $\alpha\in[(i-1)\beta,i\beta)$. Then, the sampling angle sequence $\Theta_c=\{\theta_1,\theta_2,\ldots,\theta_N\}$ corresponding to N key points is obtained. The calculation process

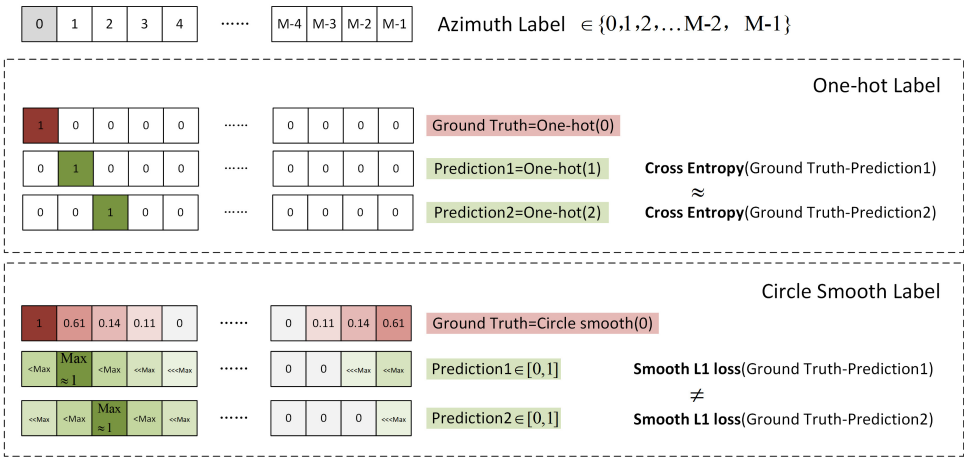


Fig. 9. Angle branch losses of the one-hot label and CSL.

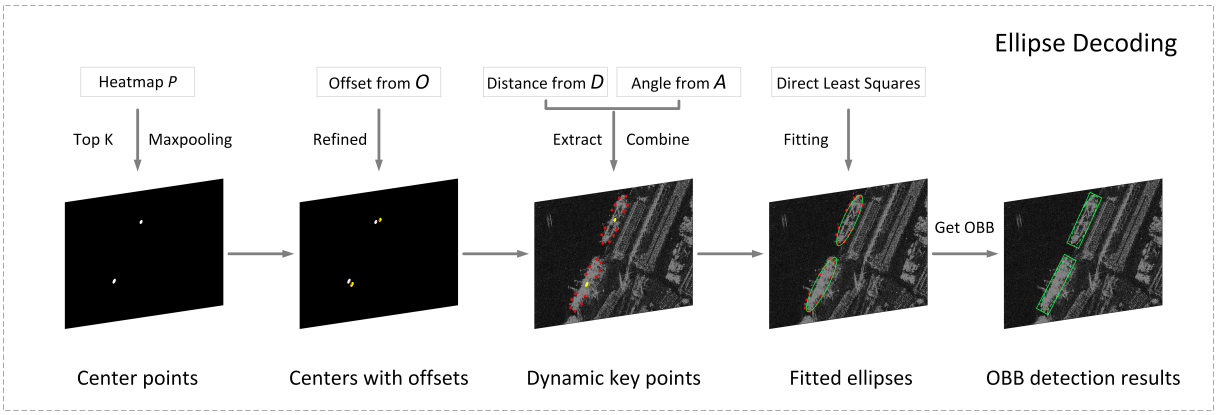


Fig. 10. Angle branch losses of the one-hot label and CSL.

can be expressed as

$$\Theta_c = \begin{cases} \left\{\left(i-\frac{N}{2}\right)\beta, \left(i-\frac{N}{2}+1\right)\beta, (i-1)\beta, i\beta, \right. \\ \quad \left.(i+1)\beta, \left(i-\frac{N}{2}+1\right)\beta\right\}, \quad i \in \left[\frac{N}{2}, M-\frac{N}{2}\right] \\ \left\{\left(i-\frac{N}{2}\right)\beta, \left(i-\frac{N}{2}+1\right)\beta, (i-1)\beta, \right. \\ \quad \left. i\beta, (i+1)\beta, 0\right\}, \qquad i = M-\frac{N}{2}+1 \\ \left\{\left(i-\frac{N}{2}\right)\beta, \left(i-\frac{N}{2}+1\right)\beta, (i-1)\beta, i\beta, 0, \beta\right\} \\ \quad i = M-1 \\ \left\{\left(M-\frac{N}{2}+1\right)\beta, (M-i-1)\beta, (i-1)\beta, \right. \\ \quad \left. i\beta, (i+1)\beta, \left(i-\frac{N}{2}+1\right)\beta\right\}, \quad i = 0 \\ \left\{(M-i-1)\beta, 0, (i-1)\beta, i\beta, \right. \\ \quad \left.(i+1)\beta, \left(i-\frac{N}{2}+1\right)\beta\right\}, \qquad i = \frac{N}{2}-1. \end{cases} \tag{6}$$

Next, the key points' set S of the OBB can be restored through the combination of distance set $D_c = \{d_{1c}, d_{2c}, \ldots, d_{Nc}\}$ and the corresponding angle sequence $\Theta_c = \{\theta_1, \theta_2, \ldots, \theta_N\}$. The process can be expressed as follows:

$$\begin{aligned} S = &\{(d_{ic}\cdot\cos\theta_i, d_{ic}\cdot\sin\theta_i), i = 1, 2, \ldots, N\} \\ &\cup\{(d_{ic}\cdot\cos(\theta_i+\pi), d_{ic}\cdot\sin(\theta_i+\pi)), i = 1, 2, \ldots, N\}. \end{aligned} \tag{7}$$

The next step is to fit the points' set S to the ellipse by direct least-squares fitting. An ellipse can be denoted as $x^2+Axy+By^2+Cx+Dy+E=0$. According to the general formula of the ellipse equation and the principle of direct least-squares fitting, the determination process of parameters $\{A, B, C, D, E\}$ is equivalent to find the minimum value of objective function $F(A, B, C, D, E)$ under the points' set S, where F can be calculated by

$$\begin{aligned} &F(A, B, C, D, E) \\ &\quad = \sum_{i=1}^{2N}\left(x_i^2 + Ax_iy_i + By_i^2 + Cx_i + Dy_i + E\right)^2. \end{aligned} \tag{8}$$

For each parameter in $\{A, B, C, D, E\}$, the value can be derived by making the partial derivative of $F(A, B, C, D, E)$ to this parameter equal to zero. After the ellipse equation is determined, (9) is used to obtained the ellipse center point (x_0, y_0), the length of the major axis a, the length of the minor

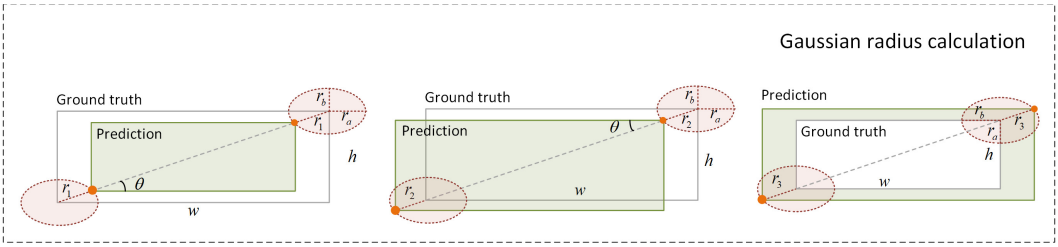


Fig. 11. Calculation of the radius of elliptical 2-D Gaussian distribution.

axis b, and the ellipse direction angle θ

$$\begin{aligned} x_0 &= \frac{2BC - AD}{A^2 - 4B} \\ y_0 &= \frac{2D - AD}{A^2 - 4B} \\ a &= \sqrt{\frac{2(ACD - BC^2 - D^2 + 4BE - A^2E)}{(A^2 - 4B)(B - \sqrt{A^2 + (1-B)^2} + 1)}} \\ b &= \sqrt{\frac{2(ACD - BC^2 - D^2 + 4BE - A^2E)}{(A^2 - 4B)(B + \sqrt{A^2 + (1-B)^2} + 1)}} \\ \theta &= \arctan\left(\text{sqrt}\left(\frac{a^2 - b^2B}{a^2B - b^2}\right)\right). \end{aligned} \tag{9}$$

For each generated ellipse, its circumscribed rectangle is taken as the predicted OBB of the ship target. The length and width of OBB are twice the major and minor axes of the ellipse, and the direction angle is the same as that of the ellipse. Through the above ellipse decoding process, we compute circumscribed rectangles of all generated ellipses as the final detection results.

D. Elliptical Heatmap and DPM

Ships are more dispersed and challenging to detect in complex inshore senses compared to ship targets in the offshore senses. Anchor-free approaches offered by CenterNet require anticipating the position of the desired center point, in contrast to anchor-based algorithms. Specifically, for each ship center in the SAR image, a 2-D Gaussian distribution is generated at the corresponding position of the target center point to generate the ground truth. The closer to the target point, the greater the activation value is. The center point branch calculates the loss through the prediction and ground truth to guide the network training. Instead of directly regressing the key point coordinates, heatmap regression can include the relationship of the target's parts in the heatmap to explicitly suppress nonkey points and provide directional advice for network training. For targets with relatively small aspect ratio in the natural scenes, the standard 2-D Gaussian distribution is currently commonly utilized to generate ground truth for the center points, whereas the standard 2-D Gaussian distribution cannot successfully incorporate the ship OBB shape information for targets with large aspect ratios, such as ships. For instance, while dealing with the densely distributed ships in the inshore scenes, since the ship-side prediction score is not sufficiently differentiated compared to the center points, the center point positions may not be accurately obtained during the screening process by max-pooling.

To combine the ship OBB aspect ratio information with the ground truth, we, therefore, use an elliptical 2-D Gaussian distribution. The score divergence between the ship-side regional predictions and the center point predictions is expanded by raising the gradient of the ship-side activation values. Yang et al. [10] and Ding et al. [11] calculate the Gaussian radius of a standard 2-D Gaussian distribution by setting a specific IOU threshold for the prediction box and the ground-truth box. In this article, we use their IOU threshold setting to calculate the long and short axes r_a and r_b in the elliptical 2-D Gaussian distribution, which satisfies $r_b/r_a = h/w$. The relationships between the three positions of the prediction box and the ground-truth box for a particular IOU threshold are shown in Fig. 11. The prediction box is represented by the green box, the ground-truth box by the gray box, and the vertex positions of the prediction box are shown by the orange dots. The distance between the vertices of the prediction box and the ground truth for each of these three positional relationships satisfies the following relationships:

$$\begin{aligned} &\frac{\sqrt{w^2+h^2} - 2r_1}{\sqrt{w^2+h^2}} = \sqrt{\text{IOU}} \Rightarrow r_1 = \frac{\sqrt{w^2+h^2}(1-\sqrt{\text{IOU}})}{2} \\ &\frac{\sqrt{w^2+h^2} - r_2}{\sqrt{w^2+h^2}} = \sqrt{\frac{2\text{IOU}}{1+\text{IOU}}} \Rightarrow r_2 = \left(1 - \sqrt{\frac{2\text{IOU}}{1+\text{IOU}}}\right) \\ &\qquad \times \sqrt{w^2+h^2} \\ &\frac{\sqrt{w^2+h^2}}{\sqrt{w^2+h^2} + 2r_3} = \sqrt{\text{IOU}} \Rightarrow r_3 = \frac{\sqrt{w^2+h^2}(1-\sqrt{\text{IOU}})}{2\sqrt{\text{IOU}}}. \end{aligned} \tag{10}$$

We choose the smaller value in $\{r_1, r_2, r_3\}$ as the distance between the two box vertices for the specific IOU threshold, that is, $r = \min\{r_1, r_2, r_3\}$. The elliptical 2-D Gaussian distribution satisfies the relationship as in (11), from which we can calculate the long axis $r_a = \sqrt{2} \cdot r(w/(w^2+h^2)^{1/2})$ and the short axis $r_b = \sqrt{2} \cdot r(h/(w^2+h^2)^{1/2})$

$$\begin{aligned} &\frac{(r \cdot \cos\theta)^2}{r_a^2} + \frac{(r \cdot \sin\theta)^2}{r_b^2} = 1 \\ &\cos\theta = \frac{w}{\sqrt{w^2+h^2}}, \quad \sin\theta = \frac{h}{\sqrt{w^2+h^2}}. \end{aligned} \tag{11}$$

After calculating the ellipse long and short axes, we generate the elliptical 2-D Gaussian heatmap, as shown in Fig. 12. The normal variance of the standard Gaussian distribution is set as $\sigma = r/3$ in [11]. The generated standard Gaussian heatmap is shown in Fig. 12(a), where the red box is the ship target OBB and the yellow bounding box denotes the Gaussian heatmap

