Shape-Adaptive Selection and Measurement for Oriented Object Detection

Liping Hou,[1] **Ke Lu,**[1,2] **Jian Xue**[1*]**, Yuqiu Li** [1]

[1] University of Chinese Academy of Sciences, Beijing 100049, China
[2] Peng Cheng Laboratory, Shenzhen 518055, China
{houliping17, liyuqiu20}@mails.ucas.ac.cn, {luk, xuejian}@ucas.ac.cn

Abstract

The development of detection methods for oriented object detection remains a challenging task. A considerable obstacle is the wide variation in the shape (e.g., aspect ratio) of objects. Sample selection in general object detection has been widely studied as it plays a crucial role in the performance of the detection method and has achieved great progress. However, existing sample selection strategies still overlook some issues: (1) most of them ignore the object shape information; (2) they do not make a potential distinction between selected positive samples; and (3) some of them can only be applied to either anchor-free or anchor-based methods and cannot be used for both of them simultaneously. In this paper, we propose novel flexible shape-adaptive selection (SA-S) and shape-adaptive measurement (SA-M) strategies for oriented object detection, which comprise an SA-S strategy for sample selection and SA-M strategy for the quality estimation of positive samples. Specifically, the SA-S strategy dynamically selects samples according to the shape information and characteristics distribution of objects. The SA-M strategy measures the localization potential and adds quality information on the selected positive samples. The experimental results on both anchor-free and anchor-based baselines and four publicly available oriented datasets (DOTA, HRSC2016, UCAS-AOD, and ICDAR2015) demonstrate the effectiveness of the proposed method.

Introduction

The detection of arbitrarily oriented objects is a fundamental yet challenging task in computer vision, and can be applied in a wide range of scenarios, such as remote sensing (RS) images and text scenes. Because using horizontal bounding boxes (HBBs) cannot accurately calibrate the position of arbitrary-oriented objects and can also easily be affected by non-maximum suppression (NMS), using oriented bounding boxes (OBBs) has become a popular positioning approach and has made significant progress. Existing oriented object methods based on a deep convolutional neural network (CNN) mostly focus on the problems of the huge scale variations of objects and complex backgrounds but pay little attention to the large aspect ratio variations of objects.

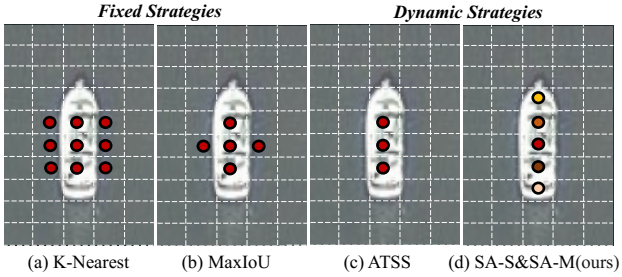

Figure 1: Different strategies for the selection and measurement of positive samples. Circles represent the positive samples. (a) K-Nearest uses a static center distribution for sampling. (b) MaxIoU uses a fixed IoU threshold for sampling. (c) ATSS assigns a dynamic IoU threshold for an object. (d) The SA-S strategy uses a dynamic IoU threshold, and then the SA-M strategy adds quality information for selected positive samples, which is indicated by different colors.

As reported in ATSS (Zhang et al. 2020), the selection of positive and negative samples plays a critical role in detection performance, and is also the essential difference between anchor-based and anchor-free methods. Existing sample selection strategies are mainly divided into fixed and dynamic strategies. Although the simplicity and intuitiveness of fixed label assignment strategies make them a popular choice, they ignore the actual shape and content of the intersecting region of objects, particularly for oriented object detection. As shown in Fig. 1 (a) and (b), anchor-based methods (e.g., RetinaNet (Lin et al. 2017b)) use the max intersection over union (IoU) value (MaxIoU, for simplicity) between proposals and objects, and anchor-free method (e.g., FCOS (Tian et al. 2020)) uses k-nearest distance (K-Nearest for simplicity) between a point and the object's center for sample selection. The developers of ATSS proposed a sample selection strategy using dynamic IoU thresholds, which is shown in Fig. 1 (c). In other recent works, researchers have proposed dynamic sample selection or anchor learning strategies (Yang et al. 2018a; Kim and Lee 2020; Zhang et al. 2020). Although these strategies are more efficient than fixed assignment strategies, they have the following problems: (1) the oriented object shape information (e.g., large aspect ratio) is overlooked; (2) selected positive

*Corresponding author.

samples are processed in a uniform manner without considering their quality; and (3) there are limited applications to insertable architectures, for example, some dynamic anchor learning or selection strategies cannot be used for anchor-free architectures. Therefore, to avoid the aforementioned problems and further optimize the entire process, two shape-adaptive strategies for arbitrary-oriented object detection are proposed in this study for dynamically selecting samples and evaluating the quality of positive samples.

Specifically, for the purpose of adaptively selecting samples, the shape-adaptive selection (SA-S) strategy is proposed, which uses the object shape information effectively, and the object shape information is focused on the aspect ratio, which is calculated as the ratio of the long edge to the short edge in this study. The proposed SA-S strategy is designed to calculate optimal IoU thresholds for the objects with different shapes, so it is suitable for both anchor-based and anchor-free methods which adopt IoU threshold to assign labels.

Furthermore, considering that the selected positive samples have different qualities and potentials, a shape-adaptive measurement (SA-M) strategy is designed to add quality information to them. The SA-M strategy measures the positive sample's quality using a new concept, that is, the normalized shape distance, which combines the center and shape of the object to calculate the distance of the sample point relative to the object. Additionally, a boundary-center loss function is elaborated based on an anchor-free architecture for keypoint learning. The main contributions of this study are as follows.

1. A novel dynamic SA-S strategy is proposed, which selects positive samples according to the shape and characteristics distribution of objects.
2. A new SA-M strategy is proposed, which evaluates the quality of the selected positive samples. Additionally, the new concept of the normalized shape distance is designed, which eliminates the effect of the object shape on the estimation of relative object distances.
3. Sufficient experiments were conducted to prove that the proposed dynamic sample selection and measurement strategies can be embedded into both anchor-free and anchor-based methods to achieve significant improvements in detection performance.

The experimental results demonstrated that the proposed method was superior to other state-of-the-art methods on the benchmark datasets DOTA (Xia et al. 2018), UCAS-AOD (Zhu et al. 2015), HRSC2016 dataset (Liu et al. 2017), and ICDAR2015 (Karatzas et al. 2015).

