Arbitrary-Oriented Ellipse Detector for Ship Detection in Remote Sensing Images

Kexue Zhou ⓘ, Min Zhang ⓘ, Honghui Zhao, Rui Tang, Sheng Lin ⓘ, Xi Cheng ⓘ, and Hai Wang ⓘ

Abstract—At present, in arbitrary-oriented object detection, the angular periodicity problem of rotated bounding box described by angle causes an object to have different numerical representations, which leads to uncertainty of rotated bounding box regression. To eliminate the angular periodicity problem, in this article, we propose a novel and simple ellipse parameters representation method for arbitrary-oriented object, which hides the angle of the object in the focal vector of ellipse to avoid direct angle prediction. Moreover, the proposed representation method can enable the arbitrary-oriented object to have only one numerical representation, which is beneficial to alleviate the uncertainty of bounding box regression. In order to adapt the proposed ellipse parameters representation method, we adopt 2-D Gaussian distribution label assign for coarse samples selection, then the Kullback–Leibler divergence loss and SimOTA are used to refine the coarse samples to obtain the best positive samples. We extend the YOLOX with medium parameters as an oriented ship detector according to the proposed ellipse parameters representation method, and conduct the experiments on HRSC2016, RSDD-SAR, and RHRSID to demonstrate the effectiveness of the proposed method. The experimental results show that the proposed representation method achieves impressive results compared with state-of-the-art arbitrary-oriented object detection methods.

Index Terms—Anchor free, arbitrary-oriented ship detection, ellipse parameters, remote sensing images, single stage.

I. INTRODUCTION

OBJECT detection, is one of the basic tasks in computer vision, which can recognize the category of object and locate the object by predicting a horizontal bounding box [1], [2], [3], [4]. However, in the remote sensing images, the orientation of object is arbitrary, which is different from the object in natural images. Therefore, there is another form of object detection method, which is arbitrary-oriented object detection implemented by predicting a bounding box with angular

Kexue Zhou, Min Zhang, Sheng Lin, Xi Cheng, and Hai Wang are with the School of Aerospace Science and Technology, Xidian University, Xi'an 710126, China (e-mail: kxzhou@stu.xidian.edu.cn; minzhang@xidian.edu.cn; linsheng @stu.xidian.edu.cn; cheng_xi@stu.xidian.edu.cn; wanghai@mail.xidian.edu. cn).

Honghui Zhao is with the Xingtu Zhihua (Xi'an) Digital Technology Co., Ltd., Xi'an 710000, China (e-mail: zhaohonghui@envmap.cn).

Rui Tang is with the Shaanxi Academy of Aerospace Technology Application Co., Ltd., Xi'an 710199, China (e-mail: tangrui@tgis.top).

Digital Object Identifier 10.1109/JSTARS.2023.3267240

orientation. The arbitrary-oriented object detection can obtain more accurate object information unlike generic object detection which introduces too much background information. Because of the rapid development of deep learning, arbitrary-oriented object detection also has made great progress [5], [6], [7], [8] in recent years. Among these methods, the five-parameter regression-based methods occupy the mainstream, which are implemented by predicting an extra angle parameter on the basis of original horizontal bounding box parameters. However, predicting the angle of object directly will lead to the issues of boundary discontinuity and angular periodicity [5], [9], [10], [11].

Hiding the angle in a vector is one of the ways to solve the problems caused by direct angle prediction. Yi et al. [12] described an oriented object via the vectors from the center point to the four sides, which provides a new idea for arbitrary-oriented object detection, and achieve the competitive detection performance. However, there are two disadvantages, 1) to handle corner cases, the oriented bounding box is grouped into two categories: a) Horizontal bounding box; and b) rotational bounding box, which increases the complexity of processing oriented bounding box; and 2) there are at least ten parameters to predict, which has too many parameters to predict compared with five-parameter regression-based methods. In [13], He et al. also adopted the idea of hiding angle in a vector, however, according to the idea of [13], the order of the four vertices is based on the values of x_i, i = 1, 2, 3, 4, i.e., $(x_1 \leq x_2 < x_3 \leq x_4)$, which has the obvious problem that the order of the four vertices cannot be determined only based on the value of x_i, because x_2 can also be equal to x_3 in practice. When $x_2 = x_3$, the order of x_2 and x_3 cannot be determined, which leads to the ambiguity of oriented bounding box representation.

In order to solve the problems of [12] and [13] and avoid the problems cased by predicting angle directly, in this article, we propose a new oriented bounding box representation method for arbitrary-oriented object detection by hiding the angle in the focal vector of an ellipse. Specifically, we predict the center point (x, y), the absolute value $(|u|, |v|)$ of focal vector (u, v), the difference (m) between long-axis length and length of focal vector, and a binary value (α) to indicate the orientation of object belongs to the first and third quadrants or the second and fourth quadrants. The definition of binary value (α) eliminates the corner cases in [13], which simplify the judgment of the oriented bounding box. We predict the sign (α) by regarding it as a classification task, therefore, there are totally seven parameters to describe the oriented bounding box, which is less than the parameters of BBAVetors [12]. Moreover, according to

the proposed method, we give a simpler and complete method to determine the long axis and short axis instead of the incomplete constraint on four vertices in [13]. The proposed ellipse parameters representation method makes an object have only one numerical representation, which can reduce the uncertainty of predicted parameters regression in the training phase. Finally, we summarize our contributions as follows.

1) We propose a new arbitrary-oriented object representation method to avoid problems caused by predicting angle directly, which hides the angle in a vector and makes the object have only one numerical representation.

2) We determine the long axis and short axis of oriented object by judging the lengths of two side vectors in oriented bounding box, which has no ambiguity to describe a oriented object instead of depending on the order of the four vertices like [13].

3) We employ the three-stage samples selection strategy to assign positive for ground truths. First, We adopt 2-D Gaussian distribution label assign [14] to generate the coarse positive samples. Second, we calculate Kullback–Leibler divergence (KLD) loss [15] between the coarse positive samples and ground truths. Finally, according to the obtained KLD losses, we use SimOTA [16] to choose optimal positive samples for each ground truth.

4) We extend the YOLOX [16] to detect the oriented object. To distinguish it from YOLOX, we named it arbitray-oriented ellipse detector (AEDet), which is a single stage and anchor free detector. Oure AEDet achieves the state-of-the-art results on HRSC2016 [17], RSDD-SAR [18], and RHRSID [19] datasets.

The rest of this article is organized as follows. The related works are introduced in Section II. A detailed description of the proposed method is presented in Section III. The results and analysis are presented in Section IV. Finally, Section V concludes this article.

II. RELATED WORK

A. Angle Prediction-Based Methods

The angle prediction-based methods are implemented by adding an extra angle parameter on the basis of the general vanilla detectors [3], [20], [21], [22]. In order to solve the inconsistency between classification and location tasks in object detection, Han et al. [23] proposed the S^2ANet, in which, the detection head is divided into two separate parts, one for predicting the object category, another for predicting the location and angle of object. To handle the angles of the objects, a new detector with newly defined rotatable bounding box (RBox) is proposed, which is denoted as DRBox [24]. The DRBox achieves the rotation invariant property to better learn the correct angle information of the objects. Ming et al. [25] proposed a critical feature capturing network by extracting the discriminative features, which can extract more discriminative features, generate more precise anchors, and make more reasonable label assignments. The periodicity and discontinuity of orientation angle regression lead the abnormal loss of detection networks in the training phase, therefore, Yang et al. [5], [9] treated the regression of angle as classification to overcome the periodicity and discontinuity of direct angle regression. The authors of [10], [15], and [26] optimized the arbitrary-oriented object detection networks from the perspective of designing better loss functions, which improve the detection performance efficiently.

