Learning Modulated Loss for Rotated Object Detection

Wen Qian[1,2] Xue Yang[3,4] Silong peng[1,2,*] Junchi Yan[3,4] Yue Guo[1,2]

{qianwen2018, silong.peng, guoyue2013}@ia.ac.cn, {yangxue-2019-sjtu, yanjunchi}@sjtu.edu.cn,

[1]Institute of Automation, Chinese Academy of Sciences

[2]University of Chinese Academy of Sciences

[3]Department of Computer Science and Engineering, Shanghai Jiao Tong University

[4] MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University

Introduction

- **Rotation Object Detection**

Task: Solving the discontinuity of loss which is caused by the contradiction between the definition of the rotated bounding box and the loss function.

➢ Challenges

- **Parameterization of rotated bounding box**. two mainstream protocols for bounding box parameterization i.e. the five- parameter and eight-parameter models.
- **Discontinuity of Loss**. The case exists both in the five-parameter and eight-parameter models cased by the contradiction between the definition of the rotated bounding box and the loss function.

- **Our main contributions**

➢ Formulate the important while relatively ignored rotation sensitivity error (RSE) for region-based rotation detectors, which refers to the loss discontinuity.

➢ For the traditionally widely used five-parameter system and eight-parameter system, we devise a special treatment to ensure the loss continuity. The new loss is termed by L_{mr}

➢ Based on L_{mr}, we respectively extend it to the one- stage and two-stage detection frameworks, which show state-of-the-art performance on DOTA and UCAS-AOD benchmarks.

➢ Codes: https://github.com/Mrqianduoduo/RSDet-8P-4R

Proposed Approach

- **Overview**

➢ In this section, we firstly present two mainstream protocols for bounding box parameterization i.e. the five- parameter and eight-parameter models. Then we formally determine the loss discontinuity in the five-parameter and eight-parameter methods. We call such issues collectively as rotation sensitivity error (RSE) and propose a modulated rotation loss to achieve more smooth learning.

- **Parameterization of Rotated Bounding Box**

➢ Our five-parameter definition is in line with that in OpenCV.

➢ The definition of eight-parameter is more simple: starting from the lower left corner, four clockwise vertices (a, b, c, d) of the rotated bounding box are used to describe its location.

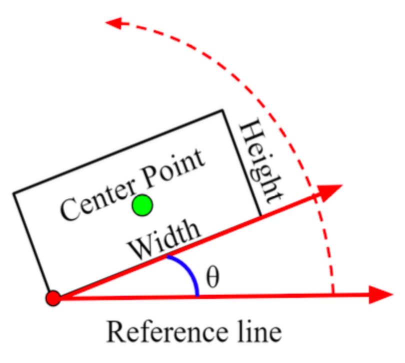


Fig1: The five-parameter definition

- **RSE in Five-parameter Methods**

➢ RSE is mainly caused by two reasons: i) The adoption of the angle param- eter and the exchange between width and height contribute to the sudden loss change (increase) in the boundary case. ii) Regression inconsistency of measure units exists in the five-parameter model.

➢ The angle parameter causes the loss discontinuity. To obtain the predicted box that coin- cides with the ground truth box, the horizontal reference box is rotated counterclockwise, as shown in Fig.2a.

➢ Different measurement units of five parameters make regression inconsistent. However, the impact of such artifacts is still unclear and has been rarely studied in the literature. Relationships among all the parameters and IoU are empirically studied in Fig. 3.

- **RSE in Eight-parameter Methods**

➢ The discontinuity of loss still exists in the eight- parameter regression model. Therefore, consider the situation of an eight-parameter regression in the boundary case, as shown in Fig. 2b.

- **The Proposed Modulated Rotation Loss**

$$\ell_{mr} = \min \begin{cases} \ell_1(para.) \\ \ell_1(modulated - para.). \end{cases} \tag{1}$$

➢ Five-parameter Modulated Rotation Loss

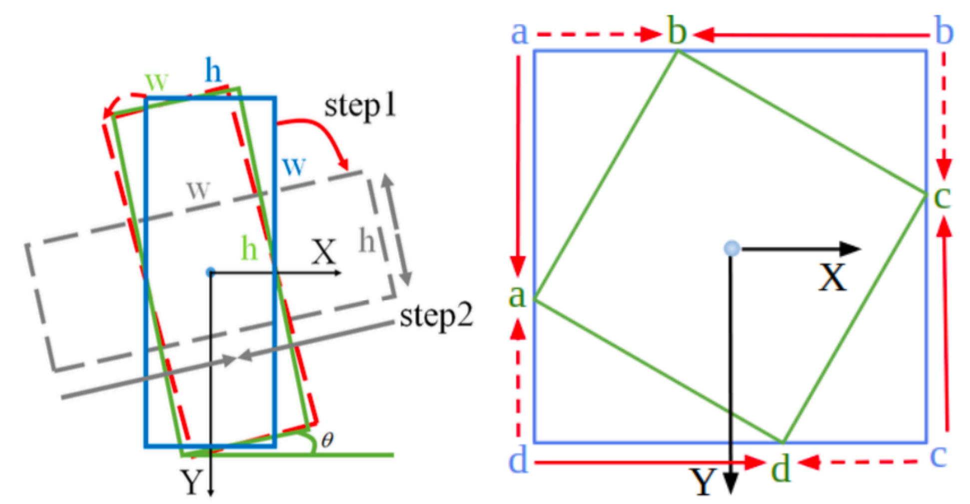
In this paper, we devise the following boundary constraints to modulate the loss as termed by modulated rotation loss L_{mr} :

