PP-YOLOE-R: An Efficient Anchor-Free Rotated Object Detector

Xinxin Wang, Guanzhong Wang, Qingqing Dang, Yi Liu, Xiaoguang Hu, Dianhai Yu
Baidu Inc.
{wangxinxin08, wangguanzhong, dangqingqing, liuyi22, huxiaoguang, yudianhai} @baidu.com

Abstract

Arbitrary-oriented object detection is a fundamental task in visual scenes involving aerial images and scene text. In this report, we present PP-YOLOE-R, an efficient anchor-free rotated object detector based on PP-YOLOE. We introduce a bag of useful tricks in PP-YOLOE-R to improve detection precision with marginal extra parameters and computational cost. As a result, PP-YOLOE-R-l and PP-YOLOE-R-x achieve 78.14 and 78.28 mAP respectively on DOTA 1.0 dataset with single-scale training and testing, which outperform almost all other rotated object detectors. With multi-scale training and testing, PP-YOLOE-R-l and PP-YOLOE-R-x further improve the detection precision to 80.02 and 80.73 mAP. In this case, PP-YOLOE-R-x surpasses all anchor-free methods and demonstrates competitive performance to state-of-the-art anchor-based two-stage models. Further, PP-YOLOE-R is deployment friendly and PP-YOLOE-R-s/m/l/x can reach 69.8/55.1/48.3/37.1 FPS respectively on RTX 2080 Ti with TensorRT and FP16-precision. Source code and pre-trained models are available at PaddleDetection[1], which is powered by PaddlePaddle[2].

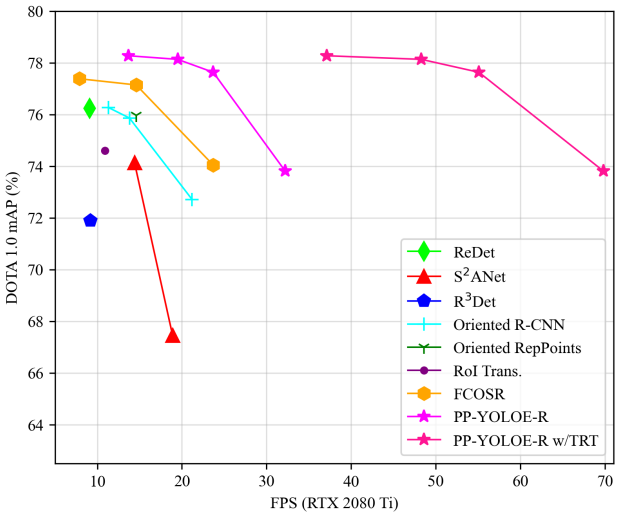

Figure 1: Comparison of PP-YOLOE-R and other state-of-the-art models with single-scale training and testing on DOTA 1.0 dataset. PP-YOLOE-R-s/m/l/x achieves 73.82/77.64/78.14/78.28 mAP respectively at the speed of 69.8/55.1/48.3/37.1 FPS on RTX 2080 Ti with TensorRT and FP16-precision.

1. Introduction

Detecting arbitrary-oriented objects is significant to understand remote sensing images and has attracted increasing attention. Due to massive variations in the scale and orientation of objects, rotated object detection still remains challenging. Benefiting from the rapid development of the horizontal object detection, more and more rotated object detectors[2, 5, 7, 24, 6, 16, 13, 10, 8, 11, 20] have emerged gradually, which are mainly derived from corresponding horizontal object detectors[14, 17, 18, 28]. Among these rotated object detectors, the representation of oriented objects can be roughly divide into three ways, which are rotated bounding boxes with five parameters, quadrangles with eight parameters and a set of key points. Currently, rotated object detectors based on five-parameter representation dominate this research area. Although having achieved promising result, direct five-parameter regression still has some theoretical problem such as boundary discontinuity problem[22, 23, 25]. The boundary discontinuity problem is mainly caused by periodicity of angle and exchange ability of edges while the latter is relative to specific definition of rotated bounding boxes such as long-edge definition. There are a lot of works proposed to resolve the boundary discontinuity problem such as [15, 25, 26, 27, 22, 23]. [15, 25, 26, 27] model rotated bounding box as Gaussian distribution and propose computing-friendly IoU-based loss as a substitute for differentiable SkewIoU loss to avoid direct angle regression. [22, 23] consider angular prediction as classification and design smooth label to avoid boundary discontinuity problem. Fully draw on the excellent ideas of

