ECCV'20
ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

www.eccv2020.eu

Arbitrary-Oriented Object Detection with Circular Smooth Label

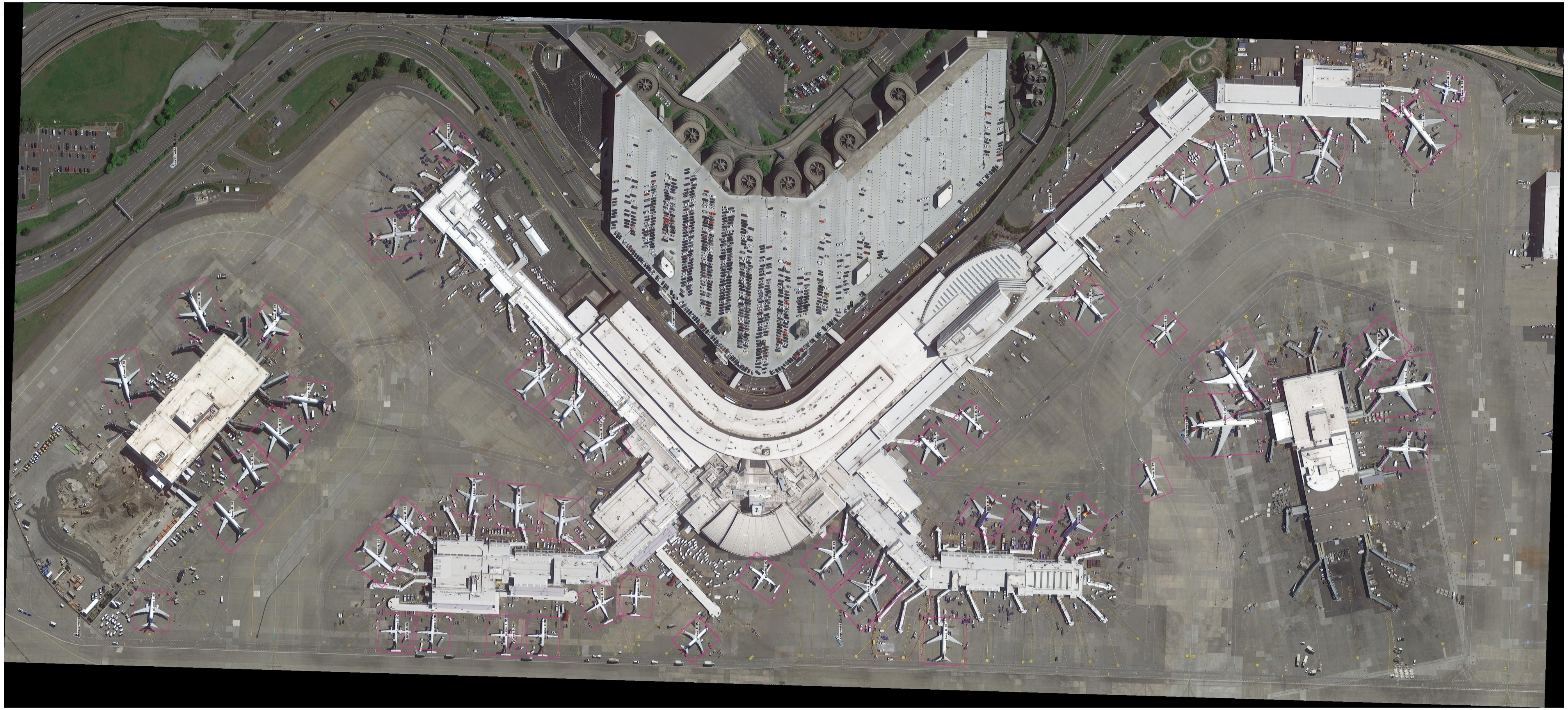
Xue Yang - Shanghai Jiao Tong University