$$\ell_{cp} = |x_1 - x_2| + |y_1 - y_2|, \tag{2}$$

$$\ell^{5p}_{mr} = \min \begin{cases} \ell_{cp} + |w_1 - w_2| + |h_1 - h_2| + |\theta_1 - \theta_2| \\ \ell_{cp} + |w_1 - h_2| + |h_1 - w_2| \\ \quad + |90 - |\theta_1 - \theta_2|| , \end{cases} \tag{3}$$

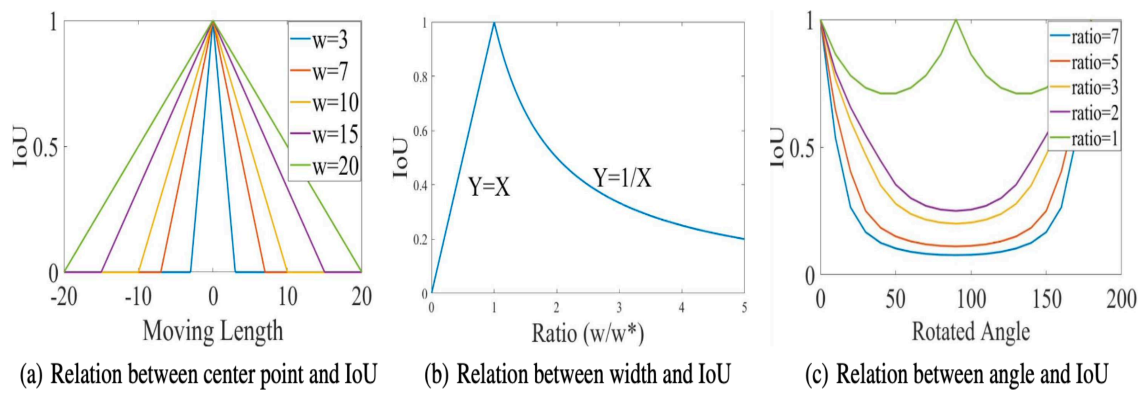
➢ Eight-parameter Modulated Rotation Loss

$$\ell^{8p}_{mr} = \min \begin{cases} \sum_{i=0}^{3}\left(\frac{|x_{(i+3)\%4} - x_i^*|}{w_a} + \frac{|y_{(i+3)\%4} - y_i^*|}{h_a}\right) \\ \sum_{i=0}^{3}\left(\frac{|x_i - x_i^*|}{w_a} + \frac{|y_i - y_i^*|}{h_a}\right) \\ \sum_{i=0}^{3}\left(\frac{|x_{(i+1)\%4} - x_i^*|}{w_a} + \frac{|y_{(i+1)\%4} - y_i^*|}{h_a}\right) \end{cases}$$



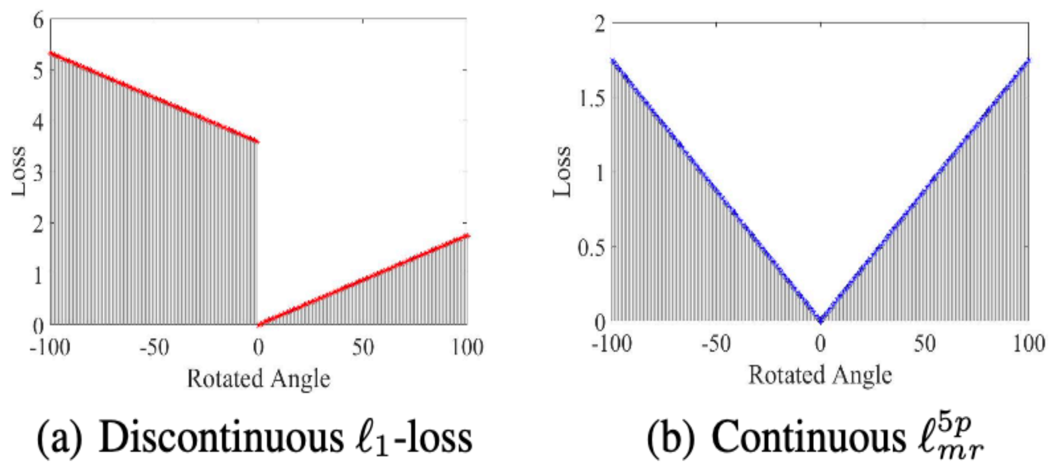
(a) 5-parameter regression w/ two steps in boundary condition procedure (b) Eight-parameter regression

Fig2: Boundary discontinuity analysis of five-parameter regression and eight-parameter regression. The red solid ar- row indicates the actual regression process, and the red dot- ted arrow indicates the ideal regression process.