[1] https://github.com/PaddlePaddle/PaddleDetection
[2] https://github.com/PaddlePaddle/Paddle

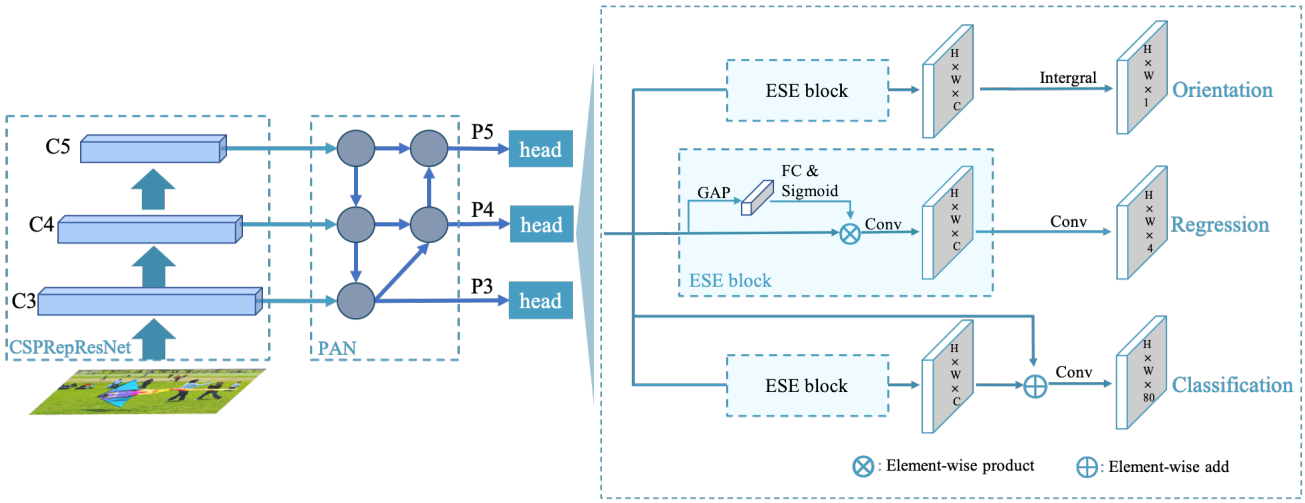

Figure 2: The overall architecture of PP-YOLOE-R. The structure of PP-YOLOE-R is similar to that of PP-YOLOE, except that a decoupled angle prediction head is introduced into PP-YOLOE-R.

advanced horizontal and oriented detectors, we propose PP-YOLOE-R, an efficient anchor-free rotated object detector based on PP-YOLOE[21]

Compared with PP-YOLOE, the main changes of PP-YOLOE-R can be attributed to four aspects: (1) we introduce ProbIoU loss[15] like [13] as regression loss to avoid boundary discontinuity problem. (2) we introduce Rotated Task Alignment Learning to be suitable for rotated object detection on the basis of Task Alignment Learning[4] (3) we design a decoupled angle prediction head and directly learn the General distribution of angle through DFL loss[12] for a more accurate angle prediction. (4) we make a slight modification to the re-parameterization mechanism[3] by adding a learnable gating unit to control the amount of information from the previous layer. As a result, PP-YOLOE-R achieves state-of-the-art performance in terms of speed and accuracy trade-off on DOTA 1.0 dataset. Specifically, PP-YOLOE-R-l and PP-YOLOE-R-x achieve 78.14 and 78.28 mAP respectively with single-scale training and testing. With multi-scale training and testing, PP-YOLOE-R-l and PP-YOLOE-R-x further improve the detection precision to 80.02 and 80.73 mAP respectively. While maintaining high precision, PP-YOLOE-R-l can achieve the speed of 48.3 FPS at 1024×1024 resolution with TensorRT and FP16-precision. Moreover, PP-YOLOE-R-s and PP-YOLOE-R-m also have excellent performance and are suitable for edge devices with relatively low computing power. Our code can be available at PaddleDetection[1].

