Article

Rotated Object Detection with Circular Gaussian Distribution

Hang Xu [1], Xinyuan Liu [2], Yike Ma [2], Zunjie Zhu [1,3], Shuai Wang [1,3], Chenggang Yan [1,3] and Feng Dai [2,*]

[1] School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China; hxu@hdu.edu.cn (H.X.); zunjiezhu@gmail.com (Z.Z.); shuaiwang.tai@gmail.com (S.W.); cgyan@hdu.edu.cn (C.Y.)
[2] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100086, China; liuxinyuan21s@ict.ac.cn (X.L.); ykma@ict.ac.cn (Y.M.)
[3] Lishui Institute ofHangzhou Dianzi University, Lishui 323000, China
[*] Correspondence: fdai@ict.ac.cn

Abstract: Rotated object detection is a challenging task due to the difficulties of locating the rotated objects and separating them effectively from the background. For rotated object prediction, researchers have explored numerous regression-based and classification-based approaches to predict a rotation angle. However, both paradigms are constrained by some flaws that make it difficult to accurately predict angles, such as multi-solution and boundary issues, which limits the performance upper bound of detectors. To address these issues, we propose a circular Gaussian distribution (CGD)-based method for angular prediction. We convert the labeled angle into a discrete circular Gaussian distribution spanning a single minimal positive period, and let the model predict the distribution parameters instead of directly regressing or classifying the angle. To improve the overall efficiency of the detection model, we also design a rotated object detector based on CenterNet. Experimental results on various public datasets demonstrated the effectiveness and superior performances of our method. In particular, our approach achieves better results than state-of-the-art competitors, with improvements of 1.92% and 1.04% in terms of AP points on the HRSC2016 and DOTA datasets, respectively.

Keywords: rotated object detection; circular Gaussian; CenterNet; CGD

1. Introduction

Rotated object detection has emerged as a fundamental component for visual analysis across various types of images, including aerial images [1,2], panoramic images [3–5] and scene text [6]. It is a more general approach compared to traditional horizontal object detection [7,8]. As conventional Horizontal Bounding Boxes (HBBs) can not tightly enclose oriented objects, Rotated Bounding Boxes (RBBs) have been introduced in recent works [9,10]. For rotated object detection, the rotation angle is a sensitive parameter, and even a small deviation in angle can lead to a significant drop of Intersection over Union (IoU) between predicted boxes and ground truth. This effect is especially pronounced when the aspect ratio of an object is large. Therefore, accurate angular prediction is crucial to improve the performance of oriented object detectors.

The rotation angle is a numerical attribute with intrinsic periodicity, which gives rise to the multi-solution issue and the boundary issue [11], as illustrated in Figure 1a. Accurate prediction of this attribute is challenging. A rotated bounding box (RBB) is produced by rotating a horizontal bounding box (HBB) around its center, and the period of its rotation angle is 180° (long-edge representation). The RBBs rotated with the angles offset by several periods (e.g., 1° and 181°) are exactly identical. This means that there exists multi-solution in angular space, which increases the uncertainty of the model optimization. When only considering a single period, the RBBs rotated with the larger angle and the smaller angle (e.g., 88° and −88°) are actually similar in the pose. This may cause the model to take long detours during the optimization, as noted in the analysis of GWD [12].
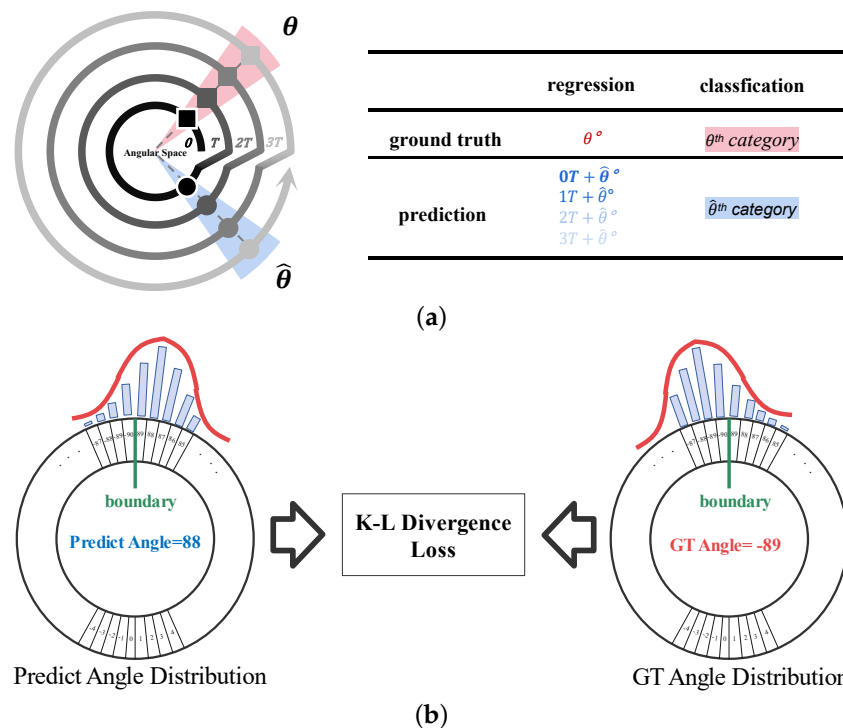
Figure 1. (a) The angular space is a $1-$D periodic space with multi-solution issue (marks of the same shapes on the different rings represent equivalent angles) and boundary issue (adjacent marks of different shapes on the same ring are far away in this space). Regression paradigms directly predict angles across multiple periods, while classification paradigms predict the equivalence classes to which angles belong. (b) Our proposed circular Gaussian distribution (CGD)-based method for angular prediction.