X. Yang, J. Yan. “Arbitrary-Oriented Object Detection with Circular Smooth Label.” In ECCV 2020.

Glasgow, Scotland, UK. 2020. 7

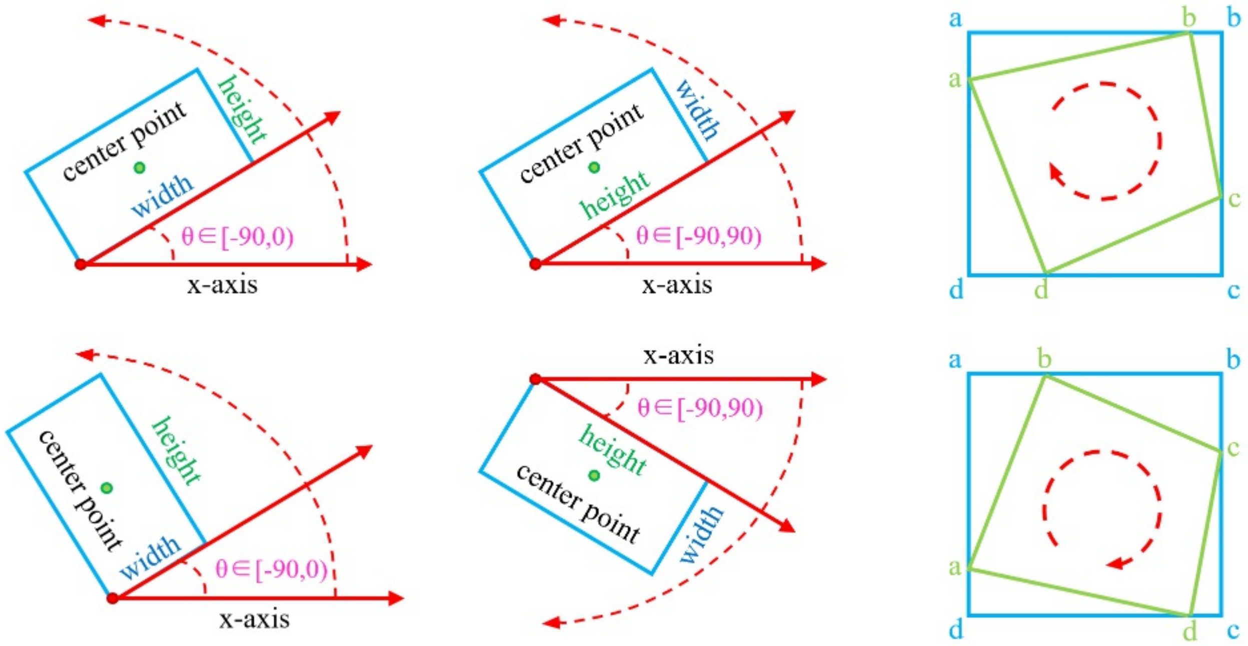
Arbitrary-Oriented Object Detection

Arbitrary-oriented object detection finds bounding box with orientation.



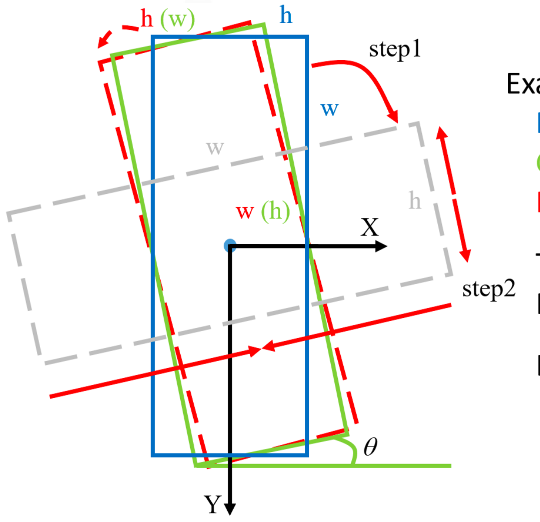
Arbitrary-Oriented Object

- Three representations for arbitrary-oriented object
  - Opencv Definition (x,y,w,h,θ)
  - Long Edge Definition (x,y,h,w,θ)
  - Quadrilateral Definition (x1,y1,…,x4,y4)



Boundary Problem

- The boundary discontinuity problem often makes the model's loss value suddenly increase at the boundary situation.
  - periodicity of angular **(PoA)**
  - exchangeability of edges **(EoE)**



Example:

Proposal: (0, 0, 100, 25, -pi/2)

Ground-Truth: (0, 0, 25, 100, -pi/8)

Predict box: (0, 0, 100, 25, -5pi/8)