Related Work

Oriented Object Detection

Representation of object in object detection has been dominated by HBBs for several years, whose cornerstone is the horizontal anchor (Ren et al. 2015; Lin et al. 2017b). With the growing demand for the detection of objects with arbitrary orientation, such as text and targets in remote sensing scenes, oriented object detection methods (Yang et al. 2018b, 2021c,d) have attracted much attention. There are five types of detection methods for oriented object detection: (1) generating oriented region proposals directly (Azimi et al. 2018; Ding et al. 2019); (2) regressing the angle parameters based on horizontal region proposals (Yang et al. 2019a; Zhang, Lu, and Zhang 2019; Yang et al. 2020b); (3) using the mask prediction of the mask branch to locate the object area (Li et al. 2020); (4) regressing the angle parameters (Yang et al. 2021b; Han et al. 2021); and (5) predicting angles using a classification method (Yang and Yan 2020; Yang et al. 2021a; Yang, Yan, and He 2020). Although the anchor-based methods mentioned above have obtained promising detection results, some limitations remain, such as too many hyperparameters, overlapping calculations, and complex post-processing.

To overcome the shortcomings of anchor-based methods, anchor-free methods have become a new research focus in recent years. Horizontal object representations based on anchor-free methods can be summarized as keypoint-based methods (Zhou, Zhuo, and Krahenbuhl 2019; Duan et al. 2019), pixel-based methods (Tian et al. 2020), and point set-based methods (Yang et al. 2019b). Many excellent studies have emerged recently that explore the effective representation using anchor-free methods for oriented object detection. O^2-Det (Wei et al. 2020) detects a pair of corresponding middle lines. PolarDet (Zhao et al. 2021), and P-RSDet (Zhou et al. 2020) represent the oriented objects using the polar method in the polar coordinate system.

Sample Selection for Object Detection

Classical anchor-based detectors, for example, RetinaNet (Lin et al. 2017b), select positive and negative samples based on the fixed MaxIoU matching strategy, which adopts the IoU value (anchor with ground-truth box) as a matching metric. Many excellent dynamic sample selection strategies have been proposed recently. MetaAnchor (Yang et al. 2018a) is a type of anchor function that generates adaptive anchors from arbitrary customized prior boxes. DAL (Ming et al. 2021b) dynamically assigns anchors according to a defined matching degree, which can comprehensively evaluate the localization potential of the anchors. FreeAnchor (Zhang et al. 2021) is a learning-to-match approach that allows objects to dynamically select anchors under the maximum likelihood principle. PAA (Kim and Lee 2020) is a novel anchor assignment strategy that adaptively separates anchors into positive samples and negative samples for a ground-truth bounding box in a probabilistic manner.

Although these adaptive strategies achieve dynamic sample selection, most of them ignore object shape information and the differentiation between the selected positive and can only be applied to either the anchor-free or anchor-based methods.

The Proposed Method

The oriented anchor-free method RepPoints is used as a baseline example to introduce the proposed method, and the pipeline is illustrated in Fig. 2. RepPoints is constructed with a backbone network, initial detection head, and refinement

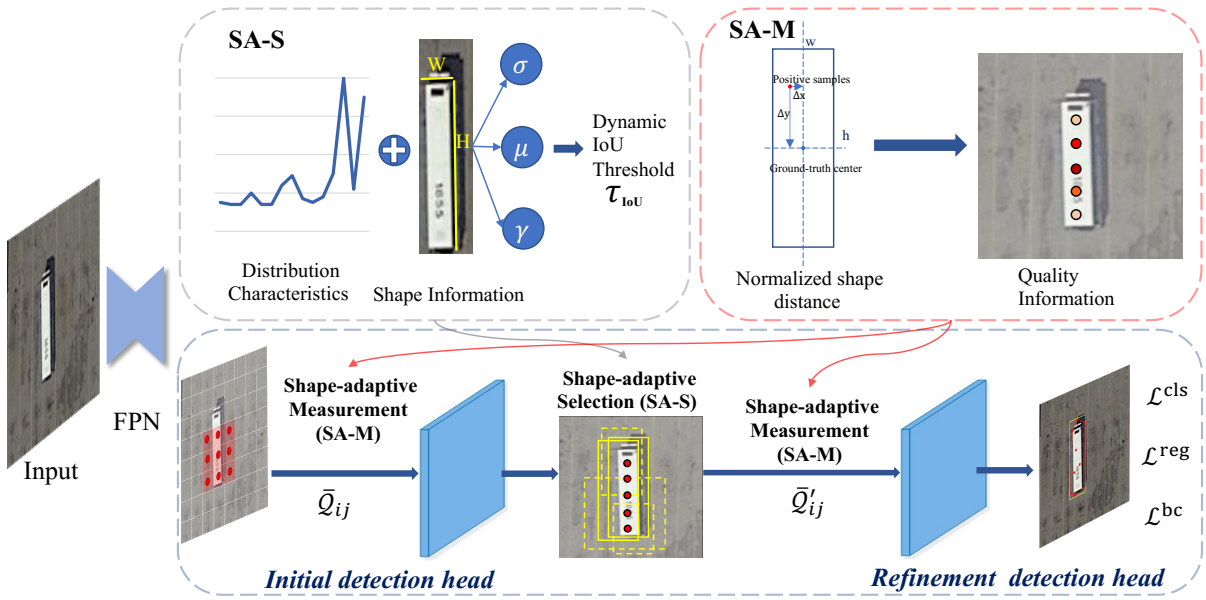

Figure 2: Pipeline of the proposed method. The SA-S strategy dynamically selects samples based on the shape of the object for the refinement detection head, and then SA-M strategy measures the quality of positive samples in the initial detection head and refinement detection head. Boxes represent the predicted point sets for a clear visualization, where solid boxes and dashed boxes represent positive and negative samples, respectively. μ and σ denote the mean and standard deviation of the IoU values between the proposal predictions and the ground-truth box. γ is the aspect ratio of the object.