2. Related work

Anchor-Based Rotated Object Detectors. Anchor-based rotated object detectors have one-stage and two-stage approches similar to horizontal object detectors. RoI Transformer[2] proposes a RRoI learner to predict the offsets of Rotated Ground Truths (RGTs) relative to predicted HRoI. Oriented R-CNN[20] designs a lightweight oriented RPN to generate high-quality oriented proposals. ReDet[7] introduces rotation-equivariant CNN (ReCNN) to obtain rotation-equivariant feature maps and Rotation-invariant RoI Align (RiRoI Align) to extract features of RRoIs. S^2ANet[6] and R^3Det[24] both adopt the refined single-stage framework to detect oriented objects. S^2ANet proposes Alignment Convolution Layer (ACL) based on Deformable Convolution Network while R^3Det designs Feature Refinement Module (FRM) based on interpolation to alleviate feature misalignment.

Anchor-Free Rotated Object Detectors. Anchor-free rotated object detectors are mainly based on center point or a set of key points. DAFNe[10] proposes oriented centerness and center-to-corner prediction strategy while FCOSR[13] focuses on the label assignment strategy based on FCOS[18] to improve detection performance. CFA[5] and Oriented RepPoints[11] indirectly predict oriented bounding boxes by predicting nine representative points based on RepPoints[28].

Label Assignment. The purpose of label assignment is to distinguish positive and negative samples. Label assignment can be divide into static and dynamic strategies. During training, dynamic label assignment utilizes the output of model as a basis for selecting positive and negative samples while static label assignment determines positive and negative samples according to ground truths and predefined rules. FCOSR[13] proposes ellipse center sampling method, fuzzy sample assignment strategy and multi-level sampling module to ease insufficient sampling prolem. [8]

Model	mAP(%)	Parameters(M)	GFLOPs
baseline	75.61	50.65	269.09
+Rotated Task Alignment Learning	77.24 (+**1.63**)	50.65	269.09
+Decoupled Angle Prediction Head	77.78 (+**0.54**)	52.20	272.72
+Angle Prediction with DFL	78.01 (+**0.23**)	53.29	281.65
+Learnable Gating Unit for RepVGG	**78.14** (+**0.13**)	53.29	281.65

Table 1: Ablation study of PP-YOLOE-R-l on DOTA 1.0 dataset with single-scale training and testing. All parameters and GFLOPs are calculated after re-parameterization and the resolution of the input image is 1024×1024

proposes Shape-Adaptive Selection (SA-S) to adjust IoU threshold according to the shape of samples. G-Rep[9] substitutes IoU with normalized Gaussian distribution distance as an assignment indicator. DAL[16] introduces a matching degree considering a prior of spatial matching and feature alignment ability to dynamic select positive samples. Similarly, Oriented RepPoints[11] designs a quality measure to assign samples.

Loss. Due to periodicity of angle and exchange ability of edges, direct regression-based rotated object detectors suffer boundary discontinuity problem. CSL[23] and DCL[22] predict angle in a classification way. To avoid boundary discontinuity problem, GWD[25], ProbIoU[15], KLD[26] and KFIoU[27] converts the rotated bounding box to a 2D Gaussian distribution and constructs a distance metric of two Gaussian distributions to measure the similarity of two rotated bounding boxes. GWD[25] utilizes Gaussian Wasserstein distance to approximate SkewIoU while ProbIoU[15] utilizes Bhattacharyya Coefficient to measure the similarity of two rotated bounding boxes. KLD[26] calculates Kullback-Leibler Divergence (KLD) between two Gaussian distributions as the regression loss. Moreover, KFIoU[27] achieves trend-level alignment with SkewIoU by adopting Kalman filter to mimic SkewIoU according to its definition.