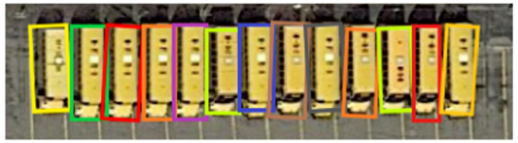
Target offset: (0, 0, log(1/4), log(4), 3pi/8)

Predict offset: (0, 0, 0, 0, -pi/8)

Loss=Smooth-L1(Predict - Target)

= loss(**PoA**) + loss(**EoE**) >> 0

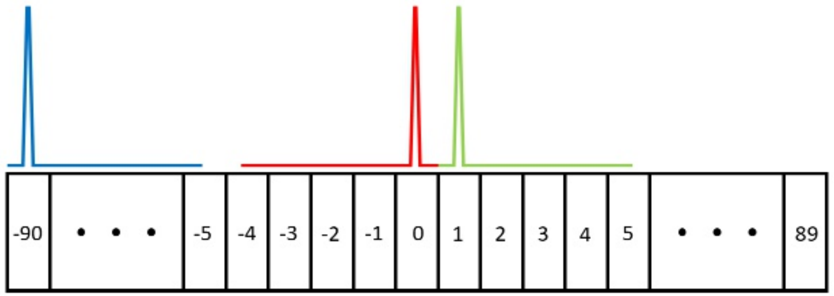
- The root cause of boundary problems based on regression methods is that the ideal predictions are beyond the defined range.



Yang, Xue, et al. "Scrdet: Towards more robust detection for small, cluttered and rotated objects." Proceedings of the IEEE International Conference on Computer Vision. 2019.

Vanilla Angular Classification

- Consider the prediction of the object angle as a classification problem to better limit the prediction results.



- Vanilla Angular Classification
  - EoE problem still exists
  - agnostic to the angle

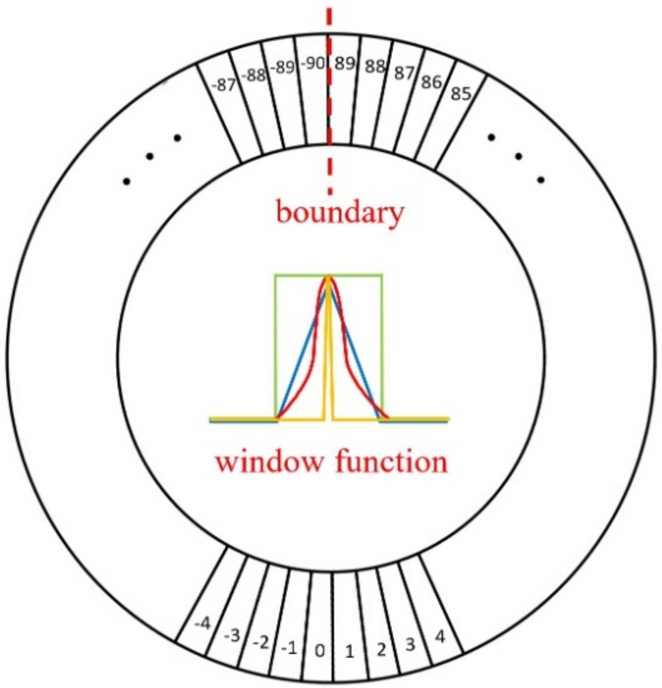
ground truth = one-hot(0)
predict1 ≈ one-hot(1)
predict2 ≈ one-hot(-90)
FL(gt – predict1) ≈ FL(gt – predict2)

Circular Smooth Label (CSL)