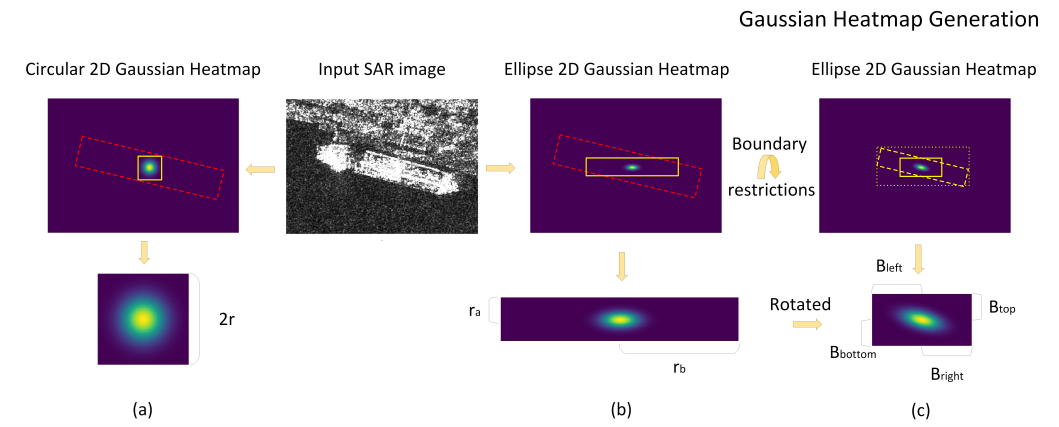


Fig. 12. Elliptical 2-D Gaussian heatmap generation process. (a) Circular 2-D Gaussian heatmap. (b) Ellipse 2-D Gaussian heatmap. (c) Ellipse 2-D Gaussian heatmap with boundary restrictions.

area generated for this target. In this article, we set $\sigma_a = r_a/3$ and $\sigma_b = r_b/3$ in the elliptical Gaussian distribution. In this case, the 2-D Gaussian distribution without the OBB angle information can be expressed as (12), where m is the value of the horizontal and vertical coordinates of the Gaussian map (x, y) position, μ is the OBB center point coordinates, and S is the variance matrix. The heatmap is shown in Fig. 12(b)

$$\begin{aligned} N(m|\mu, S) &= \exp\left(-\frac{1}{2}(m-\mu)^T S^{-1}(m-\mu)\right) \\ m &= (x, y), \quad \mu = (x_c, y_c) \\ S^{\frac{1}{2}} &= \begin{pmatrix} \sigma_a & 0 \\ 0 & \sigma_b \end{pmatrix}. \end{aligned} \tag{12}$$

In our method, the Gaussian distribution is rotated using (13) to make sure that it is consistent with the OBB direction, where θ is the OBB orientation angle in the range of $[-90°, 90°)$, R is the rotation matrix, and the elliptical 2-D Gaussian distribution after rotation can be expressed as $N(m|\mu, \Sigma)$. The box surrounded by yellow dashed lines in Fig. 12(c) shows the rotated Gaussian region distribution

$$\begin{aligned} N(m|\mu, \Sigma) &= \exp\left(-\frac{1}{2}(m-\mu)^T \Sigma^{-1}(m-\mu)\right) \\ \Sigma &= RS^{\frac{1}{2}}R^T \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \sigma_a & 0 \\ 0 & \sigma_b \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \sigma_a\cos^2\theta + \sigma_b\sin^2\theta & (\sigma_a - \sigma_b)\sin\theta\cos\theta \\ (\sigma_a - \sigma_b)\sin\theta\cos\theta & \sigma_a\sin^2\theta + \sigma_b\cos^2\theta \end{pmatrix}. \end{aligned} \tag{13}$$

In the rotation process, the Gaussian distributions at the center points of the densely distributed ships may cover each other, so we design a scheme illustrated in (14) to constrain the rotation area. We first calculate the circumscribed rectangular of the rotated Gaussian heatmap region and use half of the outer rectangular length and width as the border value of the final Gaussian heatmap region

$$\begin{aligned} &\mathrm{B}_{\mathrm{left}} \\ &= \mathrm{abs}\left\lfloor \frac{1}{2}\mathrm{Min}(-\sigma_a\cos\theta + \sigma_b\sin\theta, -\sigma_a\cos\theta - \sigma_b\sin\theta) \right\rfloor \\ &\mathrm{B}_{\mathrm{right}} \\ &= \mathrm{abs}\left\lfloor \frac{1}{2}\mathrm{Max}(\sigma_a\cos\theta + \sigma_b\sin\theta, \sigma_a\cos\theta - \sigma_b\sin\theta) \right\rfloor \\ &\mathrm{B}_{\mathrm{top}} \\ &= \mathrm{abs}\left\lfloor \frac{1}{2}\mathrm{Min}(-\sigma_a\sin\theta - \sigma_b\cos\theta, \sigma_a\sin\theta - \sigma_b\cos\theta) \right\rfloor \\ &\mathrm{B}_{\mathrm{bottom}} \\ &= \mathrm{abs}\left\lfloor \frac{1}{2}\mathrm{Max}(\sigma_a\sin\theta + \sigma_b\cos\theta, -\sigma_a\sin\theta + \sigma_b\cos\theta) \right\rfloor. \end{aligned} \tag{14}$$

In the process of predicting the center points, Ding et al. [11] first apply a max-pooling operation to the center point prediction map. Then, a set of predictions for the ship centers is obtained by finding the points whose positions in the pooled map are equal to those in the original prediction map. Standard max-pooling can give an effective estimate of the ship centers for targets in the natural scenes, effectively replacing the nonmaximum suppression (NMS) postprocessing procedure. However, for densely distributed ship targets, because the proximity of the target centers in the prediction map is not a standard 2-D Gaussian distribution, strong response points between adjacent target centers may occur. This causes the true center point responses to be suppressed in the max-pooling procedure, resulting in the missed detection of the corresponding targets. For this reason, we recommend that (15) van be used to perform two rounds of maximal screening to reduce missed detections, and this method is called DPM

$$\begin{aligned} \mathrm{Peak}_1 &= (\mathrm{Maxpool2d}(F_p) == F_p) \\ F_p' &= F_p - F_p \cdot \mathrm{Peak}_1 \\ \mathrm{Peak}_2 &= \left(\mathrm{Maxpool2d}\left(F_p'\right) == F_p'\right) \\ \mathrm{Peak} &= \mathrm{Peak}_1 + \mathrm{Peak}_2. \end{aligned} \tag{15}$$

In (15), F_p is the center point prediction map, Maxpool2d represents the 3×3 max-pooling, Peak_1 represents the peak values obtained from the first pooling, F_p' denotes the prediction map after removing the current cluster of peak points, and Peak_2 represents the peak values obtained from the second pooling process. The concatenation of Peak_1 and Peak_2 serves as the final set of predicted center points. The visualization

TABLE II
DETAILED INFORMATION OF THE RSSDD AND RSDD-SAR

Dataset	Satellite	Polarization	Imaging	Average size	Average target number	Images number	Imaging area	Scenes
RSSDD	RadarSat-2,Sentinel-1,TerraSAR-X	HH,VV,HV,VH	Interferometric, wide swath, Spotlight, Strip	481×331	2.12	1160	Yantai, China, Visakhapat-nam, India	Inshore, Off-shore
RSDD-SAR	GF-3,TerraSAR-X	HH,VV,HV,VH	Ultra-fine, Fine, Standard, Quad-pol	512×512	1.47	7000	Yantai, China, Athens, Greece	Inshore, Off-shore

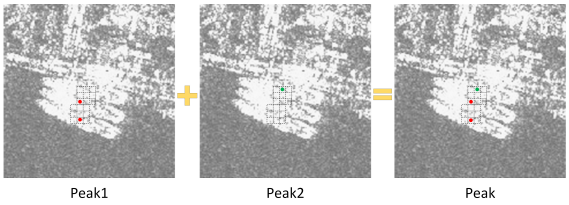


Fig. 13. Visualization process of DPM.

process of DPM is shown in Fig. 13, where the red points represent $Peak_1$ and the green points denote $Peak_2$. By setting the score threshold, DPM can significantly improve the recall metrics in the inshore scenes with little increase in false alarms, and a detailed comparison of the effectiveness of standard max-pooling and DPM for key points' detection will be presented in the subsequent ablation experiments.

IV. EXPERIMENTS

This section presents in-depth results from experiments performed on the RSSDD and RSDD-SAR datasets. First, the RSSDD and RSDD-SAR datasets are introduced. Next, experimental visualization results to compare our method and other OBB-based detection methods are presented to verify the effectiveness of our method. Next, we carried out ablation experiments on the RSSDD for the elliptical 2-D Gaussian distribution heatmap and DPM, respectively. In the discussion section, comparative experimental results of Gaussian window radius for the CSL, and the OBB and OBB inscribed ellipse sampling continuity analysis are given.

A. Dataset Description and Experimental Setup

We conduct experiments on the benchmark RSSDD and RSDD-SAR datasets to verify the effectiveness of our proposed method compared to state-of-the-art approaches. Details of the two benchmark datasets are shown in Table II.

The dataset RSSDD includes 1160 SAR images with multiple resolutions, multiple polarization modes, and multiple sea surface conditions. There are 2540 SAR ship targets in this dataset, with each SAR image containing an average of 2.12 ship targets. The multiresolution coverage ensures adequate detection performance of the training model for multiscale targets. In the experiments conducted in this study, the training set and the test set are divided according to an 8:2 ratio. The RSSDD contains a variety of scenes, including the inshore scenes with more land clutter than the offshore scenes, making it more difficult to detect the ship targets, with the number of images also being smaller. In order to better distinguish the detection performance of our model in different scenes, the test set is divided into two parts: inshore and offshore, comprising 39 and 193 SAR images, respectively.

The dataset RSDD-SAR contains 127 scenes of data, including 84 scenes of HSPA-3 data, 41 scenes of TerraSAR-X data slices, two scenes of uncropped big images, and 7000 slices representing various imaging modes, polarization techniques, and resolutions, along with 10 263 ship targets. Rich scenes, including typical scenes, such as harbor, waterway, offshore low resolution, and offshore high resolution, can be found in the RSDD-SAR dataset. Several slices of rich scenes in the RSDD-SAR dataset are shown in Fig. 14. In summary, the dataset features arbitrary rotation direction, a large aspect ratio, a high percentage of small targets, and rich scenes. In the experiments reported in this article, the training set and the test set are divided according to a 5:2 ratio. A total of 159 inshore scene slices and 1841 offshore scene slices make up the test dataset, which is used to assess how well each model performs in various scenes.

The details of our experimental setup remain the same for both datasets. During the training process, the parameters of the feature extractor ResNet-101 are initialized through a pretrained model. We adopt an adaptive moment estimation (Adam) method [43] to optimize the training process with the weight decay set to 0.0005. We set the initial learning rate as 1.25×10^{-4}. The learning rate is reduced by a factor of 10 at the 100th epoch. We set the number of small batch samples for random gradient descent to 6, for a total of 120 training epochs. Our code for the experiments is implemented using the Pytorch framework [44], and the comparison methods are implemented using OBB detection

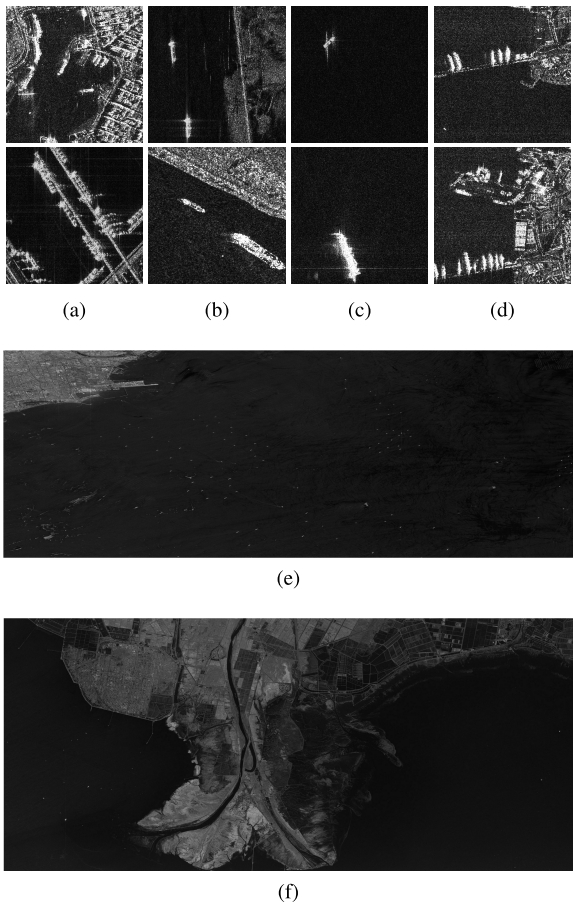


Fig. 14. Rich scenes in the RSDD-SAR datasets. (a) Harbor. (b) Waterway. (c) Offshore. (d) Arranged. (e) Offshore large-scene image. (f) Inshore large-scene image.

and MMrotate based on MMDet [45]. All experiments are conducted on an Ubuntu 18.04 system with 16-GB RAM and a Tesla P100 GPU.