Researchers have explored numerous regression-based [3–5,12–14] and classification-based [15,16] approaches to predict rotation angle. The naive regression predicts a continuous value and measures the numerical deviation of the angle directly. It fails to handle either multi-solution issues or boundary issues by itself. To remedy the natural imperfections, some periodic methods, such as trigonometric functions and modulus operators, have been introduced into the loss calculation. However, the angular values output by the model can still lie in any period, and these methods only decorate angular predictions by projecting them into a single period. Other integrated methods, such as GWD [12], ingeniously transform the OBB into a continuous Gaussian distribution, but it is essentially equivalent to applying trigonometric functions to optimization in some sophisticated way, which never completely avoids the previous shortcoming. Another appealing approach, classification, predicts a category of fine-grained angular ranges that can naturally aggregate angular values across periods into their equivalence classes to fix the multi-solution issue. However, it still suffers from boundary issues, as a result of indiscrimination treatment for different angle ranges, where the near-far relation between angle ranges/categories is completely ignored. Some improved versions, such as CSL [15] and VGL [16], utilize window functions to smooth category labels, but they do not completely eliminate the defects of classification. In a word, it is difficult to accurately predict angles through the above paradigms due to various issues, which limits the performance upper bound of detectors.

To address the above issues, we propose a circular Gaussian distribution (CGD)-based method for angular prediction, as shown in Figure 1b. Specifically, we convert the labeled angle into a discrete circular Gaussian distribution spanning a single minimal positive period as new ground truth, and let the model predict the distribution parameters instead of directly regressing or classifying the angle. Then, we calculate the loss in the way to measure the Kullback–Leibler divergence between the predicted distribution and the ground-truth distribution. Here, the Gaussian distribution can reasonably reflect the

adjacency between the angles, and each angle is assigned a certain probability based on its offset from the actual angle of the object. In this way, the angular distribution as a whole effectively overcomes the disadvantage of classifying each angular bin in isolation. The circularization of Gaussians solves the boundary issue, and discretization avoids the multi-solution issue. Additionally, we design a rotated object detector based on CenterNet [17,18] to improve the overall efficiency of detection.

In summary, the main contributions of this paper are as follows:

1. We propose a new paradigm for angular prediction, namely CGD. It effectively avoids the shortcomings of previous approaches.
2. We design a rotated object detector, based on CenterNet, which can improve the overall efficiency of the detection model.
3. We conduct extensive experiments on various public datasets to verify the effectiveness and superior performances of our approaches.

2. Related Work

In this section, we introduce related works, including horizontal object detection, rotated object detection, and loss function for rotated object detection.

2.1. Horizontal Object Detection Method

Due to the development of CNNs, significant progress has been made in the past few years in horizontal box-based object detection. Currently, object detectors can generally be classified into two paradigms: two-stage detectors and one-stage detectors.

The RCNN [19] was the first two-stage detector method, where the first stage is used to generate candidate boxes through a selective search algorithm, and the second stage uses CNNs to extract features from the candidate boxes. Fast-RCNN [20] improved on RCNN by extracting region-of-interest (RoI) from feature maps, saving the computational cost of sharing backbone networks in RCNN. Faster-RCNN [7] introduced the region proposal network (RPN) for candidate boxes generation. Therefore, the entire network is trained end-to-end. To detect objects of different scales, the feature pyramid network (FPN) [21] was proposed to establish a pyramid hierarchical structure, which can effectively improve the performance of the detector.

One-stage detectors remove the ROI extraction process in two-stage detectors and directly perform the bounding box regression and classification. Early representative one-stage detectors include SSD and YOLOv1. SSD densely places anchor boxes on input images, while YOLOv1 divides input images into grids of different sizes. To address class imbalance issues, Focal loss is designed in RetinaNet [8] to dynamically adjust the weight of each anchor box. FCOS further improves RetinaNet by removing predefined anchor boxes and directly regressing and classifying reference points. CenterNet [17] proposes to regress the width and height of bounding boxes from the center of objects. During inference, CenterNet does not use the NMS algorithm, thus improving the inference speed of the detector.