AP γ_m / $\mathcal{T}_{IoU}$	PL	RA	BC	SP	SV	TC	LV	HA
	1.0	1.0	1.12	1.22	1.72	1.89	3.45	3.92
0.5, 0.4	**87.78**	**71.26**	72.36	68.85	75.35	89.60	55.72	48.38
0.4, 0.3	87.48	67.68	**82.54**	**70.36**	78.12	90.78	58.72	60.93
0.3, 0.2	80.81	67.16	76.88	69.01	**78.47**	**90.84**	58.69	60.25
0.1, 0.1	79.07	66.82	77.95	65.66	63.12	90.07	**59.64**	**67.58**

Table 1: Pilot experiments for the relationship between predefined IoU threshold and the aspect ratio of the object on DOTA.

detection head. The initial detection head generates coarse point sets, which are converted into convex hulls using the Jarvis March (Jarvis 1973) algorithm and further refined in the refinement detection head inspired by (Guo et al. 2021). The proposed SA-S strategy is adopted in the refinement detection head to dynamically select positive samples. The SA-M strategy is used in both heads to evaluate the quality of the selected positive samples. A boundary-center loss function is designed for the anchor-free pipeline.

The Motivation

First, the motivation for conducting pilot experiments to show the relationship between a reasonable IoU threshold and the aspect ratio of the object is presented, and experimental results are listed in Table 1. $\mathcal{T}_{IoU}$ represents the predefined IoU threshold containing positive and negative thresholds, γ_m represents the mean aspect ratio of all objects in one category, and AP is the average precision. It can be seen in Table 1 that while the aspect ratio of the object is larger, the performance is better with a low IoU threshold. This may be because the IoU value has different sensitivities to localization errors under different shapes.

Therefore, because the traditional IoU-based sample selection strategy uses the same predefined IoU threshold for all objects, mining high-quality samples for multi-class object detection is ineffective, particularly when a wide variety of object shapes exists. Motivated by the above experimental results and analysis, two shape-adaptive strategies are proposed, SA-S and SA-M, for dynamically selecting and measuring the samples, respectively.

Shape-Adaptive Selection

The IoU-based selection strategy is used in the refinement detection head of RepPoints. However, the IoU-based strategy ignores the object shape and processes all objects using the same fixed rule, which may be applicable to most objects, but some objects with special shapes are ignored.

To optimize the selection process, an SA-S strategy is proposed that adaptively adjusts the IoU threshold according to the shape and characteristics distribution of objects to select samples. Inspired by the ATSS (Zhang et al. 2020), the mean and standard deviation of objects are adopted to dynamically calculate the IoU threshold. For the i-th ground-truth box, the IoU threshold $\mathcal{T}_i^{\text{IoU}}$ for selecting samples is calculated as:

$$\mathcal{T}_i^{\text{IoU}} = f(\gamma_i) * (\mu + \sigma)\,, \tag{1}$$

where

$$\mu = \frac{1}{J}\sum_{j=1}^{J}\mathcal{I}_{i,j}\ , \quad \sigma = \sqrt{\frac{1}{J}\sum_{j=1}^{J}(\mathcal{I}_{i,j}-\mu)^2},$$

J is the number of candidate samples, and $\mathcal{I}_{i,j}$ is the IoU value between the i-th ground-truth box and the j-th prediction. γ_i represents the aspect ratio of the ground-truth box corresponding to the prediction and is calculated as the ratio of the long edge to the short edge. According to the above analysis, the weight should decrease as the aspect ratio increases so that elongated objects are assigned a low IoU threshold. A monotonic decreasing function is designed for the weighting factor that depends on the object's aspect ratio. $f(\gamma_i)$ is the weighting factor function of the object and is calculated as:

$$f(\gamma_i) = e^{-\frac{\gamma_i}{\omega}}, \tag{2}$$

where ω is a weighted parameter that defaults empirically to 4. A larger ω usually achieves better performance when a dataset contains a large number of elongated objects. Positive samples are selected using a general assignment strategy, which selects candidates whose IoUs are greater than or equal to the threshold $\mathcal{T}_i^{\mathrm{IoU}}$.

Shape-Adaptive Measurement

Compared with the points located inside the object, the points located near the boundaries of the object contain more information about the clutter background, and even nearby objects. Therefore, the points located inside the object, particularly the points located around the center of the object, are more representative of the object's features than the points located close to the boundaries of the object. Processing all positive samples in the same manner would lead to the misjudgment of some high-quality samples. Points inside the object that are far from the center of the object are likely to be suppressed by background points that are closer to the center of the object. The above analysis leads to the conclusion that the detection potential of each point is strongly related to the shape of the object and not only the distance of each point from the the object center.

To optimize this process, an SA-M strategy is proposed to evaluate and add quality information to each positive sample. The quality of a positive sample is estimated using its position relative to the object, which is called the normalized shape distance in this study. A function is elaborated to calculate the normalized shape distance using the distance from the sample to the corresponding object's center and the shape information of the object. Specifically, each ground-truth box is described as five parameters (x, y, w, h, θ), where (x, y), w, and h denote the center coordinates, width, and height of the ground-truth box, respectively, and θ represents the angle of box following (Han et al. 2021). The normalized selection of (w, h) on the x or y-axis is determined by the angle. The normalized shape distance Δd_{ij} from the j-th sample point to the i-th object's center is calculated as:

$$\Delta d_{ij} = \begin{cases} \sqrt{\frac{(x_i-x_j)^2}{w_i} + \frac{(y_i-y_j)^2}{h_i}} & \text{if} \quad 0 \le \theta_i \le \pi/2 \\ \sqrt{\frac{(x_i-x_j)^2}{h_i} + \frac{(y_i-y_j)^2}{w_i}} & \text{otherwise} \end{cases}. \tag{3}$$

Then, the quality $\bar{\mathcal{Q}}_{ij}$ of the positive sample is calculated after obtaining the normalized shape distance:

$$\bar{\mathcal{Q}}_{ij} = e^{-\Delta d_{ij}}. \tag{4}$$

For the selected positive samples, a distinction is established in terms of quality and the influence of inappropriate process for selecting positive samples is eliminated.

Loss Functions

Boundary-Center Loss An isolated point with a large deviation greatly affects the quality of the convex hull (calculated from predicted point set) and has a negative influence on precise localization. To address this problem, a boundary-center loss is proposed in this study. The left-most, right-most, top-most and bottom-most point are selected from the point set, and a mean center point is calculated by the average x and y coordinates of all the points in the point set. The five points' coordinates of prediction and ground-truth box are represented by p_i and g_i, where $i = (1, 2, ..., 5)$, is the index of the selected five points. The boundary-center loss $\mathcal{L}^{bc}$ is used to constrain the boundary and center points, and defined as:

$$\mathcal{L}^{bc} = \sum_{i=1}^{5} L_{\mathrm{smooth}}(p_i, g_i), \tag{5}$$

where L_{smooth} is the smooth L_1, which is defined as

$$\mathrm{smooth}_{l1}(t) = \begin{cases} 0.5t^2 & \text{if } |t| < 1 \\ |t| - 0.5 & \text{otherwise} \end{cases}. \tag{6}$$

The smooth L_1 distance of the five points from the predicted point set and ground-truth bounding box is calculated using $\mathrm{smooth}_{l1}(||p_i - g_i||)$, which is defined in (6), and $||p_i - g_i||$ is the L_2 distance between the two i-th points.

