SCRDet: Towards More Robust Detection for Small, Cluttered and Rotated Objects

Xue Yang[1,2,3], Jirui Yang[2], Junchi Yan[3,*], Yue Zhang[1], Tengfei Zhang[1,2] , Zhi Guo[1] , Xian Sun[1] , Kun Fu[1,2]

{yangxue-2019-sjtu, yanjunchi}@sjtu.edu.cn, {yangjirui16, zhangtengfei16}@mails.ucas.ac.cn,

zhangyue@aircas.ac.cn, fguozhi, sunxian, fukun@mail.ie.ac.cn

[1]Institute of Electronics, Chinese Academy of Sciences, Beijing , China.

[2]University of Chinese Academy of Sciences, Beijing, China.

[3]MoE Key Lab of Artificial Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University

ICCV 2019 Seoul, Korea

Introduction

- **Rotation Object Detection**

➢ Task: design a novel multi-category rotation detector for small, cluttered and rotated objects.

➢ Challenges

- **Small objects**. Aerial images often contain small objects overwhelmed by complex surrounding scenes.
- **Cluttered arrangement**. Objects for detection are often densely arranged, such as vehicles and ships.
- **Arbitrary orientations**. Objects in aerial images can appear in various orientations. It is further challenged by the large aspect ratio issue which is common in remote sensing.

- **Our main contributions**

➢ For small objects, a tailored feature fusion structure is devised by feature fusion and anchor sampling.

➢ For cluttered, small object detection, a supervised multi-dimensional attention network is developed to reduce the adverse impact of background noise.

➢ Towards more robust handling of arbitrarily-rotated objects, an improved smooth L1 loss is devised by adding the IoU constant factor, which is tailored to solve the boundary problem of the rotating bounding box regression

➢ Codes: https://github.com/DetectionTeamUCAS

Proposed Approach

- **Pipeline**

In the first stage, the feature map is expected to contain more feature information and less noise by adding SF-Net and MDA-Net. For positional sensitivity of the angle parameters, this stage still regresses the horizontal box. By the improved five-parameter regression and the rotation nonmaximum-suppression (R-NMS) operation for each proposal in the second stage, we can obtain the final detection results under arbitrary rotations.

- **Sampling and Fusion Network (SF-Net)**

➢ **Feature fusion:** SF-Net only uses C3 and C4 in Resnet for fusion to balance the semantic information and location information while ignoring other less relevant features.

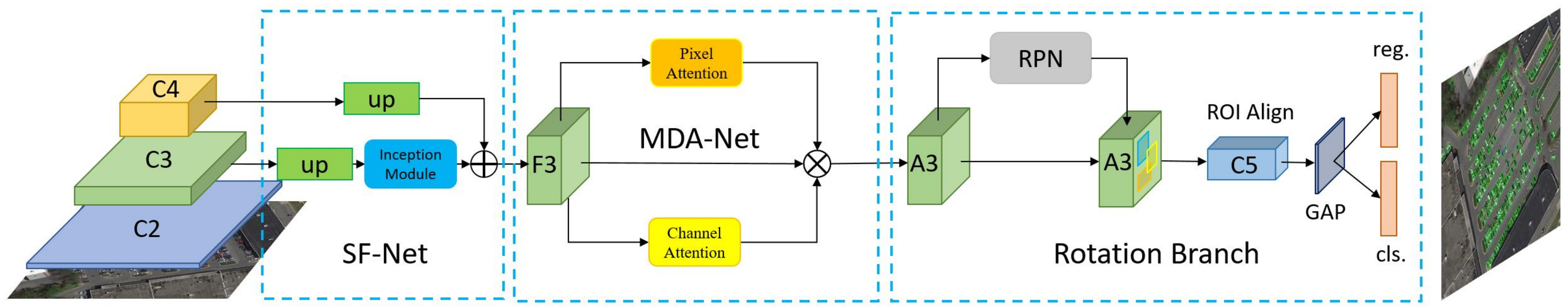

Fig: SCRDet includes SF-Net, MDA-Net against small and cluttered objects and rotation branch for rotated objects.

➢ **Finer sampling:** SF-Net solves this problem by changing the size of the feature map, making the setting of SA more flexible to allow for more adaptive sampling. In the first stage, the feature map is expected to contain.