2.2. Rotated Object Detection Method

Rotational object detection is a new task in the current field of object detection, which has been well applied in aerial image object detection and panoramic image object detection. Unlike horizontal object detectors, rotational detectors generally use rotated rectangles as bounding boxes because rotated boxes can better enclose objects. To solve the problem of mismatch in densely packed object scenarios, ROI Transformer [1] proposes spatial transformation on ROI. Oriented RCNN [22] designs an oriented RPN to directly generate high-quality oriented candidate boxes at almost no cost. S2A-Net [23] proposes a feature alignment module to obtain high-quality anchor points. Based on Faster-RCNN, CAD-NET [24] proposes a context-aware detection network to learn global and local contexts in images. SCRDET++ [25] designs a new feature map-based instance-level denoising module for detecting small, cluttered, and rotated objects.

3. Proposed Approach

3.1. Circular Gaussian Distribution Construction

In this section, we present our approach. First, we adopt the discrete circular Gaussian distribution spanning a single minimal positive period as the ground truth of the angle. The circularization of Gaussians solves the boundary problem and discretization avoids the multi-valued problem. For instance, an original label of the angle $(-89°)$ can be used to generate a circular Gaussian distribution in $[-90°, 89°]$, as shown in Figure 1b. The loss of accuracy in the rotation detection task is minimal despite the conversion from continuous angle to discrete angle, just like the analysis in CLS [15]. Next, the Gaussian distribution of the angle can be represented as a multi-dimensional vector $\boldsymbol{D}^t = \{d_{-90}, d_{-89}, \ldots, d_{89}\}$, with the l-th dimension as follows:

$$
\begin{aligned}
\overline{d_l} &= exp(\tfrac{-(l-\theta_t)^2}{2\sigma^2}), l = -90, 89, \ldots, 89 \\
d_l &= \tfrac{\overline{d_l}}{\sum_{l=-90}^{89} \overline{d_l}}, \\
\sum_{l=-90}^{89} d_l &= 1
\end{aligned}
\tag{1}
$$

where l denotes the l-th binned angle, θ_t is the binned ground-truth angle, and σ is the standard deviation of Gaussian distribution. As a hyperparameter, σ, we experimentally find that its selection is reliable within a particular range. More information will be provided in Table 1.

Table 1. Comparison between different standard deviation σ for Gaussian label on HRSC2016 dataset. The CenterNet-FPN model is used as the detector.

σ	Backbone	mAP07	mAP12
2	R-50-FPN	6.56	3.08
4	R-50-FPN	90.44	97.31
6	R-50-FPN	**90.54**	**97.76**
8	R-50-FPN	90.52	97.17
10	R-50-FPN	90.47	97.08

Then, the training set can be represented as $\{(I_i, \boldsymbol{D}_i^t), 1 \le i \le B\}$, and the objective of the model learning is to obtain a group of model parameters $\boldsymbol{\omega}$ to produce a probability distribution $\boldsymbol{F}^p(I_i; \boldsymbol{\omega})$ for the label set.

$$
\begin{aligned}
\boldsymbol{F}^p(I_i; \boldsymbol{\omega}) &= \{f(p_{-90}|I_i; \boldsymbol{\omega}), \ldots, f(p_{89}|I_i; \boldsymbol{\omega})\}, \\
\sum_{l=-90}^{89} f(p_l|I_i; \boldsymbol{\omega}) &= 1
\end{aligned}
\tag{2}
$$

Finally, the loss function quantifying the similarity between the predicted distribution $\boldsymbol{F}^p(I_i; \boldsymbol{\omega})$ and the ground-truth distribution $\boldsymbol{D}_i^t$ is constructed using the Kullback–Leibler divergence. The objective of the Label Distribution Learning is to minimize the following loss function:

$$
L_{KL} = \sum_{i=1}^{B} \sum_{l=-90}^{89} d_{I_i}^l \ln \frac{d_{I_i}^l}{f(p_l|I_i; \boldsymbol{\omega})}
\tag{3}
$$

where I is the input image, and B is the number of images in the batch.

Algorithm 1 provides the pseudo-code of circular Gaussian distribution (CGD).

Algorithm 1 Pseudocode of CGD in a Numpy-like style.

```python
# angle: angle of bounding box
# sig: standard deviation of Gaussian distribution

def Circular_Gaussian_Distribution(angle, sig=4.0):

    x = np.array(range(math.floor(-180/2), math.ceil(180/2), 1))

    d = np.exp(-(x) ** 2 / (2 * sig ** 2))

    d_left = d[math.ceil(180/2)-angle:]

    d_right = d[:math.ceil(180/2)-angle]

    CGD = np.concatenate([d_left, d_right], axis=0)

    return CGD
```