(a) Relation between center point and IoU (b) Relation between width and IoU (c) Relation between angle and IoU

Fig3: Inconsistency in five-parameter regression model. between width and IoU.



(a) Discontinuous ℓ_1-loss (b) Continuous ℓ^{5p}_{mr}

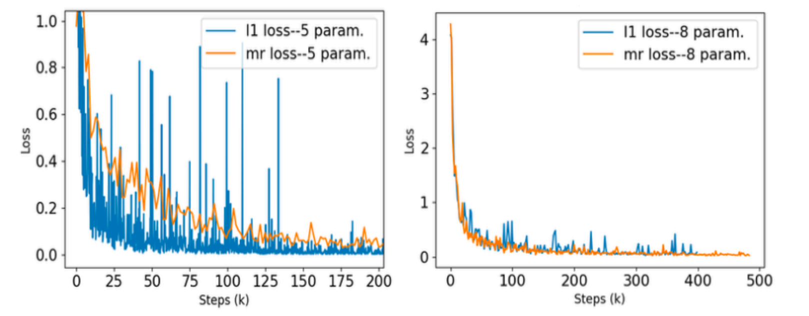
Fig4: Comparison between two loss functions.

Experiments

➢ Ablation experiments of L_{mr} on DOTA benchmark.

Backbone	Loss	Regression	mAP
resnet-50	smooth-ℓ_1	five-param.	62.14
resnet-50	ℓ_{mr}	five-param.	64.49
resnet-50	smooth-ℓ_1	eight-param.	65.59
resnet-50	ℓ_{mr}	eight-param.	**66.77**

➢ Comparisons of loss curves during training.



(a) Loss curves (five-param.) (b) Loss curves (eight-param.)

➢ Ablation study using the proposed techniques on DOTA.

Loss	Regression	mAP
smooth-ℓ_1	five-param. $[-\frac{\pi}{2}, 0)$	62.14
smooth-ℓ_1(Xia et al. 2018)	five-param. $[-\pi, 0)$	62.39
IoU-smooth-ℓ_1 (Yang et al. 2019c)	five-param. $[-\frac{\pi}{2}, 0)$	62.69
ℓ_{mr}	five-param. $[-\frac{\pi}{2}, 0)$	64.49
smooth-ℓ_1	eight-param.	65.59
ℓ_{mr}	eight-param.	**66.77**

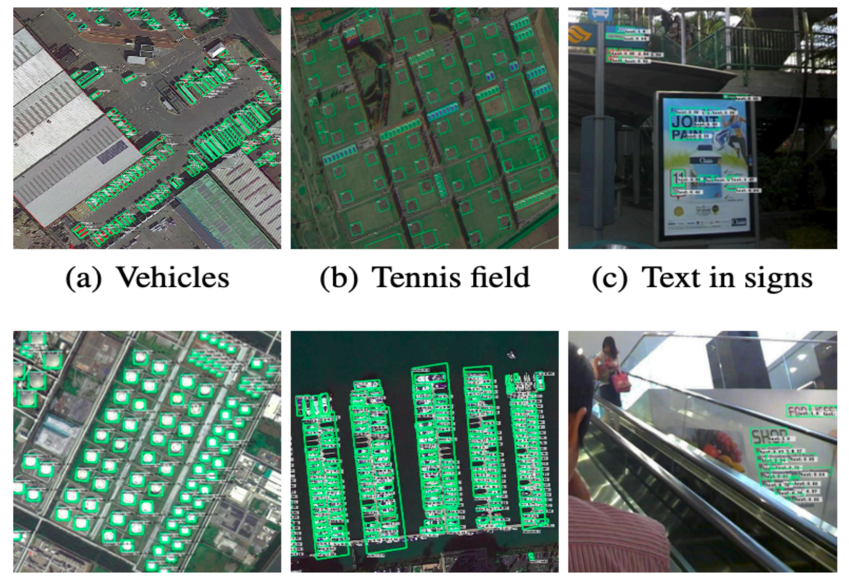
➢ Ablation experiments of backbone, data augmentation and balance.