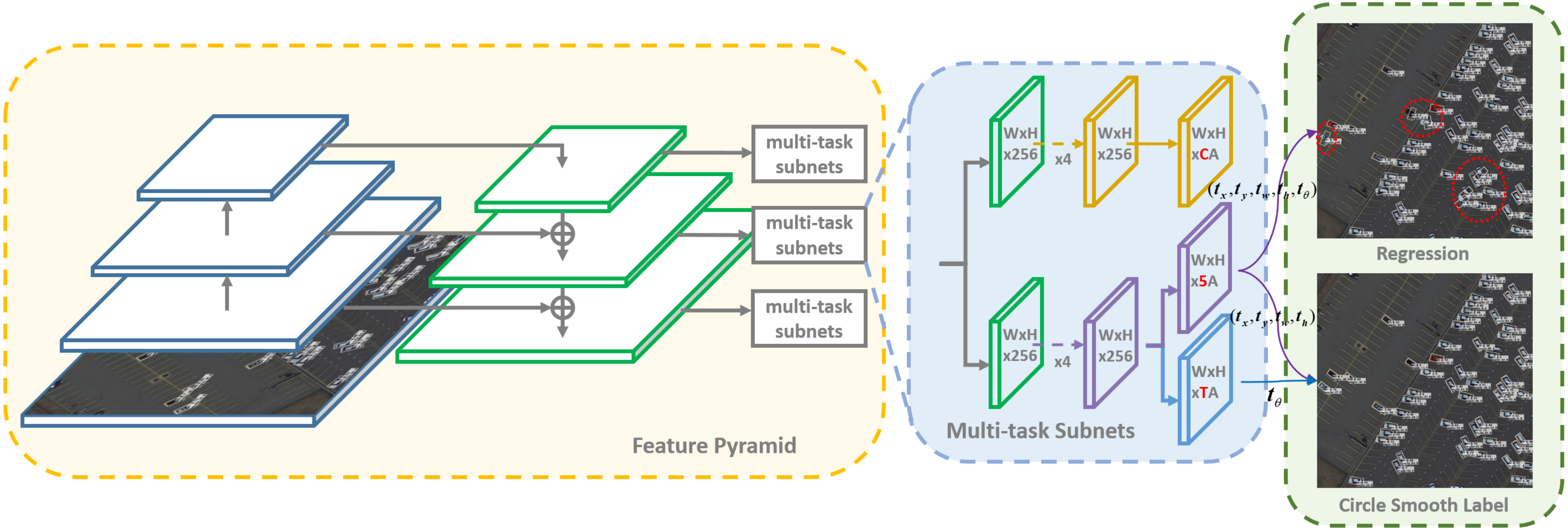
- CSL involves a circular label encoding with periodicity, and the assigned label value is smooth with a certain tolerance.

$$CSL(x) = \begin{cases} g(x), \theta - r < x < \theta + r \\ 0, \quad otherwise \end{cases}$$

- Properties:
  - Periodicity: $g(x) = g(x + kT), k \in N. \; T = 180/\omega$
  - Symmetry: $0 \leq g(\theta + \varepsilon) = g(\theta - \varepsilon) \leq 1, |\varepsilon| < r.$
  - Maximum: $g(\theta) = 1$
  - Monotonic: $0 \leq g(\theta \pm \varepsilon) \leq g(\theta \pm \varsigma) \leq 1, |\varsigma| < |\varepsilon| < r.$



Our Pipeline



Window Functions and Radius

- The Gaussian window function performs best, while the pulse function performs worst because it has not learned any orientation and scale information.