Total Loss The total loss is calculated as:

$$\mathcal{L} = \lambda_1\mathcal{L}^{\mathrm{c}} + \lambda_2\mathcal{L}^{1} + \lambda_3\mathcal{L}^{2}, \tag{7}$$

where $\mathcal{L}^{\mathrm{c}}$, $\mathcal{L}^{1}$, and $\mathcal{L}^{2}$ represent the classification loss, initial detection head loss, and refinement detection head loss, respectively. λ_1, λ_2, and λ_3 are weighting coefficients, which are empirically set to 1.0, 0.375, and 1.0, respectively. For the i-th object, the classification loss is denoted as:

$$\mathcal{L}_i^{c} = \frac{1}{N^+}\frac{1}{\sum_{p_j\in\mathbf{P}^+}\bar{\mathcal{Q}}_{ij}}\sum_{ij}\bar{\mathcal{Q}}_{ij}\mathcal{L}_{ij}^{\mathrm{cls}}, \tag{8}$$

where j, N^+, and $\mathbf{P}^+$ respectively represent the index, total number, and the set of the predicted convex hulls, respectively. The classification loss $\mathcal{L}^{\mathrm{cls}}$ adopts the focal loss (Lin et al. 2017b). $\bar{\mathcal{Q}}_{ij}$ is the quality measurement, and the scale-adaptive weight is assigned to each positive sample on the basis of $\bar{\mathcal{Q}}_{ij}$. p_j represents the predicted convex hull that is calculated using the predicted point sets.

In the initial detection head, the regression loss is defined as:

$$\mathcal{L}_i^{1} = \frac{1}{N^+}\frac{1}{\sum_{p_j\in\mathbf{P}^+}\bar{\mathcal{Q}}_{ij}}\sum_{ij}\bar{\mathcal{Q}}_{ij}\mathcal{L}_{ij}^{\mathrm{reg}} + \mathcal{L}_{ij}^{\mathrm{bc}}, \tag{9}$$

The regression loss for the convex hull $\mathcal{L}^{\text{reg}}$ adopts the GIoU loss (Rezatofighi et al. 2019) and is calculated as:

$$\mathcal{L}^{\text{reg}} = 1 - \text{GIoU}. \tag{10}$$

GIoU is an improved version of the IoU that takes the non-overlapping regions that IoU overlooks into consideration to reflect the overlap degree of the predicted box P and ground-truth box G, and is defined as:

$$\text{GIoU} = \text{IoU} - \frac{\text{area}(C \backslash (P \cup G))}{\text{area}(C)}, \tag{11}$$

where C is the smallest box enclosing P and G.

In the refinement detection head, the regression loss for the convex hull is defined as:

$$\mathcal{L}_i^2 = \frac{1}{N^+} \frac{1}{\sum_{p_j \in \mathbf{P}^+} \bar{\mathcal{Q}'}_{ij}} \sum_{ij} \bar{\mathcal{Q}'}_{ij} \mathcal{L}_{ij}^{\text{reg}} \tag{12}$$

where $\bar{\mathcal{Q}'}_{ij}$ is the quality measurement for each positive sample in the refinement detection head. Regression loss $\mathcal{L}^{\text{reg}}$ also adopts GIoU loss.

Experiments and Discussions

The results of experiments conducted on four typical publicly available datasets containing oriented objects, that is, DOTA (Xia et al. 2018), HRSC2016 (Liu et al. 2017), UCAS-AOD (Zhu et al. 2015), and ICDAR2015 (Karatzas et al. 2015) are summarized to evaluate the effectiveness of the proposed method. The details of the datasets, method implementations, evaluation metrics, and experimental results are presented in the following subsections.

Datasets

DOTA (Xia et al. 2018) is a public, large aerial image dataset for oriented object detection that contains 15 categories, and objects in a wide variety of scales, orientations, and shapes: plane (PL), baseball diamond (BD), bridge (BR), ground track field (GTF), small vehicle (SV), large vehicle (LV), ship (SH), tennis court (TC), basketball court (BC), storage tank (ST), soccer ball field (SBF), roundabout (RA), harbor (HA), swimming pool (SP), and helicopter (HC). DOTA contains 2,806 aerial images and 188,282 instances. The size of each image is in the range of 288 to 8,115 pixels in width, and 211 to 13,383 pixels in height. This dataset contains three subsets, which are the training set (1/2), validation set (1/6), and testing set (1/3), and the ground truth of the test set is not publicly accessible. All images in the training and validation sets were split into blocks of 1024×1024 pixels, with an overlap of 200 pixels for the training dataset.

HRSC2016 (Liu et al. 2017) contains 436 images for training, 181 images for validation, and 444 images for testing. The image size ranges from 300×300 to 1,500×900 pixels. The dataset contains ships with arbitrary aspect ratios and orientations. All images were resized to 800×512 for training and testing.

UCAS-AOD (Zhu et al. 2015) is an aerial image dataset for oriented car and airplane detection that contains 1,510 images with approximately 659×1280 pixels and 14,596 instances. For the experiments in this study, the dataset contained 1,057 randomly selected images for training and 302 images for testing.

ICDAR2015 (Karatzas et al. 2015) is a challenging dataset for scene text detection and recognition that contains 1,000 images for training and 500 images for testing and is used for the detection of arbitrarily oriented text.