B. Keypoints Regression-Based Methods

To address the imbalance between positive and negative samples, Yi et al. [12] extended the keypoint-based method [22] to suit arbitrary-oriented object detection task. Specifically, to obtain bounding boxes with arbitrary angles, Yi et al. detected the center of object first, then regress boundary-aware vectors of oriented bounding box. All the oriented objects are within Cartesian coordinates. Li et al. [27] adopted the adaptive points representation to extract the geometric information of object instances, which is used for aerial arbitrary-oriented object detection. Fu et al. [7] constructed a novel point-based arbitrary-oriented object detector capable of employing the spatial information explicitly, which uses a point-based representation to describe an oriented object, predict point location via fully convolutional network. Other keypoint-based arbitrary-oriented object detection methods are described in [28], [29], and [30], and they all achieve excellent detection performance.

C. Other Regression-Based Methods

To better represent an arbitrary-oriented object, Xu et al. [6] obtained the circumscribed horizontal bounding box from the four vertices of arbitrary-oriented object, then the offsets of four vertices on each corresponding side of horizontal bounding box are used to describe the arbitrary-oriented object. Xie et al. [31] proposed midpoint offset representation to describe the oriented objects, which regresses each oriented proposal by predicting its external rectangle and inferring its midpoint offset. In [32], Wu et al. proposed an arbitrary-oriented object detector ProjBB, which employs a six-parameter method to represent oriented bounding box. He et al. [13] described an oriented object via the center point, the offsets of the major semiaxis, the length of the minor semiaxis, and an orientation label. In [33], Zhou et al. adopted the polar coordinate which represents the oriented object for the first time, and the proposed that the polar remote sensing object detector reach the impressive detection accuracy. The authors of [34] and [35] also presented arbitrary-oriented objection methods with polar coordinate system.

III. PROPOSED METHOD

A. Ellipse Representation of Objects

The ellipse representation of an object is shown in Fig. 1, we can see that the object angle is decomposed into the two components (u, v) of the focal vector. Each component in (u, v) can be positive or negative, so we predict the absolute value $(|u|, |v|)$ of focal vector (u, v) by separating the orientation of the object from the focal vector. It is obvious that the separated orientation of an object can only be in the first and third quadrants or the second and fourth quadrants or coincide with the coordinate axis. When an object is in the first and third quadrants, as shown
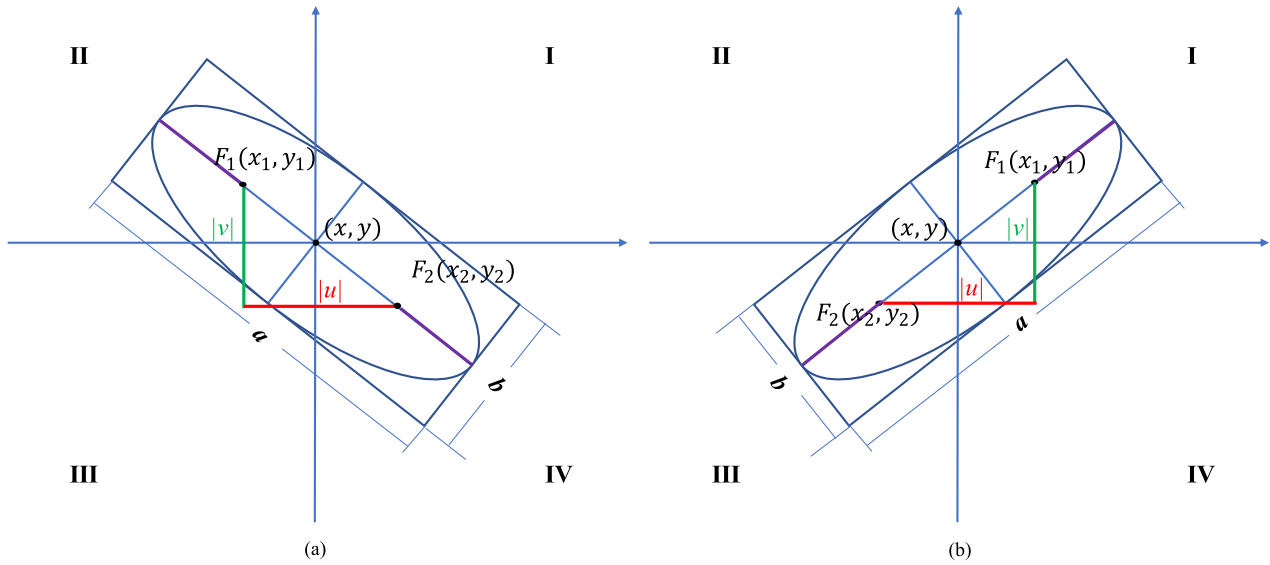
Fig. 1. Ellipse representation of an object. (a) Means the object orientation is within the second and fourth quadrants. (b) Means the object orientation is within the first and third quadrants.

in Fig. 1(a), both the horizontal component (u) and vertical component (v) are positive or negative at the same time. When an object is in the second and fourth quadrants, shown in Fig. 1(b), one of the horizontal component (u) and vertical component (v) is negative and another is positive. When the separated orientation of an object coincides with the coordinate axis, one of the $|u|$ and $|v|$ will be zero. The separated orientation $\alpha = 1$ means that the orientation of the object belongs to the first and third quadrants; $\alpha = 0$ means that the orientation of the object belongs to the second and fourth quadrants. For convenience, the separated orientation $\alpha = 1$ also means that the orientation of the object coincides with the coordinate axis. In order to represent an oriented object, the center of (x, y) is necessary. In addition, we only predict the parameters related the long axis, therefore, the difference (m) between length of long axis and length of focal vector is used as one of the key prediction parameters. Finally, we can predict the $(x, y, |u|, |v|, m, \alpha)$ to determine the location of an object.

B. Conversion Between Rectangle and Ellipse Representations

In this section, we introduce how to transform the four vertices (x_i, y_i) of oriented bounding box to the parameters $(x, y, |u|, |v|, m, \alpha)$, where $i=1, 2, 3, 4$. It should be noted that if the representation of the rectangle is (x, y, w, h, θ), then it needs to be converted to four vertices (x_i, y_i). In our method, we need to determine the long axis and short axis with no ambiguity. Therefore, we first obtain the center point $C(x, y)$ and the vectors $\overrightarrow{CD}$ and $\overrightarrow{CE}$ from center point to two adjacent vertices of four vertices, as shown in Fig. 2.

We can see from the Fig. 2 that when we know three points C, D, and E, we can obtain the two side vectors of rectangle

$$\overrightarrow{ED} = \overrightarrow{CD} - \overrightarrow{CE} \tag{1}$$

$$\overrightarrow{ME} = \overrightarrow{CD} + \overrightarrow{CE}. \tag{2}$$
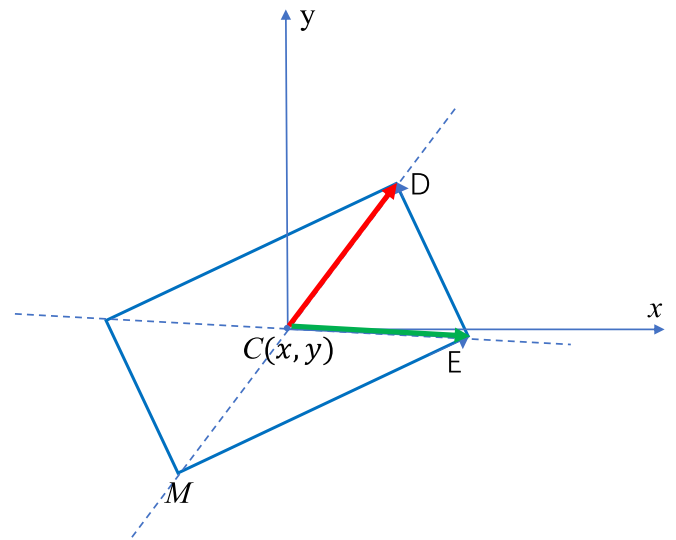

Fig. 2. Transforming the four vertices to center point and the vectors from center point to two adjacent vertices of four vertices.