3.2. Overall Architecture

Our overall framework is illustrated in Figure 2. Anchor-based detectors require the calculation of the angle for each anchor, and discretizing the angle for anchor-based detectors increases the computational complexity of the model dramatically. Therefore, we used an anchor-free detector CenterNet [17] as our baseline, which models an object as a single point (i.e., the center point of the bounding box) and predicts the center offset, object size, and angle. Specifically, the first branch is utilized to detect the center points of the bounding box and predict the offsets of the center point. The second branch is used to determine the size (i.e., width w and height h) of the bounding box for each object. The last branch predicts the angle distribution of the bounding boxes.
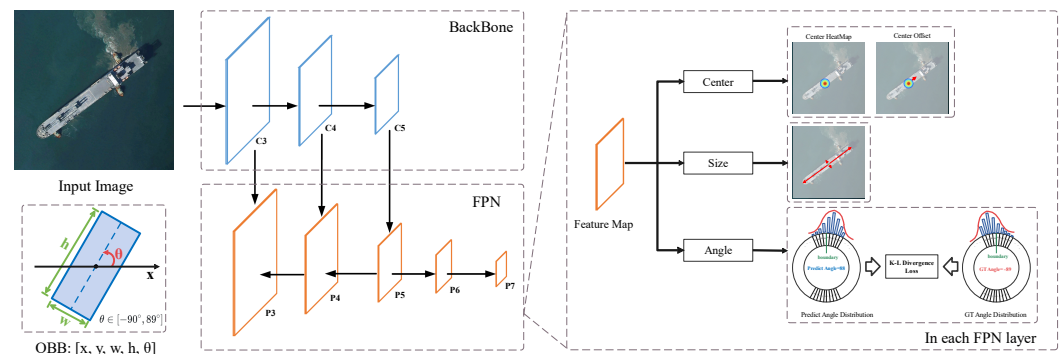

Figure 2. Overall architecture of our method to detect rotated objects in images. We propose a circular Gaussian distribution (CGD)-based method for angular prediction. (Best viewed by zooming in).

3.2.1. Backbone

Different from raw CenterNet, we used ResNet as the backbone and built a feature pyramid network (FPN). The FPN enhances a conventional convolutional network with a top-down pathway and lateral connections to effectively build a rich, multi-scale feature pyramid from a single-resolution input image.

3.2.2. Center Branch

For oriented objects in the input image, the heatmap is utilized to localize their center locations. Following the original CenterNet, a 2-D Gaussian kernel is adopted to produce a heatmap $Y \in [0,1]^{(W/4) \times (H/4) \times C}$. Any other pixel is treated as the negative sample, while the peak of the Gaussian distribution, which is also the pixel at the center of the box, is treated as the positive sample. The element-wise maximum is used when two Gaussians of the same class overlap. The training objective is a modified focal loss:

$$L_k = \frac{-1}{N} \sum_{xyc} \begin{cases} (1 - \hat{Y}_{xyc})^{\alpha} \log(\hat{Y}_{xyc}) & \text{if } Y_{xyc} = 1 \\[2mm] \begin{aligned}(1 - Y_{xyc})^{\beta}(\hat{Y}_{xyc})^{\alpha} \\ \log(1 - \hat{Y}_{xyc})\end{aligned} & \text{otherwise} \end{cases} \tag{4}$$

where α and β are hyper-parameters of the focal loss, and N is the number of keypoints in image I. We set $\alpha = 2$ and $\beta = 4$ in all our experiments, following CenterNet.

The model predicts an additional center offset $O \in [0,1]^{(W/R) \times (H/R) \times 2}$ to remove the discretization error brought on by the output stride. For the regression of the center offset, we used Smooth L1 loss.

3.2.3. Size Branch

In contrast to the raw CenterNet, which directly regressed the variables w and h, we used indirect regression $log((w/R)), log((h/R))$ to lessen the effect of the varied object aspect ratio. Then, we used Smooth L1 loss for the estimation of the $\{log((w/R)), log((h/R))\}$.

3.2.4. Angle Branch

For the object orientation, the model outputs a distribution of angle $\boldsymbol{F}^p \in [0,1]^{(W/4) \times (H/4) \times 180}$. Then, the Kullback–Leibler Divergence loss is applied to measure the difference between prediction and target distribution.

Thus, the overall training objective of our model is

$$L_{det} = L_{cls} + \lambda_{off}L_{off} + \lambda_{size}L_{size} + \lambda_{ang}L_{KL}, \tag{5}$$

where L_{cls}, L_{size}, and L_{off} are the losses of center point recognition, scale regression, and offset regression, which are the same as CenterNet; and λ_{size}, λ_{off}, and λ_{ang} are constant factors, set to 0.1 in our experiments.

4. Experiments

4.1. Datasets and Implementation Details

4.1.1. DOTA

DOTA [9] is a large dataset dedicated to rotated object detection for aerial images. The size of each image varies from 800×800 to 4000×4000. Specifically, the annotated DOTA contains 2806 aerial images, which include 188,282 instances of 15 object categories. The whole dataset is divided into 1411, 458, and 937 images for training, validation, and testing, respectively. Furthermore, the training images are cropped into patches of size 1024×1024 pixels with an overlap of 256 pixels so as to fit the limited GPU memory.

4.1.2. HRSC2016

HRSC2016 [10] is an aerial image dataset for ship detection. The dataset contains 1061 images of ships sourced from two scenarios which are the sea and the inshore at six famous harbors. The size of each image varies from 300×300 to 1500×900. HRSC2016 is divided into 436,181, and 444 images for training, validation, and testing. Furthermore, the long side of each image is resized to a fixed size (e.g., 640 px) in experiments. To keep structural information, the original aspect ratio of each image is kept.