3. Method

As shown in Fig. 2, the overall architecture of PP-YOLOE-R is similar to that of PP-YOLOE. PP-YOLOE-R improves detection performance of rotated bounding boxes at the expense of relatively small amount of parameters and computation based on PP-YOLOE. In this section, we will introduce the changes made for rotated bounding boxes in detail.

Baseline. Drawing on FCOSR[13], we introduce FCOSR Assigner and ProbIoU Loss into PP-YOLOE as our baseline. FCOSR Assigner is adopted to assign ground truth to three feature maps according to the predefined rules and ProbIoU Loss is utilized as regression loss. The backbone and neck of our baseline remain the same as PP-YOLOE while the regression branch of head is modified to predict five-parameter rotated bounding boxes (x, y, w, h, θ) directly. Our baseline achieves 75.61 mAP with single-scale training and testing as shown in Table 1.

Rotated Task Alignment Learning. Task Alignment Learning[4] is composed of a task-aligned label assignment and task-aligned loss. Task-aligned label assignment constructs a task alignment metric to select positive samples from candidate anchor points, whose coordinates fall into any ground truth boxes. The task alignment metric is calculated as follows:

$$t = s^{\alpha} \cdot \mu^{\beta} \tag{1}$$

where s denotes a predicted classification score and μ denotes the IoU value between predicted bounding box and corresponding ground truth. In Rotated Task Alignment Learning, the selection process of candidate anchor points takes advantage of the geometric properties of the ground truth bounding box and anchor points in it and the SkewIoU value of predicted and ground truth bounding box is adopted as μ. With these two simple changes, we can apply task-aligned label assignment for rotated object detection and use task-aligned loss without modification. By using Rotated Task Alignment Learning, the detection precision is further improved to 77.24 mAP as shown in Table 1.

Decoupled Angle Prediction Head. In the regression branch, most rotated object detectors predict five parameters (x, y, w, h, θ) to represent the oriented object. However, we assume that predicting θ requires different features than predicting (x, y, w, h). To verify this hypothesis, we design a decoupled angle prediction head as shown in Figure 2 to predict θ and (x, y, w, h) separately. The angle prediction head consists of a channel attention layer and a convolution layer and is very lightweight. By introducing the decoupled angle prediction head, the detection precision is improved by 0.54 mAP to 77.24 mAP as shown in Table 1.

Angle Prediction with DFL. ProbIoU Loss is adopted as regression loss to jointly optimize (x, y, w, h, θ). To calculate ProbIoU Loss, the rotated bounding box is converted to a Gaussian bounding box. When the rotated bounding box is roughly square, the orientation of the rotated bounding box cannot be determined because the orientation in Gaussian bounding box is inherited from the elliptical represen-