After we obtain the two side vectors, we judge the length of the two side vectors and select the larger one as the long axis of the ellipse to calculate the ellipse parameters of the object. Our method is complete for determining the long axis and short axis of ellipse, unlike the incomplete constraints in [13]. The detailed algorithm to implement the conversion of four vertices to ellipse parameters is shown in Algorithm 1.

We can see from the Algorithm 1 that while the orientation of the object coincides with the coordinate axis, $\alpha = 1$. In addition, the method for determining the long axis and short axis is complete and no constraints are left out.

According to the Algorithm 1, we can transform the ellipse representation to rectangle representation, as shown in Algorithm 2.

Algorithm 1: Ellipse Parameters Encoding.

\# input: $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$
\# output: $(x, y, |u|, |v|, m, \alpha)$
\# Obtain the center and two adjacent side vectors according
\# to input
$C = (x, y)$
$\overrightarrow{CD} = (p, q)$
$\overrightarrow{CE} = (h, n)$
\# Obtain the two side vectors
$\overrightarrow{svec} = \overrightarrow{ED} = \overrightarrow{CD} - \overrightarrow{CE} = (p - h, q - n)$
$\overrightarrow{lvec} = \overrightarrow{ME} = \overrightarrow{CD} + \overrightarrow{CE} = (p + h, q + n)$
\# Obtain length of two side vector
$sls2 = \overrightarrow{svec}[0] * \overrightarrow{svec}[0] + \overrightarrow{svec}[1] * \overrightarrow{svec}[1]$
$sl = math.sqrt(sls2)$
$lls2 = \overrightarrow{lvec}[0] * \overrightarrow{lvec}[0] + \overrightarrow{lvec}[1] * \overrightarrow{lvec}[1]$
$ll = math.sqrt(lls2)$
\#Obtain the long-axis vector of ellipse
if $sl > ll$ **then**
    $\overrightarrow{vec} = \overrightarrow{svec}$
    $laxis = sl$
else:
    $\overrightarrow{vec} = \overrightarrow{lvec}$
    $laxis = ll$
end
\# Obtain the length of focal vector
$fl = math.sqrt(|sls2 - lls2|)$
\# Encoding the orientation of ellipse
$s1 = 1$
$s2 = 1$
if $\overrightarrow{vec}[0] < 0$ **then**
    $\overrightarrow{vec}[0] = -\overrightarrow{vec}[0]$
    $s1 = -1$
end
if $\overrightarrow{vec}[1] < 0$ **then**
    $\overrightarrow{vec}[1] = -\overrightarrow{vec}[1]$
    $s2 = -1$
end
$\alpha = 0$ if $s1 + s2 == 0$ else 1
\# Encoding $|u|$, $|v|$ and m
$ux = \overrightarrow{vec}[0]/laxis$
$uy = \overrightarrow{vec}[1]/laxis$
$|u| = fl * ux$
$|v| = fl * uy$
$m = laxis - fl$
\# Final ellipse parameters encoding
$[x, y, |u|, |v|, m, \alpha]$

Algorithm 2: Ellipse Parameters Decoding.

\# input: $(x, y, |u|, |v|, m, \alpha)$
\# output: $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$
\# step 1. Calculate the length of focal vector
$fl = math.sqrt(|u| * |u| + |v| * |v|)$
\# step 2. Calculate the unit vectors of long axis and
\# short axis
if $\alpha == 0$:
    $\overrightarrow{ulv} = (|u|/fl, -|v|/fl)$
if $\alpha == 1$:
    $\overrightarrow{ulv} = (|u|/fl, |v|/fl)$
$\overrightarrow{usv} = (\overrightarrow{ulv}[0], -\overrightarrow{ulv}[1])$
\# step 3. Calculate the vectors of long half axis and
\# short half axis.
$al = (m + fl)/2$
$bl = math.sqrt(al * al - (fl/2) * (fl/2))$
$\overrightarrow{av} = al * \overrightarrow{ulv}$
$\overrightarrow{bv} = bl * \overrightarrow{usv}$
\# step 4. Calculate four vertices of the rectangle
$x_1 = x + \overrightarrow{av}[0] + \overrightarrow{bv}[0], y_1 = y + \overrightarrow{av}[1] + \overrightarrow{bv}[1]$
$x_2 = x + \overrightarrow{av}[0] - \overrightarrow{bv}[0], y_2 = y + \overrightarrow{av}[1] - \overrightarrow{bv}[1]$
$x_3 = x - \overrightarrow{av}[0] - \overrightarrow{bv}[0], y_3 = y - \overrightarrow{av}[1] - \overrightarrow{bv}[1]$
$x_4 = x - \overrightarrow{av}[0] + \overrightarrow{bv}[0], y_4 = y - \overrightarrow{av}[1] + \overrightarrow{bv}[1]$

We can see from the Fig. 3 that the "multitask subnets" has three main branches, where "N" represents the number of categories; "O" is constant 2, which predicts the orientation of object (α_1, α_2), α_1 means long axis of the object is within the first and third quadrants, α_2 means long axis of the object is within the second and fourth quadrants; "E" is constant 5, which predicts the center point (x, y), the absolute values $(|u|, |v|)$ of focal vector components, the difference (m) between the long axis, and norm of focal vector; and "C" is constant 1, which represents the confidence of the object. Smooth L1 loss function is used for location regression of object. Focal loss is used for classification loss, confidence loss, and orientation loss.

D. Positive and Negative Samples Selection

Because the proposed AEDet is constructed by extending the YOLOX for arbitrary-oriented object detection, we need to redesign the way of choosing positive and negative samples. According the proposed ellipse parameters representation method, we adopt a three-stage strategy of 2-D Gaussian distribution label assign [14] + KLD loss [15] + SimOTA [16] method to determine the negative and positive samples. We, first, borrow the idea of 2-D Gaussian distribution label assign, but the difference from 2-D Gaussian distribution label assign is that we select all the positions in ellipse area as the coarse positive samples instead of generating the Gaussian candidate region according to a threshhold. Second, we calculate the KLD loss between the coarse positive samples and corresponding ground truths. Finally, according to the obtained KLD losses, we use SimOTA to assign the best positive samples for corresponding ground truth dynamically.

C. Object Detector With Ellipse Parameters

We apply the ellipse parameters representation method to the single stage and anchor-free object detection network YOLOX to construct an arbitrary-oriented object detector AEDet. We give an overview of our method with ellipse parameters prediction in Fig. 3. The feature pyramid network used in YOLOX is from [36], we denote it as PAFPN.
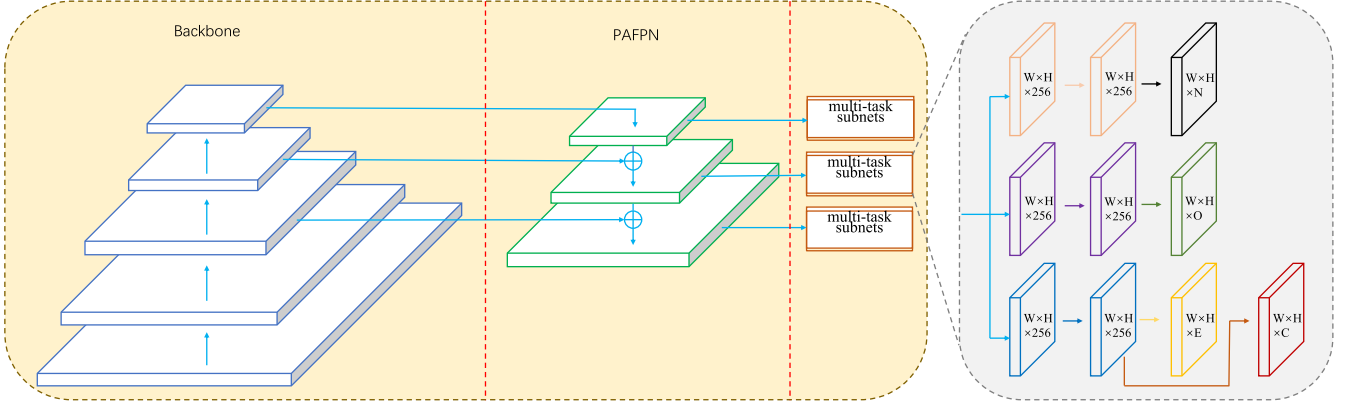
Fig. 3. Architecture of the proposed AEDet, where "N" for prediction of category, "O" for prediction of orientations, "E" for prediction of location, and "C" for prediction of confidence.