B. Evaluation Metrics

In this article, we use nine widely used evaluation metrics: precision, recall, average precision (AP), F1 score, precision–recall curve (PR curve), detection rate (P_{d}), false alarm rate (P_{fa}), detection time, and FLOPs to comprehensively evaluate the model performance.

Precision represents the percentage of the number of correct targets detected out of the total number of targets detected, and recall denotes the ratio of the number of correct targets detected to the actual total number of targets. These two metrics can be calculated by the following equation:

$$\begin{aligned} \text{precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \\ \text{recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \end{aligned} \tag{16}$$

where TP, FP, and FN denote the number of correct targets detected, the number of incorrect targets detected, and the number of correct targets not detected, respectively. The PR curve, with recall as the horizontal coordinate and precision as the vertical coordinate, measures the combined detection performance of the model under multiple thresholds. The higher the PR curve is, the better the overall performance of the model. The AP quantifies the comprehensive performance of the detector through the area enclosed by the PR curve and the x-axis, which is calculated as shown in the following equation:

$$\text{AP} = \int_0^1 P(R) dR. \tag{17}$$

Ideally, we would like to achieve both high recall and high accuracy, but, in practice, there are often tradeoffs to be made. The F1 score represents the harmonic average evaluation metric of recall and precision, taking into account the reconciled values of recall and precision. With one of them unchanged, upgrade the other one, and the F1 becomes higher. F1 can be calculated as shown in the following equation:

$$\text{F1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{18}$$

For a more intuitive comparison of the detection performance of our method and state-of-the-art methods, we evaluate these methods with two classical metrics used in the field of radar target detection, specifically the detection rate (P_{d}) and the false alarm rate (P_{fa}). P_{d} represents the proportion of the number of real targets that are correctly detected in the entire dataset, which is similar to the evaluation criterion of recall. P_{fa} denotes the proportion of incorrectly classified targets among all detected targets in the entire dataset. The two classical metrics are both influenced by the detection threshold, and the comparisons of detection performances should be under the same detection threshold. The number of false alarms generally declines as the detection threshold rises, but the number of missed detections increases. These two metrics can be calculated by (19) as follows:

$$\begin{aligned} P_{\mathrm{d}} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \\ P_{\mathrm{fa}} &= \frac{\text{FP}}{\text{TP}+\text{FP}}. \end{aligned} \tag{19}$$

The detection efficiency of the model is equally important in practical applications. In this article, the efficiency of the model is evaluated by the detection time and FLOPs. Detection time refers to the average detection time required to detect one image of the test set. FLOPs are used to measure the time complexity of the model, which solves the problem of comparing the processing time of different models on different hardware platforms. The larger the FLOPs, the higher the time complexity of the model.

C. Experiment Results

To verify the effectiveness of our method, comparison results with multiple detection metrics and visualization results in the inshore and offshore scenes between our method and several state-of-the-art OBB-based detection methods are given. In addition, in order to quantitatively evaluate the effectiveness of each module, we conduct ablation experiments on the elliptical 2-D Gaussian distribution heatmap and DPM, respectively. In the discussion section, comparison

TABLE III
DETECTION PERFORMANCE OF DIFFERENT METHODS ON THE RSSDD

Model	Backbone	Scene	Precision%	Recall%	F1%	AP%
BBAVectors	ResNet-101	offshore	93.02	93.02	93.02	89.93
		inshore	80.65	76.34	78.43	77.19
Polar Encodings	ResNet-101	offshore	94.51	94.51	94.51	**90.30**
		inshore	79.85	81.68	80.75	77.06
Gliding Vertex	ResNet-101	offshore	90.52	94.78	92.60	90.22
		inshore	72.59	74.81	73.68	70.00
ROI Transformer	ResNet-101	offshore	89.72	89.28	89.59	88.46
		inshore	63.87	75.57	69.23	68.61
FRCNN-OBB	ResNet-101	offshore	88.35	87.03	87.69	84.58
		inshore	58.39	66.41	62.14	53.00
Double Heads OBB	ResNet-101	offshore	90.27	90.27	90.50	89.14
		inshore	65.28	71.76	68.36	68.09
Rotated IOU Loss	ResNet-101	offshore	81.63	87.53	84.48	85.86
		inshore	45.81	70.99	55.69	61.89
Oriented R-CNN	ResNet-101	offshore	89.92	89.03	89.47	88.69
		inshore	73.72	77.10	75.37	73.30
Ours	ResNet-101	offshore	**95.01**	**95.01**	**95.01**	90.24
		inshore	<u>89.26</u>	<u>82.44</u>	<u>85.71</u>	<u>80.22</u>

experiments of the Gaussian window radius of CSL and the results of OBB and OBB inscribed ellipse sampling continuity analysis are given.

1) Comparison With Other OBB-Based Ship Detection Methods: In this section, we compare our method with several state-of-the-art OBB-based ship detectors. These methods are described as follows.

1) *BBAVectors:* A one-stage anchor-free OBB detection method. OBBs are classified into horizontal and rotational categories, and accurate detection of horizontal OBBs is achieved by regression of length and width.
2) *Polar Encodings:* A one-stage anchor-free OBB detection method. The OBB is encoded as an ordered set of distances of the boundary points on the OBB from the center point. The boundary discontinuity problem is avoided by training and inference directly according to the Polar Encodings.
3) *Gliding Vertex:* A two-stage anchor-based OBB detection method developed on Faster R-CNN. It first detects the HBB of the target and then determines the OBB corner points by regressing the offset of the OBB vertices relative to the HBB vertices.
4) *ROI Transformer:* A three-stage anchor-free OBB detection. It proposes a strategy to convert the horizontal ROI output from RPN into R-ROI, which can be obtained accurately without increasing the number of anchors.
5) *FRCNN-OBB:* A two-stage anchor-based OBB detection method with an OBB angle prediction branch is added in the second stage of Faster R-CNN.
6) *Double Heads OBB [46]:* A two-stage anchor-based OBB detection method. Specifically, a fully connected layer is used for classification, a convolutional layer is designed for bounding box parameters regression, and an angle prediction branch is added.
7) *Rotated IOU Loss [47]:* Rotated IOU Loss first obtains all vertices at the intersection of OBBs, then sorts them, and calculates the area by the shoelace formula to achieve differentiable IOU calculation for oriented boxes.
8) *Oriented R-CNN [48]:* A two-stage anchor-based OBB detection method. Two branches are added in the RPN to regress the midpoint bias loss and, thus, directly obtain the rotational proposals. A fixed-size feature vector is then extracted from each proposal using R-ROI alignment to perform classification and regression.

Comparison experiments are all implemented using the MMDet except for BBAVectors and Polar Encodings. ResNet-101 with 101 convolutional layers is used as the feature extractor. The training and test hyperparameters for each comparison experiment are set as suggested in DarkNet [48] or MMDet, and early stopping is adopted to mitigate the network overfitting.

Tables III and IV show the detection performance metrics of each method in the cases of inshore and offshore scenes of RSSDD and RSDD-SAR, respectively, to demonstrate a quantitative comparison of their detection effectiveness. From the comparative experimental results in Tables III and IV, it can be seen that our method achieves the highest performance in all metrics for the case of inshore scenes and has a substantial lead over other OBB-based methods on both datasets. For the RSSDD, our method outperforms the other methods by at least 5% and 3% for the F1 and AP metrics, reaching 85.71% and 80.22%, respectively. For the case of training and inferencing on the RSDD-SAR, our method has a lead of at least 0.93% and 5.89% compared to the other methods for the F1 and AP metrics, reaching 80.10% and 76.70%, respectively. Among the compared methods, BBAVectors and Polar Encodings have better overall detection performance in the inshore scenes.

TABLE IV
DETECTION PERFORMANCE OF DIFFERENT METHODS ON THE RSDD-SAR

Model	Backbone	Scene	Precision%	Recall%	F1%	AP%
BBAVectors	ResNet-101	offshore	95.71	**95.37**	95.57	90.25
		inshore	81.97	75.50	78.60	68.11
Polar Encodings	ResNet-101	offshore	95.77	95.33	95.54	90.31
		inshore	81.64	76.85	79.17	69.81
Gliding Vertex	ResNet-101	offshore	82.62	86.38	84.44	79.61
		inshore	58.53	53.92	56.11	44.94
ROI Transformer	ResNet-101	offshore	94.22	89.87	91.92	81.26
		inshore	67.29	53.43	59.54	46.65
FRCNN-OBB	ResNet-101	offshore	87.90	88.55	88.23	79.93
		inshore	52.17	55.20	53.60	44.66
Double Heads OBB	ResNet-101	offshore	77.21	87.61	82.06	79.33
		inshore	27.50	56.52	37.06	44.44
Rotated IOU Loss	ResNet-101	offshore	80.41	82.74	81.59	77.25
		inshore	46.23	51.81	48.80	42.53
Oriented R-CNN	ResNet-101	offshore	91.92	90.13	91.02	89.33
		inshore	62.96	62.48	62.72	55.01
Ours	ResNet-101	offshore	**95.96**	95.28	**95.62**	**90.31**
		inshore	<u>82.68</u>	<u>77.68</u>	<u>80.10</u>	<u>76.70</u>

Bold items denote the optimal offshore values in the columns, the underlined items represent the optimal inshore values in the columns.

For the RSSDD, Polar Encodings have an advantage over BBAVectors in several metrics; for example, a lead of 5.34% and 2.32% in recall and F1, respectively, is delivered in the inshore scenes. The two methods perform quite well on the RSDD-SAR dataset, with BBAVectors producing the highest recall of all methods for the offshore scenes (95.37%), which is 0.09% higher than our method. However, under several other evaluation metrics, both methods still have a large gap with our proposed method. This is because the two methods are both based on key point detection and solve the boundary discontinuity problem by preclassification of OBB or special encoding and decoding process. In the inshore scenes of RSSDD, the anchor-based FRCNN-OBB and Rotated IOU Loss significantly lag behind the other methods in terms of precision metric. The accuracy of Double Heads OBB is significantly lower (only 27.50%) compared to that of other methods for inshore scenes of the RSDD-SAR dataset. Also, we found from visualization experiments that this method has a high number of false alarms, and the double heads structure may not enable the network to achieve better detection results. Comparatively, experimental results that show our proposed method is capable of achieving satisfactory detection performance on a variety of datasets, demonstrating the method's efficacy and solid robustness. The enhanced performance can be mainly attributed to two factors. On the one hand, the negative impact of the boundary discontinuity problem on network training is avoided. On the other hand, the adoption of an elliptical Gaussian heatmap and DPM improves the differentiation degree of inshore ships and network prediction performance for ship centers. In the offshore scenes, the AP of our method on the RSSDD is slightly lower than that of Polar Encodings by 0.06%, while the precision, recall, and F1 metrics all exceed those of the other methods, and Gliding Vertex and BBAVectors both achieve similar results with our method on the two datasets. This is due to the less clutter, significant target location, and less distinction in the offshore scenes, which leads to all methods delivering comparable detection results. Overall, due to the full utilization of ship geometry and scattering characteristics and targeted optimization for inshore scenes, our method can effectively avoid the boundary discontinuity problem in OBB detection and achieve better detection results compared with other methods.

Fig. 15(a)–(d) shows the PR curves for each OBB-based method for the cases of inshore scenes and offshore scenes on the RSSDD and RSDD-SAR datasets, respectively. From Fig. 15(a) and (c), we can see that the PR curve of our method is higher and fuller than that of other comparison methods across a wide range. This shows that the detection performance of our method in the inshore scenes is comprehensively better and verifies its effectiveness in detecting inshore ship targets. FRCNN-OBB shows a significant decrease in accuracy with a recall of 0.65, and the curve is clearly located on the inner side, indicating poor detection performance. For the case of offshore scenes, as shown in Figs. 15(b) and (d), the PR curves of our method and Polar Encodings are located more outwardly compared to other methods. The enhanced detection performance is demonstrated by the accuracy, which is maintained in the case of a high recall rate.