Implementation Details

The baselines were an anchor-free method RepPoints (Yang et al. 2019b) and an anchor-based method S^2A-Net (Han et al. 2021) for oriented object detection. They both consisted of a backbone for feature extraction and two detection heads for predicted results refinement. FPN (Lin et al. 2017a) with ResNet50 (He et al. 2016) was used as the backbone to extract features, unless special notes were provided. The framework was trained using the SGD optimizer, where the initial learning rate, momentum, and weight decay were 0.01, 0.9, and 0.0001, respectively. The framework was trained respectively for 12, 36, 120, and 240 epochs on the DOTA, HRSC2016, UCAS-AOD, and ICDAR2015 datasets, respectively. The numbers of the points in a point set in RepPoints and the anchors at each position in S^2A-Net were set to 9 and 1, respectively. The weighted parameter ω in (2) was empirically set as 4 on DOTA, UCAS-AOD, and ICDAR2015. Considering that the HRSC2016 dataset contains a large number of elongated ships, ω was set as 14 on it. Additionally, all experiments were performed using MMDetection-1.1 (Chen et al. 2019) and PyTorch-1.3/1.2 on 2 Titan V GPUs with 11G memory and 4 Tesla V GPUs with 32G memory, while the operating system is Ubuntu 16.04. Experiments were performed more than twice and stable values were taken as final results. Data augmentation consisted of random flipping and random rotation. The experimental results for the baseline and "ours" shown in Table 5, 7, and 8 used multi-scale training and data augmentation for a fair comparison with other methods.

Ablation Study

To analyze the effectiveness of the proposed method when other conditions were fixed, a series of controlled variable comparison experiments were performed. The impact of the individual structure proposed in this paper was studied on DOTA and HRSC2016, and the results are shown in Table 2. The results demonstrate that each of the proposed structures achieved different degrees of improvement for detection performance on all datasets.

Effect of Shape-Adaptive Selection. Table 3 shows the improvements of 7.65%, 12.28%, and 15.66% were achieved for the classic large aspect ratio categories, BR, HA, and HC, respectively, which demonstrates that the dynamic SA-S strategy was effective for objects with large aspect ratios. The mAP significantly increased by 3.64% when the SA-S strategy was used, which confirms the effectiveness of the SA-S strategy.

Also, the shape-adaptive idea can be applied to other sample selection strategies to further improve detection performance. As Table 4 shows, "MaxIoU-**SA (ours)**" repre-

Dataset	BCL	SA-S	SA-M	mAP(%)	I	SI
DOTA	×	×	×	70.25		
	√	×	×	70.96	+0.71	+0.71
	√	√	×	74.60	+3.64	+4.35
	√	√	√	74.92	+0.32	+4.67
HRSC2016	×	×	×	75.11		
	√	×	×	77.38	+2.27	+2.27
	√	√	×	87.01	+9.63	+11.90
	√	√	√	88.60	+1.59	+13.49

Table 2: Ablation study results for each structure based on RepPoints on the DOTA and HRSC2016 datasets. "BCL" denotes the boundary-center loss, and "I" and "SI" indicate the individual improvement and the total improvement in mAP values for this structure compared with the baseline, respectively.

Sample Selection	BR	HA	HC	mAP(%)
MaxIoU	45.07	60.93	44.41	70.96
ATSS (Zhang et al. 2020)	50.51	63.68	51.21	72.10
SA-S (ours)	**52.72**	**73.21**	**60.07**	**74.60**

Table 3: Results of the SA-S strategy for objects with large aspect ratios based on RepPoints on DOTA.

sents the strategy that adopted the shape-adaptive idea in the MaxIoU-based strategy. In particular, the IoU threshold for an object was dynamically adjusted as follows: $e^{(-\frac{\gamma}{14})} * \mathrm{IoU}_{\mathrm{threshold}}$, where γ and $\mathrm{IoU}_{\mathrm{threshold}}$ denote the aspect ratio of the object and the predefined IoU threshold, respectively. Experiments were performed on HRSC2016, which contained a large number of various slender ships, to convincingly prove the effectiveness of the SA-S strategy.

Effect of Shape-Adaptive Measurement. As the results listed in the rows 5 and 10 of Table 2 show, the mAP performance is further improved to 74.92% and 88.60% on DOTA and HRSC2016 after the SA-M strategy was applied, respectively, which verifies the effectiveness of the shape-scale attention strategy.

The results in Table 4 show that the proposed SA-S and SA-M strategies improved the detection performance of both anchor-free and anchor-based detectors, which proves the excellent generalizability of these proposed strategies. Rows 3 and 7 in Table 4 demonstrate that the dynamic shape-

Based Method	Sample Selection	mAP(%)
RepPoints (anchor-free)	MaxIoU	75.11
	MaxIoU-**SA (ours)**	82.96
	ATSS (Zhang et al. 2020)	78.07
	SA-S & **SA-M (ours)**	**88.60** (+13.49)
S^2A-Net-D (anchor-based)	MaxIoU	80.26
	MaxIoU-**SA (ours)**	84.54
	ATSS (Zhang et al. 2020)	88.68
	SA-S & **SA-M (ours)**	**88.91** (+8.65)

Table 4: Performance of the SA-S and SA-M strategies on anchor-free and anchor-based methods on HRSC2016.

Method	Backbone	mAP
RRD (Liao et al. 2018)	VGG16	84.30
RoI-Transformer (Ding et al. 2019)	R-101	86.20
RSDet (Qian et al. 2021)	R-50	86.50
Gliding Vertex (Xu et al. 2020)	R-101	88.20
BBAVec (Yi et al. 2021)	R-101	88.6
R^3Det (Yang et al. 2021b)	R-101	89.26
CSL (FPN) (Yang and Yan 2020)	R-101	89.62
DAL (Ming et al. 2021b)	R-101	89.77
anchor-free:		
RepPoints (baseline)	R-101	85.16
Ours (RepPoints-based)	R-101	90.00
anchor-based:		
S^2A-Net (baseline)	R-101	90.17
Ours (S^2A-Net-based)	R-101	**90.27**

Table 5: Comparison of the mAP values of different rotation methods on HRSC2016.

adaptive idea could also be adopted in other fixed sample selection strategies, and further improved mAP. S^2A-Net-D denotes one of the baseline structures of S^2A-Net, which uses deformable convolution to replace the alignment convolution layer and was described in (Han et al. 2021) in detail. The results demonstrated that the proposed strategies boosted mAP on the anchor-free baseline RepPoints and anchor-based baseline S^2A-Net-D by 13.49% and 8.65%, respectively.

Effect of Boundary-Center Loss. According to the results in rows 6 and 7 in Table 2, an mAP improvement of 2.27% was obtained after the boundary-center loss was added. The boundary-center loss suppressed low-quality predicted convex hulls by constraining the boundary corner and mean center points to optimize the detection results under the guidance of spatial information.

Comparisons with State-of-the-art Detectors

Results on HRSC2016. The ship objects in HRSC2016 had large aspect ratios. Experiments performed on HRSC2016 verified the superiority of the proposed shape-adaptive method. As shown in Table 5, the object detection performance of the proposed method was superior to that of the other methods, and achieved 90.27% mAP based on S^2A-Net.