E. Bounding Box Regression

For a ground truth $(x, y, |u|, |v|, m, \alpha)$, we denote its corresponding positive sample as (x_a, y_a), the predicted box is denoted as $(x^*, y^*, |u^*|, |v^*|, m^*, \alpha^*)$. In general, the prediction parameters are nonzero, so that the more intuitive and simple bounding box regression is as follows:

$$b_x = \frac{x - x_a}{s} \tag{3}$$

$$b_y = \frac{y - y_a}{s} \tag{4}$$

$$b_u = \log(|u|/s + \text{eps}) \tag{5}$$

$$b_v = \log(|v|/s + \text{eps}) \tag{6}$$

$$b_m = \log(|m|/s + \text{eps}) \tag{7}$$

$$b_x^* = \frac{x^* - x_a}{s} \tag{8}$$

$$b_y^* = \frac{y^* - y_a}{s} \tag{9}$$

$$b_u^* = \log(|u^*|/s + \text{eps}) \tag{10}$$

$$b_v^* = \log(|v^*|/s + \text{eps}) \tag{11}$$

$$b_m^* = \log(|m^*|/s + \text{eps}) \tag{12}$$

$$b_x = \frac{x - x_a}{s} \tag{13}$$

$$b_y = \frac{y - y_a}{s} \tag{14}$$

$$b_u = \log((|u| + 1)/s) \tag{15}$$

$$b_v = \log((|v| + 1)/s) \tag{16}$$

$$b_m = \log((|m| + 1)/s) \tag{17}$$

$$b_x^* = \frac{x^* - x_a}{s} \tag{18}$$

$$b_y^* = \frac{y^* - y_a}{s} \tag{19}$$

$$b_u^* = \log((|u^*| + 1)/s) \tag{20}$$

$$b_v^* = \log((|v^*| + 1)/s) \tag{21}$$

$$b_m^* = \log((|m^*| + 1)/s). \tag{22}$$

Where the "eps" is a very small number. We can see from Algorithm 1 that one or both of $|u|$ and $|v|$ in ground truth $(x, y, |u|, |v|, m, \alpha)$ will be zero, which makes it difficult to detect the object with horizontal bounding box in training phase. The essence of the above phenomenon is that the training loss converge slows due to the large difference in the loss value of the predictor variables in the regression loss function. For solving this problem, we modify the way of bounding box regression to the following formula.

IV. EXPERIMENTS AND ANALYSIS

A. Datasets

HRSC2016 Dataset: The HRSC2016 includes 1061 images, in which, 436 images are divided as training set, 181 images for validation set, and 444 for testing set. In the training phase, we combine the training set and validation set for training.

RSDD-SAR Dataset: The RSDD-SAR dataset consists of 7000 images with size 512×512, there are a total of 10 263 ship instances in the dataset. 5000 images are divided into training dataset and 2000 images are divided into test dataset. In the test dataset, 159 images belong to inshore dataset and 1841 images belong to offshore dataset.

RHRSID Dataset: The rotation annotations of RHRSID are obtained by taking the minimum circumscribed rectangle of instance segmentation annotations. There are a total of 5604 SAR images in the HRSID dataset, in the 1962 images of test dataset, 369 images belong to inshore dataset, and 1593 images belong to offshore dataset.

B. Experimental Settings

Because the proposed AEDet is constructed based on the YOLOX with medium parameters, most of the hyperparameters are same with YOLOX. In the phase of training, we disable the multiscale training; the learning rate per image is 0.0003125 and the beta value in SmoothL1 loss is 0.2. The epochs and batch

TABLE I
COMPARISON OF RESULTS ON HRSC2016 DATASET

| Methods	Backbone	Size	mAP(07)	mAP(12)
Two-stage method				
Rotated RPN [38]	ResNet-101	800×800	79.08	85.64
R2CNN [39]	ResNet-101	800×800	73.07	79.73
RoI transformer [40]	ResNet-101	512×800	86.20	-
Gliding vertex [6]	ResNet-101	512×800	88.20	-
CenterMap-Net [41]	ResNet-50	-	-	92.80
Oriented R-CNN [31]	ResNet-50	-	90.40	96.50
One-stage method				
R3Det [42]	ResNet-101	800×800	89.26	96.01
One DAL [8]	ResNet-101	800×800	89.77	-
S^2ANet [23]	ResNet-101	512×800	90.17	95.01
RRD [43]	VGG16	384×384	84.30	-
RetinaNet-O [3]	ResNet-101	800×800	89.18	95.21
FPN-CSL [5]	ResNet-101	-	89.62	96.10
Anchor-free method				
PIoU [26]	DLA-34	512×512	89.20	-
DRN [44]	Hourglass-34	768×768	-	92.70
CenterNet-O [22]	DLA-34	-	87.89	-
AEDet	CSPDarknet-53	800×800	**90.45**	**96.90**
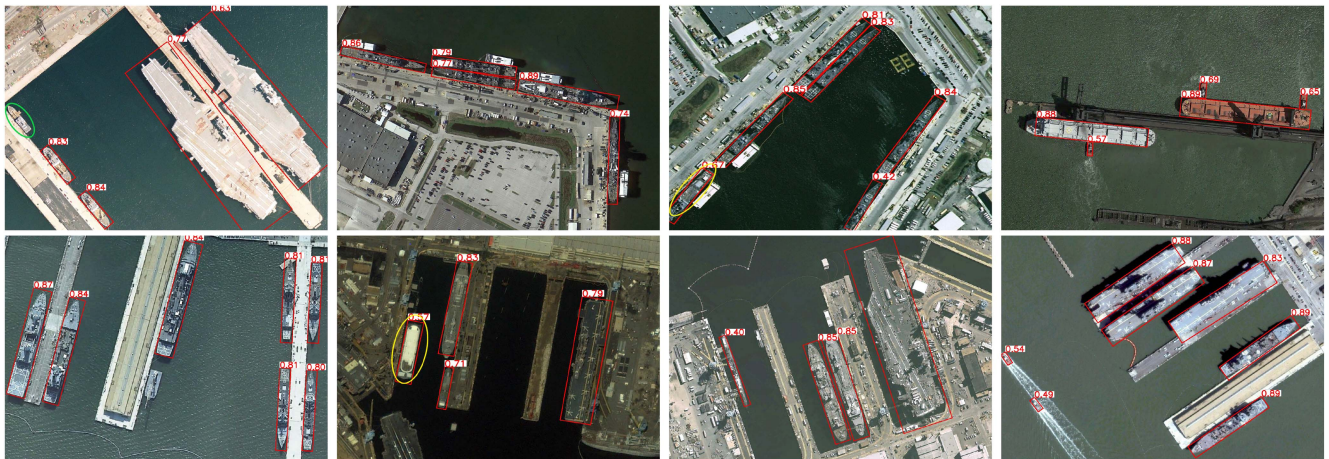
Fig. 4.   Visual results on HRSC2016 using AEDet with CSPDarknet-53 backbone, where the red oriented boxes represent the predicted boxes, the yellow oriented ellipses are the false detections, and the green oriented ellipses are the missed detections.

size are 36 and 16, respectively, for three datasets. In the testing phase, the confidence threshold is 0.05 and the NMS threshold is 0.1. In order to make the AEDet training converge quickly, we use the pretraining weights provided by the authors of YOLOX. We use two NVIDIA TITAN Xp GPUs in all experiments.