To comparatively evaluate all methods with classical metrics used in the field of radar target detection, Tables V and VI show the number of false alarms, the number of missed detections, the false alarm rate P_{fa}, and the detection rate P_{d} of

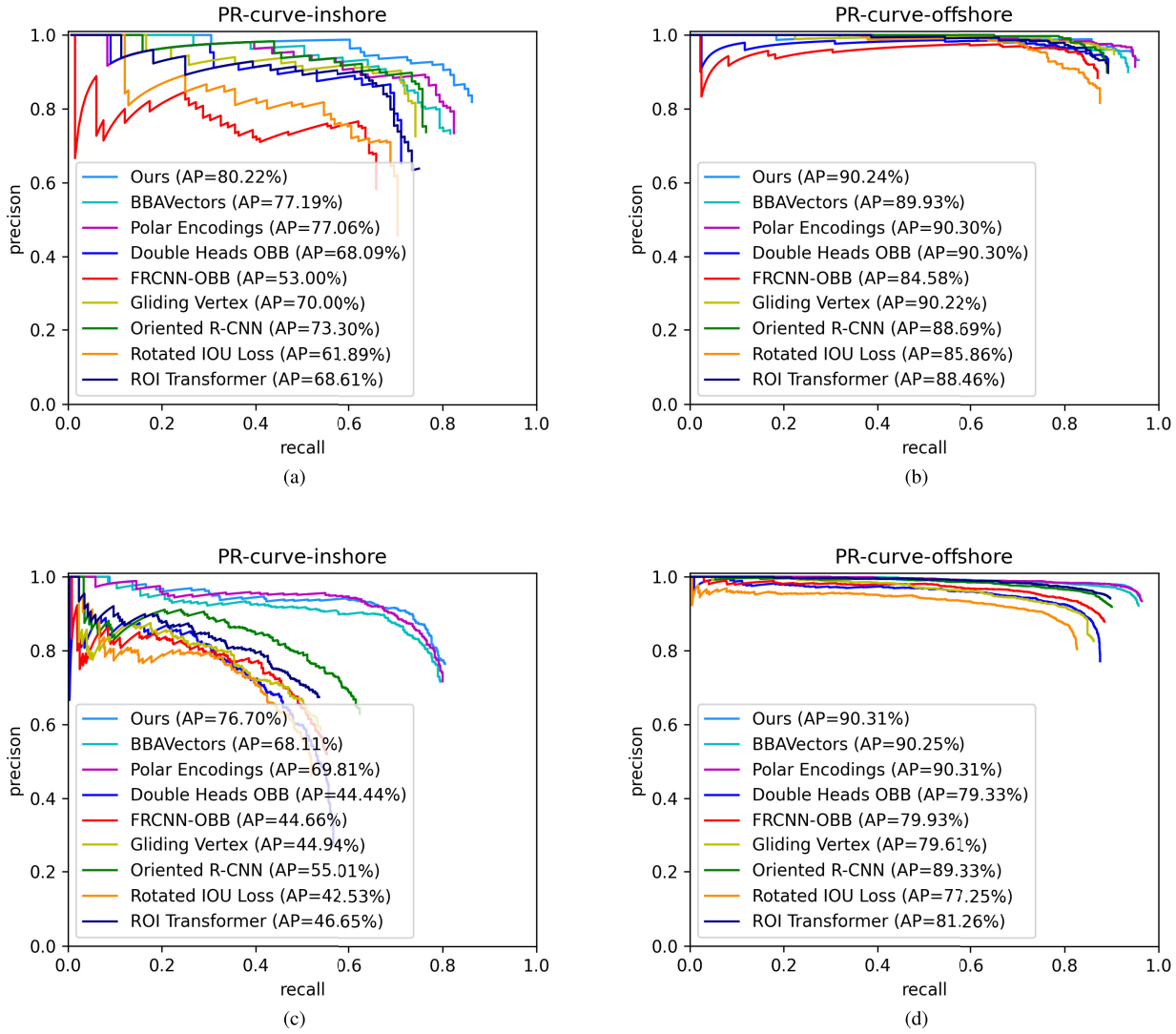


Fig. 15. PR curves for OBB-based methods. (a) PR curves in the inshore scenes of RSSDD. (b) PR curves in the offshore scenes of RSSDD. (c) PR curves in the inshore scenes of RSDD-SAR. (d) PR curves in the offshore scenes of RSDD-SAR.

each method in the inshore and offshore scenes on the RSSDD and RSDD-SAR, respectively. From the experiment results in Tables V and VI, several conclusions can be drawn.

1) Our proposed method can provide much better detection performance compared to the other state-of-the-art methods in terms of the false alarm rate P_{fa} and the detection rate P_{d} on both benchmark datasets. This demonstrates that our method can significantly improve network responses to target centers by extracting features and introducing an attention module, as well as mitigate the detrimental effects of background clutters on network performance.
2) Our method has half many missed detections as the anchor-free method Polar Encodings and BBAVectors, with only five and six missed detections in the offshore and inshore scenes of RSSDD, and 39 and 38 missed detections in the offshore and inshore scenes of RSDD-SAR, respectively. This is due to our targeted improvement in the network prediction performance of the target center point. For example, the elliptical 2-D Gaussian distribution heatmap contributes to improving the score gap between nearby ships, and the DPM can greatly enhance the detection ratio in the inshore scenes with nearly no increase in false alarms.
3) As shown in Table V, with training and inferencing on the RSSDD, the anchor-free OBB-based methods, BBAVectors, and Polar Encodings generate more false alarms while producing fewer missed detections compared to the other anchor-based methods. The number of false alarms of Polar Encodings is about 2.5 times of Gliding Vertex and ROI Transformer, five times our proposed method. The missed detection performance of the anchor-based methods deteriorates in the inshore and offshore scenes due to the poor generalization ability caused by the manually set anchor parameters and postprocessing of NMS. Among these anchor-based methods, Rotated IOU Loss generates the most missed detections, with 43 missed detections in the inshore scenes.
4) For practically all methods, the false alarm rate P_{fa} rises to variable degrees, as shown in the Table VI, on the RSDD-SAR dataset. This might be because, in contrast

TABLE V
EXPERIMENT RESULTS UNDER RADAR TARGET DETECTION METRICS OF DIFFERENT METHODS ON THE RSSDD

Model	Scene	False Alarm	Missed Detections	p_{fa}%	p_d%
BBAVectors	offshore	26	6	6.05	98.53
	inshore	27	16	19.15	87.69
Polar Encodings	offshore	12	10	2.91	97.56
	inshore	28	11	19.04	91.53
Gliding Vertex	offshore	0	25	<u>0</u>	93.90
	inshore	11	18	8.94	91.53
ROI Transformer	offshore	3	21	0.76	94.87
	inshore	14	25	11.76	80.76
FRCNN-OBB	offshore	3	19	0.76	95.37
	inshore	9	19	7.50	85.38
Double Heads OBB	offshore	5	19	1.26	95.37
	inshore	9	19	7.50	85.38
Rotated IOU Loss	offshore	2	19	0.51	95.36
	inshore	11	43	11.22	66.92
Oriented R-CNN	offshore	0	15	<u>0</u>	96.34
	inshore	3	20	2.65	84.61
Ours	offshore	5	5	**1.22**	**98.78**
	inshore	4	6	3.25	<u>95.38</u>

Bold items denote the optimal offshore values in the columns, the underlined items represent the optimal inshore values in the columns. Lower false alarm rate p_{fa} and higher detection rate p_d represent better detection performance.

to the RSSDD, the RSDD-SAR dataset contains more reefs, sea clutters, and other distracting targets and fewer ship targets in the slices of the offshore scenes. Among the anchor-based methods, oriented R-CNN yields relatively outstanding detection performance compared to our method. Double Heads OBB produces the highest P_{fa} out of all the methods, with 390 and 825 false alarms in the inshore and offshore scenes, respectively. Its performance is consistent with the corresponding precision metric in Table IV.

In order to demonstrate the detection efficiency of each method, Fig. 16 visualizes the test time and FLOPs of our method and other comparative methods. As shown in the figure, since different fully connected convolutional layers are used in the second stage for classification and regression, Double Heads OBB requires far larger FLOPs than the other methods, while the required test time is also twice as long as the FRCNN-OBB, which has the lowest detection efficiency. Our method, Polar Encodings and BBAVectors are all based on the one-stage anchor-free frame. Anchor-free methods do not involve anchor-related calculations and NMS algorithm; the required FLOPs are significantly better than other anchor-based methods, with 80.58G, 74.36G, and 76.52G, respectively. Compared with our method and BBAvectors, Polar Encodings does not require the angle branch and, therefore, requires lower FLOPs. Since our method needs to predict by combining the angle branch and the distance branch, and the process of direct least-squares fitting is performed on the CPU, the test time is slightly higher than that of some two-stage networks; however, the difference is not significant.

To visually compare our proposed method with other methods, the detection results of different methods on the inshore and offshore scenes are given in Figs. 17–20. Columns (1)–(10) represent the ground truth, the detection results of our method, BBAVectors, Polar Encodings, Gliding Vertex, ROI Transformer, FRCNN-OBB, Double Heads OBB, Rotated IOU Loss, and Oriented R-CNN, respectively.

As can be seen from Figs. 17(b) and 18(b), our method can successfully distinguish and provide specific locations and orientations for densely distributed ship targets in the inshore scenes, resulting in fewer missed detections and false alarms. False alarms often appear for slices that contain small targets. For instance, one missed detection is generated in the third image of Fig. 17(b) and the second and seventh images of Fig. 18(b). Figs. 17(c) and (d) and 18(c) and (d) illustrate that, for BBAVectors and Polar Encodings, which are key point-based methods, small ship targets and those positioned closely together in the inshore scenes frequently escape detection, while the offshore scene's reefs are frequently mistaken for true targets. For instance, in the second row of Fig. 17 and the fourth and seventh rows of Fig. 18, when multiple targets are distributed densely or the scale is very small when docked in a close-to-shore environment, the targets cannot be detected successfully. This is due to the fact that, when the standard Gaussian distribution heatmap is adopted, the scores on the ship side of adjacent ship targets are not well differentiated. In this case, the true center response will be suppressed during the max-pooling process, directly leading to missed detections of the target. As shown in Figs. 17(e) and 18(e), the anchor-based Gliding Vertex can generate accurate detection

TABLE VI
EXPERIMENT RESULTS UNDER RADAR TARGET DETECTION METRICS OF DIFFERENT METHODS ON THE RSDD-SAR

Model	Scene	False Alarm	Missed Detections	p_{fa}%	p_d%
BBAVectors	offshore	277	53	10.57	97.78
	inshore	131	62	19.70	89.59
Polar Encodings	offshore	192	66	7.61	97.24
	inshore	143	52	20.81	91.27
Gliding Vertex	offshore	309	89	3.56	91.66
	inshore	94	141	17.12	76.34
ROI Transformer	offshore	17	129	0.74	94.61
	inshore	110	233	23.25	60.91
FRCNN-OBB	offshore	43	26	1.78	99.33
	inshore	127	91	20.10	84.73
Double Heads OBB	offshore	390	67	14.43	97.20
	inshore	825	196	67.34	67.11
Rotated IOU Loss	offshore	101	33	1.33	98.62
	inshore	133	57	8.51	90.10
Oriented R-CNN	offshore	15	65	**0.64**	97.41
	inshore	42	47	7.11	92.11
Ours	offshore	19	39	0.79	**99.62**
	inshore	29	38	<u>4.94</u>	<u>93.62</u>

Bold items denote the optimal offshore values in the columns, the underlined items represent the optimal inshore values in the columns. Lower false alarm rate p_{fa} and higher detection rate p_d represent better detection performance.

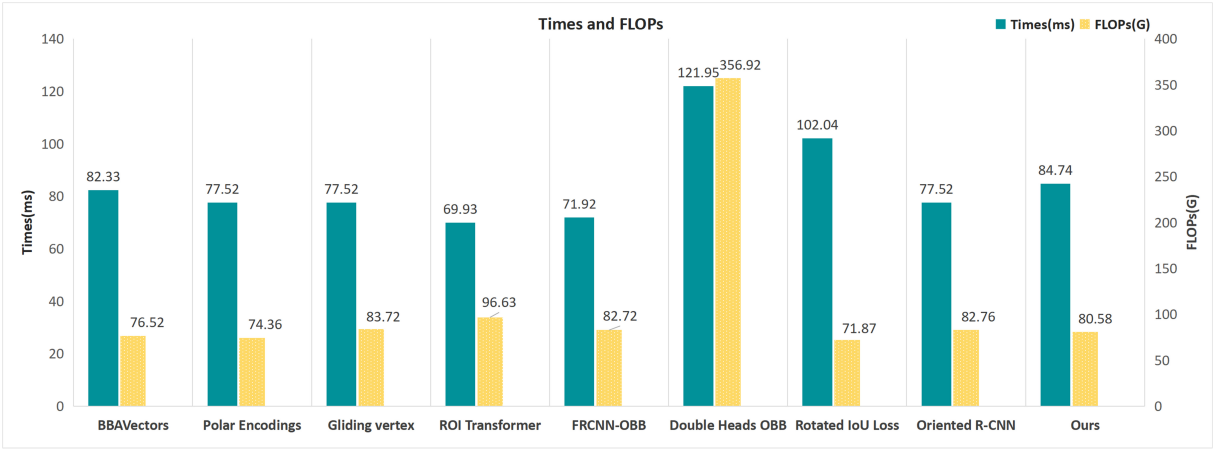


Fig. 16. Comparison of test time and FLOPs for different methods.

results for densely distributed targets of similar scales. However, for adjacent targets with large-scale differences, small targets are often missed, such as the second, fourth, and fifth rows in Fig. 17 and the bottom three rows in Fig. 18. This is because of the high overlap ratio between the two detection boxes, and the detection box of small targets is suppressed in the NMS process, resulting in missed detections. The detection results in Figs. 17(f) and 18(f) indicate that ROI Transformer also appears to detect multiple targets with large differences in adjacent scales as the same target. From Figs. 17(g) and (h) and 18(g) and (h), we can see that FRCNN-OBB and Double Heads OBB show clearly wrong angle predictions and missed detections. This is on account of both methods directly adding an angle prediction branch to the two-stage HBB detection method without any targeted improvements for boundary problems. The detection results of Rotated IOU Loss are shown in Figs. 17(i) and 18(i); it can be seen that, when using Rotated IOU Loss to guide network regression, the final detection confidence for many targets in the inshore scenes is low and not shown in the visualization. As can be seen from Figs. 17(j) and 18(j), anchor-based Oriented R-CNN can accurately detect target orientation, but there are also cases where multiple adjacent targets with large-scale differences are detected as the same target.