Results on DOTA. As shown in Table 6, the anchor-based methods remained the most popular and high-performing detectors on the DOTA dataset. The proposed method based on RepPoints achieved 74.92% mAP with the R-50-FPN (i.e., ResNet50-FPN) backbone without any tricks (e.g., data augmentation). The proposed method based on S^2A-Net with the RX-101-FPN backbone achieved the best performance with respect to the mAP values compared with all the one-stage and two-stage methods. The proposed method achieved comparable performance with other methods with fewer calculations and less inference time. Notably, because of the different backbones (e.g., R-50/101/152 (He et al. 2016), RX-101 (Xie et al. 2017), and H-104 (Newell, Yang, and Deng 2016)), different input image configurations, and

Method	BB	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP(%)
two-stage:																	
GSDet	R-101	81.12	76.78	40.78	75.89	64.50	58.37	74.21	89.92	79.40	78.83	64.54	63.67	66.04	58.01	52.13	68.28
RADet*	RX-101	79.45	76.99	48.05	65.83	65.45	**74.40**	68.86	89.70	78.14	74.97	49.92	64.63	66.14	**71.58**	62.16	69.06
RoI-T*	R-101	88.64	78.52	43.44	75.92	68.81	73.68	**83.59**	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
CAD	R-101	87.8	82.4	49.4	73.5	71.1	63.5	76.7	**90.9**	79.2	73.3	48.4	60.9	62.0	67.0	62.2	69.9
SCRDet*	R-101	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	**87.94**	**86.86**	65.02	66.68	66.25	68.24	65.21	72.61
SARD	R-101	89.93	84.11	54.19	72.04	68.41	61.18	66.00	90.82	87.79	86.59	65.65	64.04	66.68	68.84	68.03	72.95
FADet*	R-101	90.21	79.58	45.49	76.41	**73.18**	68.27	79.56	90.83	83.40	84.64	53.40	65.42	**74.17**	69.69	64.86	73.28
MFIAR*	R-152	89.62	84.03	52.41	70.30	70.13	67.64	77.81	90.85	85.40	86.22	63.21	64.14	68.31	70.21	62.11	73.49
GV	R-101	89.64	85.00	52.26	**77.34**	73.01	73.14	86.82	90.74	79.02	86.81	59.55	**70.91**	72.94	70.86	57.32	75.02
CSL-F*	R-152	**90.25**	**85.53**	**54.64**	75.31	70.44	73.51	77.62	90.84	86.15	86.69	**69.60**	68.04	73.83	71.10	**68.93**	**76.17**
one-stage:																	
P-RSDet	R-101	89.02	73.65	47.33	72.03	70.58	73.71	72.76	90.82	80.12	81.32	59.45	57.87	60.79	65.21	52.59	69.82
O^2-Det	H-104	89.31	82.14	47.33	61.21	71.32	74.03	78.62	90.76	82.23	81.36	60.93	60.17	58.21	66.98	61.03	71.04
BBAVec*	R-101	88.35	79.96	50.69	62.18	78.43	78.98	87.94	90.85	83.58	84.35	54.13	60.24	65.22	64.28	55.70	73.32
DRN*	H-104	**89.71**	82.34	47.22	64.10	76.22	74.43	85.84	90.57	**86.18**	84.89	57.65	61.93	69.30	69.63	58.48	73.23
R^3Det*	R-152	89.24	80.81	51.11	65.62	70.67	76.03	78.32	90.83	84.89	84.42	65.10	57.18	68.10	68.98	60.88	72.81
PolarDet*	R-101	89.65	**87.07**	48.14	70.97	78.53	80.34	87.45	90.76	85.63	86.87	61.64	**70.32**	71.92	73.09	67.15	76.64
DAL*	R-50	89.69	83.11	55.03	71.00	78.30	81.90	88.46	**90.89**	84.97	**87.46**	**64.41**	65.65	76.86	72.09	64.35	76.95
Ours-RP	R-50	86.42	78.97	52.47	69.84	77.30	75.99	86.72	**90.89**	82.63	85.66	60.13	68.25	73.98	72.22	62.37	74.92
Ours-RP*	RX-101	88.41	83.32	54.00	74.34	**80.87**	84.10	88.04	90.74	82.85	86.26	63.96	66.78	78.40	73.84	61.97	77.19
Ours-S*	RX-101	89.54	85.94	**57.73**	**78.41**	79.78	**84.19**	**89.25**	90.87	85.80	87.27	63.82	67.81	**78.67**	**79.35**	**69.37**	**79.17**

Table 6: Comparison of different detectors of mAP values on the OBB-based task of the DOTA-v1.0. "*" indicates that multi-scale training/testing was used in the method. "**BB**" represents "Backbone". Values with underlines indicate that the best mAP values are achieved compared to all methods. "**Ours-RP**" means the implementation of our method based on RepPoints, and "**Ours-S**" means the implementation of our method based on S^2A-Net. The references of the methods involved in the comparison are listed below: GSDet (Li, Wei, and Zhang 2021), RADet (Li et al. 2020), RoI-T (i.e. RoI-Transformer) (Ding et al. 2019), CAD (Zhang, Lu, and Zhang 2019), SCRDet (Yang et al. 2019a), SARD (Wang et al. 2019), FADet (Li et al. 2019), MFIAR(Yang et al. 2020a), GV (i.e. Gliding Vertex) (Xu et al. 2020), CSL-F (i.e. CSL FPN-based) (Yang and Yan 2020), P-RSDet (Zhou et al. 2020), O^2-Det (Wei et al. 2020), BBAVec (Yi et al. 2021), DRN (Pan et al. 2020), R^3Det (Yang et al. 2021b), PolarDet (Zhao et al. 2021), and DAL (Ming et al. 2021b).

different tricks used in each method, these results are only for reference.

Results on UCAS-AOD. To further verify the effectiveness of the proposed shape-adaptive strategies, a series of experiments were conducted on the UCAS-AOD dataset and the results are listed in Table 7. Our method based on RepPoints achieved the best AP values, 89.96 % and 90.78%, on both categories and performed better than all the state-of-the-art methods under both AP_{50} and AP_{75}, using the VOC 2007 metric proposed in (Everingham et al. 2010), where IoU thresholds for the evaluation and test were set to 0.50 and 0.75, respectively, thereby proving the superiority of the performance of the proposed method.