C. Experimental Results on HRSC2016 Dataset

We simultaneously evaluate our AEDet with two different versions (VOC2007 and VOC2012 metrics) of the VOC evaluation metric [37]. As shown in Table I, We list the results of AEDet and some other algorithms, where the proposed AEDet achieves the best results of 90.45% and 96.9% in VOC2007 and VOC2012 mAP metrics, respectively. Experimental results show that our method can achieve competitive results and outperform most state-of-the-art algorithms, which demonstrates the effectiveness of the proposed ellipse parameters representation method.

We also show some visual results, as shown in Fig. 4. It should be noted that the objects in the HRSC2016 dataset have large

aspect ratios, which is not friendly to the object detectors using angle regression, because a slight angle change will make a huge change in regression loss. However, we can see the results from the Table I and Fig. 4 that the localization of bounding boxes is fairly accurate, especially on objects with large aspect ratios, which means our ellipse parameters representation method can effectively address the shortcomings of direct angle regression methods by avoiding to predict angles directly.

D. Experimental Results on RSDD-SAR Dataset

In addition to dividing the RSDD-SAR dataset into training and test sets, where the test set is further divided into inshore set and offshore set. Therefore, we give the experimental results about the the test set, inshore set, and offshore set in the evaluation phase. We use the code from MMRotate [45] to generate the results of comparative methods. As shown in Table II, our AEDet gets the best results on the test set, inshore set, and offshore set on VOC 2007 mAP metric. Especially in the inshore scene, we obtain the 77.8% mAP, which exceeds the

TABLE II
COMPARISON OF RESULTS ON RSDD-SAR DATASET

| Methods	Backbone	Size	Inshore(mAP)	Offshore(mAP)	All(mAP)
Two-stage method					
Faster RCNN-O [21]	ResNet-50	512×512	62.2	89.9	87.7
	ResNet-101	512×512	64.6	90.1	88.5
Oriented RCNN [31]	ResNet-50	512×512	34.4	81.3	72.1
	ResNet-101	512×512	34.8	81.3	72.1
RoI transformer [40]	ResNet-50	512×512	75.6	90.5	89.7
	ResNet-101	512×512	75.0	90.5	89.7
Gliding vertex [6]	ResNet-50	512×512	64.6	90.5	88.8
	ResNet-101	512×512	65.9	90.3	88.9
One-stage method					
S^2ANet [23]	ResNet-50	512×512	67.8	90.5	89.9
	ResNet-101	512×512	68.9	90.4	89.7
RetinaNet-O [3]	ResNet-50	512×512	63.4	88.9	87.2
	ResNet-101	512×512	65.2	88.9	87.3
R3Det [42]	ResNet-50	512×512	55.0	89.6	80.3
	ResNet-101	512×512	65.7	90.2	88.2
Anchor-free method					
FCOS-O [46]	ResNet-50	512×512	55.4	90.1	80.6
	ResNet-101	512×512	66.0	90.1	88.7
Oriented reppoints [27]	ResNet-50	512×512	74.5	90.3	89.7
	ResNet-101	512×512	74.2	90.2	89.5
AEDet	CSPDarknet-53	512×512	**77.8**	**90.5**	**90.1**
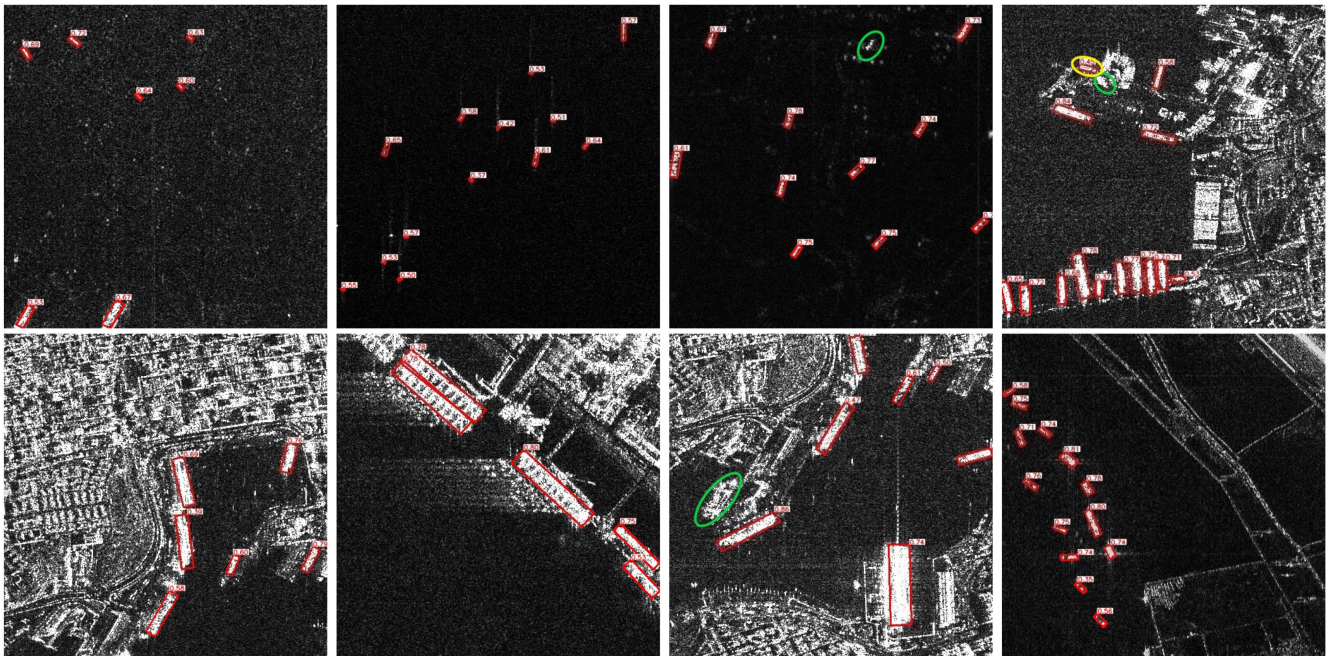
Fig. 5. Visual results on RSDD-SAR using AEDet with CSPDarknet-53 backbone, where the red oriented boxes represent the predicted boxes, the yellow oriented ellipses are the false detections, and the green oriented ellipses are the missed detections.

second-best result by 2.2%. The experimental results of AEDet on RSDD-SAR show the competitive performance compared with other algorithms, which highlights the superiority of our ellipse parameters representation method.

We show some examples of detection results in Fig. 5. In the complex inshore regions, AEDet has very few false detections and missed detections, which means the proposed AEDet can distinguish the ship objects from the complicated background well. In the simple offshore regions, AEDet has great detection performance unsurprisingly. In addition, it is worth noting that the objects in the RSDD-SAR dataset is relatively small compared to HRSC2016 dataset, which means that our ellipse

parameters representation method can describe objects of different scales efficiently.