4.1.3. Evaluation Metric

mAP is a classical metric for detection methods. Therefore, we use mAP in all our experiments to evaluate the performance. For the DOTA dataset, we can obtain the test results from the official evaluation server. For the HRSC2016 dataset, we use metrics that include the VOC07 AP and VOC12 AP with an IoU threshold of 0.5. For the PANDORA dataset, we use metrics that include mAP with an IoU threshold of 0.5.

4.1.4. Implementation Details

Our implementation is based on PyTorch and 8 NVIDIA GeForce RTX 3090 GPUs. We choose the ResNet model pretrained on ImageNet as our backbone. The models are optimized by Adam for 140 epochs with the learning rate dropped by $10\times$ at 100 and 130 epochs for all datasets. For all datasets, the batch size is set to 64, and the initial learning rates are set to 2×10^{-4} and 1.25×10^{-4}, respectively. Data augmentations are used to improve model performance. Our data augmentations include random graying, random flipping, and random rotation. ResNet-50-FPN is used for ablation studies.

4.2. Ablation Study

We create the following ablation research to eliminate the hyper-parameter's unpredictability and to confirm the viability of the suggested approach. We use CenterNet-FPN as the detection model in the ablation study. Accurate detection results are difficult to acquire because the HRSC2016 dataset comprises a high number of ships with large aspect ratios. We use HRSC2016 as our dataset for ablations.

4.2.1. Influence of Different Hyper-Parameter Value

We studied the impact of the σ (standard deviation) of the Gaussian distribution. A different constant for the σ has been established experimentally. The results are shown in Table 1. We observe that the AP only varies somewhat while adjusting σ in a specific range (from 4 to 10), which indicates that the choice of σ is robust in this range. The best performance is obtained by the model when $\sigma = 6$, as shown in Table 1. Therefore, we fix the σ to 6 in all the subsequent experiments.

4.2.2. Effectiveness of CGD

We conducted a number of baseline experiments, including direct regression-based angle prediction (Smooth L1), indirect regression-based angle prediction (trigonometric functions), CSL-based angle classification, and our CGD approach, to verify the effectiveness of our method. The above-mentioned methods share the same network structure except for the different orientation branches. The results are shown in Table 2, and using a regression method to directly predict the angle of the rotated object achieves 85.39% mAP07 and 90.25% mAP12 on the HRSC2016 test set. The indirect regression-based trigonometric functions loss enhances the direct regression approach by 2.43% mAP07 and 3.42% mAP12 by removing the discontinuous boundary problems brought on by the angular periodicity. However, the angular value output by the model can lie in any period, causing the multi-value problem. The CSL-based algorithm has a better result than the angle indirect regression-based method, because it can naturally aggregate multi-values across periods into their equivalence classes to fix the multi-value issue. The CSL-based algorithm achieves 89.98% mAP07 and 95.13% mAP12 on the HRSC2016 test set. As our CGD eliminates the discontinuous boundary issues caused by the angular periodicity and the multi-valued issues caused by out of the defined range, it achieves 90.52% mAP07 and 97.76% mAP12. Finally, in Table 2, we present the running speed of our CGD method and the baseline method. The running speed here refers to the time it takes to process four images in one run on an NVIDIA GeForce RTX 3090 GPU. As shown in Table 2, our CGD method also has an advantage in terms of speed compared to the baseline methods (0.3912 ms vs. 0.4950 ms, 0.4606 ms, 0.5910 ms). All experimental results demonstrate that the CGD's overall performance is superior to the baseline.

Table 2. Comparison between different losses for the rotated bounding box on HRSC2016 dataset. The CenterNet-FPN model is used as the detector.

Loss	Backbone	mAP07	mAP12	Speed (ms)
DR	R-50-FPN	85.39	90.25	0.4950
IR	R-50-FPN	87.82	93.67	0.4606
CSL	R-50-FPN	89.98	95.13	0.5910
CGD	R-50-FPN	**90.52**	**97.76**	0.3912

4.3. Comparisons with the State-of-the-Art Methods

To validate the effectiveness of our approach, we compared other state-of-the-art methods on the HRSC2016 and DOTA datasets. The performances of all methods were taken from a single model without cross-model test augmentation such as model ensemble by default.

The results of the HRSC2016 dataset are shown in Table 3. Our CGD obtains 90.61% and 98.14% mAP based on R-101-FPN when using the VOC07 metric and the VOC12 metric, respectively. These results are quite competitive when compared to the most recent state-of-the-art methods. We visualize the detection results in Figure 3.

Table 3. Comparison with state-of-the-art methods on the HRSC2016 dataset.