Methods	Backbone	MS	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
Anchor-based Methods																		
RoI-Trans.[2]	R101	✓	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
DAL[16]	R101		88.61	79.69	46.27	70.37	65.89	76.10	78.53	90.84	79.98	78.41	58.71	62.02	69.23	71.32	60.65	71.78
CSL[23]	R152	✓	**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
R^3Det[24]	R152	✓	89.80	83.77	48.11	66.77	78.76	83.27	87.84	90.82	85.38	85.51	65.57	62.68	67.53	78.56	72.62	76.47
DCL[22]	R152	✓	89.26	83.60	53.54	72.76	79.04	82.56	87.31	90.67	86.59	86.98	67.49	66.88	73.29	70.56	69.99	77.37
S^2ANet[6]	R50	✓	88.89	83.60	57.74	81.95	79.94	83.19	**89.11**	90.78	84.87	87.81	70.30	68.25	78.30	77.01	69.58	79.42
ReDet[7]	ReR50	✓	88.81	82.48	60.83	80.82	78.34	86.06	88.31	90.87	**88.77**	87.03	68.65	66.90	79.26	**79.71**	74.67	80.10
GWD[25]	R152	✓	89.66	84.99	59.26	82.19	78.97	84.83	87.70	90.21	86.54	86.85	**73.47**	67.77	76.92	79.22	74.92	80.23
KLD[26]	R152	✓	89.92	85.13	59.19	81.33	78.82	84.38	87.50	89.80	87.33	87.00	72.57	71.35	77.12	79.34	**78.68**	80.63
Oriented R-CNN[20]	R50	✓	89.84	85.43	61.09	79.82	79.71	85.35	88.82	90.88	86.68	87.73	72.21	70.80	**82.42**	78.18	74.11	80.87
RoI-Trans. + KFIoU[27]	Swin-Tiny	✓	89.44	84.41	**62.22**	**82.51**	**80.10**	**86.07**	88.68	**90.90**	87.32	**88.38**	72.80	**71.95**	78.96	74.95	75.27	**80.93**
Anchor-free Methods																		
BBAVectors[29]	R101	✓	88.63	84.06	52.13	69.56	78.26	80.40	88.06	90.87	87.23	86.39	56.11	65.62	67.10	72.08	63.96	75.36
CFA[5]	R152	✓	89.08	83.20	54.37	66.87	81.23	80.96	87.17	90.21	84.32	86.09	52.34	69.94	75.52	80.76	67.96	76.67
Oriented RepPoints[11]	Swin-Tiny		89.11	82.32	56.71	74.95	80.70	83.73	87.67	90.81	87.11	85.85	63.60	68.60	75.95	73.54	63.76	77.63
FCOSR-s[13]	MV2		89.09	80.58	44.04	73.33	79.07	76.54	87.28	**90.88**	84.89	85.37	55.95	64.56	66.92	76.96	55.32	74.05
FCOSR-s[13]	MV2	✓	88.60	84.13	46.85	78.22	79.51	77.00	87.74	90.85	86.84	86.71	64.51	68.17	67.87	72.08	62.52	76.11
FCOSR-m[13]	X50		88.88	82.68	50.10	71.34	81.09	77.40	88.32	90.80	86.03	85.23	61.32	68.07	75.19	80.37	70.48	77.15
FCOSR-m[13]	X50	✓	89.06	84.93	52.81	76.32	81.54	81.81	88.27	90.86	85.20	87.58	68.63	**70.38**	75.95	79.73	75.67	79.25
FCOSR-l[13]	X101		**89.50**	84.42	52.58	71.81	80.49	77.72	88.23	90.84	84.23	86.48	61.21	67.77	76.34	74.39	74.86	77.39
FCOSR-l[13]	X101	✓	88.78	**85.38**	54.29	76.81	81.52	82.76	88.38	90.80	86.61	87.25	67.58	67.03	76.86	73.22	74.68	78.80
PP-YOLOE-R-s	CRN-s		88.80	79.24	45.92	66.88	80.41	82.95	88.20	90.61	82.91	86.37	55.80	64.11	65.09	79.50	50.43	73.82
PP-YOLOE-R-s	CRN-s	✓	88.93	83.95	56.60	**79.40**	82.57	85.89	88.64	90.87	87.82	87.54	68.94	63.46	76.66	79.19	70.87	79.42
PP-YOLOE-R-m	CRN-m		89.23	79.92	51.14	72.94	81.86	84.56	88.68	90.85	86.85	87.48	59.16	68.34	73.78	81.72	68.10	77.64
PP-YOLOE-R-m	CRN-m	✓	88.63	84.45	56.27	79.12	**83.52**	**86.16**	88.77	90.81	88.01	**88.39**	**70.41**	61.44	77.65	77.70	74.30	79.71
PP-YOLOE-R-l	CRN-l		89.18	81.00	54.01	70.22	81.85	85.16	88.81	90.81	86.99	88.01	62.87	67.87	76.56	79.13	69.65	78.14
PP-YOLOE-R-l	CRN-l	✓	88.40	84.75	58.91	76.35	83.13	86.10	88.79	90.87	**88.74**	87.71	67.71	68.44	**77.92**	76.17	76.35	80.02
PP-YOLOE-R-x	CRN-x		89.49	79.70	55.04	75.59	82.40	85.20	88.35	90.76	85.69	87.70	63.17	69.52	77.09	75.08	69.38	78.28
PP-YOLOE-R-x	CRN-x	✓	88.45	84.46	**60.57**	77.70	83.34	85.36	**88.97**	90.78	88.53	87.47	69.26	65.96	77.86	**81.36**	**80.93**	**80.73**