Results on ICDAR2015. Considering that ICDAR2015 contained many text boxes with a large aspect ratio, a series of experiments was also conducted on ICDAR2015, and the results are listed in Table 8. The Precision, Recall, and F-measure were the evaluation metrics following official criteria. Precision and Recall are denoted as P and R in Table 8. Because the anchor-based method S^2A-Net had excellent performance, our method only improved the value of the F-measure by 0.8% on the baseline after careful parameter selection. The proposed method was even better than some methods designed for oriented text detection, such as RRD

Method	car	airplane	AP_{50}	AP_{75}
YOLOv3	74.63	89.52	82.08	-
RetinaNet	84.64	90.51	87.57	-
Faster-RCNN	86.87	89.86	88.36	47.08
RoI-Transformer	88.02	90.02	89.02	50.54
RIDet-Q	88.50	89.96	89.23	-
RIDet-O	88.88	90.35	89.62	-
DAL	89.25	90.49	89.87	-
anchor-based:				
S^2A-Net (baseline)	89.56	90.42	89.99	-
Ours (S^2A-Net-based)	89.49	90.53	90.00	-
anchor-free:				
RepPoints (baseline)	83.02	89.34	86.18	49.30
Ours (RepPoints-based)	**89.96**	**90.78**	**90.38**	**58.66**

Table 7: Comparison of the AP with state-of-the-art methods on UCAS-AOD. The references of the methods involved in the comparison are listed below: YOLOv3 (Redmon and Farhadi 2018), RetinaNet (Lin et al. 2017b), Faster-RCNN (Ren et al. 2015), RoI-Transformer (Ding et al. 2019), RIDet-Q (Ming et al. 2021a), RIDet-O (Ming et al. 2021a) and DAL (Ming et al. 2021b).

Method	P	R	F-measure
RRPN (Ma et al. 2018)	82.2	73.2	77.4
SCRDet (Yang et al. 2019a)	81.3	78.9	80.1
RRD (Liao et al. 2018)	85.6	79.0	82.2
DAL (Ming et al. 2021b)	84.4	80.5	82.4
anchor-based:			
S^2A-Net (baseline)	80.4	78.2	79.3
Ours (S^2A-Net-based)	81.4	78.8	80.1
anchor-free:			
RepPoints(baseline)	74.2	72.1	73.2
Ours (RepPoints-based)	**86.0**	**81.8**	**83.9**

Table 8: Comparison of the performance of different methods on ICDAR2015. "**P**" is "Precision" and "**R**" is "Recall".

(Liao et al. 2018), thereby demonstrating the robustness of the proposed method in different scenarios.

Conclusions

In this study, two novel shape-adaptive strategies were proposed, SA-S and SA-M, for oriented object detection. These strategies dynamically select samples and adaptively assign quality weights to the selected positive samples. The proposed SA-S strategy dynamically selects high-quality candidate samples as positive samples, considering the shape and characteristics distribution of objects. The SA-M strategy adds quality information to different positive samples. The shape-adaptive strategies outperformed in terms of the oriented object detection performance when there was a wide aspect ratio variation between objects. Extensive experiments were conducted on both anchor-free and anchor-based baselines and four publicly available datasets, and the results demonstrates that the proposed method achieves state-of-the-art performance and can be easily embedded into other detectors to improve detection performance. The source code of the paper will be available publicly at https://github.com/houliping/SASM.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61731022, 61871258, 61929104, 61972375), the NSFC Key Projects of International (Regional) Cooperation and Exchanges (61860206004) and the Key Project of Education Commission of Beijing Municipal (KZ201911417048).

References

Azimi, S. M.; Vig, E.; Bahmanyar, R.; Körner, M.; and Reinartz, P. 2018. Towards multi-class object detection in unconstrained remote sensing imagery. In *Asian Conference on Computer Vision*, 150–165. Springer.

Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J.; Zhang, Z.; Cheng, D.; Zhu, C.; Cheng, T.; Zhao, Q.; Li, B.; Lu, X.; Zhu, R.; Wu, Y.; Dai, J.; Wang, J.; Shi, J.; Ouyang, W.; Loy, C. C.; and Lin, D. 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv preprint arXiv:1906.07155*.

Ding, J.; Xue, N.; Long, Y.; Xia, G.-S.; and Lu, Q. 2019. Learning RoI Transformer for Oriented Object Detection in Aerial Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2849–2858.

Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; and Tian, Q. 2019. Centernet: Keypoint triplets for object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, 6569–6578.

Everingham, M.; Van Gool, L.; Williams, C. K.; Winn, J.; and Zisserman, A. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2): 303–338.

Guo, Z.; Liu, C.; Zhang, X.; Jiao, J.; Ji, X.; and Ye, Q. 2021. Beyond Bounding-Box: Convex-Hull Feature Adaptation for Oriented and Densely Packed Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8792–8801.

Han, J.; Ding, J.; Li, J.; and Xia, G.-S. 2021. Align deep features for oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–11.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Jarvis, R. A. 1973. On the identification of the convex hull of a finite set of points in the plane. *Information processing letters*, 2(1): 18–21.

Karatzas, D.; Gomez-Bigorda, L.; Nicolaou, A.; Ghosh, S.; Bagdanov, A.; Iwamura, M.; Matas, J.; Neumann, L.; Chandrasekhar, V. R.; Lu, S.; et al. 2015. ICDAR 2015 competition on robust reading. In *2015 13th International Conference on Document Analysis and Recognition*, 1156–1160. IEEE.

Kim, K.; and Lee, H. S. 2020. Probabilistic anchor assignment with iou prediction for object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, 355–371. Springer.

Li, C.; Xu, C.; Cui, Z.; Wang, D.; Zhang, T.; and Yang, J. 2019. Feature-attentioned object detection in remote sensing imagery. In *2019 IEEE International Conference on Image Processing*, 3886–3890. IEEE.

Li, W.; Wei, W.; and Zhang, L. 2021. GSDet: Object Detection in Aerial Images Based on Scale Reasoning. *IEEE Transactions on Image Processing*, 30: 4599–4609.

Li, Y.; Huang, Q.; Pei, X.; Jiao, L.; and Shang, R. 2020. Radet: Refine feature pyramid network and multi-layer attention network for arbitrary-oriented object detection of remote sensing images. *Remote Sensing*, 12(3): 389.