E. Experimental Results on RHRSID Dataset

We obtain comparative experimental results about RHRSID dataset by using the code from MMRotate [45]. The results of test set, inshore set, and offshore set are shown in Table III, where the AEDet achieves the best mAP of 88.2% on test dataset, which outperforms other comparison methods. In the inshore dataset, we achieve the best mAP of 76.5%, which is far more than the second-best result. In the offshore dataset, we also

TABLE III
COMPARISON OF RESULTS ON RHRSID DATASET

| Methods	Backbone	Size	Inshore(mAP)	Offshore(mAP)	All(mAP)
Two-stage method					
Faster RCNN-O [21]	ResNet-50	800×800	52.7	89.8	76.4
	ResNet-101	800×800	51.2	90.3	76.1
Oriented RCNN [31]	ResNet-50	800×800	41.4	90.2	62.7
	ResNet-101	800×800	41.6	90.2	62.7
RoI transformer [40]	ResNet-50	800×800	66.4	90.8	80.5
	ResNet-101	800×800	65.7	90.8	80.3
Gliding vertex [6]	ResNet-50	800×800	57.5	90.6	78.6
	ResNet-101	800×800	55.8	90.6	71.3
One-stage method					
S^2ANet [23]	ResNet-50	800×800	67.7	90.8	80.8
	ResNet-101	800×800	66.7	90.8	80.6
RetinaNet-O [3]	ResNet-50	800×800	52.8	89.9	76.3
	ResNet-101	800×800	49.8	90.0	75.9
R3Det [42]	ResNet-50	800×800	58.1	90.8	79.3
	ResNet-101	800×800	56.0	90.8	78.2
Anchor-free method					
FCOS-O [46]	ResNet-50	800×800	58.6	90.5	78.4
	ResNet-101	800×800	52.5	90.6	77.2
Orientd reppoints [27]	ResNet-50	800×800	70.2	90.7	84.7
	ResNet-101	800×800	72.8	90.7	86.3
AEDet	CSPDarknet-53	800×800	**76.5**	**90.8**	**88.2**
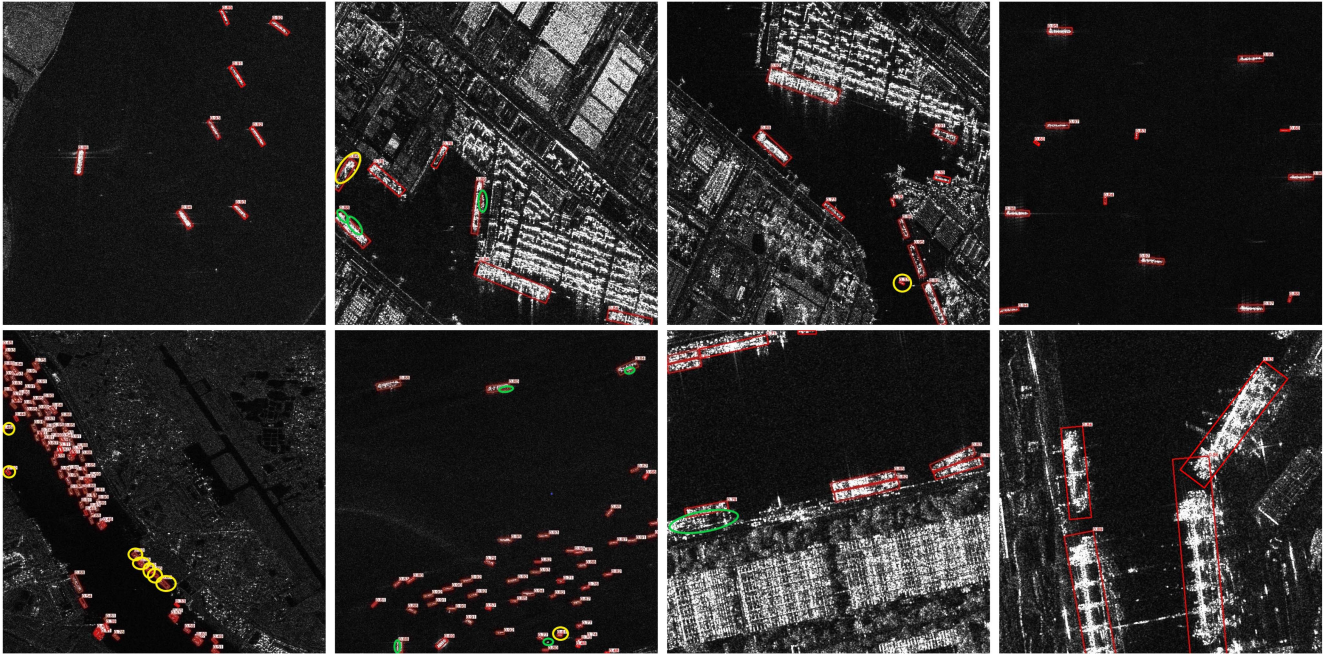
Fig. 6. Visual results on RHRSID using AEDet with CSPDarknet-53 backbone. Where the red oriented boxes represent the predicted boxes, the yellow oriented ellipses are the false detections, and the green oriented ellipses are the missed detections.

get the best result. The results on inshore datasets show that AEDet is extremely advantageous in detecting inshore object, this conclusion is also consistent with the results in RSDD-SAR. We also show some detection examples of AEDet in Fig. 6, in which, the proposed AEDet shows the remarkable capability on detecting the inshore ship objects.

F. Effect of Bounding Box Regression

As described in the previous section, when the orientation of the object coincides with the coordinate axis, one of $|u|$ and $|v|$ in ground truth $(x, y, |u|, |v|, m, \alpha)$ will be zero, which makes the formulas (3) to (12) and (13) to (22) have different performance according to theoretical analysis. In this section, we verify theoretical analysis by conducting the experiments on two datasets, one of two datasets is RHRSID. Another is STORAGETANK dataset, which is constructed by choosing the storage tank from the RDIOR [47] dataset; some example images are shown in Fig. 7. We can see that the bounding boxes of storage tanks is horizontal, which are suitable as a dataset to verify the effects on different ways of Bounding box regression.

In the experimental settings of STORAGETANK dataset, the epoch is set to 36, and the β in smooth L1 is set to 0.2, and the image size is 800×800. The other hyperparameters are
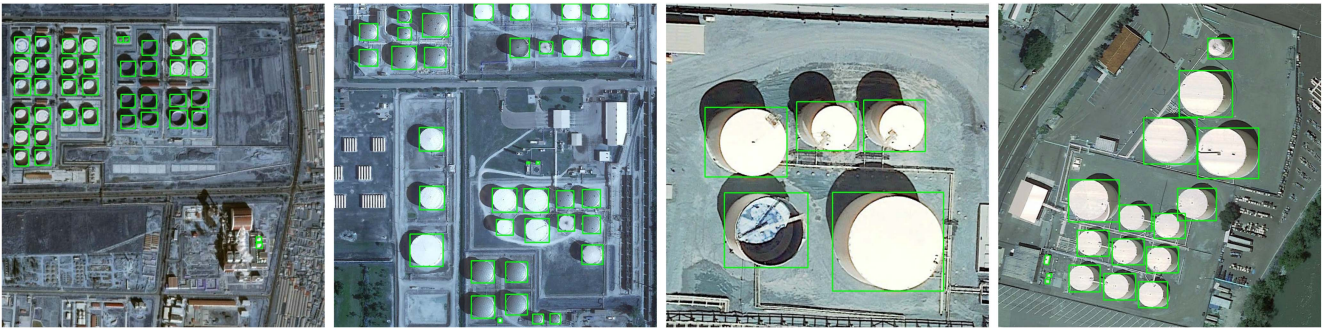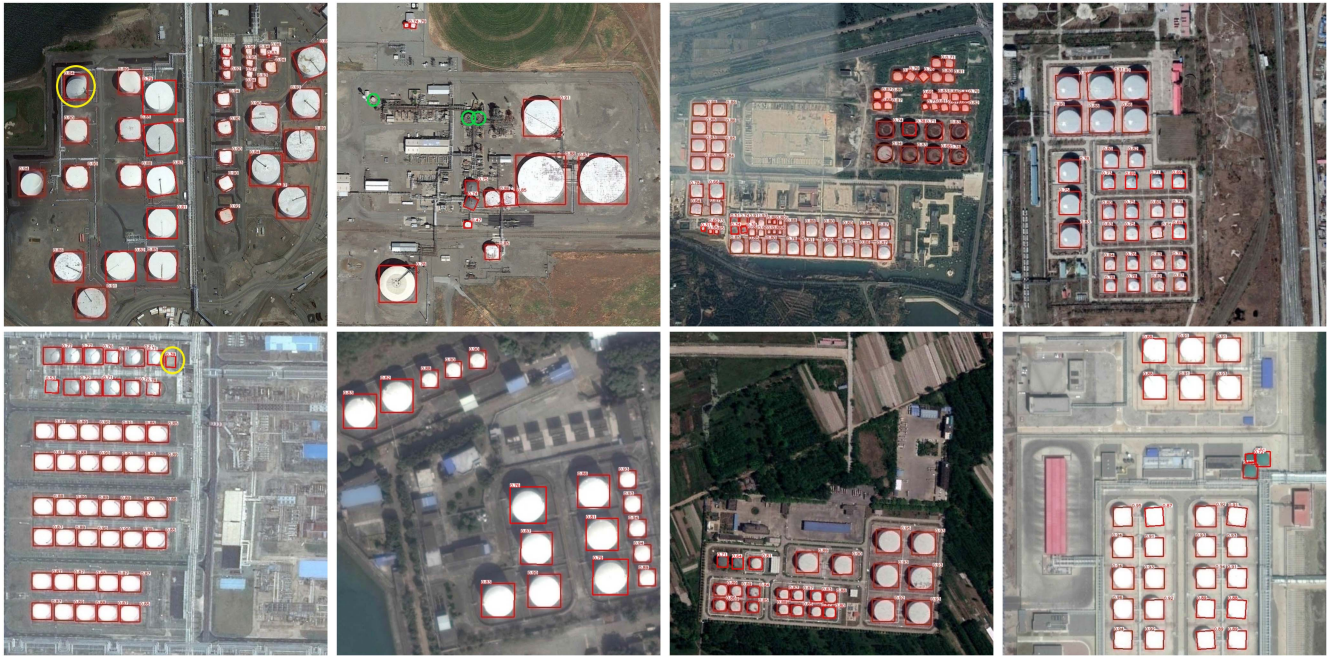
Fig. 7. Some examples of STORAGETANK dataset.