Table 2: Comparison with state-of-the-art methods on DOTA 1.0 dataset. R50 and X50 denote ResNet-50 and ResNeXt-50 (likewise for R101, R152 and X101). MV2 denotes MobileNetv2 and CRN denotes CSPRepResNet. MS means multi-scale training and testing. The bold **red** fonts indicate the best performance.

tation. To overcome this problem, we introduce Distribution Focal Loss (DFL) [12] to predict the angle. Different form l_n-norm learning the Dirac delta distribution, DFL is aimed at learning the General distribution of angle. Specifically, we discretize the angle with even intervals ω and obtain the preidcted θ in the form of integral, which can be formulated as follows:

$$\theta = \sum_{i=0}^{90} p_i \cdot i \cdot \omega \tag{2}$$

where p_i means the probability that the angle falls in each interval. In this paper, the rotated bounding box is defined in OpenCV definition and ω is set to $\pi/180$. By introducing DFL, the detection precision is improved by 0.23 mAP to 78.01 mAP.

Learnable Gating Unit for RepVGG. RepVGG proposes a multi-branch architecture composed of a 3×3 conv, a 1×1 conv and a shortcut path. The training-time information flow of RepVGG can be formulated as follows:

$$y = f(x) + g(x) + x \tag{3}$$

while $f(x)$ is a 3×3 conv and $g(x)$ is a 1×1 conv. During inference, we can re-parameterizes this architecture to an equivalent 3×3 conv. Although RepVGG is equivalent to a convolution layer, using RepVGG during training converges better. We attribute this result to that the design of RepVGG introduces useful prior knowledge. Inspired by this, we introduce a learnable gating unit in RepVGG to control the amount of information from previous layer. This design is mainly for tiny objects or dense objects to adaptively fuse the features with different receptive fields, which can be formulated as follows:

$$y = f(x) + \alpha_1 \cdot g(x) + \alpha_2 \cdot x \tag{4}$$

where α_1 and α_2 are learnable parameters. In our RepResBlock[21], the shortcut path is not used so we introduce only one parameter for each RepResBlock. During inference, the learnable parameter can be re-parameterized along with convolution layers so that neither the speed nor the amount of parameters changes. By introducing learnable gating unit, the detection precision is improved by 0.13 mAP to 78.14 mAP.

ProbIoU Loss. By modeling rotated bounding boxes as Gaussian bounding boxes, the Bhattacharyya Coefficient of two Gaussian distributions is used to measure the similarity of two rotated bounding boxes in ProbIoU[15]. GWD[25], KLD[26] and KFIoU [27] are also similarity measures based on Gaussian bounding boxes. To verify the effect of ProbIoU loss, we choose KLD loss for experiment because KLD loss is scale invariant and suitable for anchor-free methods. As shown in Table 3, replacing ProbIoU loss with KLD loss causes a significant preformance drop from

78.14 mAP to 76.03 mAP, which indicates that ProbIoU loss is more suitable for our design.

Loss	mAP(%)
ProbIoU loss[15]	**78.14**
KLD loss[26]	76.03

Table 3: Ablation study of different loss on PP-YOLOE-R-l

4. Experiment

4.1. Dataset

DOTA[19] is a large-scale remote sensing dataset for oriented object detection, which contains 15 categories: 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 is comprised of 2806 aerial images with the size about 4000 × 4000 pixels and 188,282 instances with a wide variety of scales, orientations, and shapes. Half of the aerial images are randomly selected as the training set, 1/6 as the validation set, and 1/3 as the testing set. For single-scale training and testing, we crop the original images into 1024×1024 patches with an overlap of 256 pixels. For multi-scale training and testing, the original images are resized with the scale of 0.5, 1.0 and 1.5 and then cropped into 1024×1024 patches with an overlap of 500 pixels.