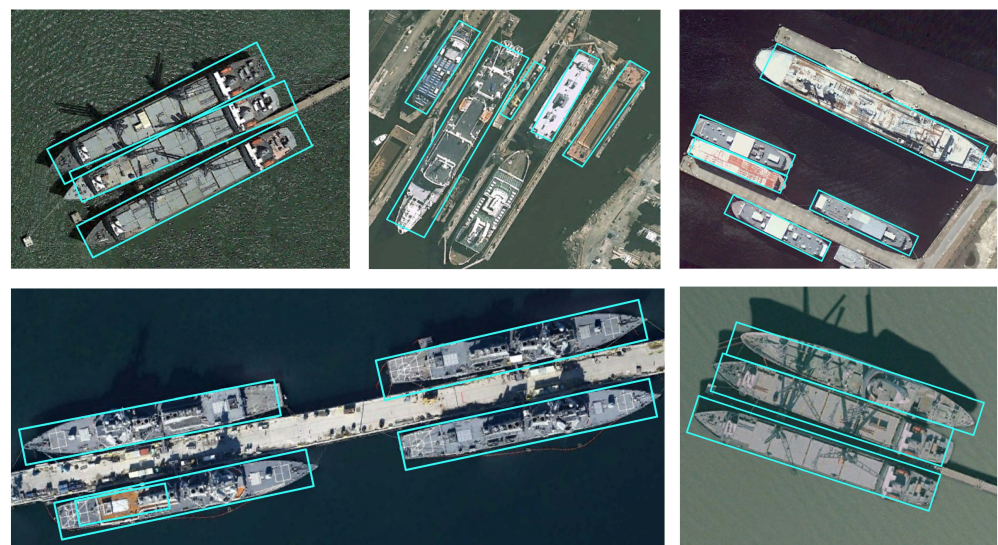| Method	Backbone	mAP07	mAP12
RRPN [26]	VGG16	79.08	85.64
R2CNN [6]	VGG16	73.07	79.73
RT [1]	R-101-FPN	86.20	-
ARG [27]	R-101-FPN	88.08	93.83
GV [2]	R-101-FPN	88.20	-
DRN [13]	Hourglass-104	-	92.70
GWD [12]	R-101-FPN	89.43	-
DAL [14]	R-101-FPN	89.77	-
VGL [16]	DLA34-DCN	89.78	-
S^2ANet [23]	R-101-FPN	90.17	95.01
AOPG [28]	R-101-FPN	90.34	96.22
CGD (Our)	R-101-FPN	**90.61**	**98.14**

Figure 3. The visualization results of our method on the HRSC2016 dataset. (Best viewed by zooming in).

There are numerous categories of complexity scenes in the DOTA dataset. We assessed the performance of state-of-the-art rotation object detection methods on DOTA dataset. We report the results of oriented detectors in Table 4. Our CGD surpasses existing advanced oriented detectors, achieving 76.41% mAP based on R-50-FPN and 77.34% mAP based on R-101-FPN. Our models also achieve the best results in some very challenging categories, such as bridge (BR), ship (SH), storage tank (ST), harbor (HA), and swimming pool (SP). We visualize some detection results and show them in Figure 4.
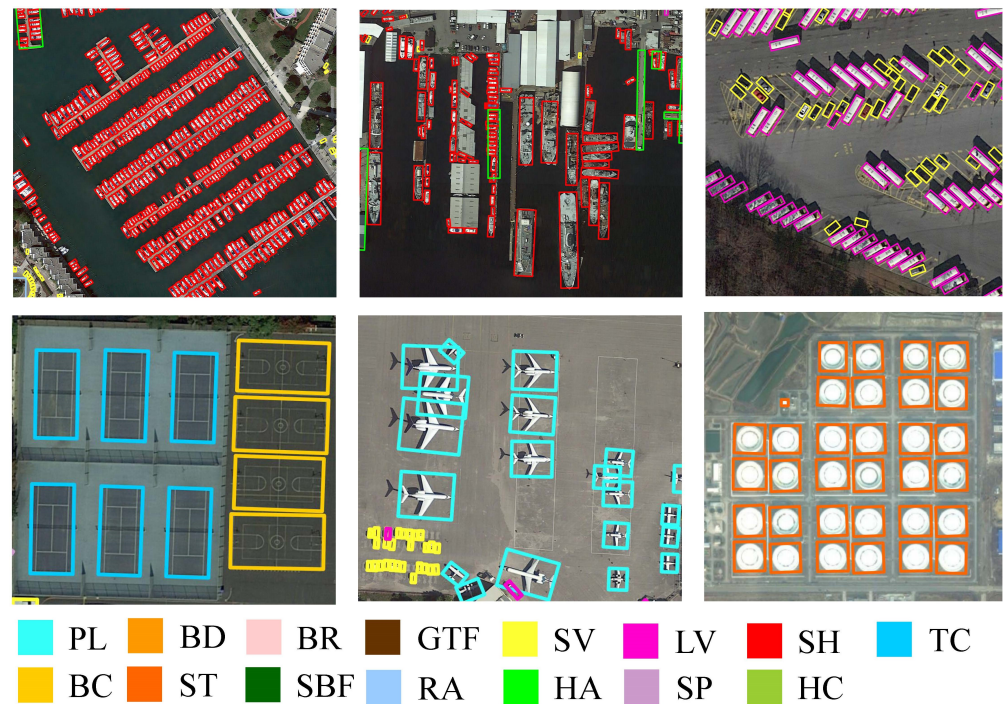

Figure 4. The visualization results of our method on the DOTA dataset. (Best viewed by zooming in).