- **Multi-Dimensional Attention Network (MDA-Net)**

➢ The supervised pixel attention network and the channel attention network are jointly explored for small and cluttered object detection by suppressing the noise and highlighting the objects feature.

- **IoU-Smooth L1 Loss**

➢ For more accurate rotation estimation, the IoU constant factor is added to the smooth L1 loss to address the boundary problem for the rotating bounding box. The IoU-Smooth L1 Loss is defined as follows:

$$L = \frac{\lambda_1}{N}\sum_{n=1}^{N} t'_n \sum_{j\in\{x,y,w,h,\theta\}} \frac{L_{reg}(v'_{nj}, v_{nj})}{|L_{reg}(v'_{nj}, v_{nj})|} |-\log(IoU)|$$

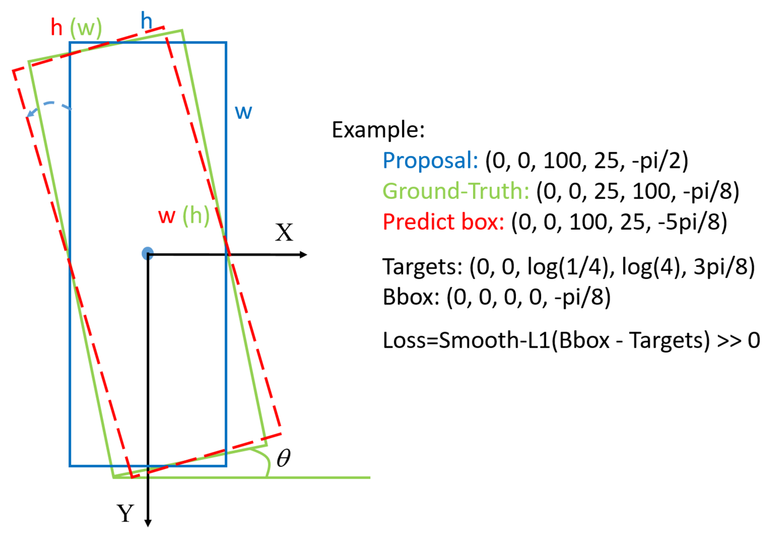

Fig: Boundary discontinuity of the rotation angle.

➢ The new regression loss can be divided into two parts, $L_{reg}(v'_j, v_j)/|L_{reg}(v'_j, v_j)|$ determines the direction of gradient propagation, $-\log(IoU)$ for the magnitude of gradient.

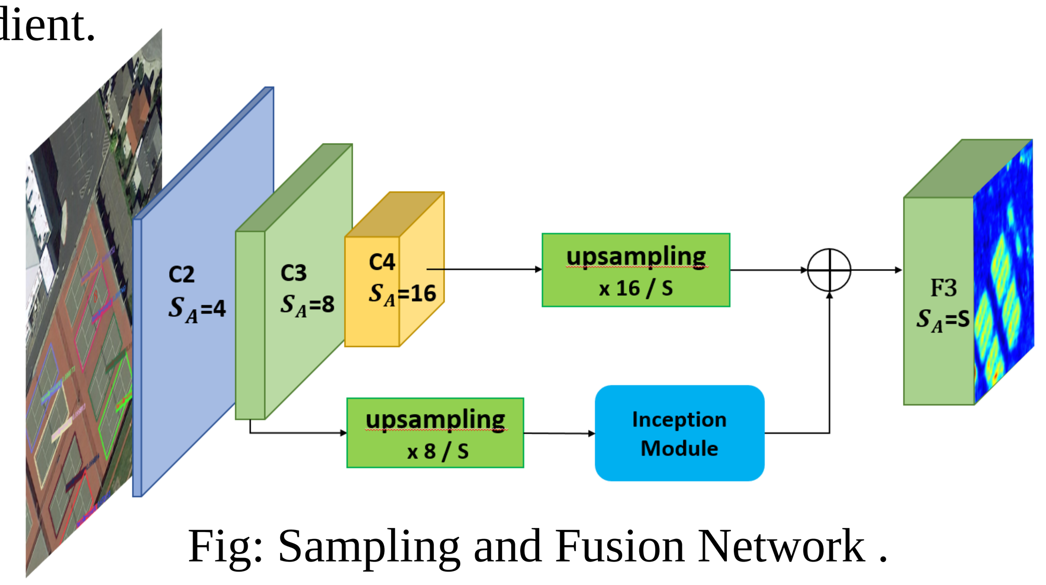

Fig: Sampling and Fusion Network .