Fig. 8. Visual results on STORAGETANK using AEDet with CSPDarknet-53 backbone. Where the red oriented boxes represent the predicted boxes, the yellow oriented ellipses are the false detections, and the green oriented ellipses are the missed detections.

TABLE IV
EXPERIMENTAL RESULTS ON DIFFERENT WAYS OF BOUNDING BOX REGRESSION

Datasets	AED_EPS	AEDet
RHRSID	80.18	88.20(+**8.02**)
STORAGETANK	59.19	76.36(+**17.17**)

TABLE V
EXPERIMENTAL RESULTS ON DIFFERENT VECTORS-BASED REPRESENTATION METHODS

Methods	HRSC2016
BBAVectors [12]	88.60
RIEDet [13]	86.15
AEDet	**90.45**

same with the settings of experiments on other datasets used in this article. The results are shown in Table IV, where the AEDet_EPS represents the bounding box regression of (3) to (12), and AEDet means the bounding box regression of (13) to (22). From the Table IV, it is obvious that the proposed AEDet has huge improvements on two datasets compared with AEDet_EPS, which demonstrate the bounding box regression of (13) to (22) is much better than the bounding box regression of (3) to (12). The results in Table IV verify the theoretical analysis about different ways of bounding box regression. We visualize the results of AEDet on STORAGETANK dataset in Fig. 8, which shows the proposed AEDet can handle well the case where the bounding box is horizontal, without needing to do it separately like [12] does.

G. Comparison of Related Work Results

In order to facilitate comparison with BBAVectors [5] and RIEDet [13], we conduct experiments on HRSC2016. As shown in Table V, AEDet gets the best result in the case of only replacing the representation method of bounding box, which verifies the effectiveness of the proposed oriented representation method.

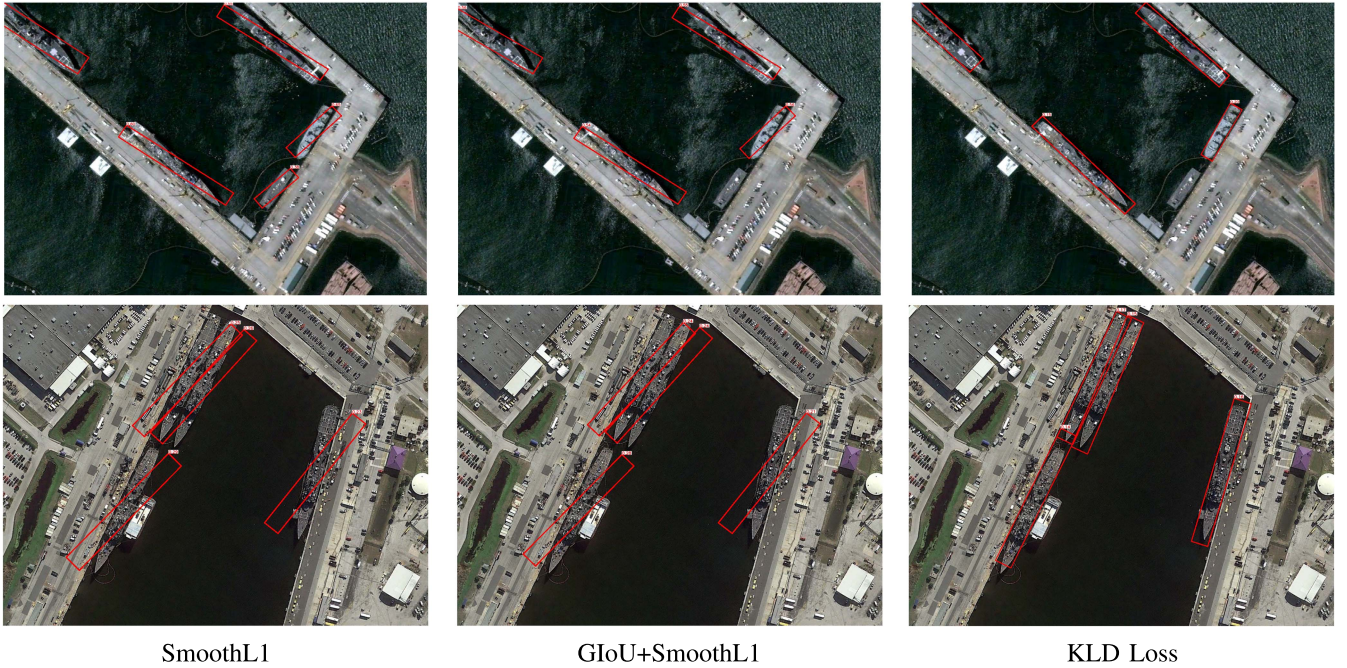| SmoothL1 | GIoU+SmoothL1 | KLD Loss |

Fig. 9. Examples of detection results using RODet with different loss functions.

TABLE VI
COMPARISON OF RESULTS BETWEEN THE PROPOSED AEDET AND
ANGLE-BASED METHOD

Methods	Loss Functions	HRSC2016
RODet	SmoothL1	44.30
	GIoU+SmoothL1	43.96
	KLD Loss	81.85
AEDet	SmoothL1	**90.45**

H. Comparison With Angle-Based Method

In order to highlight the superiority of the proposed ellipse parameters representation method over the method of directly predicting the angle without any constraints, we modify the proposed AEDet by only replacing the object representation method $(x, y, |u|, |v|, m, \alpha)$ of AEDet with the object representation method (x, y, w, h, θ). For convenience, we name the new detector with object representation (x, y, w, h, θ) as RODet. We conduct experiments on the HRSC2016 dataset. As shown in Table VI, where the loss function of RODet is SmoothL1 represents that the object representation vector (x, y, w, h, θ) is taken as a whole as input of SmoothL1 function. GIoU + SmoothL1 in RODet means that the vector (x, y, w, h) is taken as whole as input of GIoU function, and angle(θ) is the input of SmoothL1 alone. KLD loss in RODet is used to test whether there is a problem with the object representation method (x, y, w, h, θ) we implemented or not. We show some visual results of RODet according to loss functions in Fig. 9.