4.2. Implementation details

PP-YOLOE-R adopts CSPRepResNet as backbone and PAN as neck to extract P3, P4 and P5 pyramid features for rotated object detection. The stochastic gradient descent (SGD) with momentum = 0.9 and weight decay = 5e-4 is used in our training. The initial learning rate is set to 0.008 with the warming up for 1000 iterations and cosine learning rate schedule is used after warming up. We train all the models with 36 epochs for DOTA 1.0 dataset and use 4 Tesla V100 GPU devices with 32G memory for training with the total batch size of 8. During training process, the exponential moving average (EMA) strategy with decay = 0.9998 is also adopted. We adopt random flip and adopt a two step rotation augmentation method following FCOSR[13] to generate random augmentation data.

4.3. Comparsion with Other SOTA Detectors

We conduct extensive experiments on the DOTA 1.0 dataset and the experimental results are shown in Table 2. With single-scale training and testing, PP-YOLOE-R-l and PP-YOLOE-R-x achieves 78.14 and 78.28 mAP respectively, which outperform almost all rotated object detectors. With multi-scale training and testing, PP-YOLOE-R-l and PP-YOLOE-R-x further improve the detection precision to 80.02 and 80.73 mAP. PP-YOLOE-R-x outperforms all anchor-free methods and is only 0.2 mAP lower than the two-stage anchor-based model with the highest precision. Moreover, PP-YOLOE-R-s and PP-YOLOE-R-m can achieve 79.42 and 79.71 mAP with multi-scale training and testing, which are excellent results considering the parameters and GLOPS of these two models. While maintaining high precision, PP-YOLOE-R avoids using special operators, such as Deformable Convolution or Rotated RoI Align, to be deployed friendly on various hardware. As a result, PP-YOLOE-R can be easily accelerated with TensorRT whereas most other SOTA models are currently not easy to deploy using TensorRT. At the input resolution of 1024×1024, PP-YOLOE-R-s/m/l/x can reach 32.2/23.7/19.5/13.7 FPS on RTX 2080Ti. With TensorRT and FP-16 precision, PP-YOLOE-R-s/m/l/x can be further accelerated to 69.8/55.1/48.3/37.1 FPS respectively.

5. Conclusion

In this report, we propose PP-YOLOE-R, an efficient anchor-free rotated object detector based on PP-YOLOE. PP-YOLOE-R achieves high precision and real-time speed with marginal extra parameters and computational cost, surpassing all anchor-free rotated object detectors. PP-YOLOE-R is easy to deploy and has a series of models for different computing power devices, named s/m/l/x. In the future, we will conduct experiments on more datasets for rotated object detection and extend PP-YOLOE-R to related scenes.

References

[1] PaddlePaddle Authors. Paddledetection, object detection and instance segmentation toolkit based on paddlepaddle. https://github.com/PaddlePaddle/PaddleDetection, 2019. 2

[2] Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, and Qikai Lu. Learning roi transformer for oriented object detection in aerial images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2849–2858, 2019. 1, 2, 4

[3] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13733–13742, 2021. 2

[4] Chengjian Feng, Yujie Zhong, Yu Gao, Matthew R Scott, and Weilin Huang. Tood: Task-aligned one-stage object detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3490–3499. IEEE Computer Society, 2021. 2, 3

[5] Zonghao Guo, Chang Liu, Xiaosong Zhang, Jianbin Jiao, Xiangyang Ji, and Qixiang Ye. Beyond bounding-box: Convex-hull feature adaptation for oriented and densely packed object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8792–8801, 2021. 1, 2, 4

[6] Jiaming Han, Jian Ding, Jie Li, and Gui-Song Xia. Align deep features for oriented object detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–11, 2021. 1, 2, 4

[7] Jiaming Han, Jian Ding, Nan Xue, and Gui-Song Xia. Redet: A rotation-equivariant detector for aerial object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2786–2795, 2021. 1, 2, 4

[8] Liping Hou, Ke Lu, Jian Xue, and Yuqiu Li. Shape-adaptive selection and measurement for oriented object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022. 1, 2

[9] Liping Hou, Ke Lu, Xue Yang, Yuqiu Li, and Jian Xue. G-rep: Gaussian representation for arbitrary-oriented object detection. *arXiv preprint arXiv:2205.11796*, 2022. 3