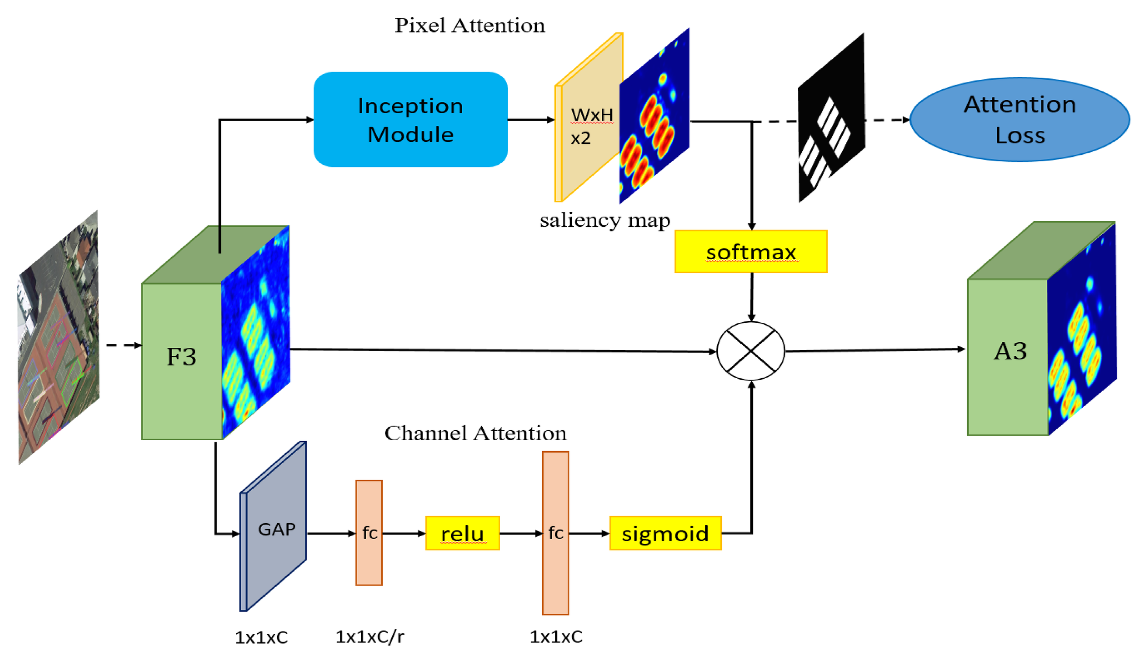

Fig: Multi-Dimensional Attention Network.

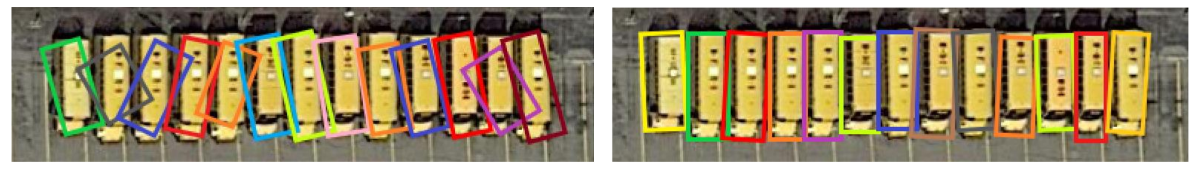

Fig: Comparison before and after using the IoU-Smooth L1 Loss.

Experiments

➢ Ablative study of each components in our proposed method on the DOTA dataset.