We can see from the Table VI and Fig. 9 that RODet with SmoothL1 and GIoU+SmoothL1 has poor detection results, but RODet with KLD loss has great performance, which proves that RODet with the object representation method (x, y, w, h, θ) we implemented is correct, and the unsatisfactory detection performance of RODet with SmoothL1 and GIoU+SmoothL1

is caused by the angle periodicity. The visual results in Fig. 9 show that RODet with SmoothL1 and GIoU+SmoothL1 can accurately predict (x, y, w, h) of object, but cannot accurately predict the angle (θ). The above phenomenon indicates that predicting angle of object directly without any constraints makes the detection performance worse. On the other hand, the proposed AEDet with ellipse parameters representation method achieves excellent result of 90.45% without any constraints, which highlights the superiority of the ellipse parameters representation method in this article.

V. CONCLUSION

In this article, in order to solve the problem of abnormal bounding box regression loss of object with multiple numerical representations caused by angle periodicity, we propose an ellipse parameters representation method to describe arbitrary-oriented objects, which has only one numerical representation for an object and can predict object by avoiding to predict angels directly. Then we propose the AEDet by applying the proposed ellipse parameters representation method to YOLOX with medium parameters. According to the proposed ellipse parameters representation method, we propose corresponding positive samples selection method and way of bounding box regression. We conduct all experiments on three arbitrary-oriented ship detection datasets: 1) HRSC2016, 2) RSDD-SAR, and 3) RHRSID. The proposed AEDet achieves the results of 90.45% mAP on HRSC2016, 90.10% mAP on RSDD-SAR, and 88.2% mAP on RHRSID, respectively. We have achieved better detection results compared to related methods, which shows the effectiveness and superiority of the AEDet and our ellipse parameters representation method.

References

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

[2] J. Redmon and A. Farhadi, "YOLOV3: An incremental improvement," 2018, *arXiv:1804.02767*.

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

[4] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.

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

[6] 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.

[7] K. Fu, Z. Chang, Y. Zhang, and X. Sun, "Point-based estimator for arbitrary-oriented object detection in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4370–4387, May 2021.

[8] Q. Ming, Z. Zhou, L. Miao, H. Zhang, and L. Li, "Dynamic anchor learning for arbitrary-oriented object detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 2355–2363.

[9] 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.*, 2021, pp. 15819–15829.

[10] 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.

[11] W. Qian, X. Yang, S. Peng, J. Yan, and Y. Guo, "Learning modulated loss for rotated object detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 2458–2466.

[12] 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/CVF Winter Conf. Appl. Comput. Vis.*, 2021, pp. 2150–2159.

[13] X. He, S. Ma, L. He, L. Ru, and C. Wang, "Learning rotated inscribed ellipse for oriented object detection in remote sensing images," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3622.

[14] Z. Huang, W. Li, X.-G. Xia, and R. Tao, "A general Gaussian heatmap label assignment for arbitrary-oriented object detection," *IEEE Trans. Image Process.*, vol. 31, pp. 1895–1910, 2022.

[15] X. Yang et al., "Learning high-precision bounding box for rotated object detection via Kullback–Leibler divergence," *Adv Neur In*, vol. 34, pp. 18381–18394, 2021.

[16] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[17] Z. Liu, H. Wang, L. Weng, and Y. Yang, "Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1074–1078, Aug. 2016.

[18] C. Xu et al., "RSDD-SAR: Rotated ship detection dataset in SAR images," *J. Radars*, vol. 11, no. 4, pp. 581–599, 2022.

[19] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.

[20] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst. Vol.*, 2015, pp. 91–99.

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

[23] J. Han, J. Ding, J. Li, and G.-S. Xia, "Align deep features for oriented object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5602511.

[24] 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.

[25] Q. Ming, L. Miao, Z. Zhou, and Y. Dong, "CFC-Net: A critical feature capturing network for arbitrary-oriented object detection in remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5605814.

[26] Z. Chen et al., "PIoU loss: Towards accurate oriented object detection in complex environments," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 195–211.

[27] W. Li, Y. Chen, K. Hu, and J. Zhu, "Oriented reppoints for aerial object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1829–1838.

[28] F. Zhang, X. Wang, S. Zhou, Y. Wang, and Y. Hou, "Arbitrary-oriented ship detection through center-head point extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5612414.

[29] Y. Sun, X. Sun, Z. Wang, and K. Fu, "Oriented ship detection based on strong scattering points network in large-scale SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5218018.

[30] Z. Guo, C. Liu, X. Zhang, J. Jiao, X. Ji, and Q. Ye, "Beyond bounding-box: Convex-hull feature adaptation for oriented and densely packed object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8792–8801.

[31] 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.*, 2021, pp. 3520–3529.

[32] Q. Wu, W. Xiang, R. Tang, and J. Zhu, "Bounding box projection for regression uncertainty in oriented object detection," *IEEE Access*, vol. 9, pp. 58768–58779, 2021.

[33] L. Zhou, H. Wei, H. Li, W. Zhao, and Y. Zhang, "Objects detection for remote sensing images based on polar coordinates," 2020, *arXiv:2001.02988*.

[34] 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.

[35] 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, 2021.

[36] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8759–8768.

[37] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.

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

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

[40] 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.*, 2019, pp. 2849–2858.

[41] J. Wang, W. Yang, H.-C. Li, H. Zhang, and G.-S. Xia, "Learning center probability map for detecting objects in aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4307–4323, May 2021.

[42] 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.*, 2021, pp. 3163–3171.

[43] M. Liao, Z. Zhu, B. Shi, G.-S. Xia, and X. Bai, "Rotation-sensitive regression for oriented scene text detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5909–5918.

[44] X. Pan et al., "Dynamic refinement network for oriented and densely packed object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11207–11216.

[45] Y. Zhou et al., "MMRotate: A rotated object detection benchmark using PyTorch," 2022, *arXiv:2204.13317*.

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

[47] G. Cheng et al., "Anchor-free oriented proposal generator for object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5625411.

Kexue Zhou received the B.E. degree in measurement and control technology and instruments from Xidian University, Xi'an, China, in 2018. He is currently working toward the Ph.D. degree in instrument science and technology with School of Aerospace Science and Technology, Xidian University, Xi'an, China.

His research interests include computer vision and machine learning, especially for object detection in remote sensing.

Min Zhang received the B.E. and Ph.D. degrees in measurement and control from Xidian University, Xi'an, China, in 2010 and 2017, respectively.

She is currently an Associate Professor with the School of Aerospace Science and Technology, Xidian University. Her research interests include time and frequency measurements, digital signal processing, remote sensing images processing, and hyperspectral images processing.

Sheng Lin recevied the B.E. degree in electrical engineering and automation and the M.S. degree in electrical engineering from the Zhengzhou University of Light Industry, Zhengzhou, China, in 2017 and 2020, respectively. He is currently working toward the Ph.D. degree in instrument science and technology with the School of Aerospace Science and Technology, Xidian University, Xi'an, China.

His research interests include hyperspectral anomaly detection and object detection.

Honghui Zhao received the B.E. degree in geographic information system from Xi'an University of Science and Technology, Xi'an, China, in 2006.

He has been engaged in remote sensing software design and development for a long time. He has worked in large foreign companies, such as ESRI in Beijing, China. He has rich experience in GIS and remote sensing software development.

Xi Cheng received the M.S. degree in control theory and control engineering from the Zhengzhou University of Light Industry, Zhengzhou, China, in 2020. He is currently working toward the Ph.D. degree in instrument science and technology with the School of Aerospace Science and Technology, Xidian University, Xi'an, China.

His research interests include hyperspectral anomaly detection and saliency object detection.

Rui Tang received the B.E. degree in remote sensing science and technology from the Chang An University, Xi'an, China, in 2010.

His research interests include object detection, tracking, image registration, image mosaic, image segmentation, and multiple view geometry.

Hai Wang received the M.S. degree in communication and information system and the Ph.D. degree in instrument science and technology from the Xidian University, Xi'an, China, in 2003 and 2007, respectively.

He is currently a Professor with the School of Aerospace Science and Technology, Xidian University, Xi'an, China. His research interests include the time and frequency measurements, frequency and phase measurements methods based on field-programmable gates array and system-on-chip, electronic instrument design, time–frequency signal processing, object detection, and hyperspectral images processing.