Table 4. Comparison with state-of-the-art methods on the DOTA dataset.

Method	Backbone	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
Two-stage																	
RRPN [26]	R-101	80.94	65.75	35.34	67.44	59.92	50.91	55.81	90.67	66.92	72.39	55.06	52.23	55.14	53.35	48.22	60.01
R2CNN [6]	R-101	80.94	65.67	35.34	67.44	59.92	50.91	55.81	90.67	66.92	72.39	55.06	52.23	55.14	53.35	48.22	60.67
RoI Transformer [1]	R-101-FPN	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
SCRDet [29]	R-101-FPN	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
Gliding Vertex [2]	R-101-FPN	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
Faster-RCNN-O [7]	R-50-FPN	88.44	73.06	44.86	59.09	73.25	71.49	77.11	90.84	78.94	83.90	48.59	62.95	62.18	64.91	56.18	69.05
AOPG [28]	R-101-FPN	89.14	82.74	51.87	69.28	77.65	**82.42**	88.08	90.89	86.26	85.13	60.60	66.30	74.05	67.76	58.77	75.39
CSL [15]	R-152-FPN	**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																	
RetinaNet-O [8]	R-50-FPN	88.67	77.62	41.81	58.17	74.58	71.64	79.11	90.29	82.18	74.32	54.75	60.60	62.57	69.67	60.64	68.43
DRN [13]	Hourglass-104	88.91	80.22	43.52	63.35	73.48	70.69	84.94	90.14	83.85	84.11	50.12	58.41	67.62	68.60	52.50	70.70
DAL [14]	R-50-FPN	88.68	76.55	45.08	66.80	67.00	76.76	79.74	90.84	79.54	78.45	57.71	62.27	69.05	73.14	60.11	71.44
RSDet [30]	R-101-FPN	89.80	82.90	48.60	65.20	69.50	70.10	70.20	90.50	85.60	83.40	62.50	63.90	65.60	67.20	68.00	72.20
R3Det [31]	R-101-FPN	88.76	83.09	50.91	67.27	76.23	80.39	86.72	90.78	84.68	83.24	61.98	61.35	66.91	70.63	53.94	73.79
S^{2}ANet [23]	R-50-FPN	89.11	82.84	48.37	71.11	**78.11**	78.39	87.25	90.83	84.90	85.64	60.36	62.60	65.26	69.13	57.94	74.12
GWD [12]	R-152-FPN	86.96	83.88	54.36	**77.53**	74.41	68.48	80.34	86.62	83.41	85.55	**73.47**	67.77	72.57	75.76	**73.40**	76.30
Ours																	
CGD	R-50-FPN	89.80	81.66	52.00	73.05	77.55	81.83	88.20	90.86	86.23	86.10	60.08	67.28	76.32	75.07	60.08	**76.41**
CGD	R-101-FPN	90.12	84.33	**55.46**	74.14	75.40	81.26	**88.92**	90.81	83.27	**87.12**	63.71	66.20	**77.13**	**80.50**	61.79	**77.34**

5. Conclusions

In this paper, we analyzed the limitations of regression-based and classification-based oriented object detectors. To address their issues, we propose a circular Gaussian distribution (CGD)-based method for angular prediction. Our method's key insight is converting the labeled angle into a discrete circular Gaussian distribution spanning a single minimal positive period as ground truth, and letting the model predict the distribution. Then, to improve the overall efficiency of the detection model, we design an oriented object detector based on the CenterNet method. The experimental results on the challenging HRSC2016 and DOTA datasets indicated that our proposed CGD could achieve superior performance over the state-of-the-art competitors. In particular, our approach outperforms state-of-the-art competitors with improvements of 1.92% and 1.04% in terms of AP points on the HRSC2016 and DOTA datasets, respectively. It is worth noting that our proposed CGD method can be applied to any task that requires angle prediction, including 3D object detection, panoramic object detection, and so on. In the future, we plan to expand the application of the CGD method to additional tasks.