Method	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
R^2CNN (baseline) [19]	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
+Pixel Attention	81.17	75.23	36.71	68.14	62.33	48.22	55.75	89.57	78.40	76.61	54.08	58.32	63.76	61.94	54.89	64.34
+MDA	84.89	77.07	38.55	67.88	61.78	51.87	56.23	89.82	75.77	76.30	53.68	63.25	63.85	65.05	53.99	65.33
+SA [45]+MDA	81.27	76.49	38.16	69.13	54.03	46.51	55.03	89.80	69.92	75.11	57.06	58.51	62.70	59.72	48.20	62.78
+SJ [45]+MDA	81.13	76.02	32.79	66.94	60.73	48.12	54.86	90.29	74.54	76.25	54.00	57.27	63.87	60.24	43.48	62.70
+BU [45] +MDA	84.63	75.34	42.84	68.47	63.11	53.69	57.13	90.70	76.93	75.28	55.63	58.28	64.57	67.10	49.19	65.53
+BUS [45]+MDA	87.50	75.60	42.41	69.48	62.45	50.89	56.10	**90.87**	78.41	75.68	58.94	58.68	63.87	67.38	52.78	66.07
+DC [45]+MDA	87.01	76.66	42.25	68.95	62.55	53.62	56.22	90.83	78.54	75.49	58.54	57.17	63.99	66.77	57.43	66.40
+SF+MDA	89.65	79.51	43.86	67.69	67.41	55.93	64.86	90.71	77.77	84.42	57.67	61.38	64.29	66.12	62.04	68.89
+SF+MDA+IoU	89.41	78.83	50.02	65.59	69.96	57.63	72.26	90.73	81.41	84.39	52.76	63.62	62.01	67.62	61.16	69.83
+SF +MDA+IoU+P	**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**

➢ Performance evaluation of OBB and HBB task on DOTA datasets.

Method	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
OBB																
FR-O [39]	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
R-DFPN [41]	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 [19]	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 [29]	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 [2]	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 [8]	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 (proposed)	**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**
HBB																
SSD [10]	44.74	11.21	6.22	6.91	2.00	10.24	11.34	15.59	12.56	17.94	14.73	4.55	4.55	0.53	1.01	10.94
YOLOv2 [30]	76.90	33.87	22.73	34.88	38.73	32.02	52.37	61.65	48.54	33.91	29.27	36.83	36.44	38.26	11.61	39.20
R-FCN [5]	79.33	44.26	36.58	53.53	39.38	34.15	47.29	45.66	47.74	65.84	37.92	44.23	47.23	50.64	34.90	47.24
FR-H [31]	80.32	77.55	32.86	68.13	53.66	52.49	50.04	90.41	75.05	59.59	57.00	49.81	61.69	56.46	41.85	60.46
FPN [23]	88.70	75.10	52.60	59.20	69.40	**78.80**	**84.50**	90.60	81.30	82.60	52.50	62.10	**76.60**	66.30	60.10	72.00
ICN [2]	90.00	77.70	53.40	**73.30**	**73.50**	65.00	78.20	90.80	79.10	84.80	57.20	62.10	73.50	70.20	58.10	72.50
SCRDet (proposed)	**90.18**	**81.88**	**55.30**	73.29	72.09	77.65	78.06	**90.91**	**82.44**	**86.39**	**64.53**	**63.45**	75.77	**78.21**	**60.11**	**75.35**

➢ Performance on NWPU VHR-10.

Method	mAP
R-P-Faster R-CNN [14]	76.50
SSD512 [26]	78.40
DSSD321 [10]	78.80
DSOD300 [33]	79.80
Deformable R-FCN [40]	79.10
Deformable Faster R-CNN [32]	84.40
RICADet [22]	87.12
RDAS512 [3]	89.50
Multi-Scale CNN [13]	89.60
SCRDet (proposed)	**91.75**

➢ Effectiveness of the proposed structure on generic datasets

Dataset	Model	Backbone	mAP/F1
COCO	FPN*	Res50	36.1
	FPN*+IoU-Smooth	Res50	36.2
	FPN*+MDA-Net	Res50	**36.8**
VOC2007	FPN*	Res101	76.14
	FPN*+MDA-Net	Res101	**78.36**
ICDAR2015	R^2CNN-4*	Res101	77.23
	SCRDet-R^2CNN	Res101	**80.08**

➢ Visualization on different datasets.

Recent Works

➢ We have designed a fast and accurate single-stage based refined rotation detector that solves the problem of feature misalignment. [arXiv:1808.03766]