From the visualization results from Figs. 19(b)–(j) and 20(b)–(j), we can draw the following conclusions: all

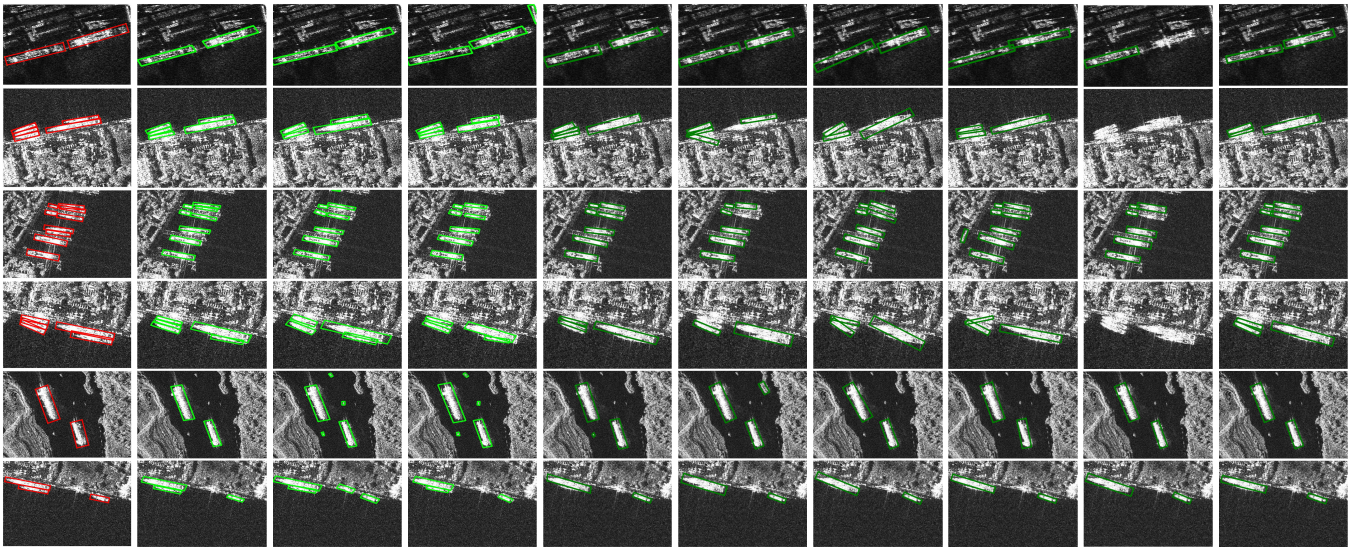


Fig. 17. Visualization of detection results of different methods in the inshore scenes of RSSDD. (a) Ground truth. (b) Our method. (c) BBAVectors. (d) Polar Encodings. (e) Gliding Vertex. (f) ROI Transformer. (g) FRCNN-OBB. (h) Double Heads OBB. (i) Rotated IOU Loss. (j) Oriented R-CNN.

methods are effective in providing the true location and orientation of the ship targets in the offshore scenes for the case of less clutter and significant target locations. However, BBAVectors and Polar Encodings based on center point detection produce more obvious false alarms compared to other anchor-based methods, and the surface reefs are wrongly detected as ship targets in the three given rows shown in Fig. 19 and the bottom three rows in Fig. 20. Our method generates only one false alarm in the third image due to the addition of the spatial attention module to suppress background clutter and enhance the target saliency response during feature fusion. In general, it can be seen from all visual detection results in Figs. 17–20 that our proposed method achieves more accurate detection results for the case of both inshore and offshore scenes compared to state-of-the-art methods.

We further conducted detection experiments on two large-scene images from the RSDD-SAR dataset in order to more thoroughly assess the generalization capability of our model. Nine representative regions, including slices of ports (slice 2), docks (slice 3), channels (slice 4), sea surfaces with low and high resolution (slices 1, 5, and 7–9), and dense distributed (slice 2), were chosen from the two large-scene images. Fig. 21 illustrates the detection results, including where the slices are cropped. It can be seen from the comparative detection results that our model can effectively detect multiscale targets in both inshore and offshore scenes although there are still some missed detections in scenes with densely packed targets. It can be concluded that our model demonstrates excellent detection performance and generalizability for the case of both the slice detection results of the two datasets mentioned above and the detection results of the two large-scene images of RSDD-SAR.

Fig. 22 shows several detection results of our proposed method in the inshore and offshore scenes, where the red points denote the regressed dynamic key points, the yellow points are the ship centers, and the green boxes represent the final detection results by finding the circumscribed rectangle of the ellipse obtained by direct least-squares fitting. The following can be observed from Fig. 22.

1) The key points sampled on the ellipse can be accurately restored by the prediction results of the distance branch and the angle branch. It indicates the feasibility of the encoding and decoding process, and verifies that accurate classification of the bow angle can be achieved using CSL.
2) The key points are accurately included in the strong scattering region of the ship bow. It not only illustrates that using the OBB inscribed ellipse to fit the ship contour for encoding is sensible but also shows that using these strong scattering points with rich feature information as sampling key points can achieve effective OBB regression.
3) When sampling dynamic key points in the bow region, the key points can fit the bow contour well for each aspect ratio of the ship target, and the predicted OBBs can completely enclose the ship target. In summary, for the ship targets in the inshore and offshore scenes, the representation method based on the OBB inscribed ellipse encoding proposed in this article can effectively localize the ship targets, reflecting the effectiveness of our method.

2) Ablation Experiments: The proposed elliptical Gaussian distribution heatmap and DPM are subjected to ablation experiments on the RSSDD and RSDD-SAR datasets to show their utility and statistically evaluate their impact on network performance.

Tables VII and VIII show the ablation experiment results of the elliptical Gaussian distribution heatmap. By comparing the detection precision, recall, F1, and AP metrics of the network when generating heatmaps using a standard Gaussian distribution and elliptical Gaussian distribution, the following conclusions can be drawn.

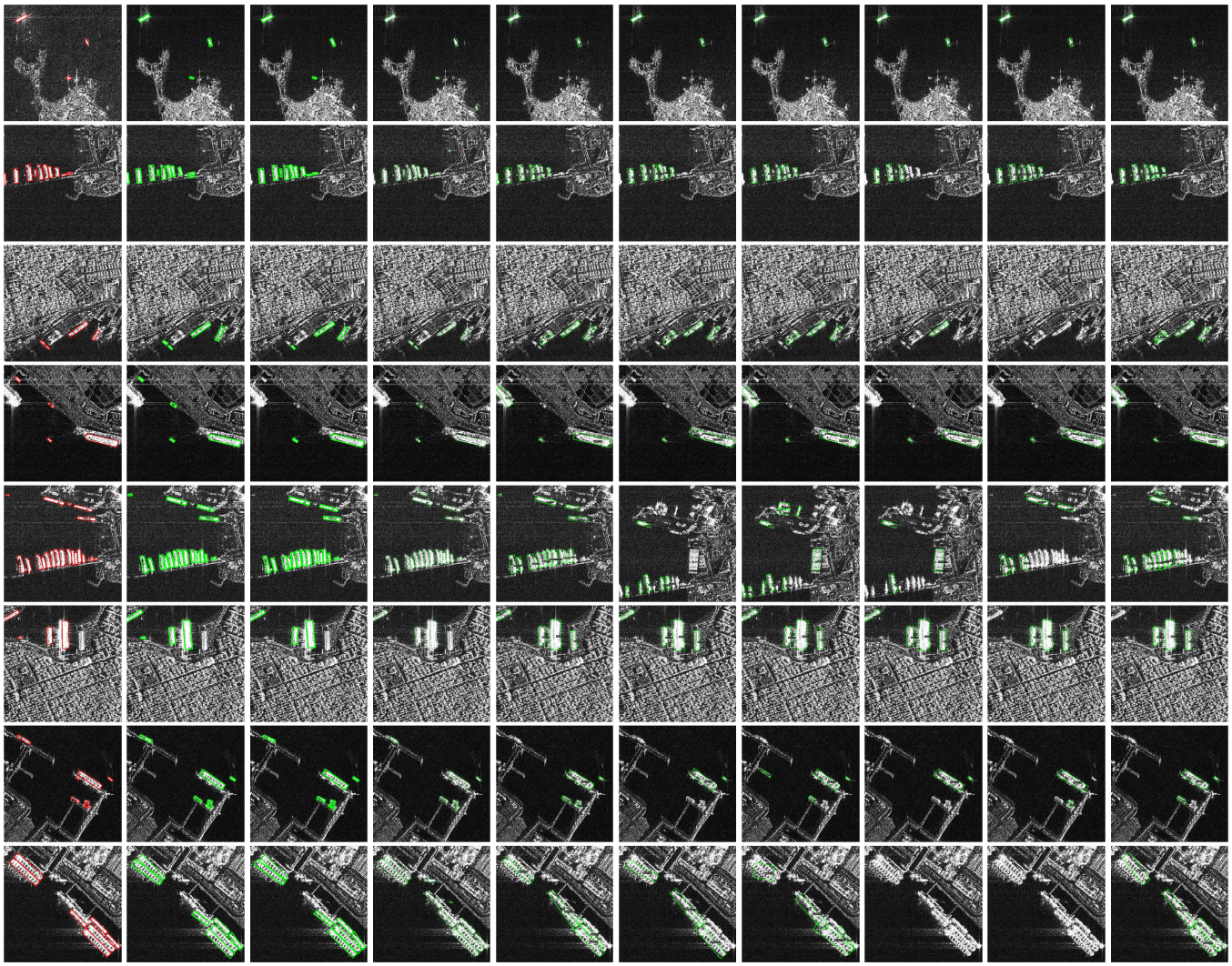


Fig. 18. Visualization of detection results of different methods in the inshore scenes of RSDD-SAR. (a) Ground truth. (b) Our method. (c) BBAVectors. (d) Polar Encodings. (e) Gliding Vertex. (f) ROI Transformer. (g) FRCNN-OBB. (h) Double Heads OBB. (i) Rotated IOU Loss. (j) Oriented R-CNN.