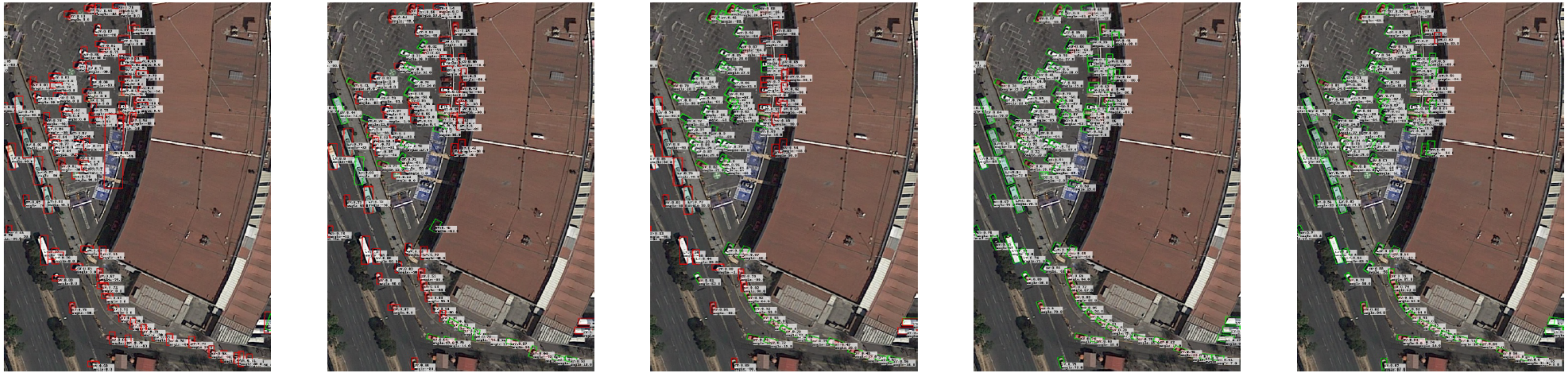
Based Method	Angle Range	EoE	Label Mode	BR	SV	LV	SH	HA	5-mAP	mAP
RetinaNet-H (CSL-Based)	90	✓	Pulse	9.80	28.04	11.42	18.43	23.35	18.21	39.52
	90	✓	Rectangular	37.62	54.28	48.97	62.59	50.26	50.74	58.86
	90	✓	Triangle	37.25	54.45	44.01	60.03	52.20	49.59	60.15
	90	✓	Gaussian	**41.03**	**59.63**	**52.57**	**64.56**	**54.64**	**54.49**	**63.51**
	180		Pulse	13.95	16.79	6.50	16.80	22.48	15.30	42.06
	180		Rectangular	36.14	60.80	50.01	65.75	53.17	53.17	61.98
	180		Triangle	32.69	47.25	44.39	54.11	41.90	44.07	57.94
	180		Gaussian	**41.16**	**63.68**	**55.44**	**65.85**	**55.23**	**56.21**	**64.50**

- The radius of the window function is very important.

Based Method	Angle Range	Label Mode	r=0	r=2	r=4	r=6	r=8
RetinaNet-H(CSL-Based)	180	Gaussian	40.78	59.23	62.12	**64.50**	63.99
FPN-H(CSL-Based)	180	Gaussian	48.08	70.18	70.09	**70.92**	69.75

Radius of Window Functions

Visualization of detection results under different radius. The red bounding box indicates that no orientation and scale information has been learned, and the green bounding box is the correct detection result.