Backbone	Data Aug	Balance	mAP
resnet-50			66.77
resnet-50	✓		70.79
resnet-50	✓	✓	71.22
resnet-101	✓	✓	72.16
resnet-152	✓	✓	**73.51**

➢ Detection accuracy on on DOTA.

Method	Backbones	Regression Methods	MS	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
Two-stage methods																			
FR-O (Xia et al. 2018)	ResNet101	5-para		79.09	69.12	17.17	63.49	34.20	37.16	36.20	89.19	69.60	58.96	49.4	52.52	46.69	44.80	46.30	52.93
IENet (Lin, Feng, and Guan 2019)	ResNet101	6-para	✓	80.20	64.54	39.82	32.07	49.71	65.01	52.58	81.45	44.66	78.51	46.54	56.73	64.40	64.24	36.75	57.14
R-DFPN (Yang et al. 2018b)	ResNet101	5-para		80.92	65.82	33.77	58.94	55.77	50.94	54.78	90.33	66.34	68.66	48.73	51.76	55.10	51.32	35.88	57.94
R^2CNN (R22)	ResNet101	5-para		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
RRPN (Ma et al. 2018)	ResNet101	5-para		88.52	71.20	31.66	59.30	51.85	56.19	57.25	90.81	72.84	67.38	56.69	52.84	53.08	51.94	53.58	61.01
ICN (Azimi et al. 2018)	ResNet101	5-para	✓	81.40	74.30	47.70	70.30	64.90	67.80	70.00	90.80	79.10	78.20	53.60	62.90	67.00	64.20	50.20	68.20
RoI Transformer (Ding et al. 2019)	ResNet101	5-para	✓	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-Net (Zhang, Lu, and Zhang 2019)	ResNet101	5-para		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 (Yang et al. 2019c)	ResNet101	5-para	✓	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 (Xu et al. 2020)	ResNet101	9-para		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
Mask OBB (Wang et al. 2019)	ResNet101	pixel-based	✓	89.56	**85.95**	54.21	72.90	76.52	74.16	85.63	89.85	83.81	86.48	54.89	**69.64**	73.94	69.06	63.32	75.33
FFA (Fu et al. 2020)	ResNet101	5-para	✓	90.1	82.7	54.2	75.2	71.0	**79.9**	83.5	90.7	83.9	84.6	61.2	68.0	70.7	76.0	63.7	75.7
APE (Zhu, Du, and Wu 2020)	ResNet101	5-para		89.96	83.62	53.42	76.03	74.01	77.16	79.45	90.83	87.15	84.51	**67.72**	60.33	74.61	71.84	65.55	75.75
CenterMap OBB (Wang et al. 2020)	ResNet101	5-para		89.83	84.41	54.60	70.25	**77.66**	78.32	**87.19**	90.66	84.89	85.27	56.46	69.23	74.13	71.56	66.06	76.03
RSDet-II (Ours)	ResNet152	8-para	✓	89.93	84.45	53.77	74.35	71.52	78.31	78.12	91.14	87.35	**86.93**	65.64	65.17	**75.35**	**79.74**	63.31	**76.34**
Single-stage methods																			
P-RSDet (Zhou et al. 2020)	ResNet101	3-para (polar)	✓	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-DNet (Wei et al. 2019)	Hourglass104	10-para	✓	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
DRN (Pan et al. 2020)	Hourglass104	5-para	✓	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 (Yang et al. 2019b)	ResNet152	5-para		89.49	81.17	50.53	66.10	70.92	78.66	78.21	90.81	85.26	84.23	61.81	63.77	68.16	69.83	**67.17**	73.74
RSDet-I (Ours)	ResNet152	8-para	✓	**90.01**	83.97	**54.72**	69.90	70.62	79.61	75.44	**91.20**	**88.03**	85.64	65.21	69.16	67.04	70.23	64.61	**75.03**

➢ Detection results on DOTA and ICDAR15.



(a) Vehicles (b) Tennis field (c) Text in signs (d) Storage tank (e) Harbor and ship (f) Text in elevator

➢ Performance on UCAS-AOD .

Method	Plane	Car	mAP
YOLOv2 (Redmon et al. 2016)	96.60	79.20	87.90
R-DFPN (Yang et al. 2018b)	95.90	82.50	89.20
DRBox (Liu, Pan, and Lei 2017)	94.90	85.00	89.95
S^2ARN (Bao et al. 2019)	97.60	92.20	94.90
RetinaNet-H (Yang et al. 2019b)	97.34	93.60	95.47
ICN (Azimi et al. 2018)	-	-	95.67
FADet (Li et al. 2019a)	98.69	92.72	95.71
R^3Det (Yang et al. 2019b)	98.20	94.14	96.17
Ours (RSDet)	**98.04**	**94.97**	**96.50**