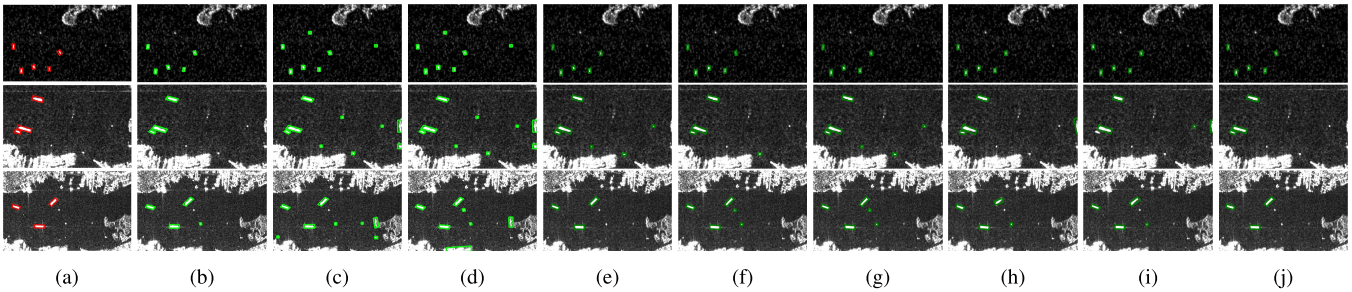
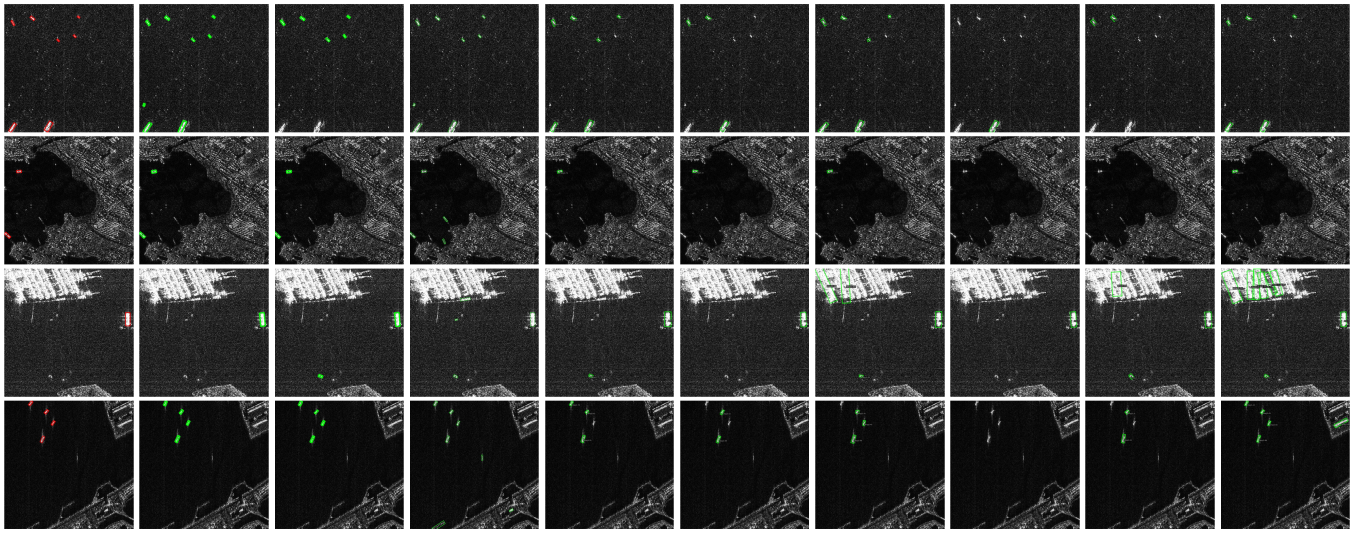


Fig. 19. Visualization of detection results of different methods in the offshore scenes of RSSDD. (a) Ground truth. (b) Our method. (c) BBAVectors. (d) Polar Encodings. (e) Gliding Vertex. (f) ROI Transformer. (g) FRCNN-OBB. (h) Double Heads OBB. (i) Rotated IOU Loss. (j) Oriented R-CNN.

1) When the elliptical Gaussian distribution is used to generate the heatmap, the network achieves better performance in both inshore and offshore scenes in all metrics compared to the use of a standard Gaussian distribution.
2) The improvement in the inshore scenes is particularly significant, with a 3.19% improvement in precision and a 1.54% improvement in recall on the RSSDD, and a 1.24% improvement in precision and a 0.70% improvement in recall on the RSDD-SAR. This indicates that the degree of distinction of densely distributed ships can be effectively improved by increasing the ship-side score distance and verifies the feasibility of using the elliptical Gaussian distribution to generate the heatmap.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 20. Visualization of detection results of different methods in the offshore scenes. (a) Ground truth. (b) Our method. (c) BBAVectors. (d) Polar Encodings. (e) Gliding Vertex. (f) ROI Transformer. (g) FRCNN-OBB. (h) Double Heads OBB. (i) Rotated IOU Loss. (j) Oriented R-CNN.

TABLE VII
ABLATION EXPERIMENT RESULT OF ELLIPTICAL GAUSSIAN HEATMAP ON THE RSSDD

Heatmap	Center Points	Scene	Precision%	Recall%	F1%	AP%
Standard 2-D Gaussian	DPM	offshore	93.56	94.51	93.93	90.14
		inshore	86.07	80.92	83.00	79.52
Elliptical 2-D Gaussian	DPM	offshore	**95.01**	**95.01**	**95.01**	**90.24**
		inshore	89.26	82.44	85.71	80.22

Bold items denote the optimal offshore values in the columns, the underlined items represent the optimal inshore values in the columns.

TABLE VIII
ABLATION EXPERIMENT RESULT OF ELLIPTICAL GAUSSIAN HEATMAP ON THE RSDD-SAR

Heatmap	Center Points	Scene	Precision%	Recall%	F1%	AP%
Standard 2-D Gaussian	DPM	offshore	95.82	95.21	95.51	90.19
		inshore	81.44	76.98	79.15	73.72
Elliptical 2-D Gaussian	DPM	offshore	**95.96**	**95.28**	**95.62**	**90.31**
		inshore	82.68	77.68	80.10	76.70

Bold items denote the optimal offshore values in the columns, the underlined items represent the optimal inshore values in the columns.

TABLE IX
ABLATION EXPERIMENT RESULT OF DPM ON THE RSSDD

Heatmap	Center Points	Scene	Precision%	Recall%	F1%	AP%
Elliptical 2-D Gaussian	3x3 Max-pooling	offshore	94.00	94.26	94.03	90.12
		inshore	86.29	81.68	83.92	80.19
Elliptical 2-D Gaussian	DPM	offshore	**95.01**	**95.01**	**95.01**	**90.24**
		inshore	89.26	82.44	85.71	80.22

Bold items denote the optimal offshore values in the columns, the underlined items represent the optimal inshore values in the columns.

Tables IX and X show the ablation experiment results of DPM. By comparing the detection precision, recall, F1, and AP of the network when extracting the ship centers from the center point prediction map using a standard

TABLE X
ABLATION EXPERIMENT RESULT OF DPM ON THE RSDD-SAR

Heatmap	Center Points	Scene	Precision%	Recall%	F1%	AP%
Elliptical 2-D Gaussian	3x3 Max-pooling	offshore	95.79	95.14	95.46	90.13
		inshore	81.67	76.88	79.20	74.35
Elliptical 2-D Gaussian	DPM	offshore	**95.96**	**95.28**	**95.62**	**90.31**
		inshore	<u>82.68</u>	<u>77.68</u>	<u>80.10</u>	<u>76.70</u>

Bold items denote the optimal offshore values in the columns, the underlined items represent the optimal inshore values in the columns.

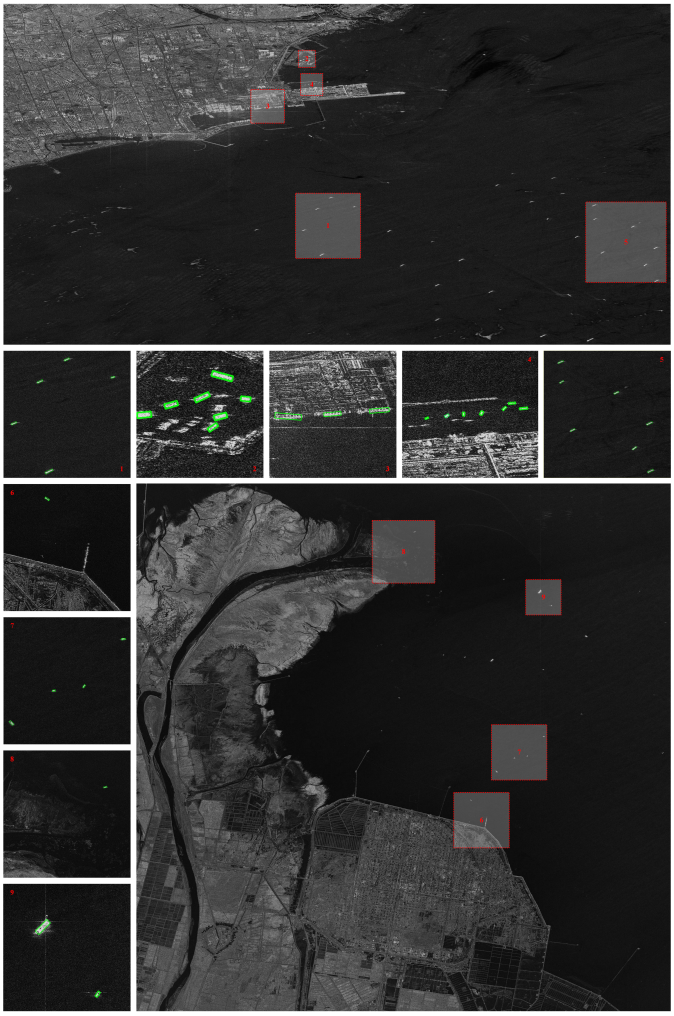


Fig. 21. Detection results on the large-scene image.

3 × 3 max-pooling and DPM, the following conclusions can be drawn. Using DPM to extract center points effectively improves the detection precision and recall of the network in the inshore and offshore scenes, with a 2.97% and 0.76% precision improvement, respectively, in the inshore scenes of RSSDD, and 1.01% and 0.60% recall improvements, respectively, in the inshore scenes of RSSDD. This indicates that the DPM designed in our method avoids missed detections of densely distributed targets to a certain extent by extraction of twice the maximum values. At the same time, when combined with the score threshold, it does not introduce more false alarms to the detection results, which can guarantee high precision while improving recall.

From the ablation experiment results, it can be seen that: 1) compared to DPM, the elliptical Gaussian distribution heatmap improves the network performance more significantly and 2) when the two methods are used together, the network detection performance improves even more.

V. DISCUSSION

In this section, first, we conduct experiments on RSSDD to determine the effect of the Gaussian window radius of CSL on the comprehensive performance of the network and the classification performance of the angle branch. Then, we perform a detailed comparison with the sampling method of Polar Encodings to demonstrate the superiority of sampling on the inscribed ellipse of the OBB compared to sampling directly on the OBB.

A. Gaussian Window Radius of CSL

CSL smooths the one-hot label using a Gaussian window function, and the window radius Sigma is critical to angle classification performance. The Gaussian window function converges to the impulse function when the window radius Sigma is too small, which is equivalent to a direct hard classification of the angle category. Angle classification is also challenging when the window radius Sigma is too large since the scores of nearby angle groups are not sufficiently distinct. In order to assure the angle prediction branch's great classification performance, it is vital to choose the proper window radius Sigma. Yang and Yan [17] classify the target angle into 180 categories and experimentally demonstrate that the best results can be achieved on the DOTA [8] dataset with Sigma $=$ 6. Our method needs to classify $M = 18$ bow angle intervals in the case of key points $N = 6$. The thickness of the classification branch is much smaller than that of [32], so experiments are needed to rechoose the most suitable Gaussian window radius Sigma. Fig. 23 shows the network precision, recall, F1, and AP under different Gaussian window radii of CSL. From Fig. 23(a), it can be seen that the network achieves a good combined detection performance under the given several window radii in the offshore scenes, where the network achieves the best detection performance for Sigma $=$ 1. This is because there is less surface clutter and more emphasis on the ship targets' directional properties, allowing for relatively simple correct categorization of the bow

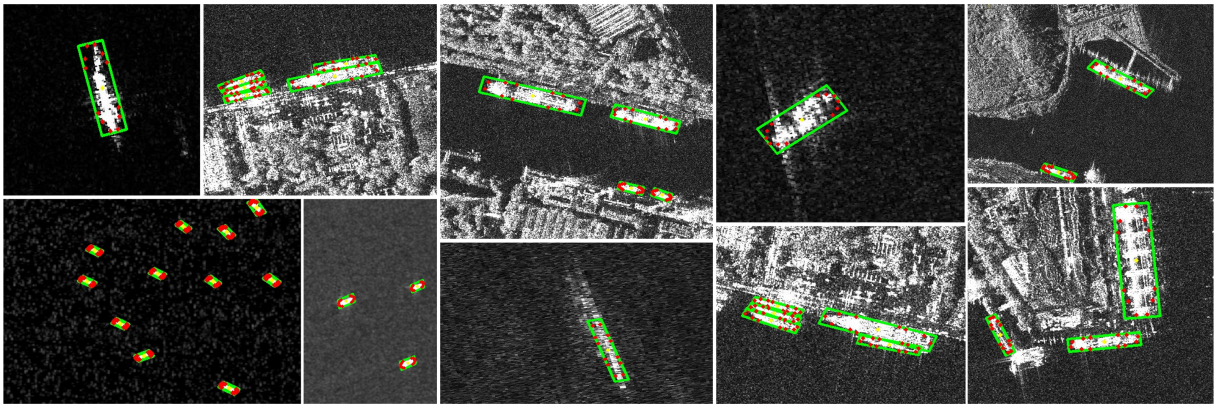
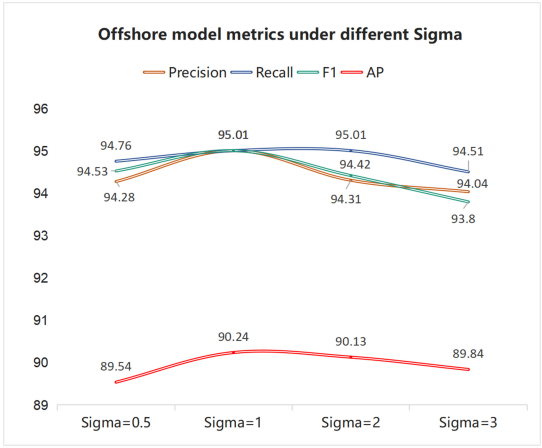