(a) radius=0 (b) radius=2 (c) radius=4 (d) radius=6 (e) radius=8

ECCV'20 ONLINE 23-28 AUGUST 2020

CSL-Based VS Regression-Based

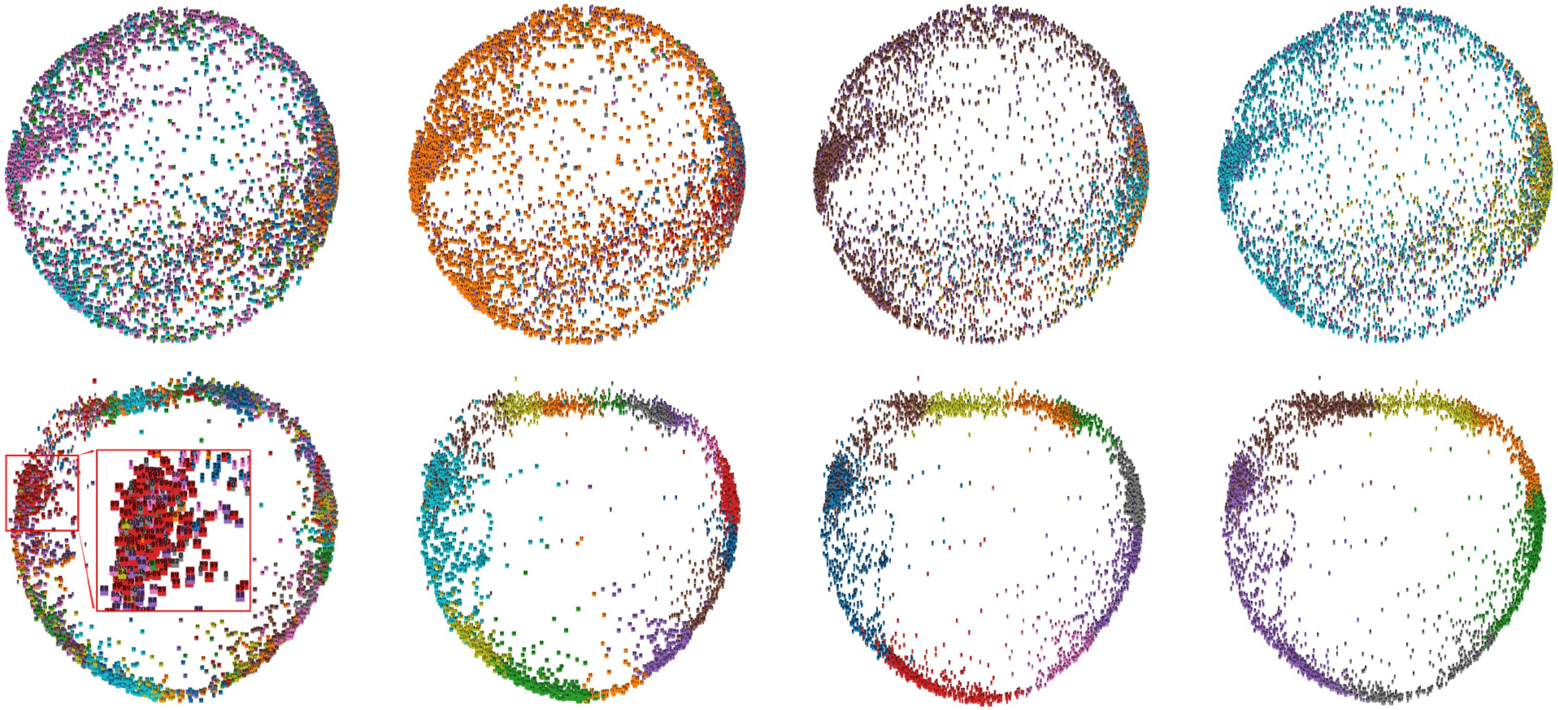
CSL has better detection ability for objects with large aspect ratios and more boundary conditions.

Based Method	Angle Range	Angle Pred.	PoA	EoE	Label Mode	BR	SV	LV	SH	HA	5-mAP	mAP
RetinaNet-H	90	regression-based	✓	✓	-	41.15	53.75	48.30	55.92	55.77	50.98	63.18
	90	CSL-based		✓	Gaussian	41.03	59.63	52.57	64.56	54.64	54.49 (**+3.51**)	63.51 (**+0.33**)
	180	regression-based	✓		-	38.47	54.15	47.89	60.87	53.63	51.00	64.10
	180	CSL-based			Gaussian	41.16	63.68	55.44	65.85	55.23	56.21 (**+5.21**)	64.50 (**+0.40**)
RetinaNet-R	90	regression-based	✓	✓	-	32.27	64.64	71.01	68.62	53.52	58.01	62.76
	90	CSL-based		✓	Gaussian	35.14	63.21	73.92	69.49	55.53	59.46 (**+1.45**)	65.45 (**+2.69**)
FPN-H	90	regression-based	✓	✓	-	44.78	70.25	71.13	68.80	54.27	61.85	68.25
	90	CSL-based		✓	Gaussian	45.46	70.22	71.96	76.06	54.84	63.71 (**+1.86**)	69.02 (**+0.77**)
	180	regression-based	✓		-	45.88	69.37	72.06	72.96	62.31	64.52	69.45
	180	CSL-based			Gaussian	47.90	69.66	74.30	77.06	64.59	66.70 (**+2.18**)	70.92 (**+1.47**)

Method	ICDAR2015			MLT			HRSC2016	
	Recall	Precision	Hmean	Recall	Precision	Hmean	mAP (07)	mAP (12)
FPN-regression-based	81.81	83.07	82.44	56.15	80.26	66.08	88.33	94.70
FPN-CSL-based	83.00	84.30	83.65 (**+1.21**)	56.72	80.77	66.64 (**+0.56**)	89.62 (**+1.29**)	96.10 (**+1.40**)

Visualizations

Angular feature visualization of the 90-CSL-FPN detector on the DOTA dataset. Each point represents a RoI of the test set with a index of the bin it belongs to.



pulse function

gaussian function

(a) bin=90 (b) bin=15 (c) bin=9 (d) bin=6

Comparison with the State-of-the-Art

TABLE V

DETECTION ACCURACY ON EACH OBJECT (AP) AND OVERALL PERFORMANCE (MAP) ON DOTA. NOTE O^2-DNET USES HOURGLASS104 [39] AS BACKBONE.