Author Contributions: Methodology, H.X. and X.L.; Software, H.X.; Validation, H.X.; Writing—original draft preparation, H.X. and X.L.; Writing—review and editing, Z.Z. and S.W.; Supervision, Y.M.; Project administration, C.Y. and F.D.; Funding acquisition, F.D. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data used to support the findings of this study can be found freely at https://captain-whu.github.io/DOTA/dataset.html and https://paperswithcode.com/dataset/hrsc2016 (accessed on 19 June 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ding, J.; Xue, N.; Long, Y.; Xia, G.S.; Lu, Q. Learning RoI Transformer for Oriented Object Detection in Aerial Images. In Proceedings of the International IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–19 June 2019; pp. 2849–2858.
2. Xu , Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.S.; Bai, X. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 1452–1459. [CrossRef] [PubMed]
3. Xu , H.; Zhao, Q.; Ma, Y.; Li, X.; Yuan, P.; Feng, B.; Yan, C.; Dai, F. PANDORA: A Panoramic Detection Dataset for Object with Orientation. In Proceedings of the European Conference on Computer Vision (ECCV), Tel Aviv, Israel, 23–27 October 2022; pp. 237–252.
4. Liu, X.; Xu, H.; Chen, B.; Zhao, Q.; Ma, Y.; Yan, C.; Dai, F. Sph2Pob: Boosting Object Detection on Spherical Images with Planar Oriented Boxes Methods. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), Macao, China, 19–25 August 2023.
5. Xu , H.; Liu, X.; Zhao, Q.; Ma, Y.; Yan, C.; Dai, F. Gaussian Label Distribution Learning for Spherical Image Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 1033–1042.
6. Jiang, Y.; Zhu, X.; Wang, X.; Yang, S.; Li, W.; Wang, H.; Fu, P.; Luo, Z. R2cnn: Rotational region cnn for orientation robust scene text detection. *arXiv* **2017**, arXiv:1706.09579.
7. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149.
8. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
9. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3974–3983.

10. Liu, Z.; Yuan, L.; Weng, L.; Yang, Y. 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, Porto, Portugal, 24–26 February 2017; Volume 2, pp. 324–331.

11. Xu, H.; Liu, X.; Xu, H.; Ma, Y.; Zhu, Z.; Yan, C.; Dai, F. Rethinking Boundary Discontinuity Problem for Oriented Object Detection. *arXiv* **2023**, arXiv:2305.10061.

12. Yang, X.; Yan, J.; Qi, M.; Wang, W.; Xiaopeng, Z.; Qi, T. Rethinking Rotated Object Detection with Gaussian Wasserstein Distance Loss. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021.

13. Pan, X.; Ren, Y.; Sheng, K.; Dong, W.; Yuan, H.; Guo, X.; Ma, C.; Xu, C. Dynamic refinement network for oriented and densely packed object detection. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11207–11216.

14. Ming, Q.; Zhou, Z.; Miao, L.; Zhang, H.; Li, L. Dynamic Anchor Learning for Arbitrary-Oriented Object Detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 2355–2363.

15. Yang, X.; Yan, J. Arbitrary-Oriented Object Detection with Circular Smooth Label. In Proceedings of the European Conference on Computer Vision. Springer, Tel Aviv, Israel, 23–27 October 2020; pp. 677–694.

16. Zhao, T.; Liu, N.; Celik, T.; Li, H.C. An Arbitrary-Oriented Object Detector Based on Variant Gaussian Label in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]

17. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as Points. *arXiv* **2019**, arXiv:1904.07850.

18. Dai, F.; Chen, B.; Xu, H.; Ma, Y.; Li, X.; Feng, B.; Yan, C.; Zhao, Q. Unbiased IoU for Spherical Image Object Detection. In Proceedings of the AAAI, Virtual, 22 February–1 March 2022.

19. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014.

20. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.

21. Lin, T.Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

22. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 3520–3529.

23. Han, J.; Ding, J.; Li, J.; Xia, G.S. Align Deep Features for Oriented Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]

24. Zhang, G.; Lu, S.; Zhang, W. CAD-Net: A Context-Aware Detection Network for Objects in Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 10015–10024. [CrossRef]

25. Yang, X.; Yan, J.; Liao, W.; Yang, X.; Tang, J.; He, T. SCRDet++: Detecting Small, Cluttered and Rotated Objects via Instance-Level Feature Denoising and Rotation Loss Smoothing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 2384–2399. [CrossRef] [PubMed]

26. Ma, J.; Shao, W.; Ye, H.; Wang, L.; Wang, H.; Zheng, Y.; Xue, X. Arbitrary-oriented scene text detection via rotation proposals. *IEEE Trans. Multimedia* **2018**, *20*, 3111–3122. [CrossRef]

27. Zhang, C.; Xiong, B.; Li, X.; Kuang, G. Aspect-Ratio-Guided Detection for Oriented Objects in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]

28. Cheng, G.; Wang, J.; Li, K.; Xie, X.; Lang, C.; Yao, Y.; Han, J. Anchor-Free Oriented Proposal Generator for Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [CrossRef]

29. Yang, X.; Yang, J.; Yan, J.; Zhang, Y.; Zhang, T.; Guo, Z.; Sun, X.; Fu, K. Scrdet: Towards more robust detection for small, cluttered and rotated objects. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–19 June 2019; pp. 8232–8241.

30. Qian, W.; Yang, X.; Peng, S.; Yan, J.; Guo, Y. Learning Modulated Loss for Rotated Object Detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 2458–2466.

31. Yang, X.; Yan, J.; Feng, Z.; He, T. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021; Volume 35, pp. 3163–3171.