Liao, M.; Zhu, Z.; Shi, B.; Xia, G.-s.; and Bai, X. 2018. Rotation-sensitive regression for oriented scene text detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5909–5918.

Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017a. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2117–2125.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017b. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988.

Liu, Z.; Yuan, L.; Weng, L.; and Yang, Y. 2017. A high resolution optical satellite image dataset for ship recognition and some new baselines. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, volume 2, 324–331.

Ma, J.; Shao, W.; Ye, H.; Wang, L.; Wang, H.; Zheng, Y.; and Xue, X. 2018. Arbitrary-oriented scene text detection via rotation proposals. *IEEE Transactions on Multimedia*, 20(11): 3111–3122.

Ming, Q.; Miao, L.; Zhou, Z.; Yang, X.; and Dong, Y. 2021a. Optimization for Arbitrary-Oriented Object Detection via Representation Invariance Loss. *IEEE Geoscience and Remote Sensing Letters*.

Ming, Q.; Zhou, Z.; Miao, L.; Zhang, H.; and Li, L. 2021b. Dynamic Anchor Learning for Arbitrary-Oriented Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2355–2363.

Newell, A.; Yang, K.; and Deng, J. 2016. Stacked hourglass networks for human pose estimation. In *European Conference on Computer Vision*, 483–499. Springer.

Pan, X.; Ren, Y.; Sheng, K.; Dong, W.; Yuan, H.; Guo, X.; Ma, C.; and Xu, C. 2020. Dynamic refinement network for oriented and densely packed object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11207–11216.

Qian, W.; Yang, X.; Peng, S.; Yan, J.; and Guo, Y. 2021. Learning Modulated Loss for Rotated Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2458–2466.

Redmon, J.; and Farhadi, A. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, 91–99.

Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; and Savarese, S. 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 658–666.

Tian, Z.; Shen, C.; Chen, H.; and He, T. 2020. Fcos: A simple and strong anchor-free object detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Wang, Y.; Zhang, Y.; Zhang, Y.; Zhao, L.; Sun, X.; and Guo, Z. 2019. SARD: Towards scale-aware rotated object detection in aerial imagery. *IEEE Access*, 7: 173855–173865.

Wei, H.; Zhang, Y.; Chang, Z.; Li, H.; Wang, H.; and Sun, X. 2020. Oriented objects as pairs of middle lines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169: 268–279.

Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; and Zhang, L. 2018. DOTA: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3974–3983.

Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; and He, K. 2017. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1492–1500.

Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.-S.; and Bai, X. 2020. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE transactions on pattern analysis and machine intelligence*, 43(4): 1452–1459.

Yang, F.; Li, W.; Hu, H.; Li, W.; and Wang, P. 2020a. Multi-scale feature integrated attention-based rotation network for object detection in VHR aerial images. *Sensors*, 20(6): 1686.

Yang, T.; Zhang, X.; Li, Z.; Zhang, W.; and Sun, J. 2018a. Metaanchor: Learning to detect objects with customized anchors. *arXiv preprint arXiv:1807.00980*.

Yang, X.; Hou, L.; Zhou, Y.; Wang, W.; and Yan, J. 2021a. Dense label encoding for boundary discontinuity free rotation detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 15819–15829.

Yang, X.; Sun, H.; Fu, K.; Yang, J.; Sun, X.; Yan, M.; and Guo, Z. 2018b. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sensing*, 10(1): 132.

Yang, X.; and Yan, J. 2020. Arbitrary-Oriented Object Detection with Circular Smooth Label. In *European Conference on Computer Vision*, 677–694. Springer.

Yang, X.; Yan, J.; Feng, Z.; and He, T. 2021b. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 3163–3171.

Yang, X.; Yan, J.; and He, T. 2020. On the Arbitrary-Oriented Object Detection: Classification based Approaches Revisited. *arXiv preprint arXiv:2003.05597*.

Yang, X.; Yan, J.; Ming, Q.; Wang, W.; Zhang, X.; and Tian, Q. 2021c. Rethinking rotated object detection with gaussian wasserstein distance loss. In *International Conference on Machine Learning*, 11830–11841. PMLR.

Yang, X.; Yan, J.; Yang, X.; Tang, J.; Liao, W.; and He, T. 2020b. SCRDet++: Detecting Small, Cluttered and Rotated Objects via Instance-Level Feature Denoising and Rotation Loss Smoothing. *arXiv preprint arXiv:2004.13316*.

Yang, X.; Yang, J.; Yan, J.; Zhang, Y.; Zhang, T.; Guo, Z.; Sun, X.; and Fu, K. 2019a. Scrdet: Towards more robust detection for small, cluttered and rotated objects. In *Proceedings of the IEEE International Conference on Computer Vision*, 8232–8241.

Yang, X.; Yang, X.; Yang, J.; Ming, Q.; Wang, W.; Tian, Q.; and Yan, J. 2021d. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. *Advances in Neural Information Processing Systems*, 34.

Yang, Z.; Liu, S.; Hu, H.; Wang, L.; and Lin, S. 2019b. Reppoints: Point set representation for object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, 9657–9666.

Yi, J.; Wu, P.; Liu, B.; Huang, Q.; Qu, H.; and Metaxas, D. 2021. Oriented object detection in aerial images with box boundary-aware vectors. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2150–2159.

Zhang, G.; Lu, S.; and Zhang, W. 2019. Cad-net: A context-aware detection network for objects in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12): 10015–10024.

Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; and Li, S. Z. 2020. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9759–9768.

Zhang, X.; Wan, F.; Liu, C.; Ji, X.; and Ye, Q. 2021. Learning to match anchors for visual object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhao, P.; Qu, Z.; Bu, Y.; Tan, W.; and Guan, Q. 2021. Polardet: A fast, more precise detector for rotated target in aerial images. *International Journal of Remote Sensing*, 42(15): 5821–5851.

Zhou, L.; Wei, H.; Li, H.; Zhao, W.; Zhang, Y.; and Zhang, Y. 2020. Arbitrary-Oriented Object Detection in Remote Sensing Images Based on Polar Coordinates. *IEEE Access*, 8: 223373–223384.

Zhou, X.; Zhuo, J.; and Krahenbuhl, P. 2019. Bottom-up object detection by grouping extreme and center points. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 850–859.

Zhu, H.; Chen, X.; Dai, W.; Fu, K.; Ye, Q.; and Jiao, J. 2015. Orientation robust object detection in aerial images using deep convolutional neural network. In *2015 IEEE International Conference on Image Processing*, 3735–3739. IEEE.