Method	Backbone	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
FR-O [33]	ResNet101	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 [40]	ResNet101	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 [5]	ResNet101	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
TOSO [41]	ResNet101	80.17	65.59	39.82	39.95	49.71	65.01	53.58	81.45	44.66	78.51	48.85	56.73	64.40	64.24	36.75	57.92
R^2CNN [8]	ResNet101	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 [9]	ResNet101	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
Axis Learning [42]	ResNet101	79.53	77.15	38.59	61.15	67.53	70.49	76.30	89.66	79.07	83.53	47.27	61.01	56.28	66.06	36.05	65.98
ICN [4]	ResNet101	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
RADet [43]	ResNeXt101	79.45	76.99	48.05	65.83	65.46	74.40	68.86	89.70	78.14	74.97	49.92	64.63	66.14	71.58	62.16	69.09
RoI-Transformer [2]	ResNet101	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
P-RSDet [44]	ResNet101	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
CAD-Net [45]	ResNet101	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
O^2-DNet [46]	Hourglass104	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
AOOD [47]	ResNet101	89.99	81.25	44.50	73.20	68.90	60.33	66.86	90.89	80.99	86.23	64.98	63.88	65.24	68.36	62.13	71.18
Cascade-FF [48]	ResNet152	89.9	80.4	51.7	77.4	68.2	75.2	75.6	90.8	78.8	84.4	62.3	64.6	57.7	69.4	50.1	71.8
SCRDet [3]	ResNet101	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 [49]	ResNet101	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
GLS-Net [50]	ResNet101	88.65	77.40	51.20	71.03	73.30	72.16	84.68	90.87	80.43	85.38	58.33	62.27	67.58	70.69	60.42	72.96
DRN [51]	Hourglass104	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
FADet [52]	ResNet101	90.21	79.58	45.49	76.41	73.18	68.27	79.56	90.83	83.40	84.68	53.40	65.42	74.17	69.69	64.86	73.28
MFIAR-Net [53]	ResNet152	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
R^3Det [1]	ResNet152	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 [20]	ResNet152	90.1	82.0	53.8	68.5	70.2	**78.7**	73.6	**91.2**	87.1	84.7	64.3	68.2	66.1	69.3	63.7	74.1
Gliding Vertex [27]	ResNet101	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 [54]	ResNeXt-101	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 [55]	ResNet101	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 [56]	ResNeXt-101	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
CSL (FPN based)	ResNet152	**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**

DOTA dataset

TABLE VI

ACCURACY AND SPEED ON HRSC2016. 07 (12) MEANS USING THE 2007(2012) EVALUATION METRIC. TICK MEANS THE MODULE IS ENABLED.

Method	Backbone	mAP (07)	mAP (12)
R^2CNN [8]	ResNet101	73.07	79.73
RC1 & RC2 [36]	VGG16	75.7	–
RRPN [9]	ResNet101	79.08	85.64
R^2PN [57]	VGG16	79.6	–
RetinaNet-H	ResNet101	82.89	89.27
RRD [10]	VGG16	84.3	–
RoI-Transformer [2]	ResNet101	86.20	–
Gliding Vertex [27]	ResNet101	88.20	–
DRN [51]	Hourglass104	–	92.70
SBD [30]	ResNet50	–	93.70
RetinaNet-R [1]	ResNet101	89.18	95.21
R^3Det [1]	ResNet101	89.26	96.01
CSL (FPN based)	ResNet101	**89.62**	**96.10**

HRSC2016 dataset

Thank you!

- Paper: https://arxiv.org/abs/2003.05597
- Code: https://github.com/Thinklab-SJTU/CSL_RetinaNet_Tensorflow

- Contact:

  - Xue Yang: yangxue-2019-sjtu@sjtu.edu.cn
  - Junchi Yan: yanjunchi@sjtu.edu.cn

- Homepage of our lab:
  - http://thinklab.sjtu.edu.cn/

ECCV'20
ONLINE
23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON
COMPUTER VISION

WWW.ECCV2020.EU