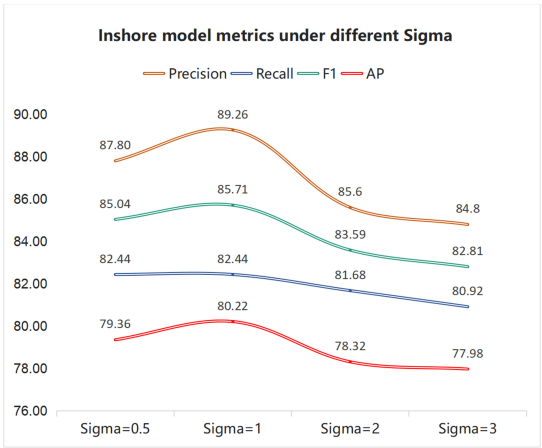
Fig. 22. Detection results of our proposed method. The red points denote the regressed dynamic key points, the yellow points are the ship centers, and the green boxes represent the final detection results.

angle interval. As can be seen in Fig. 23(b), the window radius Sigma has a significant impact on the comprehensive detection performance of the network in the inshore scenes. Specifically, the overall performance shows a trend of the first improvement and then decreases as the window radius Sigma increases, and the best detection precision, recall, F1, and AP metrics are achieved simultaneously at Sigma $=$ 1. Considering the performance metrics of the network under different window radii in the inshore and offshore scenes, in our method, the Gaussian window radius Sigma $=$ 1 is selected.

The impact of various window radii on the performance of the angle branch is then more specifically compared. In this article, the principal component analysis (PCA) [49] is used to first show the angle characteristics learned from RSSDD, as shown in Fig. 24. Fig. 24(a)–(d) shows the visualization of the angle features learned by this branch at window radius Sigma $=$ 0.5, Sigma $=$ 1, Sigma $=$ 2, and Sigma $=$ 3, respectively. Fig. 24(a) illustrates that this branch does not learn the angle information well when Sigma $=$ 0.5; some of the angle features are indistinguishable after dimension reduction and are reflected in the visualization as marker points overlapping each other. At this point, the Gaussian window function converges to an impulse function, which is equivalent to a hard classification of the angle interval directly. Fig. 24(b) shows that, when Sigma $=$ 1, the angle features learned by the prediction branch present as a ring structure with uniform distribution. The adjacent angle features are close to each other and partially overlap, and this feature reflects the principle that CSL can calculate angle distances without introducing boundary problems compared to a one-hot label. As can be seen in Fig. 24(c) and (d), the angle branch learns the ring angle features equally well when Sigma $=$ 2 and Sigma $=$ 3. However, at this point, more randomly distributed points show up in the circle's center, and the overlap degree of angle characteristics grows noticeably. This is because, as the window radius increases, the scoring distinction between neighboring angles decreases, and the classification of angles becomes somewhat more challenging. Second, for all ship targets in both scenes of the RSSDD dataset, we count the angle classification results of the angle prediction branch and compare them with the true values of the bow angle interval.



(a)

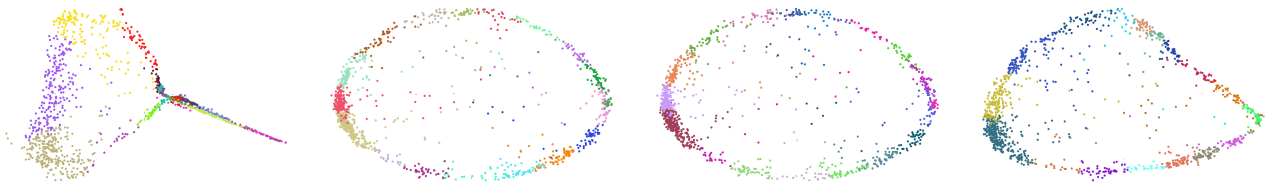


(b)

Fig. 23. PR curves for OBB-based methods. (a) PR curves in the offshore scenes. (b) PR curves in the inshore scenes.

Accuracy is then computed to evaluate more precisely the effects of different Gaussian window radii on angle branch prediction performance. Accuracy is defined as follows:

$$\text{Accuracy} = \frac{N_c}{N_c + N_w} = \frac{N_c}{N_{\text{all}}} \tag{20}$$



(a) (b) (c) (d)

Fig. 24. Visualization features of angle branch under the RSSDD dataset. (a) Sigma = 0.5. (b) Sigma = 1. (c) Sigma = 2. (d) Sigma = 3.

TABLE XI
CLASSIFICATION ACCURACY OF THE ANGLE BRANCH UNDER DIFFERENT WINDOW RADII

		Sigma = 0.5	**Sigma = 1**	**Sigma = 2**	**Sigma = 3**
Accuracy%	offshore	96.21	99.53	98.81	98.34
	inshore	94.44	98.59	94.44	90.28

Bold items denote the optimal offshore values in the columns, the underlined items represent the optimal inshore values in the columns.

where N_c denotes the number of targets correctly classified, N_w denotes the number of misclassified targets, and N_{all} represents the number of all targets. In the RSSDD test set divided in our method, there are 193 inshore scenes containing 422 ship targets and 39 inshore scenes containing 71 targets. The classification accuracy of the angle interval prediction branch under different window radii is shown in Table XI.

Combining the four parts in Fig. 24 and the experiment results in Table XI, these conclusions can be derived: 1) the experiment findings of the classification accuracy of the angle branch are consistent with the visualization of the angle features learned under various window radii and 2) the angle branch's classification accuracy has a direct impact on the network's detection performance, and both remain consistent. When the window radius Sigma = 1, the network has the best angle classification and detection performance, which verifies the appropriateness of choosing Sigma = 1 as the window radius of CSL in our method.

B. OBB Inscribed Ellipse and OBB Sampling Continuity Analysis

Polar Encodings encodes the OBB as an ordered set of distances of the boundary points on the OBB from the center point, effectively solving the problems of PoA and EoE. However, the scattering characteristics and geometry of the ship targets are not effectively exploited, and the continuity of the sampling distance is affected by the number of sampling points N. In order to visualize the continuity of the sampling distance during the sampling process, Fig. 25(a) and (b) gives the sampling distance $d(\varphi)$ when sampling on the OBB and the OBB inscribed ellipse, respectively. In the OBB sampling process, we set the number of sampling points to 8 in the case of Polar Encodings, and the OBB inscribed ellipse sampling process adopts this work's normal setting of six sampling points. Given the target OBB B_0 with height h and width w, the OBB obtained by rotating it clockwise by θ is B_θ. The corresponding inscribed ellipses of B_0 and B_θ are E_0 and E_θ, respectively. When sampling in the range of $[-\pi, \pi]$, the distances of the sampling points on the OBB and the OBB inscribed ellipse from the center of the OBB are shown in Fig. 25(a) and (b). The yellow curve represents the distance from the center as a function of the sampling angle on B_0 and E_0, and the blue curve represents the distance from the center as a function of sampling angle on B_θ and E_θ.

We can find the following from Fig. 25(a) and (b).

1) The sampling distance function $d(\varphi)$ is a periodic function with a period of π either for OBB or the OBB inscribed ellipse. The maximum sampling distances are $(w^2 + h^2)^{1/2}/2$ and $w/2$, respectively.
2) The rotation angle θ, i.e., the angle error, is to make $d(\varphi)$ shift θ.
3) When sampling on OBB and the sampling angle is around the diagonal angle $\arctan(h/w)$, the sampling distance on B_θ tends to increase and then decrease more sharply. At this point, the sampling distance error will spike at $\arctan(h/w)$ due to the smooth decrease of the sampling distance on B_0.

In order to visually compare the sampling distance error $s(\varphi)$ generated on OBB and the OBB inscribed ellipse in the case of angular error is equal to θ, $s(\varphi)$ can be calculated by $\sum_{i=1}^{N} |\mathrm{d}_0(\varphi_i) - \mathrm{d}_\theta(\varphi_i)|/N$, where N denotes the sampling points. The sampling points are shown as blue and yellow points in Fig. 25. Fig. 26(a) and (b) shows the $s(\varphi)$ when sampling on an OBB inscribed ellipse and an OBB with the aspect ratio of two. Fig. 26(c) illustrates the $s(\varphi)$ for an OBB with the aspect ratio of one. $s(\varphi)$ at different sampling points is given in each graph. From Fig. 26(a) and (b), it can be seen that the following holds.

1) For the target with the aspect ratio of two, when the angle error varies within the range of $[0, \pi]$, $s(\varphi)$ generated by sampling on the OBB and OBB inscribed ellipse tends to increase first and then decrease, which is consistent with the actual situation. The target overlap is lowest, and $s(\varphi)$ is highest when the angle error equals $\pi/2$.
2) As the sampling points increase, $s(\varphi)$ curve gradually smooths out.
3) $s(\varphi)$ from sampling on the OBB inscribed ellipse is much smaller than that on the OBB using the same or fewer sampling points, and the error curve is much smoother.

As can be seen from Fig. 26(c), for a target with the aspect ratio of one, the distance error period generated by direct

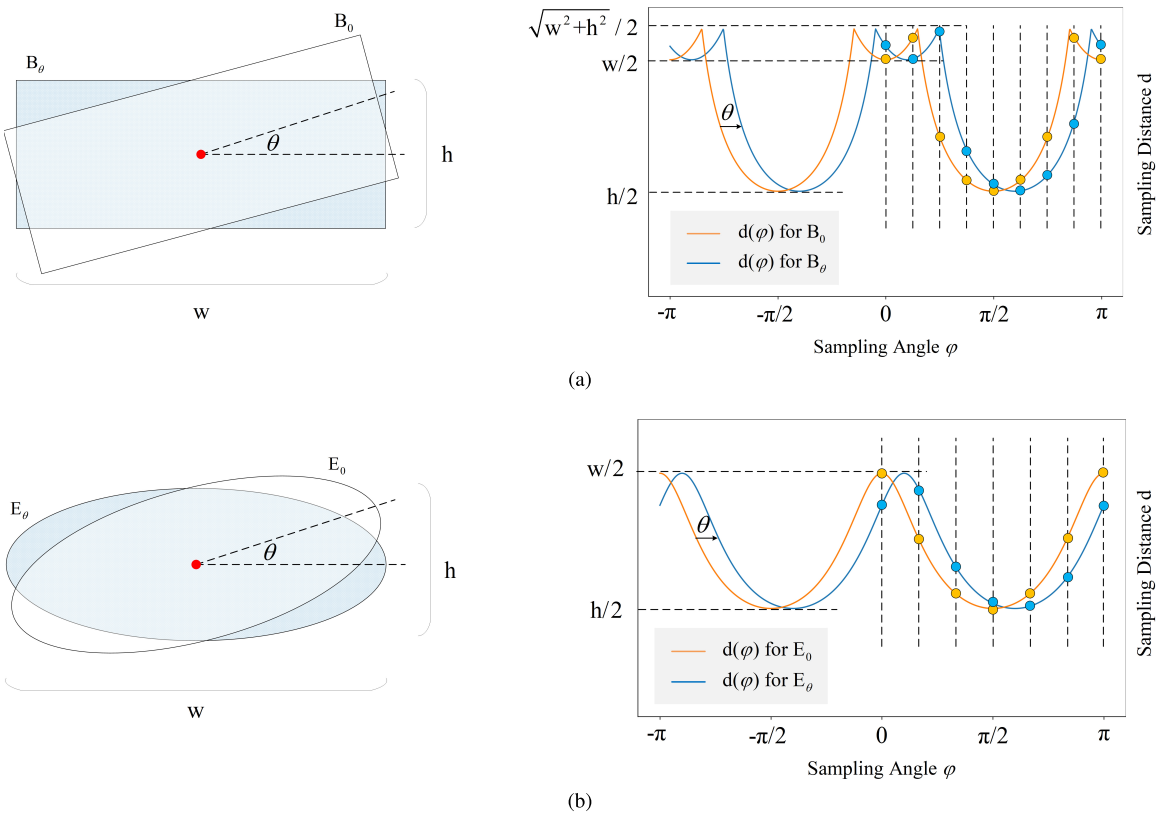


Fig. 25. Sampling distance $d(\varphi)$ on OBB and OBB inscribed ellipse. (a) $d(\varphi)$ when sampling on OBB. (b) $d(\varphi)$ when sampling on OBB inscribed ellipse.