[10] Steven Lang, Fabrizio Ventola, and Kristian Kersting. Dafne: A one-stage anchor-free deep model for oriented object detection. *arXiv preprint arXiv:2109.06148*, 2021. 1, 2

[11] Wentong Li, Yijie Chen, Kaixuan Hu, and Jianke Zhu. Oriented reppoints for aerial object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1829–1838, 2022. 1, 2, 3, 4

[12] Xiang Li, Wenhai Wang, Lijun Wu, Shuo Chen, Xiaolin Hu, Jun Li, Jinhui Tang, and Jian Yang. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Advances in Neural Information Processing Systems*, 33:21002–21012, 2020. 2, 4

[13] Zhonghua Li, Biao Hou, Zitong Wu, Licheng Jiao, Bo Ren, and Chen Yang. Fcosr: A simple anchor-free rotated detector for aerial object detection. *arXiv preprint arXiv:2111.10780*, 2021. 1, 2, 3, 4, 5

[14] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 1

[15] Jeffri M Llerena, Luis Felipe Zeni, Lucas N Kristen, and Claudio Jung. Gaussian bounding boxes and probabilistic intersection-over-union for object detection. *arXiv preprint arXiv:2106.06072*, 2021. 1, 2, 3, 4, 5

[16] Qi Ming, Zhiqiang Zhou, Lingjuan Miao, Hongwei Zhang, and Linhao Li. Dynamic anchor learning for arbitrary-oriented object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 2355–2363, 2021. 1, 3, 4

[17] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 1

[18] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9627–9636, 2019. 1, 2

[19] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3974–3983, 2018. 5

[20] Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao, and Junwei Han. Oriented r-cnn for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3520–3529, 2021. 1, 2, 4

[21] Shangliang Xu, Xinxin Wang, Wenyu Lv, Qinyao Chang, Cheng Cui, Kaipeng Deng, Guanzhong Wang, Qingqing Dang, Shengyu Wei, Yuning Du, et al. Pp-yoloe: An evolved version of yolo. *arXiv preprint arXiv:2203.16250*, 2022. 2, 4

[22] Xue Yang, Liping Hou, Yue Zhou, Wentao Wang, and Junchi Yan. Dense label encoding for boundary discontinuity free rotation detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15819–15829, 2021. 1, 3, 4

[23] Xue Yang and Junchi Yan. Arbitrary-oriented object detection with circular smooth label. In *European Conference on Computer Vision*, pages 677–694. Springer, 2020. 1, 3, 4

[24] Xue Yang, Junchi Yan, Ziming Feng, and Tao He. R3det: Refined single-stage detector with feature refinement for rotating object. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 3163–3171, 2021. 1, 2, 4

[25] Xue Yang, Junchi Yan, Qi Ming, Wentao Wang, Xiaopeng Zhang, and Qi Tian. Rethinking rotated object detection with gaussian wasserstein distance loss. In *International Conference on Machine Learning*, pages 11830–11841. PMLR, 2021. 1, 3, 4

[26] Xue Yang, Xiaojiang Yang, Jirui Yang, Qi Ming, Wentao Wang, Qi Tian, and Junchi Yan. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. *Advances in Neural Information Processing Systems*, 34:18381–18394, 2021. 1, 3, 4, 5

[27] Xue Yang, Yue Zhou, Gefan Zhang, Jitui Yang, Wentao Wang, Junchi Yan, Xiaopeng Zhang, and Qi Tian. The kfiou loss for rotated object detection. *arXiv preprint arXiv:2201.12558*, 2022. 1, 3, 4

[28] Ze Yang, Shaohui Liu, Han Hu, Liwei Wang, and Stephen Lin. Reppoints: Point set representation for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9657–9666, 2019. 1, 2

[29] Jingru Yi, Pengxiang Wu, Bo Liu, Qiaoying Huang, Hui Qu, and Dimitris Metaxas. Oriented object detection in aerial images with box boundary-aware vectors. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2150–2159, 2021. 4