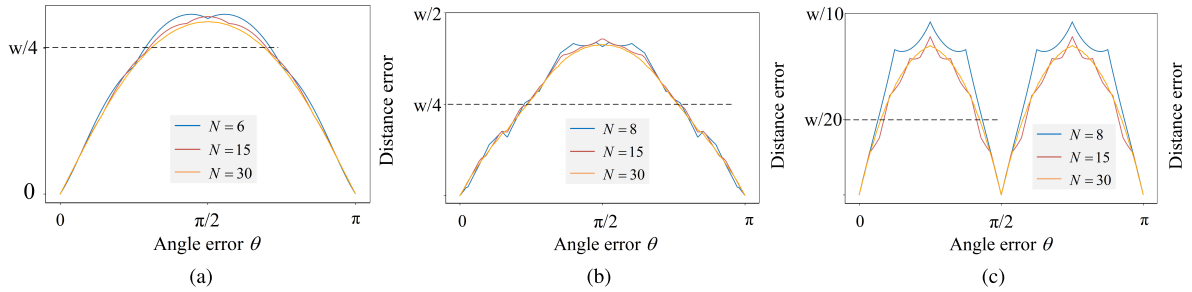


Fig. 26. Sampling distance error $s(\varphi)$. (a) $s(\varphi)$ when sampling on an OBB inscribed ellipse with the aspect ratio of two. (b) $s(\varphi)$ when sampling on an OBB with the aspect ratio of two. (c) $s(\varphi)$ when sampling on an OBB with the aspect ratio of one.

sampling on the OBB is $\pi/2$, and the maximum $s(\varphi)$ is generated around the angular error $\pi/4$. Since the long and short axes of the inscribed ellipse of OBB are equal to the standard circle at this time, the distance of each sampling point equals the radius, and $s(\varphi)$ is zero, which is, thus, not shown in the figure.

In summary, sampling on the OBB inscribed ellipse not only generates a periodic loss function to overcome the boundary discontinuity problem but also has significant advantages over sampling directly on the OBB. On the one hand, there is no sudden change in sampling distance. On the other hand, the overall generated average distance error is smaller, and the variation is smoother, which can guide the network regression more effectively.

VI. Conclusion

In this article, we proposed an OBB detector based on dynamic key points' detection for arbitrary-oriented ship detection in SAR images. To mitigate the negative impact of the boundary discontinuity problem on network performance and to effectively utilize the geometric and scattering characteristics of ship targets, the ship contour is fit by an OBB inscribed ellipse and encoded as a distance set of the dynamic key points on the bow from the ship center. The predicted OBB is obtained by a combination of the bow angle interval classification and a special decoding process. To improve the network performance for the detection of densely arranged ship targets in the inshore scenes, we further designed an elliptical Gaussian distribution heatmap and a pooling approach termed DPM. A series of experiments conducted on the RSSDD and RSDD-SAR datasets demonstrate that our method can accurately predict the target position and orientation, and effectively deal with the boundary discontinuity problem. Ablation experiment results demonstrated that the elliptical Gaussian distribution heatmap and DPM can further improve the inshore detection performance. Overall,

our method outperformed other state-of-the-art OBB-based methods, demonstrating its effectiveness as a new benchmark resource for ship target detection in SAR images. Future work will carry out a more exhaustive comparison with other state-of-the-art approaches, including complexity and real-time implementation challenges. We will also explore innovative semisupervised approaches building on our work in [50].

ACKNOWLEDGMENT

The authors are indebted to the anonymous reviewers for their insightful comments and suggestions that helped improve the quality of this article.

REFERENCES

[1] X. Ma, S. Hou, Y. Wang, J. Wang, and H. Wang, "Multiscale and dense ship detection in SAR images based on key-point estimation and attention mechanism," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.

[2] F. Ma, F. Zhang, D. Xiang, Q. Yin, and Y. Zhou, "Fast task-specific region merging for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.

[3] M. Messina, M. Greco, L. Fabbrini, and G. Pinelli, "Modified Otsu's algorithm: A new computationally efficient ship detection algorithm for SAR images," in *Proc. Tyrrhenian Workshop Adv. Radar Remote Sens. (TyWRRS)*, Sep. 2012, pp. 262–266.

[4] F. Gao, T. Huang, J. Wang, J. Sun, A. Hussain, and E. Yang, "Dual-branch deep convolution neural network for polarimetric SAR image classification," *Appl. Sci.*, vol. 7, no. 5, p. 447, Apr. 2017.

[5] K. El-Darymli, P. McGuire, D. Power, and C. Moloney, "Target detection in synthetic aperture radar imagery: A state-of-the-art survey," *J. Appl. Remote Sens.*, vol. 7, no. 1, Mar. 2013, Art. no. 071598.

[6] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[7] Y. Jiang et al., "R2CNN: Rotational region CNN for orientation robust scene text detection," 2017, *arXiv:1706.09579*.

[8] G.-S. Xia et al., "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3974–3983.

[9] J. Ma et al., "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3111–3122, Mar. 2018.

[10] X. Yang, J. Yan, Z. Feng, and T. He, "R3Det: Refined single-stage detector with feature refinement for rotating object," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 4, May 2021, pp. 3163–3171.

[11] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, "Learning RoI transformer for oriented object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2849–2858.

[12] Y. Zhu, J. Du, and X. Wu, "Adaptive period embedding for representing oriented objects in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7247–7257, Oct. 2020.

[13] P. Zhao, Z. Qu, Y. Bu, W. Tan, and Q. Guan, "PolarDet: A fast, more precise detector for rotated target in aerial images," *Int. J. Remote Sens.*, vol. 42, no. 15, pp. 5831–5861, Aug. 2021.

[14] Y. Xu et al., "Gliding vertex on the horizontal bounding box for multi-oriented object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 4, pp. 1452–1459, Apr. 2021.

[15] J. Yi, P. Wu, B. Liu, Q. Huang, H. Qu, and D. Metaxas, "Oriented object detection in aerial images with box boundary-aware vectors," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2150–2159.

[16] W. Qian, X. Yang, S. Peng, J. Yan, and Y. Guo, "Learning modulated loss for rotated object detection," in *Proc. AAAI Conf. Arti. Intell.*, vol. 35, no. 3, May 2021, pp. 2458–2466.

[17] X. Yang and J. Yan, "Arbitrary-oriented object detection with circular smooth label," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 677–694.

[18] X. Yang and J. Yan, "On the arbitrary-oriented object detection: Classification based approaches revisited," *Int. J. Comput. Vis.*, vol. 130, no. 5, pp. 1340–1365, May 2022.

[19] X. Yang et al., "SCRDet: Towards more robust detection for small, cluttered and rotated objects," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 8232–8241.

[20] X. Yang, L. Hou, Y. Zhou, W. Wang, and J. Yan, "Dense label encoding for boundary discontinuity free rotation detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 15819–15829.

[21] Y. He, F. Gao, J. Wang, A. Hussain, E. Yang, and H. Zhou, "Learning polar encodings for arbitrary-oriented ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3846–3859, 2021.

[22] X. Yang, J. Yan, Q. Ming, W. Wang, X. Zhang, and Q. Tian, "Rethinking rotated object detection with Gaussian Wasserstein distance loss," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 11830–11841.

[23] T. Zhang et al., "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sens.*, vol. 13, no. 18, p. 3690, Sep. 2021.

[24] C. Xu et al., "RSDD-SAR: Rotated ship detection dataset in SAR images," *J. Radars*, vol. 11, no. 4, pp. 581–599.

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–14.

[26] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[27] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.

[28] 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. (CVPR)*, Jun. 2016, pp. 779–788.

[29] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9627–9636.

[30] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 734–750.

[31] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.

[32] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery," *Remote Sens.*, vol. 11, no. 5, p. 531, 2019.

[33] Y.-L. Chang, A. Anagaw, L. Chang, Y. Wang, C.-Y. Hsiao, and W.-H. Lee, "Ship detection based on YOLOv2 for SAR imagery," *Remote Sens.*, vol. 11, no. 7, p. 786, Apr. 2019.

[34] T. Zhang, X. Zhang, J. Shi, and S. Wei, "High-speed ship detection in SAR images by improved Yolov3," in *Proc. 16th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process.*, Dec. 2019, pp. 149–152.

[35] F. Gao, Y. He, J. Wang, A. Hussain, and H. Zhou, "Anchor-free convolutional network with dense attention feature aggregation for ship detection in SAR images," *Remote Sens.*, vol. 12, no. 16, p. 2619, 2020.

[36] J. Wang, C. Lu, and W. Jiang, "Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression," *Sensors*, vol. 18, no. 9, p. 2851, 2018.

[37] Q. An, Z. Pan, L. Liu, and H. You, "DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8333–8349, Nov. 2019.

[38] Z. Pan, R. Yang, and Z. Zhang, "MSR2N: Multi-stage rotational region based network for arbitrary-oriented ship detection in SAR images," *Sensors*, vol. 20, no. 8, p. 2340, Apr. 2020.

[39] C. Chen, C. He, C. Hu, H. Pei, and L. Jiao, "MSARN: A deep neural network based on an adaptive recalibration mechanism for multiscale and arbitrary-oriented SAR ship detection," *IEEE Access*, vol. 7, pp. 159262–159283, 2019.

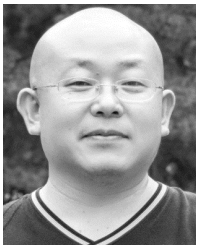
[40] Z. Sun, X. Leng, Y. Lei, B. Xiong, K. Ji, and G. Kuang, "BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images," *Remote Sens.*, vol. 13, no. 21, p. 4209, Oct. 2021.

[41] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 483–499.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[44] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–12.
[45] K. Chen et al., "MMDetection: Open MMLab detection toolbox and benchmark," 2019, *arXiv:1906.07155*.
[46] Y.-W. Chao, S. Vijayanarasimhan, B. Seybold, D. A. Ross, J. Deng, and R. Sukthankar, "Rethinking the faster R-CNN architecture for temporal action localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1130–1139.
[47] D. Zhou et al., "IoU loss for 2D/3D object detection," in *Proc. Int. Conf. 3D Vis. (3DV)*, Sep. 2019, pp. 85–94.
[48] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, "Oriented R-CNN for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3520–3529.
[49] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Comput. Statistic*, vol. 2, no. 4, pp. 433–459, Jul./Aug. 2010.
[50] Z. Yue et al., "A novel semi-supervised convolutional neural network method for synthetic aperture radar image recognition," *Cognit. Comput.*, vol. 13, no. 4, pp. 795–806, 2021.



Fei Gao received the B.S. degree in electrical automation and the M.S. degree in electromagnetic measurement technology and instrument from the Xi'an Petroleum Institute, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degree in signal and information processing from Beihang University, Beijing, China, in 2005.

He is currently a Professor with the School of Electronic and Information Engineering, Beihang University. His research interests include target detection and recognition, image processing, and deep learning for applications in remote sensing.

Yiyang Huo received the B.S. degree in electronic and information engineering from Beihang University, Beijing, China, in 2020, where he is currently pursuing the M.E. degree in information and communication engineering.

His research activities include target detection, instance segmentation, and remote sensing image processing.

Jinping Sun (Member, IEEE) received the M.Sc. and Ph.D. degrees from Beihang University (BUAA), Beijing, China, in 1998 and 2001, respectively.

He is currently a Professor with the School of Electronic and Information Engineering, BUAA. His research interests include statistical signal processing, high-resolution radar signal processing, target tracking, image understanding, and robust beamforming.

Tao Yu received the B.E. degree in photoelectric engineering from the Electronic Engineering Institute of PLA, Hefei, China, in 2013.

His research activities include aerospace telemetry and space exterior ballistics measurement.

Amir Hussain received the B.Eng. and Ph.D. degrees in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 1992 and 1997, respectively.

Following post-doctoral and senior academic positions at the University of the West of Scotland, Paisley, U.K., from 1996 to 1998, the University of Dundee, Dundee, U.K., from 1998 to 2000, and the University of Stirling, Stirling, U.K., from 2000 to 2018, respectively, he joined Edinburgh Napier University, Edinburgh, U.K., as the Founding Head of the Cognitive Big Data and Cybersecurity (CogBiD) Research Laboratory and the Centre for AI and Data Science. His research interests include cognitive computation, machine learning, and computer vision.

Huiyu Zhou received the B.Eng. degree in radio technology from the Huazhong University of Science and Technology, Wuhan, China, in 1990, the M.S. degree in biomedical engineering from the University of Dundee, Dundee, U.K., in 2002, and the Ph.D. degree in ratio technology, biomedical engineering, and computer vision from Heriot-Watt University, Edinburgh, U.K., in 2006.

He is currently a Professor with the School of Computing and Mathematical Sciences, University of Leicester, Leicester, U.K. His research interests include medical image processing, computer vision, intelligent systems, and data mining.